



(10) **DE 20 2016 107 480 U1** 2017.07.13

(12) **Gebrauchsmusterschrift**

(21) Aktenzeichen: **20 2016 107 480.0**

(22) Anmeldetag: **29.12.2016**

(47) Eintragungstag: **02.06.2017**

(45) Bekanntmachungstag im Patentblatt: **13.07.2017**

(51) Int Cl.: **G06F 3/16 (2006.01)**

(30) Unionspriorität:  
**15/057,453 01.03.2016 US**

(73) Name und Wohnsitz des Inhabers:  
**GOOGLE INC., Mountain View, Calif., US**

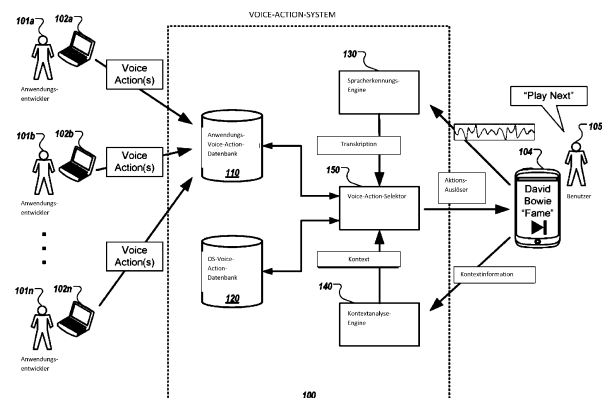
(74) Name und Wohnsitz des Vertreters:  
**Maikowski & Ninnemann Patentanwälte  
Partnerschaft mbB, 10707 Berlin, DE**

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

(54) Bezeichnung: **Voice-Action-Systeme für Entwickler**

(57) Hauptanspruch: Computersystem, umfassend:  
das Empfang von Daten durch ein Voice-Action-System, die eine neue Voice Action für eine Software-Anwendung spezifizieren, die vom besagten Voice-Action-System verschieden ist, wobei die Daten eine oder mehrere Operationen zum Ausführen der neuen Voice Action und einen oder mehrere Auslöserbegriffe zur Auslösung der neuen Voice Action umfassen.

das Erstellen einer kontextuellen Voice-Action-Absicht für die Software-Anwendung durch das Voice-Action-System auf Grundlage von mindestens den empfangenen Daten, wobei die kontextuelle Voice-Action-Absicht Daten umfasst, die, wenn sie von der Software-Anwendung empfangen werden, anfordern, dass die Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt; Verknüpfung, durch das Voice-Action-System, der kontextuellen Voice-Action-Absicht mit dem einen oder den mehreren Auslöserbegriffen für die neue Voice Action; wobei das Voice-Action-System dazu konfiguriert ist: eine Angabe einer Benutzeräußerung, die von einem Gerät erhalten wurde, auf dem die Software-Anwendung installiert ist, zu empfangen, festzustellen, dass eine Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit der kontextuellen Voice-Action-Absicht verknüpft sind; und dem Gerät die kontextuelle Voice-Action-Absicht bereitzustellen, in Reaktion auf das Feststellen, und dadurch anzufordern, dass die auf dem Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt.



**Beschreibung**

## SACHGEBIET

**[0001]** Diese Spezifikation bezieht sich auf Voice Actions, und eine bestimmte Implementierung bezieht sich auf die Konfiguration von Voice Actions.

## ALLGEMEINER STAND DER TECHNIK

**[0002]** Eine Aufgabe in einer Anwendung kann eine oder mehrere in definierte Aktivitäten in Software beinhalten, mit der ein Benutzer interagiert, um eine bestimmte Arbeit auszuführen. Aktivität ist eine Kategorie, die den Lebenszyklus der Ausführung der Aufgabe steuert, so dass die verschiedenen Aktivitäten, die innerhalb der Aufgabe ausgeübt werden, dem Benutzer gestatten, die Arbeit auszuführen. In einigen Implementierungen kann ein Befehl mit einer Aktivität oder einer mit einer Aktivität verbundenen Aktion verknüpft sein, so dass eine Benutzereingabe des Befehls die Aktivität oder die Aktion auslösen kann. Eine spezifische Absicht kann ausgelöst werden, um den Beginn der Aktivität oder der Ausführung der Aktion auszulösen.

## KURZDARSTELLUNG

**[0003]** Diese Spezifikation beschreibt eine Plattform, die es einem Anwendungsentwickler erlaubt, neue Voice Actions für zuvor installierte Software-Anwendungen auszurollen. Im Allgemeinen kann der Anwendungsentwickler unter Nutzung der Plattform Informationen eingeben, die die neue Voice Action definieren, wobei die Information die Anwendung spezifiziert, eine Aktion, die die neue Voice Action auslösen soll, einen Auslöserbegriff zum Auslösen der Aktion und einen Kontext, in dem der Auslöserbegriff anwendbar sein soll, um die Aktion in der Anwendung auszulösen.

**[0004]** Der vom Anwendungsentwickler eingegebene Kontext kann einen Benutzergerätestatus spezifizieren, in dem die Voice Action anwendbar sein soll, um die Aktion in der Anwendung auszulösen. Ein Gerätestatus kann beispielsweise beinhalten, welche Anwendung(en) im Vordergrund (d. h. gegenwärtig auf der Benutzeroberfläche des Geräts aktiv) oder im Hintergrund (d. h. gegenwärtig auf der Benutzeroberfläche des Geräts versteckt) läuft/laufen, oder kann spezifische Information zu Anwendungen beinhalten, wie beispielsweise die gegenwärtig laufenden Aktivitäten, den Status laufender Aktivitäten und so weiter.

**[0005]** Wenn der Anwendungsentwickler Informationen eingibt, die die neue Voice Action definieren, kann der Dienst oder das Werkzeug die Information, die die neue Voice Action definiert, validieren um festzustellen, ob die neue Voice Action mit der Anwendung kompatibel ist oder um auf andere Art festzu-

stellen, ob die neue Voice Action implementiert werden kann. Wenn die neue Voice Action gültig ist, kann eine neue Instanz einer passiven Datenstruktur, die Absicht genannt wird und ein spezifisches Format besitzt, für die Voice Action durch Einfügung der Information, die die neue Voice Action definiert, erstellt werden. Die Absicht kann einen Teil oder die gesamte Information, die die neue Voice Action definiert, spezifizieren, wie beispielsweise die Anwendung, den Auslöserbegriff, die Aktivität oder Aktion, die auf die Erkennung des Auslöserbegriffs hin ausgelöst werden soll, sowie den nötigen Kontext für die Auslösung der Aktivität oder Aktion durch den Auslöserbegriff. Die Voice Action kann dann so für die Anwendung ausgerollt werden, dass die Voice Action eine freigegebene Voice Action für die Anwendung wird, ohne dass weitere Änderungen am Anwendungscode vorgenommen werden, um die Voice Action zu unterstützen.

**[0006]** Sobald die Voice Action ausgerollt ist, kann ein Benutzer, der ein Benutzergerät bedient, eine Spracheingabe vornehmen. Das Benutzergerät kann Kontextinformation für das Benutzergerät übermitteln, und die Kontextinformation und eine Transkription der Spracheingabe können genutzt werden, um eine Absicht zu identifizieren und die identifizierte Absicht auszulösen. Sobald die Absicht ausgelöst wird, werden Daten zur Ausführung der Aktivität oder Aktion bestimmt und an das Benutzergerät übertragen, um die Aktivität oder Aktion auf die Spracheingabe hin auszuführen.

**[0007]** Ein Anwendungsentwickler einer Mediaplayer-Anwendung kann beispielsweise eine neue Voice Action zum Springen zu einem nächsten Lied definieren, die den Auslöserbegriff „play next“ enthält. Der Anwendungsentwickler kann einen Kontext spezifizieren, in dem die Voice Action „play next“ freigegeben sein soll. Der Anwendungsentwickler kann beispielsweise spezifizieren, dass die Voice Action „play next“ den Mediaplayer nur dann dazu bringen soll, zum nächsten Lied zu springen, wenn die Mediaplayer-Anwendung im Vordergrund läuft und sich in einem Modus befindet, der die Mediaplayer-Anwendung dazu veranlasst, als Audioplayer zu funktionieren. Nachdem die Voice Action vom Anwendungsentwickler ausgerollt wurde, könnte ein Benutzer, der zuvor die Mediaplayer-Anwendung auf seinem Benutzergerät installiert hat, die Spracheingabe „play next“ vornehmen. Das Benutzergerät kann Kontextinformation übermitteln, die einen Status des Benutzergeräts oder von auf dem Benutzergerät installierten Anwendungen angibt, wie beispielsweise Kontextinformation, die angibt, dass die Mediaplayer-Anwendung im Vordergrund läuft und sich im Audioplayer-Modus befindet. Auf den Empfang der Spracheingabe „play next“ und der Information, die den Kontext spezifiziert, hin, können Daten an das Benutzergerät übertragen werden, die die Mediaplayer-An-

wendung dazu bringen, zum nächsten Lied zu springen. Wenn im Gegensatz dazu die Spracheingabe als „play next“ bestimmt wird, die Kontextinformation aber anzeigt, dass eine Anwendung eines sozialen Netzwerks statt der Mediaplayer-Anwendung im Vordergrund des Benutzergeräts läuft, ist es möglich, dass die Spracheingabe „play next“ keinen Effekt hat, oder dazu führt, dass auf dem Benutzergerät eine andere Operation ausgeführt wird.

**[0008]** Innovative Aspekte des in dieser Spezifikation beschriebenen Gegenstands können in computerimplementierten Systemen ausgeführt sein, die die folgenden Aktionen beinhalten: Empfang von Daten über ein Voice-Action-System, die eine neue Voice Action für eine Software-Anwendung spezifizieren, die vom besagten Voice-Action-System verschieden ist, wobei die Daten eine oder mehrere Operationen zur Ausführung der neuen Voice Action sowie einen oder mehrere Auslöserbegriffe zur Auslösung der neuen Voice Action umfassen, Erstellung einer kontextuellen Voice-Action-Absicht für die Software-Anwendung durch das Voice-Action-System, die zumindest auf den empfangenen Daten beruht, wobei die kontextuelle Voice-Action-Absicht Daten umfasst, die, wenn sie von der Software-Anwendung empfangen werden, anfordern, dass die Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt, Verknüpfung der kontextuellen Voice-Action-Absicht mit dem einen oder den mehreren Auslöserbegriffen für die neue Voice Action durch das Voice-Action-System, wobei das Voice-Action-System dazu konfiguriert ist: eine Angabe der Benutzeräußerung zu empfangen, die von einem Gerät erhalten wird, auf dem die Software-Anwendung installiert ist, festzustellen, dass eine Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit der kontextuellen Voice-Action-Absicht verknüpft sind und, auf die Feststellung hin, dem Gerät die kontextuelle Voice-Action-Absicht bereitzustellen und dadurch anzufordern, dass die Software-Anwendung, die auf dem Gerät installiert ist, die eine oder die mehreren Operationen der neuen Voice-Action ausführt.

**[0009]** Diese und andere Ausführungsformen können jeweils optional eine oder mehrere der folgenden Funktionen beinhalten. In verschiedenen Beispielen spezifizieren die empfangenen Daten einen Kontext, wobei der Kontext einen Gerätestatus oder einen Status der Software-Anwendung spezifiziert, bei deren Vorliegen die neue Voice Action freigegeben ist; der Kontext spezifiziert, dass die Software-Anwendung im Vordergrund eines Geräts läuft, auf dem die Software-Anwendung installiert ist; der Kontext spezifiziert, dass die Software-Anwendung im Hintergrund eines Geräts läuft, auf dem die Software-Anwendung installiert ist; der Kontext spezifiziert, dass die Software-Anwendung eine spezifische Aktivität ausführt;

der Kontext spezifiziert, dass eine von der Software-Anwendung ausgeführte spezifische Aktivität sich in einem bestimmten Aktivitätszustand befindet.

**[0010]** In anderen Beispielen können diese und andere Ausführungsformen jeweils optional Funktionen enthalten, umfassend: Verknüpfung der kontextuellen Voice-Action-Absicht mit dem Kontext für die neue Voice Action, Empfang von Kontextinformation durch das Voice-Action-System, die einen Status eines bestimmten Geräts, auf dem die Software-Anwendung installiert ist, oder der Software-Anwendung, die auf dem bestimmten Gerät installiert ist, angibt, Feststellung, dass die Kontextinformation die Kontextanforderungen für die neue Voice Action erfüllt, und auf die Feststellung hin, dass eine Transkription einer vom bestimmten Gerät erhaltenen Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit der kontextuellen Voice-Action-Absicht verknüpft sind, und dass die Kontextinformation die Kontextanforderungen erfüllt, die mit der kontextuellen Voice-Action-Absicht verknüpft sind, Bereitstellung der kontextuellen Voice-Action-Absicht durch das Voice-Action-System an das bestimmte Gerät, und dadurch die Anfrage, dass die Software-Anwendung, die auf dem bestimmten Gerät installiert ist, die eine oder die mehreren Operationen der neuen Voice Action ausführt; Empfang der Kontextinformation, die den Status des bestimmten Geräts oder der Software-Anwendung, die auf dem bestimmten Gerät installiert ist, angibt, umfasst: Stellung einer Anfrage nach bestimmter Kontextinformation durch das Voice-Action-System an das bestimmte Gerät und Empfang der bestimmten Kontextinformation auf die Anfrage hin.

**[0011]** In weiteren Beispielen können diese und andere Ausführungsformen jeweils eine oder mehrere Funktionen beinhalten, umfassend: Feststellung, dass die Kontextinformation die Kontextanforderungen für eine zweite Voice Action erfüllt, und dass eine Transkription einer vom bestimmten Gerät erhaltenen Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit einer Absicht für die zweite Voice Action verknüpft sind, wobei die Absicht für die zweite Voice Action eine oder mehrere Operationen zur Ausführung der zweiten Voice Action spezifiziert, auf die Feststellung hin, Auswahl einer Voice Action zwischen der neuen Voice Action und der zweiten Voice Action und Bereitstellung der Absicht, die mit der ausgewählten Voice Action verknüpft ist, durch das Voice-Action-System an das bestimmte Gerät und dadurch die Anfrage, dass die Software-Anwendung, die auf dem bestimmten Gerät installiert ist, die eine oder die mehreren Operationen der ausgewählten Voice Action ausführt, Auswahl der ausgewählten Voice Action zwischen der neuen Voice Action und der zweiten Voice Action umfasst die Auswahl der ausgewählten Voice Action auf den Empfang von Daten hin, die eine Benutzeraus-

wahl zwischen der neuen Voice Action und der zweiten Voice Action angeben; die Auswahl der ausgewählten Voice Action zwischen der neuen Voice Action und der zweiten Voice Action umfasst: Zuweisung eines Werts für sowohl die neue Voice Action als auch die zweite Voice Action und die Auswahl der ausgewählten Voice Action zumindest auf Grundlage des Werts, der jeweils der neuen Voice Action und der zweiten Voice Action zugewiesen wurde; die Auswahl der ausgewählten Voice Action zwischen der neuen Voice Action und der zweiten Voice Action umfasst die Auswahl der ausgewählten Voice Action auf die Feststellung hin, dass die Software-Anwendung, die mit der ausgewählten Voice Action verknüpft ist, im Vordergrund läuft.

**[0012]** In anderen Beispielen können diese und andere Ausführungsformen jeweils eine oder mehrere Funktionen beinhalten, umfassend: Erstellung der kontextuellen Voice-Action-Absicht für die die Software-Anwendung umfasst die Feststellung, dass die eine oder die mehreren Operationen der neuen Voice Action von der Software-Anwendung ausführbar sind; Feststellung, dass eine Transkription einer Benutzeräußerung, die von einem bestimmten Gerät erhalten wird, auf dem die Software-Anwendung installiert ist, Ähnlichkeit mit dem einen oder den mehreren Auslöserbegriffen aufweist, die mit der kontextuellen Voice-Action-Absicht verknüpft sind, auf die Feststellung hin, Bereitstellung von Daten durch das Voice-Action-System an das bestimmte Gerät, die eine Anfrage nach Benutzereingaben angeben, die bestätigen, ob die Benutzeräußerung den einen oder die mehreren Auslöserbegriffe spezifiziert haben oder das Ziel hatten, die Software-Anwendung dazu zu veranlassen, die neue Voice Action auszuführen, auf die Anfrage hin, Empfang, durch das Voice-Action-System vom bestimmten Gerät, von Daten, die eine Bestätigung angeben und auf den Empfang der Daten hin, die die Bestätigung angeben, Bereitstellung, durch das Voice-Action-System an das bestimmte Gerät, der kontextuellen Voice-Action-Absicht, und dadurch die Anfrage, dass die Software-Anwendung, die auf dem bestimmten Gerät installiert ist, die eine oder die mehreren Operationen der neuen Voice Action ausführt.

**[0013]** In anderen Beispielen können diese und andere Ausführungsformen jeweils eine oder mehrere Funktionen beinhalten, umfassend: Empfang, durch das Voice-Action-System von einem mit der Software-Anwendung verbundenen Entwickler, einer Anfrage, die neue Voice Action für die Software-Anwendung auszurollen, und auf die Anfrage hin, Ausrollen der neuen Voice Action für die Software-Anwendung auf die Anfrage hin, wobei das Ausrollen der neuen Voice Action für die Software-Anwendung die neue Voice Action für die Software-Anwendung aktiviert; Empfang, durch das Voice-Action-System von einem mit der Software-Anwendung verbundenen Entwick-

ler, einer Anfrage, das Ausrollen der neuen Voice Action für die Software-Anwendung rückgängig zu machen, und auf die Anfrage hin, Rückgängigmachen des Ausrollens der neuen Voice Action für die Software-Anwendung auf die Anfrage hin, wobei das Rückgängigmachen des Ausrollens der neuen Voice Action die neue Voice Action für die Software-Anwendung deaktiviert.

**[0014]** In anderen Beispielen können diese und andere Ausführungsformen jeweils eine oder mehrere Funktionen beinhalten, umfassend: Empfang, durch das Voice-Action-System von einem mit der Software-Anwendung verbundenen Entwickler, einer Anfrage, eine Erprobung der neuen Voice Action freizugeben, wobei die Anfrage ein oder mehrere Geräte spezifiziert, für die die neue Voice Action aktiviert werden soll, und auf die Anfrage hin, Aktivierung der neuen Voice Action für das eine oder die mehreren spezifizierten Geräte, wobei die neue Voice Action für Geräte, die nicht in den spezifizierten Geräten enthalten sind, deaktiviert ist; Empfang, durch das Voice-Action-System, einer Angabe einer Benutzeräußerung, die von einem bestimmten Gerät, auf dem die Software-Anwendung installiert ist, erhalten wird, Feststellung, durch das Voice-Action-System, dass eine Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit der kontextuellen Voice-Action-Absicht verknüpft sind, und, auf die Feststellung hin, Bereitstellung, durch das Voice-Action-System an das bestimmte Gerät, der kontextuellen Voice-Action-Absicht, und dadurch die Anfrage, dass die Software-Anwendung, die auf dem bestimmten Gerät installiert ist, die eine oder die mehreren Operationen der neuen Voice Action ausführt.

**[0015]** Die Details einer oder mehrerer Ausführungsformen des Gegenstands, der in dieser Spezifikation beschrieben wird, werden in den beigefügten Zeichnungen und in der nachstehenden Beschreibung dargestellt. Andere Merkmale, Aspekte und Vorteile des Gegenstands werden aus der Beschreibung, den Zeichnungen und den Ansprüchen deutlich.

#### KURZBESCHREIBUNG DER ZEICHNUNGEN

**[0016]** Fig. 1 zeigt ein beispielhaftes System für ein System und einen Dienst für die Voice-Action-Entwicklung.

**[0017]** Fig. 2 zeigt ein beispielhaftes System zur Erstellung neuer Voice Actions unter Nutzung eines Systems und Dienstes für die Voice-Action-Entwicklung.

**[0018]** Fig. 3 zeigt ein beispielhaftes System zur Nutzung von Voice Actions.

**[0019]** Fig. 4 ist ein Flussdiagramm eines beispielhaften Prozesses, der mit einem System und Dienst für die Voice-Action-Entwicklung verknüpft ist.

**[0020]** Gleiche Bezugszeichen in unterschiedlichen Zeichnungen geben gleiche Elemente an.

#### DETAILLIERTE BESCHREIBUNG

**[0021]** Fig. 1 veranschaulicht ein Beispiel für ein Voice-Action-System **100**. Voice-Action-System **100** bietet eine Plattform und einen Dienst, durch die ein Anwendungsentwickler neue Voice Actions für eine Anwendung, die zuvor auf den Geräten anderer Benutzer installiert worden ist, anlegen kann. Danach können Benutzer der Anwendung, die die Anwendung auf ihren Geräten installiert haben, Sprachbefehle nutzen, um die Anwendung dazu zu bringen, eine spezifische Operation wie von der neuen Voice Action spezifiziert auszuführen.

**[0022]** Voice-Action-System **100** empfängt Daten von den Terminals **102a–102n**, die neue Voice Actions definieren, die von den Anwendungsentwicklern **101a–101n** übermittelt werden, wobei die neuen Voice Actions für eine oder mehrere verschiedene Anwendungen vorgesehen sind. Voice-Action-System **100** kann feststellen, ob jede einzelnen neue Voice Action eine gültige Voice Action ist. Jede neue Voice Action, die als gültig bestimmt wird, kann eingefügt werden, um eine Absicht zu generieren, die der neuen Voice Action entspricht, und die Absicht, die der neuen Voice Action entspricht, kann in einer Anwendungs-Voice-Action-Datenbank **110** des Voice-Action-Systems **100** gespeichert werden.

**[0023]** Voice-Action-System **100** kann beispielsweise eine Validierungseingabe besitzen, die die Daten, die die neue Voice Action definieren, empfängt und validiert. Die Validierung der Daten kann die Feststellung beinhalten, dass das Format der Daten, die die Voice Action definieren die Anforderungen erfüllt, dass die neue Voice Action mit der spezifizierten Anwendung kompatibel ist oder dass eine Voice Action auf Grundlage der vom Anwendungsentwickler **101a–101n** empfangenen Daten angelegt werden kann. Eine Kennzeichnungseingabe des Voice-Action-Systems **100** kann die validierten Daten empfangen, die die neue Voice Action definieren und kann die Daten einfügen, um die der neuen Voice Action entsprechende Absicht zu erstellen. Die Absicht kann dann in der Anwendungs-Voice-Action-Datenbank **110** gespeichert werden.

**[0024]** Die Anwendungs-Voice-Action-Datenbank **110** kann Voice Actions speichern, die Aktivitäten oder Aktionen betreffen, die innerhalb zahlreicher verschiedener Software-Anwendung ausgeführt werden können. Die Voice Actions, die in der Anwendungs-Voice-Action-Datenbank **110** enthalten sind,

können sowohl eingebaute Voice Actions beinhalten, die von Anwendungsentwicklern **101a–101n** während der Erstellung der Anwendung eingefügt wurden, als auch Voice Actions, die von Anwendungsentwicklern **101a–101n** nach der Erstellung der Anwendung für von der Anwendung unterstützte Operationen eingefügt wurden.

**[0025]** Die Voice Actions können auch anwendungsspezifische Voice Actions beinhalten, die standardmäßig vom Betriebssystem unterstützt werden. Diese vorunterstützten Voice Actions können Voice Actions sein, die in Anwendungen implementierbar sind, ohne von einem Anwendungsentwickler **101a–101n** in das Voice-Action-System **100** eingefügt werden zu müssen. Beispielsweise kann die Voice Action „Anwendung verlassen“, um eine im Vordergrund laufende Aufgabe oder Aktivität zu verlassen, für jedwede Anwendung automatisch verfügbar sein und kann für jede Anwendung in der Anwendungs-Voice-Action-Datenbank **110** enthalten sein, ohne dass ein Anwendungsentwickler **101a–101n** Information in das Voice-Action-System **100** einfügen muss, um die Voice Action „Anwendung verlassen“ zu definieren.

**[0026]** Zusätzlich zur Anwendungs-Voice-Action-Datenbank **110**, kann das Voice-Action-System **100** eine Betriebssystem(OS)-Voice-Action-Datenbank **120** beinhalten, die Voice Actions speichert, die nicht mit einer bestimmten Anwendung oder einem bestimmten Kontext verknüpft sind. Eine Voice Action „Telefon sperren“, die ein Gerät in einen gesperrten Zustand versetzt, kann beispielsweise eine Voice Action sein, die in der OS-Voice-Action-Datenbank **120** als Voice Action spezifiziert ist, die nicht mit einer bestimmten Anwendung verknüpft ist oder die universell zugänglich ist, d. h. unabhängig vom Status eines Geräts, wenn die „Gerät sperren“-Voice-Action festgestellt wird. Im Allgemeinen sind die in der OS-Voice-Action-Datenbank **120** gespeicherten Voice Actions Voice Actions, die nicht auf Grundlage von Anwendungsentwicklern **101a–101n** definierten Voice-Actions erstellt werden, da die in der OS-Voice-Action-Datenbank **120** gespeicherten Voice Actions typisch für eine Betriebsumgebung eines Benutzergeräts sind und nicht für irgendeine bestimmte Anwendung, die von dritten Anwendungsentwicklern **101a–101n** erstellt wurde.

**[0027]** Voice-Action-System **100** gibt einem Benutzer **105**, der sich im Besitz eines Benutzergeräts **104** befindet, die Möglichkeit, Spracheingaben an seinem Benutzergerät **104** vorzunehmen, um die Ausführung von Aktionen durch das Benutzergerät **104** oder durch Anwendungen, die auf seinem Benutzergerät **104** betrieben werden, auszulösen. Benutzer **105**, der sich im Besitz eines Benutzergeräts **104** befindet, betreibt beispielsweise eine Mediaplayer-Anwendung auf seinem Benutzergerät **104** und nimmt die Spracheingabe „play next“ vor, während der Me-

dioplayer im Vordergrund des Benutzergeräts **104** läuft.

**[0028]** Audiodaten, die der Spracheingabe und der Kontextinformation entsprechen, die einen Status des Benutzergeräts **104** spezifiziert, inklusive des Status von Anwendungen, die auf dem Benutzergerät **104** laufen, werden an das Voice-Action-System **100** übertragen. Insbesondere für das gezeigte Beispiel in **Fig. 1**, werden die Audiodaten, die der Spracheingabe „play next“ entsprechen, von einer Spracherkennungseingabe **130** des Voice-Action-Systems **100** empfangen, und die Kontextinformation, die den Status des Benutzergeräts **104** angibt, wird von einer Kontextanalyse-Engine **140** des Voice-Action-Systems **100** empfangen. Die Kontextinformation kann angeben, dass die Mediaplayer-Anwendung im Vordergrund des Benutzergeräts **104** läuft, dass die Mediaplayer-Anwendung sich derzeit im Audio-player-Modus befindet, und kann andere Informationen zum Status des Benutzergeräts **104** und von Anwendungen, die auf dem Benutzergerät **104** installiert sind, angeben.

**[0029]** Die Spracherkennungseingabe **130** empfängt die Audiodaten, die der Spracheingabe entsprechen, erstellt eine Transkription der Spracheingabe und stellt die Transkription der Spracheingabe dem Voice-Action-Selektor **150** bereit. Die Kontextanalyse-Engine **140** empfängt die Kontextinformation vom Benutzergerät **104** und verarbeitet die Kontextinformation, um relevante Kontextinformation zu bestimmen. Kontextanalyse-Engine **140** kann beispielsweise die Kontextinformation analysieren, um Anwendungen, die auf dem Benutzergerät **104** installiert sind, zu identifizieren, und Metadaten, die mit jeder dieser Anwendungen verknüpft sind, können verfügbare Voice Actions für die Anwendung spezifizieren, sowie Kontextinformation, die relevant für die Feststellung ist, welche der Voice Actions für einen gegebenen Gerätestatus aktiviert werden können. In einigen Fällen kann die Kontextanalyse-Engine **140** auf Grundlage der Analyse feststellen, dass zusätzliche Kontextinformation nötig ist um zu identifizieren, welche Voice Actions für einen bestimmten Status des Benutzergeräts **104** aktiviert sein sollten, wodurch die Kontextanalyse-Engine **140** zusätzliche Kontextinformation vom Benutzergerät **104** anfordern kann. Die Kontextanalyse-Engine **140** leitet die verarbeitete Kontextinformation an den Voice-Action-Selektor **150** weiter.

**[0030]** Wie zum Beispiel in **Fig. 1**, kann die Spracherkennungseingabe **130** die Audiodaten empfangen, die der Spracheingabe „play next“ entsprechen und kann eine Transkription der Spracheingabe erhalten. Die Kontextanalyse-Engine **140** empfängt Kontextinformation vom Benutzergerät **104**, die angibt, dass die Mediaplayer-Anwendung im Vordergrund läuft, dass die Mediaplayer-Anwendung in einem Au-

dioplayer-Modus läuft, und die andere Informationen zum Benutzergerät **104** spezifiziert. Die Kontextinformation kann z. B. auch angeben, dass eine Anwendung eines sozialen Netzwerks im Hintergrund des Benutzergeräts **104** läuft, und dass die Batterieladung des Benutzergeräts **104** derzeit 50% beträgt. Die Kontextanalyse-Engine **150** kann Information erhalten, die angibt, dass auf dem Benutzergerät **104** sowohl der Mediaplayer als auch Anwendungen von sozialen Netzwerken installiert sind, und kann auf Grundlage von Kontextinformation, die einen Batterieladestatus des Geräts angibt, bestimmen, dass weder die Mediaplayer-Anwendung noch die Anwendungen von sozialen Netzwerken Voice Actions zulassen. Deshalb kann die Kontextanalyse-Engine **140** die Kontextinformation filtern, um nur anzugeben, dass die Mediaplayer-Anwendung im Vordergrund des Benutzergeräts **104** läuft, dass die Mediaplayer-Anwendung sich in einem Audio-player-Modus befindet und dass die Anwendung des sozialen Netzwerks im Hintergrund des Benutzergeräts **104** läuft. Die gefilterte Kontextinformation und die Transkription der Spracheingabe „play next“ können dann dem Voice-Action-Selektor **150** bereitgestellt werden.

**[0031]** Voice-Action-Selektor **150** empfängt die Transkription der Spracheingabe von der Spracherkennungseingabe **130** und einen Kontext von der Kontextanalyse-Engine **140**, der verarbeitete Kontextinformation beinhaltet. Voice-Action-Selektor **150** nutzt die Transkription und den Kontext, um eine bestimmte Absicht zu identifizieren, die mit einer bestimmten Voice Action verknüpft ist, die auf dem Benutzergerät **104** auszulösen ist. Voice-Action-Selektor **150** kann z. B. auf die Anwendungs-Voice-Action-Datenbank **110** und auf die OS-Voice-Action-Datenbank **120** zugreifen, um einen Satz von möglichen Voice Actions zu identifizieren, die für den gegenwärtigen Kontext des Benutzergeräts **104** verfügbar sind. Die Identifizierung des Satzes von möglichen Voice Actions, kann die Identifizierung eines Satzes von Absichten beinhalten, die in der Anwendungs-Voice-Action-Datenbank **110** oder der OS-Voice-Action-Datenbank **120** gespeichert sind und die Kontexte spezifizieren, die mit dem vom Voice-Action-Selektor **150** erhaltenen Kontext übereinstimmen.

**[0032]** Sobald der Satz von möglichen Voice Actions, die für den gegenwärtigen Kontext verfügbar sind, identifiziert ist, kann der Voice-Action-Selektor **150** die Transkription der Spracheingabe mit einem oder mehreren Auslösersätzen vergleichen, die jeweils mit den verfügbaren Voice Actions verknüpft sind. Im Allgemeinen kann ein Auslösersatz einen oder mehrere Auslöserbegriffe beinhalten und operiert als Kennung für eine Voice Action, dergestalt dass die Erkennung des einen oder der mehreren Begriffe eines bestimmten Auslösersatzes zur Identifizierung und Auslösung der Voice Action führt, die dem bestimmten Auslösersatz entspricht. Voice-Action-

tion-Selektor **150** kann beispielsweise die Transkription mit entsprechenden Auslösersätzen vergleichen, die mit den Absichten der verfügbaren Voice Actions verknüpft sind. Voice-Action-Selektor **150** identifiziert eine bestimmte Absicht für eine bestimmte Voice Action auf Grundlage der Feststellung, dass die Transkription mit einem Auslöserbegriff übereinstimmt, der von der Absicht, die mit der bestimmten Voice Action verknüpft ist, spezifiziert wird.

**[0033]** Voice-Action-Selektor **150** kann z. B. die Transkription der Spracheingabe „play next“ und den Kontext für das Benutzergerät **104** empfangen und kann den empfangenen Kontext dazu nutzen, mögliche Voice Actions zu identifizieren, die auf dem Benutzergerät **104** initiiert sind, d. h. Voice Actions, die auf dem Benutzergerät **104** auf Grundlage des gegenwärtigen Status des Benutzergeräts **104** initiiert werden können. Nach der Bestimmung des Satzes von möglichen Voice Actions für den Kontext des Benutzergeräts **104** vergleicht der Voice-Action-Selektor **150** die Transkription der Spracheingabe „play next“ mit Auslösersätzen, die von den Absichten der möglichen Voice Actions spezifiziert werden. Voice-Action-Selektor **150** kann feststellen, dass die Transkription „play next“ mit dem Auslösersatz übereinstimmt, der von der Absicht der „play next“-Voice-Action für die Mediaplayer-Anwendung spezifiziert wird.

**[0034]** Eine Transkription wird aufgrund der Feststellung, dass zumindest ein Teil der Transkription mit einem oder mehreren Begriffen des Auslösersatzes übereinstimmt, als mit einem Auslösersatz übereinstimmend identifiziert. Eine Übereinstimmung zwischen einer Transkription und einem Auslösersatz kann beispielsweise auf Grundlage dessen identifiziert werden, dass jeder der Begriffe der Transkription mit jedem der Begriffe des Auslösersatzes übereinstimmen. In einigen Implementierungen kann eine Übereinstimmung auf Grundlage der Feststellung identifiziert werden, dass weniger als alle Begriffe der Transkription mit Begriffen des Auslösersatzes übereinstimmen. Eine Transkription kann beispielsweise auch dann mit einem Auslösersatz übereinstimmen, wenn einige der Begriffe des Auslösersatzes in der Transkription abwesend sind, wenn die Transkription im Vergleich zum Auslösersatz zusätzliche Begriffe enthält oder wenn einer oder mehrere der Begriffe der Transkription sich von Begriffen des Auslösersatzes unterscheiden. Somit kann jedwede der Transkriptionen „neue E-Mail“, „neue E-Mail senden“ oder „jetzt neue E-Mail öffnen“ als mit dem Auslösersatz „neue E-Mail öffnen“ übereinstimmend identifiziert werden, obwohl keine der Transkriptionen exakt mit dem Auslösersatz übereinstimmt. In einigen Fällen kann eine Transkription als mit einem Auslösersatz übereinstimmend identifiziert werden, wenn die Transkription einige oder alle Begriffe aus dem Auslösersatz in einer anderen Reihenfolge enthält als derjenigen, die vom Auslösersatz vorgegeben wird. Die Transkrip-

tion „Sperrung Telefon“ kann beispielsweise als mit dem Auslösersatz „Telefon Sperrung“ übereinstimmend identifiziert werden. In einigen Fällen kann eine Transkription auf Grundlage der Feststellung als mit einem Auslösersatz übereinstimmend identifiziert werden, dass die Transkription eines oder mehrere Synonyme oder alternative Begriffe für Begriffe des Auslösersatzes beinhaltet. Somit können die Transkriptionen „neue E-Mail verfassen“ oder „neue E-Mail öffnen“ als mit dem Auslösersatz „neue E-Mail schreiben“ übereinstimmend identifiziert werden, basierend darauf, dass „verfassen“ ein Synonym von „schreiben“, und „öffnen“ ein identifizierter alternativer Begriff für „schreiben“ ist.

**[0035]** Sobald die bestimmte Voice Action identifiziert ist, stellt der Voice-Action-Selektor **150** dem Benutzergerät **104** aktionsauslösende Daten bereit, die die Initiierung der Aktivität oder Aktion, die mit der bestimmten Voice Action verknüpft ist, hervorrufen. Zu diesem Zweck kann der Voice-Action-Selektor **150** Information zur Steuerung des Benutzergeräts **104** erhalten, um die Aktion oder Aktivität, die mit der ausgewählten Voice Action verknüpft ist, auszuführen. In einigen Fällen kann die Steuerung des Benutzergeräts **104** zur Ausführung der Aktion oder Aktivität, die mit der ausgewählten Voice Action verknüpft ist, das Senden einer Absicht der ausgewählten Voice Action beinhalten. Das Senden der Absicht der ausgewählten Voice Action kann dazu führen, dass Information für die Initiierung der Aktivität oder Aktion, die mit der ausgewählten Voice Action verknüpft ist, dem Benutzergerät **104** bereitgestellt wird.

**[0036]** Die Absicht für die ausgewählte Voice Action kann beispielsweise Daten beinhalten, die die Ausführung von Aktionen oder Aktivitäten, die mit der Voice Action verknüpft sind, durch die mit der Voice Action verknüpfte Anwendung herbeiführen. Die ausgewählte Absicht kann an das Benutzergerät **104** übertragen werden, dergestalt dass der Empfang der Absicht durch das Benutzergerät **104** als Anfrage agieren kann oder die Ausführung der Aktionen oder Aktivitäten, die mit der Voice Action verknüpft sind, auslösen kann. Alternativ dazu kann der Voice-Action-Selektor **150** andere Daten dazu bestimmen, als Aktionsauslöserdaten zu operieren, die die ausgewählte Voice Action dazu bringen, auf dem Benutzergerät **104** ausgeführt zu werden und die Information an das Benutzergerät **104** zu übertragen. Eine Absicht kann beispielsweise lediglich die Aktionen oder Aktivitäten definieren, die durch das Benutzergerät **104** auszuführen sind, um die Voice Action auszuführen, und der Voice-Action-Selektor **150** kann aktionsauslösende Daten bestimmen, die die Anwendung auf dem Benutzergerät **104** steuern können, um die Aktionen und Aktivitäten, die mit der Voice Action verknüpft sind, auszuführen. Die bestimmten aktionsauslösenden Daten können so auf das Benutzergerät **104** übertragen werden, dass die aktionsauslösen-

den Daten die Ausführung der Aktionen oder Aktivitäten herbeiführen, die mit der Voice Action verknüpft sind.

**[0037]** Nach der Auswahl der „play next“-Voice-Action aus dem Satz von möglichen Voice Actions überträgt der Voice-Action-Selektor **150** beispielsweise die Absicht für die „play next“-Voice-Action an das Benutzergerät **104** oder erhält andere Informationen zur Steuerung der Mediaplayer-Anwendung, die auf dem Benutzergerät **104** läuft, um zum nächsten Lied zu springen. Voice-Action-Selektor **150** überträgt die Daten zur Steuerung der Mediaplayer-Anwendung zum Springen zum nächsten Lied an das Benutzergerät **104**, welches wiederum einen Prozess zum Springen zum nächsten Lied für die Mediaplayer-Anwendung auf dem Benutzergerät **104** initiiert, beispielsweise zum Springen zum Lied „Fame“ von David Bowie.

**[0038]** Kurz gefasst, wie besprochen, beinhaltet das System aus **Fig. 1** eines oder mehrere Terminals **102a–102n**, die einem oder mehreren Anwendungsentwicklern **101a–101n** entsprechen. Terminals **102a–102n** können jeweils mit dem Voice-Action-System **100** kommunizieren, beispielsweise über ein oder mehrere kabelbasierte oder drahtlose Netzwerke. Jedes der Terminals **102a–102n** kann ein Mobilgerät sein, wie beispielsweise ein Mobiltelefon, Smartphone, Tablet-Computer, Laptop-Computer, Personal Digital Assistant (PDA), Netbook-Computer oder ein anderes mobiles Computergerät, oder es kann jedwedes stationäre Computergerät sein, so wie Desktopcomputer oder andere stationäre Computer. Ein mit einem Benutzer **105** verknüpftes Benutzergerät **104** kann auch über ein oder mehrere kabelbasierte oder drahtlose Netzwerke mit dem Voice-Action-System **100** in Kommunikation stehen, wobei das Benutzergerät **104** auch ein mobiles oder stationäres Computergerät sein kann, so wie ein Mobiltelefon, Smartphone, Tablet-Computer, Netbook, PDA, Laptop-Computer, Desktopcomputer oder ein anderes Computergerät.

**[0039]** Das beispielhafte Voice-Action-System **100** aus **Fig. 1** beinhaltet eine Anwendungs-Voice-Action-Datenbank **110**, eine OS-Voice-Action-Datenbank **120**, eine Spracherkennungseingabe **130**, eine Kontextanalyse-Engine **140** und einen Voice-Action-Selektor **150**. Jede der Komponenten des Voice-Action-Systems **100**, inklusive der Anwendungs-Voice-Action-Datenbank **110**, der OS-Voice-Action-Datenbank **120**, der Spracherkennungseingabe **130**, der Kontextanalyse-Engine **140** und des Voice-Action-Selektors **150**, kann mit einer oder mehreren Komponenten des Voice-Action-Systems **100** über einen oder mehrere kabelbasierte oder drahtlose Datenwege, die den Austausch elektronischer Kommunikation ermöglichen, in Kommunikation stehen. In einigen Implementierungen können eine oder mehrere

re Komponenten des Voice-Action-Systems **100** so kombiniert sein, dass ihre Funktionen von einer einzigen Komponente ausgeführt werden oder können von zwei oder mehr Komponenten dargestellt werden, so dass ihre Funktionen sich unter den zwei oder mehr Komponenten aufteilen. Die Komponenten des Voice-Action-Systems **100** können auf einem einzelnen Computergerät implementiert sein oder sie können auf mehreren Computergeräten implementiert sein, die miteinander über einen oder mehrere kabelbasierte oder drahtlose Datenwege in Kommunikation stehen, die den Austausch von elektronischer Kommunikation zwischen den Komponenten ermöglichen.

**[0040]** **Fig. 2** veranschaulicht ein beispielhaftes System, dass es einem Anwendungsentwickler ermöglicht, neue Voice Actions für eine Anwendung zu erstellen, die auf einem oder mehreren Benutzergeräten installiert ist. Beispielsweise kann der Anwendungsentwickler, nachdem eine Anwendung veröffentlicht und auf eines oder mehrere Benutzergeräte heruntergeladen worden ist, unter Nutzung eines Voice-Action-Entwicklungsdienstes eine neue Voice Action definieren. Der Voice-Action-Entwicklungsdienst kann die neue Voice Action ausrollen, um die neue Voice Action für die Anwendung zu aktivieren, ohne dass dazu Modifikationen an der Anwendung selbst nötig wären.

**[0041]** Kurz gefasst, **Fig. 2** zeigt ein Entwickler-Voice-Action-System **200**, das Voice-Action-Entwicklungsdienste für einen Anwendungsentwickler **201** bietet, der mit einem Terminal **202** verknüpft ist. Terminal **202** kann mit dem Entwickler-Voice-Action-System **200** über eine oder mehrere kabelbasierte oder drahtlose Netzwerkverbindungen in Kommunikation stehen, wie z. B. eine kabelbasierte oder drahtlose Internetverbindung, die das Terminal **202** in die Lage versetzt, elektronische Kommunikation mit dem Entwickler-Voice-Action-System **200** auszutauschen. Entwickler-Voice-Action-System **200** beinhaltet eine Validierungseingabe **210**, die mit einer Validierungskriterien-Datenbank **215** in Kommunikation steht. Entwickler-Voice-Action-System **200** beinhaltet des Weiteren eine Grammatikinduktions-Engine **220**, die mit einem Anwendungsmetadaten-Backend **225** und einer Anwendungs-Voice-Action-Datenbank **230** in Kommunikation steht. Die Anwendungs-Voice-Action-Datenbank **230** ähnelt möglicherweise der Anwendungs-Voice-Action-Datenbank **110** aus **Fig. 1** und kann Absichten speichern, die aus Grammatiken erstellt werden, die von Anwendungsentwicklern eingegeben werden, um neue Voice Actions für Anwendungen zu definieren oder die auf Information basieren, die von Anwendungsentwicklern eingegeben wird und neue Voice Actions definiert. In einigen Implementierungen ist das Entwickler-Voice-Action-System **200**, das den Dienst bereitstellt, mit dem Anwendungsentwickler neue Voice Actions anlegen,



möglicherweise ein Teil des Voice-Action-Systems **100** aus **Fig. 1**. In einigen Implementierungen können von Komponenten des Entwickler-Voice-Action-Systems **200** ausgeführte Operationen von verschiedenen Komponenten des Entwickler-Voice-Action-Systems **200** ausgeführt werden. Entwickler-Voice-Action-System **200** kann mehr, weniger oder andere Komponenten umfassen, um die Operationen auszuführen, die als vom Entwickler-Voice-Action-System **200** ausgeführt beschrieben werden.

**[0042]** Wie gezeigt in **Fig. 2**, kann der Anwendungsentwickler **201** unter Nutzung des Terminals **202** Information in das Entwickler-Voice-Action-System **200** einfügen, die eine neue Voice Action definiert. Der Terminal **202** kann beispielsweise eine Anwendung für einen Entwickler-Voice-Action-Dienst beinhalten oder der Terminal **202** kann in der Lage sein, auf den Entwickler-Voice-Action-Dienst über eine Netzwerkverbindung fernzugreifen, beispielsweise indem auf einer Website über das Internet auf den Entwickler-Voice-Action-Dienst zugegriffen wird. In anderen Implementierungen kann der Entwickler-Voice-Action-Dienst von einem Entwickler-Voice-Action-System **200** dergestalt gehostet werden, dass das Terminal **202** über die Netzwerkverbindung zwischen Terminal **202** und Entwickler-Voice-Action-System **200** auf den Entwickler-Voice-Action-Dienst im Entwickler-Voice-Action-System **200** zugreifen kann.

**[0043]** Um eine neue Voice Action für eine bestimmte Anwendung zu implementieren, kann der Anwendungsentwickler **201** eine neue Grammatik in das Entwickler-Voice-Action-System **200** einfügen, die die neue Voice Action definiert, oder er kann andere Informationen einfügen, die die neue Voice Action für das Entwickler-Voice-Action-System **200** definieren. Im Beispiel, dargestellt in **Fig. 2**, definiert der Anwendungsentwickler **201**, um eine neue Voice Action zu implementieren, zunächst eine neue Voice Action für eine Mediaplayer-Anwendung, die bereits auf einem oder mehreren Benutzergeräten installiert ist. Zu diesem Zweck fügt der Anwendungsentwickler **201** eine Grammatik ein, die die neue Voice Action für die Mediaplayer-Anwendung definiert. Die vom Anwendungsentwickler **201** eingefügte Grammatik spezifiziert eine Anwendung, nämlich die Mediaplayer-Anwendung, einen Auslösersatz „play next“, der mit der neuen Voice Action zu verknüpfen ist, eine Aktion, die auf die Erkennung des Auslösersatzes „play next“ hin auszuführen ist, nämlich zum nächsten Lied zu springen, und einen Kontext für die Aktivierung der neuen Voice Action, nämlich dass die Mediaplayer-Anwendung im Vordergrund eines Benutzergeräts laufen muss und dass die Mediaplayer-Anwendung sich im Audioplayer-Modus befinden muss.

**[0044]** Die vom Anwendungsentwickler **201** eingefügte Grammatik kann ein spezifisches Format besitzen, dass es dem Anwendungsentwickler **201** er-

möglicht, problemlos die neue Voice Action zu definieren, ohne mit einem spezifischen Datenformat, das vom Voice-Action-System **200** für Voice-Action-Absichten benutzt wird, vertraut zu sein. In anderen Implementierungen kann der Anwendungsentwickler **201** andere Informationen einfügen, die von der Validierungsengine **210** oder einer anderen Komponente des Entwickler-Voice-Action-Systems **200** empfangen und übersetzt werden können, um eine Grammatik zu erstellen, die die neue Voice Action definiert. Dem Anwendungsentwickler **201** kann beispielsweise ein Formular angezeigt werden, das Felder beinhaltet, die der Anwendungsentwickler **201** ausfüllen kann, um eine neue Voice Action zu definieren. Das Formular kann dem Anwendungsentwickler **201** die Möglichkeit geben, Information über die neue Voice Action zu spezifizieren, z. B. eine Anwendung, einen Auslösersatz für die neue Voice Action und eine Aktion oder Aktivität, die durch die Anwendung oder diese betreffend auf den Auslösersatz hin ausgeführt werden soll. Das Formular kann es dem Anwendungsentwickler **201** ermöglichen, einen Kontext anzugeben, der einen spezifischen Status eines Benutzergeräts oder von einer oder mehreren Anwendungen, die auf dem Benutzergerät laufen, erfordert, damit die neue Voice Action aktiviert wird, das heißt, damit die Voice Action eine mögliche Voice Action ist, die ein Benutzer durch Verwenden des Auslösersatzes auslösen kann. Andere Informationen, die nötig sind oder optional benutzt werden, um eine neue Voice Action zu definieren, können in das Formular eingegeben werden. Die Eingaben in den Feldern des Formulars, die vom Anwendungsentwickler **201** vorgenommen werden, können in eine Grammatik zur Eingabe in das Entwickler-Voice-Action-System **200** umgewandelt werden, oder die Eingaben können vom Entwickler-Voice-Action-System **200** empfangen werden und im Entwickler-Voice-Action-System **200** in eine Grammatik umgewandelt werden. Alternativ kann der Anwendungsentwickler **201** die neue Voice Action durch die Eingabe der nötigen Information unter Nutzung des Grammatikformats, das in das Voice-Action-Entwicklersystem **200** eingegeben wird, oder unter Nutzung eines anderen Formats, so wie eines Delimiter-Separated-Value(DSV)-Formats, definieren.

**[0045]** In einigen Implementierungen kann eine neue Voice Action dazu entwickelt werden, eine oder mehrere Operationen in einer Software-Anwendung, die auf einem Benutzergerät läuft, auszuführen, eine oder mehrere Operationen, die das Benutzergerät im Allgemeinen betreffen (d. h., die Hardware des Benutzergeräts zu steuern), eine Anwendung zu steuern, die extern zum Benutzergerät ist (d. h., eine Anwendung, die auf einem anderen Gerät läuft), oder um ein anderes Gerät zu steuern, das extern zum Benutzergerät ist. Der Anwendungsentwickler **201** kann beispielsweise eine Grammatik angeben, die eine neue Voice Action definiert, um eine spezifische An-

wendung auf einem Benutzergerät zu steuern (z. B. eine Anwendung eines sozialen Netzwerks oder eine Mediaplayer-Anwendung), um das Benutzergerät zu steuern (z. B. um die Ausrichtung eines Bildschirms des Geräts zu ändern oder um das Gerät auszuschalten), um eine Anwendung zu steuern, die auf einem anderen Gerät läuft (z. B. um eine Kabelfernseh-Software zu steuern, die auf einem Kabelfernsehreceiver läuft, der extern zum Benutzergerät ist), oder um ein anderes Gerät zu steuern, das vom Benutzergerät getrennt ist (z. B. um ein Garagentor durch eine Spracheingabe zu steuern, die dem Benutzergerät bereitgestellt wird). In einigen Implementierungen kann eine Voice Action so definiert werden, dass die Auslösung der Voice Action zur Ausführung von zwei oder mehr Aktionen oder Aktivitäten führt, die in/auf zwei oder mehr verschiedenen Anwendungen oder Geräten ausgeführt werden.

**[0046]** Ein von einem Anwendungsentwickler **201** spezifizierter Kontext identifiziert eine oder mehrere Bedingungen, die erfüllt werden müssen, damit eine Voice Action aktiviert wird, d. h., so dass die Voice Action nicht von einem mit der Voice Action verknüpften Auslösersatz ausgelöst wird, solange der Kontext nicht auch erfüllt ist. Anwendungsentwickler **201** kann einen Kontext dazu definieren, eine oder mehrere verschiedene Anforderungen oder Bedingungen zu besitzen.

**[0047]** In einigen Fällen kann ein von einem Anwendungsentwickler **201** eingegebener Kontext einen Status oder Attribute eines Benutzergeräts spezifizieren, so wie ob das Benutzergerät eingeschaltet oder gesperrt ist, ob das Benutzergerät eine Kamera, ein Gyroskop, ein Barometer oder eine andere Komponente oder Funktion aufweist. Somit kann beispielsweise eine Voice Action zur Aufnahme eines Fotos unter Nutzung einer auf dem Benutzergerät installierten Anwendung möglicherweise nicht aktiviert sein, wenn vom Benutzergerät empfangene Kontextinformation angibt, dass das Benutzergerät keine Kamera besitzt.

**[0048]** In anderen Beispielen kann ein Kontext erfordern, dass spezifische Anwendungen im Vordergrund oder Hintergrund eines Benutzergeräts laufen. Wie beispielsweise gezeigt in **Fig. 2**, kann die „play next“-Voice-Action nur dann aktiviert sein, wenn die Mediaplayer-Anwendung im Vordergrund von Benutzergerät **204** läuft, aber nicht, wenn die Mediaplayer-Anwendung geschlossen ist oder im Hintergrund von Benutzergerät **204** läuft. In anderen Beispielen kann eine Voice Action mit einem Kontext definiert sein, der die Voice Action aktiviert, wenn eine Anwendung im Hintergrund läuft. Eine Anwendung eines sozialen Netzwerks kann beispielsweise eine verknüpfte Voice Action besitzen, um eine Anfrage zur Verbindung mit einem anderen Benutzer zu akzeptieren, und ein Benutzer kann in der Lage sein, die Voice Ac-

tion zur Annahme der Anfrage, sich mit einem anderen Benutzer zu verbinden, auszulösen, auch wenn die Anwendung des sozialen Netzwerks nur im Hintergrund des Benutzergeräts läuft.

**[0049]** Ein Kontext kann zusätzlich oder alternativ erfordern, dass eine bestimmte Anwendung in einem bestimmten Modus läuft, damit eine Voice Action aktiviert wird. Ein Modus kann eine bestimmte Aktivität oder Aufgabe sein, die die Anwendung ausführt. Eine E-Mail-Anwendung kann beispielsweise darauf festgelegt sein, in einem Schreibmodus zu sein, eine Mediaplayer-Anwendung kann sich in einem Audioplayer-Modus befinden, oder eine Kameraanwendung kann sich in einem Kameramodus oder einem Bildbetrachtungsmodus für Fotoalben befinden. Das Benutzergerät kann in der Lage sein festzustellen, in welchem der Modi eine bestimmte Anwendung läuft und kann diese Information der Kontextinformation beifügen, die genutzt wird um zu bestimmen, ob eine bestimmte Voice Action aktiviert ist.

**[0050]** Zusätzlich kann ein Kontext in einigen Fällen erfordern, dass ein Anwendungsmodus einen gewissen Status besitzt. Ein Kontext kann beispielsweise angeben, dass eine Voice Action zum „Pausieren“ eines Films nur dann aktiviert ist, wenn eine Mediaplayer-Anwendung sich im Film-Abspielmodus befindet und wenn der Status der Anwendung im Film-Abspielmodus der ist, dass die Anwendung derzeit den Film abspielt. In einigen Fällen kann es sein, dass ein Benutzergerät nicht in der Lage ist, den Status einer Anwendung zu bestimmen, die in einem bestimmten Modus läuft. In diesen Fällen kann es notwendig sein, die Anwendung dazu zu konfigurieren, Information betreffend des Status der Anwendung in einem bestimmten Modus bereitzustellen. Somit kann die Spezifizierung einer solchen Bedingung in einem Kontext für eine bestimmte Anwendung Modifikationen an der Anwendung selbst erforderlich machen, damit die vom Benutzergerät bereitgestellte Kontextinformation die notwendige Statusinformation beinhaltet.

**[0051]** In einigen Beispielen kann ein für eine Voice Action spezifizierter Kontext erfordern, dass bestimmte Objekte oder Objekttypen auf einem Bildschirm eines Benutzergeräts dargestellt werden, damit die Voice Action aktiviert wird. Ein Kontext kann beispielsweise spezifizieren, dass eine „Bild speichern“-Voice-Action nur dann aktiviert ist, wenn ein Bild auf dem Display des Benutzergeräts ausgegeben wird, oder kann spezifizieren, dass eine „Markieren“-Voice-Action nur dann verfügbar ist, wenn auf dem Display Text angezeigt wird. In einem weiteren Beispiel kann eine Voice Action zur Auswahl eines bestimmten Listeneintrags, so wie die Voice Action „das erste auswählen“, nur dann verfügbar sein, wenn Kontextinformation angibt, dass eine Nummer „1“ oder ein Aufzählungszeichen auf dem Display des

Benutzergeräts angezeigt wird, oder wenn eine Liste von Items auf dem Display angezeigt wird. In einigen Fällen kann es sein, dass das Benutzergerät nicht dazu in der Lage ist festzustellen, welche Information zu einem gegebenen Zeitpunkt auf seinem Display dargestellt wird, z.B. wenn eine Anwendung das Display steuert. Somit kann es, wenn ein Kontext spezifiziert, dass gewisse Informationen oder Informationstypen auf dem Display ausgegeben werden müssen, damit eine Voice Action aktiviert wird, notwendig sein, die Anwendung zu modifizieren, um diese Information bereitzustellen. Das Benutzergerät kann dann die Information, die angibt, was auf dem Display des Benutzergeräts ausgegeben wird, der Kontextinformation beifügen, die genutzt wird um festzustellen, ob eine spezifische Voice Action aktiviert sein sollte.

**[0052]** In anderen Fällen kann ein Kontext erfordern, dass eine Benachrichtigung oder ein bestimmter Typ von Benachrichtigung auf einem Benutzergerät ausgegeben wurde. Zum Beispiel kann eine E-Mail-Anwendung, die Benutzern Pop-up-Benachrichtigungen bereitstellt, wenn eine neue E-Mail empfangen wird, eine verknüpfte Voice Action „E-Mail lesen“ besitzen, die aktiviert ist, wenn eine neue E-Mail-Benachrichtigung von der Anwendung auf dem Display des Benutzergeräts ausgegeben wird. Somit kann die Kontextinformation angeben, ob eine Benachrichtigung auf dem Display des Benutzergeräts ausgegeben wird, und diese Information kann genutzt werden um festzustellen, ob die Voice Action „E-Mail lesen“ aktiviert sein sollte.

**[0053]** In einigen Beispielen kann ein Kontext erfordern, dass ein Benutzergerät sich an einem bestimmten geografischen Standort befindet, damit eine Voice Action aktiviert wird. Ein Benutzergerät kann dazu in der Lage sein, seinen geografischen Standort unter Nutzung von Mobilfunkmasten-Triangulation, auf Grundlage des Zugriffs auf eine Internetverbindung, die mit einem bestimmten geografischen Standort verknüpft ist, mit Hilfe von Global-Positioning-System(GPS)-Technologie oder mit anderen Mitteln zu bestimmen. Das Benutzergerät kann Information, die seinen geografischen Standort angibt, der Kontextinformation beifügen, und eine Voice Action kann auf Grundlage des geografischen Standorts aktiviert werden, wenn dieser die Standortbedingungen eines Kontextes erfüllt. Beispielsweise kann ein Kontext, der mit einer Voice Action für eine Einzelhändleranwendung verknüpft ist, spezifizieren, dass eine gewisse Voice Action nur dann verarbeitet werden sollte, wenn ein Benutzergerät, auf dem die Einzelhändleranwendung installiert ist, innerhalb der Reichweite eines der Geschäfte des Einzelhändlers ist. Die Voice Action für die Einzelhändleranwendung kann auf Grundlage von Kontextinformation vom Benutzergerät aktiviert werden, die angibt, dass der geografische Standort des Benutzergeräts einem geografischen Standort entspricht, der sich innerhalb der

Reichweite eines der Geschäfte des Einzelhändlers befindet.

**[0054]** Ein von einem Anwendungsentwickler definierter Kontext kann auch angeben, dass eine Voice Action, die mit dem Kontext verknüpft ist, nur zu bestimmten Uhrzeiten, Daten, Monaten oder Jahreszeiten aktiviert ist, oder wenn andere zeitliche Bedingungen erfüllt sind. Ein Kontext kann beispielsweise spezifizieren, dass eine bestimmte Voice Action nur dann aktiviert ist, wenn die von einem Benutzergerät oder einer anderen Informationsquelle empfangene Kontextinformation eine zeitliche Bedingung erfüllt. Eine Voice Action für eine Banking-Anwendung kann beispielsweise mit einem Kontext verknüpft sein, der eine zeitliche Bedingung beinhaltet, dergestalt dass die Voice Action nur aktiviert wird, wenn Kontextinformation angibt, dass der Zeitpunkt der Eingabe der Voice Action zwischen 6:00 morgens und 10:00 abends liegt.

**[0055]** Ein Kontext kann andere Anforderungen für die Aktivierung einer Voice Action spezifizieren. Entwickler-Voice-Action-System **200** kann zum Beispiel Zugriff auf Information aus Quellen haben, die von einem Benutzergerät verschieden sind, oder kann in der Lage sein, Information von einem Benutzergerät zu empfangen, die aus einer oder mehreren Informationsquellen erhalten wurde. Die von den anderen Quellen empfangene Information kann erforderliche Kontextinformation für die Aktivierung einer Voice Action sein. Solche Information kann beispielsweise Wetterinformation, emotionale Information für einen Benutzer, Nachrichteninformation, Börseninformation oder andere Information beinhalten. Entwickler-Voice-Action-System **200** kann beispielsweise in der Lage sein, über eine oder mehrere kabelbasierte oder drahtlose Netzwerkverbindungen, z. B. eine Internetverbindung oder eine andere Netzwerkverbindung zu einem Serversystem, auf andere Informationsquellen zuzugreifen. Ein Kontext für eine Voice Action kann spezifizieren, dass die Voice Action nur dann aktiviert ist, wenn Wetterinformation für ein geografisches Gebiet, das dem Standort eines relevanten Benutzergeräts entspricht, angibt, dass es an diesem Standort regnet. Um zu überprüfen, ob die Kontextbedingung erfüllt ist, kann das Entwickler-Voice-Action-System **200** in der Lage sein, über die eine oder die mehreren Netzwerkverbindungen auf Wetterinformation für einen bekannten Standort eines Benutzergeräts zuzugreifen.

**[0056]** Die Grammatik oder andere Information, die die neue vom Anwendungsentwickler **201** eingegebene Voice Action definiert, kann über ein Netzwerk vom Terminal **202** aus an das Entwickler-Voice-Action-System **200** übermittelt werden. Die eingegebene Grammatik wird im Entwickler-Voice-Action-System **200** von der Validierungseingine **210** empfangen. Wenn der Anwendungsentwickler **201** Information in

das Entwickler-Voice-Action-System **200** eingibt, die die Voice Action definiert, aber nicht im Grammatikformat ist, kann das Entwickler-Voice-Action-System **200** die Information, die die neue Voice Action definiert, in Grammatik umwandeln. Validierungseingabe **210** analysiert die Grammatik um festzustellen, ob die eingegebene Grammatik in eine Absicht für eine neue Voice Action eingefügt werden kann. Die Validierungseingabe **210** kann beispielsweise in der Lage sein festzustellen, ob ein von der Grammatik spezifizierter Auslösersatz ein gültiger Auslösersatz ist, kann feststellen, ob die von der Grammatik spezifizierte Aktivität von der spezifizierten Anwendung ausgeführt werden kann, kann feststellen, ob das Format der eingegebenen Grammatik syntaktisch gültig ist, kann feststellen, ob der von der Grammatik spezifizierte Kontext gültig ist oder kann anderweitig feststellen, ob die eingegebene Grammatik eine gültige Grammatik ist, die eingefügt werden kann, um eine Absicht für die neue Voice Action zu erstellen. Um die Validierung auszuführen, kann die Validierungseingabe **210** auf eine Validierungskriterien-Datenbank **215** zugreifen, die Regeln beinhaltet, die benutzt werden, um die Grammatik zu validieren.

**[0057]** Die Validierungseingabe **210** empfängt beispielsweise die Grammatik, die die neue Voice Action für die Mediaplayer-Anwendung spezifiziert. Die Grammatik spezifiziert den Auslösersatz „play next“, dass der Auslösersatz „play next“ die Mediaplayer-Anwendung dazu bringen soll, zu einem nächsten Lied zu springen, und einen Kontext, der spezifiziert, dass die neue Voice Action aktiviert sein soll, wenn die Mediaplayer-Anwendung im Vordergrund läuft und sich im Audioplayer-Modus befindet. Validierungseingabe **210** kann in der Validierungskriterien-Datenbank **215** auf Validierungsregeln zugreifen und die empfangene Grammatik auf der Grundlage der Regeln validieren. Validierungseingabe **210** kann beispielsweise feststellen, ob die Mediaplayer-Anwendung eine bestehende Anwendung ist, dergestalt dass neue Voice Actions für die spezifizierte Anwendung erstellt werden können. Validierungseingabe **210** kann auf eine Regel zugreifen, die spezifiziert, dass ein Auslösersatz mehr als eine gesprochene Silbe lang sein muss, und kann den Auslösersatz „play next“ validieren, basierend darauf, dass der Auslösersatz „play next“ länger als eine gesprochene Silbe ist. Die Validierungsregeln können eine Liste von möglichen Aktionen spezifizieren, die die Mediaplayer-Anwendung ausführen kann, und kann die spezifizierte Aktion zum nächsten Lied zu springen validieren, basierend auf der Feststellung, dass das Springen zu einem nächsten Lied eine Operation ist, die die Validierungseingabe **210** ausführen kann. Validierungseingabe **210** kann den von der Grammatik spezifizierten Kontext validieren um sicherzustellen, dass der Kontext keine Widersprüche enthält. Die Validierungseingabe kann beispielsweise sicherstellen, dass der spezifizierte Kontext es nicht erfor-

dert, dass die Anwendung im Vordergrund läuft und auch im Hintergrund läuft, damit die neue Voice Action aktiviert wird, da nur eine dieser zwei Bedingungen erfüllt werden kann. Andere Validierungskriterien können auf eine empfangene Grammatik angewendet werden um festzustellen, ob die vom Anwendungsentwickler **201** eingefügte Grammatik gültig ist.

**[0058]** Andere Validierungsformen können von der Validierungseingabe **210** ausgeführt werden. Validierungseingabe **210** kann beispielsweise feststellen, ob der Anwendungsentwickler **201** autorisiert ist, neue Voice Actions für die spezifizierte Anwendung zu erstellen, kann feststellen, ob eine notwendige Gebühr zur Erstellung der neuen Voice Action bezahlt worden ist, oder kann anderweitig feststellen, ob eine von einem Anwendungsentwickler **201** eingefügte Grammatik in die Erstellung einer Absicht für die neue Voice Action münden kann.

**[0059]** Auf Grundlage des Ergebnisses der von der Validierungseingabe **210** durchgeführten Validierung kann dem Anwendungsentwickler **201** Feedback gegeben werden, das angibt, ob die Grammatik gültig ist. Wenn Validierungseingabe **210** feststellt, dass die Grammatik ungültig ist, kann Validierungseingabe **210** das Terminal **202** dazu veranlassen, Information auszugeben, die angibt, dass die Grammatik ungültig ist, und die anfordert, dass der Anwendungsentwickler **201** die Grammatik korrigiert, oder kann auch Arten vorschlagen, auf die der Anwendungsentwickler **201** die Grammatik korrigieren kann. Wenn beispielsweise festgestellt wird, dass eine Grammatik ungültig ist, weil sie eine Aktion spezifiziert, die von der spezifizierten Anwendung nicht ausgeführt werden kann (z. B. wenn die Grammatik spezifiziert hat, dass die neue Voice Action die Mediaplayer-Anwendung dazu bringen soll, eine neue E-Mail zu öffnen), kann Validierungseingabe **210** Information am Terminal **202** ausgeben lassen, die angibt, dass die Grammatik eine Aktion spezifiziert, die mit der spezifizierten Anwendung inkompatibel ist, und kann angeben, dass der Anwendungsentwickler **201** die Grammatik durch die Änderung der spezifizierten Aktion oder der spezifizierten Anwendung korrigieren kann. Wenn die eingefügte Grammatik gültig ist, kann die Validierungseingabe **210** einen Hinweis bereitstellen, dass die eingefügte Grammatik gültig ist.

**[0060]** Zusätzlich kann die Validierungseingabe **210** zusätzliche Eingaben vom Anwendungsentwickler **201** anfordern. Wenn beispielsweise festgestellt wird, dass die Grammatik gültig ist, kann die Validierungseingabe **210** dem Anwendungsentwickler **201** die Option bieten, mit der Erstellung des Sprachbefehls fortzufahren, dergestalt dass keine Absicht für den neuen Sprachbefehl aus der Grammatik erstellt wird, solange der Anwendungsentwickler **201** nicht bestätigt, dass er mit der Erstellung des neuen Sprachbefehls fortfahren will. Dies kann einem Anwendungs-

entwickler **201** die Möglichkeit geben festzustellen, ob er Grammatiken richtig in das Entwickler-Voice-Action-System **200** eingibt, ohne mit der Grammatikinduktion oder dem Ausrollen der neuen Voice Action fortfahren zu müssen.

**[0061]** Auf Grundlage der Validierung der vom Anwendungsentwickler **201** übermittelten Grammatik und, optional, der Bereitstellung einer Eingabe durch den Anwendungsentwickler **201**, die angibt, dass der Voice-Action-Erstellungsprozess fortgeführt werden soll, kann die Grammatik durch die Validierungseingabe **210** der Grammatikinduktions-Engine **220** bereitgestellt werden. Die Grammatikinduktions-Engine **220** kann die empfangene Grammatik einfügen, um eine Absicht für die neue Voice Action zu erstellen. Die erstellte Absicht kann Information spezifizieren, die in der Grammatik enthalten ist, die die neue Voice Action definiert. Die Grammatikinduktions-Engine **220** kann eine empfangene Grammatik einfügen, um eine Instanz einer Absichts-Datenstruktur zu erstellen, die eines oder mehrere aus Anwendung, Auslösersatz, auf die Erkennung des Auslösersatz hin auszuführende Aktion oder Aktivität und für die neue Voice Action spezifiziertem Kontext spezifiziert. Grammatikinduktions-Engine **220** kann die Absicht für die neue Voice Action in der Anwendungs-Voice-Action-Datenbank **230** speichern.

**[0062]** Zum Beispiel, wie in **Fig. 2**, kann Grammatikinduktions-Engine **220** die vom Anwendungsentwickler **201** eingefügte Grammatik empfangen, die von der Validierungseingabe **210** validiert worden ist. Grammatikinduktions-Engine **220** kann eine neue Absicht, „Absicht Y“ für die neue Voice Action auf Grundlage der empfangenen Grammatik erstellen. Um die Absicht zu erstellen, kann die Grammatikinduktions-Engine **220** eine neue Instanz einer Absichts-Datenstruktur erstellen, und die Absichts-Datenstruktur mit Information, die in der empfangenen Grammatik enthalten ist, bestücken, inklusive von Information, die die Mediaplayer-Anwendung spezifiziert, Information, die den Auslösersatz „play next“ spezifiziert, Information, die spezifiziert, dass die Voice Action „play next“ mit dem Springen zu einem nächsten Lied in der Mediaplayer-Anwendung verknüpft sein soll, und Information, die einen Kontext spezifiziert, der angibt, dass die Voice Action „play next“ aktiviert sein sollte, wenn die Mediaplayer-Anwendung im Vordergrund läuft und sich im Audioplayer-Modus befindet.

**[0063]** Die neue „Absicht Y“ kann von der Grammatikinduktions-Engine **220** in der Anwendungs-Voice-Action-Datenbank **230** gespeichert werden, die andere Absichten für andere Voice Actions beinhaltet, so wie „Absicht X“ und „Absicht Z“. Jede der Absichten „Absicht X“ und „Absicht Z“ kann mit anderen Voice Actions für andere Anwendungen verknüpft sein, oder mit anderen Voice Actions für die Me-

diaplayer-Anwendung. Sobald die neue „Absicht Y“ in der Anwendungs-Voice-Action-Datenbank **230** gespeichert wurde, kann die neue Voice Action so ausgerollt werden, dass die neue Voice Action von Benutzern, die die Mediaplayer-Anwendung auf ihren Geräten gespeichert haben, benutzt werden kann. Das bedeutet, dass Benutzer der Mediaplayer-Anwendung, sobald sie ausgerollt ist, die neue Voice Action „play next“ nutzen können, ohne dass sie ihre Anwendungen aktualisieren müssen.

**[0064]** In einigen Implementierungen kann die Einfügung der Grammatik zur Erstellung der Absicht auch die Erstellung eines oder mehrerer alternativer Auslöserbegriffe, die zumindest teilweise auf dem Auslöserbegriff basieren, der vom Anwendungsentwickler **201** eingefügt wurde, und die Einpflegung der alternativen Auslöserbegriffe in die für die neue Voice Action erstellte Absicht betreffen. Zum Beispiel können einen oder mehrere Erweiterungsregeln, Synonymregeln, Optionalisierungsregeln, Ersetzungsregeln, Rewriteregeln oder andere Regeln auf den vom Anwendungsentwickler **201** eingefügten Auslöserbegriff angewendet werden, um alternative Auslöserbegriffe zu bestimmen. Grammatikinduktions-Engine **220** kann die alternativen Auslöserbegriffe in die Absicht für die neue Voice Action einfügen, dergestalt dass die Erkennung des vom Anwendungsentwickler **201** eingefügten Auslöserbegriffs oder eines alternativen Auslöserbegriffs die Auslösung der Voice Action nach sich ziehen kann.

**[0065]** Zusätzlich kann in einigen Implementierungen die Grammatikinduktions-Engine **220** Informationen in Bezug auf eine neue Absicht getrennt von der Absichts-Datenstruktur speichern. Beispielsweise können Information, die einen Kontext spezifizieren, der erforderlich ist, damit die neue Voice Action aktiviert ist, oder ein Auslöserbegriff, der mit der neuen Voice Action verknüpft ist, getrennt von der Absicht für die neue Voice Action gespeichert werden. In solchen Fällen kann Kontextinformation, die von einem Benutzergerät empfangen wird, oder ein Auslöserbegriff, der vom Benutzergerät empfangen wird, mit dem Kontext und dem Auslöserbegriff verglichen werden, die mit der Absicht verknüpft sind, ohne dass auf die Daten in der Absicht zugegriffen werden muss. Die Absicht kann ausgelöst werden, basierend auf der Feststellung, dass die empfangene Kontextinformation und/oder der empfangene Auslöserbegriff mit dem entsprechenden Kontext und/oder Auslöserbegriff übereinstimmen, die mit der Absicht verknüpft sind. Auf diese Weise muss nur Information, die zur Auslösung der Aktivität oder Aktion, die mit der neuen Voice Action verknüpft ist, in die Absicht eingefügt werden, und andere Information, die genutzt wird, um die auszulösende Absicht auszuwählen, oder andere Information, kann extern von der Absicht gespeichert werden. Entwickler-Voice-Action-System **200** kann beispielsweise eine Zuord-

nungstabelle, eine verlinkte Liste oder andere Daten speichern, die einen Kontext oder Auslöserbegriff mit spezifischen Absichten verbindet, dergestalt dass die Identifikation einer bestimmten Paarung eines Kontexts und Auslöserbegriffs die mit dieser Paarung verknüpfte Absicht auslösen kann.

**[0066]** In einigen Implementierungen kann das Entwickler-Voice-Action-System **200** auch Möglichkeiten für den Anwendungsentwickler **201** bieten, die neue Voice Action zu testen, bevor sie für den öffentlichen Gebrauch ausgerollt wird. Entwickler-Voice-Action-System **200** kann beispielsweise Möglichkeiten für den Anwendungsentwickler **201** bieten, die neue Voice Action unter Nutzung des Terminals **202** zu testen, ohne dass die Voice Action für andere Benutzer verfügbar ist. Entwickler-Voice-Action-System **200** kann dem Anwendungsentwickler **201** auch Möglichkeiten bieten, eines oder mehrere Benutzergeräte zu spezifizieren, für die die neue Voice Action zum Test aktiviert sein soll. Anwendungsentwickler **201** kann beispielsweise spezifizieren, dass die neue Voice Action für eine Beta-Test-Gruppe von bekannten Benutzergeräten ausgerollt werden soll, um den ordnungsgemäßen Betrieb der neuen Voice Action vor dem Ausrollen der neuen Voice Action für den allgemeinen öffentlichen Gebrauch sicherzustellen.

**[0067]** In einigen Implementierungen kann das Entwickler-Voice-Action-System **200** die Erprobung der neuen Voice Action freigeben, indem es eine mit der Absicht verknüpfte Einstellung ändert, dergestalt dass die Absicht identifiziert werden kann und auf Grundlage des Empfangs eines Sprachbefehls von einem der zur Erprobung spezifizierten Benutzergeräte ausgelöst wird. Entwickler-Voice-Action-System **200** kann beispielsweise eine Einstellung für jede der in der Anwendungs-Voice-Action-Datenbank **230** gespeicherten Absichten speichern, wobei die Einstellung angibt, ob eine Absicht für eine bestimmte Voice Action allgemein für den öffentlichen Gebrauch ausgerollt wurde, oder es kann spezifische Benutzergeräte identifizieren, auf denen die Absicht zu Testzwecken ausgerollt wurde. Nach der Speicherung einer Absicht in der Anwendungs-Voice-Action-Datenbank **230** kann das Entwickler-Voice-Action-System **200** Information vom Terminal **202** empfangen, die angibt, dass die Absicht für einen spezifischen Satz von einem oder mehreren Benutzergeräten zu Testzwecken freigegeben werden soll. Daraufhin kann die Einstellung für die Absicht aktualisiert werden, um die Absicht auf dem spezifizierten Satz von Benutzergeräten so auszurollen, dass die mit der Absicht verknüpfte Voice Action ausgelöst werden kann. Nach der Testfreigabe kann eine Spracheingabe, die von einem Benutzergerät im Satz empfangen wird, die Absicht auslösen, wodurch der Test der Voice Action ermöglicht wird.

**[0068]** In anderen Ausführungsformen kann das Entwickler-Voice-Action-System **200** die Erprobung der neuen Voice Action freigeben, bevor die Absicht für die neue Voice Action in der Anwendungs-Voice-Action-Datenbank **230** gespeichert wird. Anwendungsentwickler **201** kann beispielsweise angeben, dass er die neue Voice Action testen möchte, bevor diese ausgerollt wird, und kann eine Anzahl von Benutzergeräten oder Benutzerkonten angeben, für die die neue Voice Action temporär zur Erprobung freigegeben werden soll. Nach der Erstellung der Absicht für die neue Voice Action kann das Entwickler-Voice-Action-System **200** die Voice Action für die spezifizierten Benutzergeräte oder Benutzerkonten freigeben, dergestalt dass ein spezifiziertes Benutzergerät oder ein mit einem spezifizierten Benutzerkonto verknüpftes Benutzergerät die neue Voice Action auslösen kann, wenn ein Benutzer den Auslösersatz, der mit der neuen Voice Action verknüpft ist, als Eingabe an einem der Benutzergeräte vornimmt.

**[0069]** In noch weiteren Implementierungen kann das Entwickler-Voice-Action-System **200** Testmöglichkeiten für neue Voice Actions bieten, bevor die Grammatik für die neue Voice Action von der Grammatikinduktions-Engine **220** eingefügt wird. Nach der Validierung der empfangenen Grammatik kann beispielsweise die Validierungsengine **210** oder eine andere Komponente des Entwickler-Voice-Action-Systems **200** in der Lage sein, den Betrieb der durch die Grammatik definierten neuen Voice Action zu simulieren, um dem Anwendungsentwickler **201** oder Benutzern anderer Geräte die Möglichkeit zu bieten, die neue Voice Action zu testen. In einigen Fällen kann das Simulieren der neuen Voice Action erfordern, dass das Terminal **202** oder andere für das Testen benutzte Benutzergeräte einen Simulator oder Testmodus benutzen, der vom Entwickler-Voice-Action-System **200** zur Simulation des Betriebs der neuen Voice Action bereitgestellt wird. Die Bereitstellung von Testmöglichkeiten vor der Einfügung der neuen Grammatik kann ein schnelleres iteratives Testen von neuen Voice Actions durch den Anwendungsentwickler **201** erlauben und dadurch den Voice-Action-Entwicklungsprozess insgesamt beschleunigen.

**[0070]** Die Möglichkeit Voice Action auszurollen und das Ausrollen von Voice Actions rückgängig zu machen kann auch durch das Entwickler-Voice-Action-System **200** zur Verfügung gestellt werden. Nachdem zum Beispiel eine Grammatik, die eine neue Voice Action definiert, durch den Anwendungsentwickler **201** ins Entwickler-Voice-Action-System **200** eingefügt wurde, validiert wurde, eingefügt wurde, um eine neue Absicht zu erstellen, und die neue Absicht in der Anwendungs-Voice-Action-Datenbank **230** gespeichert wurde, kann das Entwickler-Voice-Action-System **200** abfragen, ob der Anwendungsentwickler **201** die neue Voice Action ausrollen möchte. Anwendungsentwickler **201** kann entscheiden, die neue

Voice Action auszurollen, wobei das Ausrollen der neuen Voice Action die Auslösung der neuen Voice Action auf Benutzergeräten der breiten Öffentlichkeit ermöglicht, auf denen die Anwendung, die mit der neuen Voice Action verknüpft ist, installiert ist. Anwendungsentwickler **201** kann eine Eingabe am Terminal **202** vornehmen, um die neue Voice Action auszurollen. Die neue Voice Action kann dann dergestalt ausgerollt werden, dass die neue Voice Action in der Lage ist, auf anderen Benutzergeräten, auf denen die mit der neuen Voice Action verknüpfte Anwendung installiert ist, ausgelöst zu werden. Das heißt, nachdem die neue Voice Action ausgerollt wurde, führt die Eingabe des Auslösersatzes für die neue Voice Action auf einem Benutzergerät, auf dem die Anwendung installiert ist, dazu, dass die Absicht, die mit der neuen Voice Action verknüpft ist, ausgelöst wird. Das Auslösen der mit der neuen Voice Action verknüpften Absicht führt dazu, dass die Absicht oder andere aktionsauslösende Daten an das Benutzergerät übertragen werden, so dass Daten in der Absicht oder die aktionsauslösenden Daten dafür sorgen, dass die Aktionen oder Aktivitäten der Voice Action vom Benutzergerät ausgeführt werden. Wie besprochen kann das Ausrollen der Absicht lediglich erfordern, dass das Entwickler-Voice-Action-System **200** eine Einstellung aktualisiert, die mit der Absicht der neuen Voice Action verknüpft ist, um die Absicht zu einer möglichen Absicht zu machen, die auf eine Spracheingabe hin auszulösen ist.

**[0071]** Nachdem eine Voice Action ausgerollt worden ist, kann Anwendungsentwickler **201** das Ausrollen der neuen Voice Action rückgängig machen, so dass die neue Voice Action nicht mehr auf Benutzergeräten ausgelöst werden kann, auf denen die mit der neuen Voice Action verknüpfte Anwendung installiert ist. Anwendungsentwickler **201** kann beispielsweise eine Eingabe im Entwickler-Voice-Action-System **200** vornehmen, die die Voice Action identifiziert und seinen Wunsch angibt, das Ausrollen der Voice Action rückgängig zu machen. Daraufhin kann das Entwickler-Voice-Action-System **200** das Ausrollen der Voice Action rückgängig machen, indem es beispielsweise eine Einstellung aktualisiert, die mit der Absicht für die Voice Action verknüpft ist, so dass die Absicht für die Voice Action nicht ausgelöst werden kann. Da das Ausrollen oder die Rücknahme eine Voice Action lediglich erfordert, dass das Entwickler-Voice-Action-System **200** auf dem Entwickler-Voice-Action-System **200** gespeicherte Information aktualisiert, ohne die auf Benutzergeräten gespeicherte Anwendung zu modifizieren, kann der Anwendungsentwickler **201** leicht Voice Actions ausrollen oder das Ausrollen rückgängig machen.

**[0072]** In einigen Implementierungen kann das Entwickler-Voice-Action-System **200** auch Auffindbarkeitsbeispiele für eine neue Voice Action erstellen. Ein Auffindbarkeitsbeispiel kann Information enthal-

ten, die einem Benutzer eines Benutzergeräts angezeigt werden kann, die den Benutzer informiert, dass bestimmte Voice Actions verfügbar sind, wobei möglicherweise neue Voice Actions enthalten sind, die von einem Entwickler eingefügt wurden, nachdem die Anwendung installiert wurde. In einigen Fällen kann das Auffindbarkeitsbeispiel in Reaktion darauf ausgegeben werden, dass das Benutzergerät einen Status besitzt, der einem Kontext entspricht, der mit der Voice Action verknüpft ist, oder als Reaktion auf die Frage eines Benutzers, welche Voice Actions verfügbar sind.

**[0073]** Zum Beispiel kann Grammatikinduktions-Engine **220** oder eine andere Komponente des Entwickler-Voice-Action-Systems **200** auf Grundlage des Empfangs einer Grammatik, die eine neue Voice Action definiert, Auffindbarkeitsbeispiele für die von der Grammatik definierte Voice Action erstellen und die Auffindbarkeitsbeispiele im Anwendungs-Metadaten-Backend **225** speichern. Die Auffindbarkeitsbeispiele können Benutzergeräten bereitgestellt werden, auf denen die mit der neuen Voice Action verknüpfte Anwendung installiert ist. Die Metadaten der Anwendung können beispielsweise periodisch aktualisiert werden, um eine aktuelle Liste von Auffindbarkeitsbeispielen für die Anwendung zu beinhalten, die im Anwendungs-Metadaten-Backend **225** gespeichert sind. Alternativ kann auf Grundlage der Angabe von Kontextinformation durch ein Benutzergerät, die dessen aktuellen Status angibt, Information auf das Benutzergerät übertragen werden, die dazu führt, dass Auffindbarkeitsbeispiele für den Status des Benutzergeräts aktivierte Voice Actions auf dem Benutzergerät ausgegeben werden.

**[0074]** Beispielsweise kann das Entwickler-Voice-Action-System **200** zusätzlich zu der Erstellung der neuen „Absicht Y“ für die Voice Action „play next“ ein oder mehrere Auffindbarkeitsbeispiele für die Voice Action „play next“ erstellen. Auffindbarkeitsbeispiele können Information beinhalten, die auf Benutzergeräten angezeigt wird, auf denen die Mediaplayer-Anwendung installiert ist. Das Auffindbarkeitsbeispiel kann auf einem Benutzergerät angezeigt werden, basierend auf der Erfüllung des Kontextes, der mit der Voice Action „play next“ verknüpft ist, oder basierend auf anderen Bedingungen. Das Auffindbarkeitsbeispiel für die Voice Action „play next“ kann eine Nachricht „Versuchen Sie, 'play next' zu sagen“ sein, die angezeigt wird, wann immer die Mediaplayer-Anwendung im Vordergrund eines Benutzergeräts läuft und sich im Audioplayer-Modus befindet. In einigen Implementierungen kann ein Auffindbarkeitsbeispiel auf andere Ereignisse oder Bedingungen hin angezeigt werden. Zum Beispiel kann die Auffindbarkeitsbeispiel-Nachricht „Versuchen Sie, 'play next' zu sagen“ in Reaktion auf die Vornahme einer Eingabe durch einen Benutzer des Benutzergeräts angezeigt werden, indem dieser eine Taste oder ein Steuerele-

ment betätigt oder sagt „Welche Voice Actions kann ich benutzen?“

**[0075]** Im Allgemeinen können Auffindbarkeitsbeispiele für Anwendungs-Voice-Actions Benutzern zur Verfügung gestellt werden, ohne dass die Anwendungen selbst Handlungen ausführen müssen. Anwender-Voice-Action-System **200** kann beispielsweise Daten inklusive Auffindbarkeitsbeispiele einem Benutzergerät bereitstellen, und das Benutzergerät kann dafür sorgen, dass die Auffindbarkeitsbeispiele angezeigt werden, ohne dass Anwendungen, die in Bezug zu den Auffindbarkeitsbeispielen stehen, Operationen ausführen müssen. In einigen Implementierungen kann ein Betriebssystem, das auf einem Benutzergerät läuft, in der Lage sein, Auffindbarkeitsbeispiele zu empfangen oder auf Anwendungsmetadaten zuzugreifen, die Auffindbarkeitsbeispiele spezifizieren und die Auffindbarkeitsbeispiele auf dem Benutzergerät anzeigen zu lassen. Somit können Auffindbarkeitsbeispiele für Voice Actions, die im Zusammenhang mit einer bestimmten Anwendung stehen, ohne Modifikationen an der bestimmten Anwendung zur Anzeige bereitgestellt werden.

**[0076]** In einigen Implementierungen können Auffindbarkeitsbeispiele für Anwendungs-Voice-Actions dergestalt ausgegeben werden, dass ein Benutzer bemerken würde, dass die Auffindbarkeitsbeispiele nicht von der Anwendung selbst bereitgestellt werden. Zum Beispiel kann ein Benutzergerät Auffindbarkeitsbeispiele dergestalt anzeigen, dass es offensichtlich ist, dass die Auffindbarkeitsbeispiele vom Entwickler-Voice-Action-System **200** oder einem Betriebssystem bereitgestellt werden und nicht von den spezifischen Anwendungen, die im Zusammenhang mit den Voice Actions stehen. Ein Benutzergerät kann beispielsweise eine spezifische Computerstimme benutzen, die als mit dem Betriebssystem des Benutzergeräts verknüpft wiedererkennbar ist, wenn es Auffindbarkeitsbeispiele bereitstellt. Alternativ können Auffindbarkeitsbeispiele dergestalt ausgegeben werden, dass klar ist, dass die Anwendung die Auffindbarkeitsbeispiele nicht bereitstellt. Beispielsweise kann ein Auffindbarkeitsbeispiel ausgegeben werden, das lautet „Versuchen Sie, 'play next' oder 'Pause' zu sagen, und ich werde die Mediaplayer-Anwendung dazu bringen, ihre Anfrage zu bearbeiten“, um die Rolle des Betriebssystems von Aktionen zu unterscheiden, die von der Anwendung selbst ausgeführt werden. Andere Arten anzuzeigen, dass Auffindbarkeitsbeispiele von einem separaten System oder einer separaten Anwendung bereitgestellt werden, können auch implementiert werden.

**[0077]** Fig. 3 veranschaulicht ein beispielhaftes Voice-Action-System **300** zur Verarbeitung von Voice Actions, die von einem Benutzer eines Benutzergeräts empfangen werden. In einigen Implementierungen können die Voice Actions, die vom Voice-Action-

System **300** verarbeitet werden, Voice Actions sein, die ein Anwendungsentwickler für eine Anwendung erstellt hat, nachdem die Anwendung auf das Benutzergerät heruntergeladen wurde. Die vom Anwendungsentwickler erstellte Voice Action kann für die Anwendung erstellt worden sein unter Nutzung des Entwickler-Voice-Action-Systems **200** aus Fig. 2.

**[0078]** Kurz gefasst, befindet sich Voice-Action-System **300** über ein oder mehrere kabelbasierte oder drahtlose Netzwerke, wie eine oder mehrere kabelbasierte oder drahtlose Internet- oder Handydaten-Verbindungen, die das Benutzergerät **304** in die Lage versetzen, elektronische Kommunikation mit dem Voice-Action-System **300** auszutauschen, in Kommunikation mit einem Benutzergerät **304**, das einem Benutzer **305** gehört. Voice-Action-System **300** beinhaltet eine Spracherkennungs-Engine **330** und eine Kontextanalyse-Engine **340**, die Information verarbeiten, die vom Benutzergerät **304** über die eine oder die mehreren Netzwerkverbindungen empfangen wurde. Das Voice-Action-System beinhaltet einen Matcher **350**, der sich mit der Spracherkennungs-Engine **330** und der Kontextanalyse-Engine **340** in Kommunikation befindet und der auch Zugriff auf eine Anwendungs-Voice-Action-Datenbank **310** und eine Betriebssystem-Voice-Action-Datenbank **320** besitzt. Der Matcher **350** kommuniziert mit einem Aktionsauslöser-Generator **360**, der Zugriff auf eine Anwendungsdaten-Datenbank **370** hat. Eine Disambiguierungs-Engine **380**, die Zugriff auf einen Benutzeraktivitäts-Verlauf **390** hat, kann auch elektronische Kommunikationen mit dem Benutzergerät **304** über die eine oder die mehreren Netzwerkverbindungen senden oder empfangen.

**[0079]** Um eine vom Benutzer **305** eingegebene Spracheingabe zu verarbeiten, werden Audiodaten, die der Spracheingabe entsprechen und Kontextinformationen, die einen Status des Benutzergeräts **304** und/oder von Anwendungen, die auf dem Benutzergerät **304** laufen, angeben, dem Voice-Action-System **300** bereitgestellt. In einigen Fällen kann das Benutzergerät **304** die Spracheingabe erhalten, weil ein Benutzer **305** ein Steuerelement auswählt, das das Benutzergerät **304** dazu bringt, Sprechakte des Benutzers aufzunehmen, weil der Benutzer **305** einen Satz äußert, dessen Erkennung das Benutzergerät **304** dazu bringt, nachfolgende Spracheingaben vom Benutzer **305** zu erhalten, oder kann anderweitig bestimmen, die Spracheingabe des Benutzers **305** aufzunehmen. Eine Anfrage, spezifische Eingaben vorzunehmen, um die Aufnahme von Sprechakten des Benutzers auszulösen, kann verhindern, dass das Benutzergerät **304** unnötigerweise Sprechakte des Benutzers verarbeitet, d. h. durch die Reduktion der Menge an Audiodaten, die verarbeitet werden müssen, oder dass so viel von den Sprechakten des Benutzers aufgenommen wird, dass die Privatsphäre des Benutzers **305** verletzt wird, d. h. durch die exklu-



sive Aufnahme spezifischer Sprechhandlungen, die der Benutzer **305** an das Benutzergerät **304** richtet.

**[0080]** Im Beispiel, dargestellt in **Fig. 3**, kann der Benutzer **305** beispielsweise ein Steuerelement am Benutzergerät **304** auswählen, um das Benutzergerät **304** dazu zu bringen, Spracheingaben des Benutzers **305** aufzunehmen, und nachdem der Benutzer **305** das Steuerelement angewählt hat, kann der Benutzer **305** eine Spracheingabe vornehmen, indem er „play next“ sagt. Benutzergerät **304** kann die der Spracheingabe entsprechenden Audiodaten über eines oder mehrere Netzwerke an das Voice-Action-System **300** übertragen. Zusätzlich kann Benutzergerät **304** Kontextinformation für das Benutzergerät **304** oder Anwendungen, die auf dem Benutzergerät **304** laufen, bestimmen, und Kontextinformation für das Benutzergerät **304** kann über das eine oder die mehreren Netzwerke an das Voice-Action-System **300** übertragen werden.

**[0081]** In einigen Fällen beinhaltet die Kontextinformation, die vom Benutzergerät **304** an das Voice-Action-System **300** übertragen wird, spezifische Information, die durch Informationen, die auf dem Benutzergerät **304** gespeichert ist, identifiziert wird. Benutzergerät **304** kann beispielsweise Information speichern, die spezifische Statusinformation angibt, die als Kontextinformation an das Voice-Action-System **300** zu übertragen ist. Die Kontextinformation, die dem Voice-Action-System **300** durch das Benutzergerät **304** bereitgestellt wird, kann einen standardisierten Körper von Kontextinformation beinhalten, dergestalt dass alle Benutzergeräte die gleiche Kontextinformation an das Voice-Action-System **300** schicken, oder die Kontextinformation, die durch das Benutzergerät **304** bereitgestellt wird, kann ein spezifischer Satz von Kontextinformation sein, der für das Benutzergerät **304** bestimmt wird. In einigen Implementierungen kann das Voice-Action-System **300** gewisse Kontextinformationen vom Benutzergerät **304** anfordern, und die Kontextinformation, die vom Benutzergerät übertragen wird, kann Information sein, die auf die Anfrage hin dem Voice-Action-System **300** bereitgestellt wird. In noch weiteren Fällen kann die mit auf dem Benutzergerät **304** installierten Anwendungen verknüpfte Information, wie beispielsweise Metadaten, Kontextinformation spezifizieren, die dem Voice-Action-System **300** bereitzustellen sind, und das Benutzergerät **304** kann Kontextinformation an das Voice-Action-System **300** übertragen, die die durch die Anwendungsdaten spezifizierte Information beinhaltet.

**[0082]** Beispielsweise im Beispiel, gezeigt in **Fig. 3**, kann die vom Benutzergerät **304** an das Voice-Action-System **300** übertragene Kontextinformation Information beinhalten, die angibt, dass eine Mediaplayer-Anwendung derzeit im Vordergrund des Benutzergeräts **304** läuft, dass die Mediaplayer-Anwen-

dung in einem Audioplayer-Modus betrieben wird, und dass die derzeitige Batterieladung des Benutzergeräts **304** 50% beträgt.

**[0083]** Die vom Benutzergerät **304** übertragenen Audiodaten können von der Spracherkennungseingine **330** des Voice-Action-Systems **300** empfangen werden, und die Spracherkennungseingine **330** kann eine Transkription der Audiodaten erstellen. Zum Beispiel, wie besprochen im Hinblick auf die Spracherkennungseingine **130** aus **Fig. 1**, kann die Spracherkennungseingine **330** in der Lage sein, die Audiodaten zu empfangen und eine Transkription der Audiodaten zu erstellen, die der Spracheingabe des Benutzers **305** entspricht. Nachdem sie die Transkription der Audiodaten erstellt hat, kann die Spracherkennungseingine **330** die Transkription dem Matcher **350** bereitstellen.

**[0084]** In einigen Implementierungen kann die Spracherkennungseingine **330**, zusätzlich zum Empfang der Audiodaten, auch Kontextinformation empfangen, die dem Voice-Action-System **300** vom Benutzergerät **304** bereitgestellt wird. Spracherkennungssystem **330** kann die Spracherkennung auf Grundlage der empfangenen Kontextinformation beeinflussen. In einigen Implementierungen kann die Beeinflussung der Spracherkennung auf Grundlage von Kontextinformation beinhalten, die Wahrscheinlichkeit anzupassen, dass gewisse Wörter in Audiodaten erkannt werden. Spracherkennungssystem **330** kann beispielsweise die Wahrscheinlichkeit anpassen, dass gewisse Wörter erkannt werden, basierend auf ihrer Relevanz für einen von der Kontextinformation angegebenen Kontext.

**[0085]** Beispielsweise kann die Spracherkennungseingine **330** Audiodaten empfangen, die der Aussage „play next“ entsprechen, sowie Kontextinformation, die angibt, dass eine Mediaplayer-Anwendung derzeit im Vordergrund des Benutzergeräts **304** läuft. Spracherkennungseingine **330** kann die Spracherkennung dergestalt auf Grundlage der Kontextinformation beeinflussen, dass die Spracherkennungseingine **330** die Transkription „play next“ statt der Transkription „Planets“ erstellt, basierend darauf, dass die Spracherkennungseingine **330** eine Wahrscheinlichkeit erhöht, dass das Wort „play“ oder der Satz „play next“ in den Audiodaten erkannt wird oder die Wahrscheinlichkeit vermindert, dass das Wort „Planets“ erkannt wird, basierend auf der Feststellung, dass diese Wörter Relevanz oder keine Relevanz für die Mediaplayer-Anwendung besitzen, die im Vordergrund des Benutzergeräts **304** läuft.

**[0086]** Kontextanalyse-Engine **340** des Voice-Action-Systems empfängt die vom Benutzergerät **304** übertragene Kontextinformation und verarbeitet die Kontextinformation, um einen Kontext des Benutzergeräts **304** oder von Anwendungen, die auf dem

Benutzergerät **304** laufen, festzustellen. Ähnlich zur Kontextanalyse-Engine **140** aus **Fig. 1**, kann die Kontextanalyse-Engine **340** des Voice-Action-Systems **300** empfangene Kontextinformation verarbeiten, um Teile der empfangenen Kontextinformation zu bestimmen, die relevant für die Identifikation einer auszulösenden Voice Action sind, oder kann die empfangene Kontextinformation verarbeiten, um festzustellen, ob eine andere Kontextinformation existiert, die vom Benutzergerät **304** benötigt wird, um eine auszulösende Voice Action zu identifizieren. Wenn zusätzliche Kontextinformation benötigt wird, um eine auszuführende Voice Action zu bestimmen, kann die Kontextanalyse-Engine **340** die zusätzliche Kontextinformation vom Benutzergerät **304** anfordern, dergestalt dass das Benutzergerät **304** mit der zusätzlichen Kontextinformation auf die Anfrage reagieren kann. In einigen Implementierungen kann die Kontextanalyse-Engine **340** das Format der empfangenen Kontextinformation modifizieren, um einen Kontext zu erstellen, der vom Voice-Action-System **300** dazu genutzt werden kann, eine auszulösende Voice Action zu identifizieren. Kontextanalyse-Engine **340** stellt die verarbeitete Kontextinformation dem Matcher **350** bereit.

**[0087]** Im Beispiel, dargestellt in **Fig. 3**, empfängt die Kontextanalyse-Engine **340** beispielsweise die Kontextinformation vom Benutzergerät **304**. Die empfangene Kontextinformation kann angeben, dass eine Mediaplayer-Anwendung im Vordergrund des Benutzergeräts **304** läuft, dass die Mediaplayer-Anwendung im Audioplayer-Modus läuft und dass die Batterieladung des Benutzergeräts **304** 50% beträgt. Kontextanalyse-Engine **340** kann feststellen, dass die Information zum Batterieladezustand dafür irrelevant ist zu bestimmen, welche Voice Action auf die Spracheingabe des Benutzers **305** hin ausgelöst wird, und somit kann die Kontextanalyse-Engine **340** dem Matcher **350** einen Kontext bereitstellen, der lediglich angibt, dass die Mediaplayer-Anwendung im Vordergrund des Benutzergeräts **304** läuft und sich im Audioplayer-Modus befindet.

**[0088]** Matcher **350** empfängt den Kontext von der Kontextanalyse-Engine **340** und die Transkription von der Spracherkennungs-Engine **330** und nutzt den empfangenen Kontext und die empfangene Transkription, um eine oder mehrere mögliche auszulösende Voice Actions zu identifizieren. Matcher **350** kann beispielsweise den von der Kontextanalyse-Engine **340** empfangenen Kontext mit den Kontexten abgleichen, die mit einer oder mehreren Absichten verknüpft sind, die in der Anwendungs-Voice-Action-Datenbank **310** und/oder der OS-Voice-Action-Datenbank **320** enthalten sind. Auf Grundlage der Feststellung, dass der empfangene Kontext mit Kontexten übereinstimmt, die mit einem oder mehreren der Absichten verknüpft sind, kann der Matcher **350** diese Absichten als mögliche auszulösende Absichten

identifizieren. Nachdem auf Grundlage des Abgleichs des empfangenen Kontextes mit den Kontexten, die mit den möglichen Absichten verknüpft sind, mögliche Absichten identifiziert sind, kann der Matcher **350** den Satz von möglichen Absichten einengen, indem er die empfangene Transkription mit Auslösersätzen vergleicht, die mit den möglichen Absichten verknüpft sind. Matcher **350** kann Information, die den eingengten Satz von möglichen Absichten identifiziert, dem Aktionsauslöser-Generator **360** bereitstellen.

**[0089]** In einigen Fällen kann der Matcher **350** mehrere mögliche Transkriptionen einer Spracheingabe empfangen und kann auf Grundlage der Feststellung, dass jede der möglichen Absichten mit einem Kontext, der mit dem von der Kontextanalyse-Engine **340** empfangenen Kontext übereinstimmt, und einer oder mehreren der möglichen Transkriptionen der von der Spracherkennungs-Engine **330** empfangenen Spracheingabe verknüpft ist, einen eingengten Satz von möglichen Absichten identifizieren. Spracherkennungs-Engine **330** kann beispielsweise mehrere mögliche Transkriptionen für die Spracheingabe bestimmen und kann Information bereitstellen, die jede der möglichen Transkriptionen dem Matcher **350** angibt. Matcher **350** kann dann einen eingengten Satz von möglichen Absichten identifizieren, die jeweils einen Kontext spezifizieren, der mit dem von Matcher **350** empfangenen Kontext kompatibel ist, die auch einen Auslösersatz spezifizieren, der mit einer möglichen Transkription der Spracheingabe übereinstimmt.

**[0090]** Im Beispiel, dargestellt in **Fig. 3**, empfängt der Matcher **350** einen Kontext, der angibt, dass auf dem Benutzergerät **304** die Mediaplayer-Anwendung im Vordergrund und im Audioplayer-Modus läuft. Matcher **350** empfängt auch zwei Transkriptionen der Audiodaten, nämlich die Transkriptionen „play next“ und „Planets“. Matcher **350** identifiziert auf Grundlage dieser Information mögliche auszulösende Absichten, indem er den empfangenen Kontext und die Transkriptionen mit dem Kontext und den Auslösersätzen vergleicht, die mit Absichten verknüpft sind, die in der Anwendungs-Voice-Action-Datenbank **310** und der OS-Voice-Action-Datenbank **320** gespeichert sind. Auf Grundlage des Vergleichs identifiziert der Matcher **350** zwei mögliche Absichten, inklusive einer Absicht für eine Voice Action „play next“, die die Mediaplayer-Anwendung dazu bringt, zu einem nächsten Lied zu springen, und einer Absicht für eine „Planets“, die eine Anwendung namens „Planets“ öffnet. Matcher **350** kann dem Aktionsauslöser-Generator **360** Information bereitstellen, die die zwei möglichen Absichten identifiziert, oder der Matcher **350** kann dem Aktionsauslöser-Generator **360** eine Kopie der möglichen Absichten bereitstellen.

**[0091]** Aktionsauslöser-Generator **360** empfängt Information, die eine oder mehrere mögliche Absichten, die auf die Spracheingabe hin auszulösen sind, und vom Benutzergerät **304** empfangene Kontextinformation. Aktionsauslöser-Generator **360** erstellt Aktionsauslöser für eine oder mehrere bestimmte Absichten, wobei jeder Aktionsauslöser Daten umfasst, die in der Lage sind, das Benutzergerät **304** zu steuern, damit dieses die mit der entsprechenden Absicht verknüpfte Voice Action ausführt. In einigen Implementierungen erstellt der Aktionsauslöser-Generator **360** einen Aktionsauslöser für jede vom Matcher **350** identifizierte mögliche Absicht. In anderen Implementierungen kann der Aktionsauslöser-Generator **360** eine Angabe einer bestimmten auszulösenden Absicht aus den möglichen Absichten erhalten, und kann einen Aktionsauslöser nur für die bestimmte Absicht erstellen. In beiden Implementierungen kann der Aktionsauslöser-Generator **360** einen Aktionsauslöser erstellen, basierend auf der vom Matcher **350** empfangenen Information und der Information auf die in der Anwendungsdaten-Datenbank **370** zugegriffen wird.

**[0092]** Beispielsweise kann der Aktionsauslöser-Generator **360** eine Absicht für eine mögliche Voice Action empfangen, und der Aktionsauslöser-Generator **360** kann auf Information in der Anwendungsdaten-Datenbank **370** zugreifen, die dazu führen kann, dass die Voice Action auf dem Benutzergerät **304** ausgeführt wird. Zum Beispiel kann die Information, auf die in der Anwendungsdaten-Datenbank **370** zugegriffen wird, Information zur Ausführung oder Auslösung der Aktivität oder Aktion beinhalten, die mit einer Voice Action verknüpft ist, so wie Information über das Benutzergerät **304**, Information über die mit der Voice Action verknüpfte Anwendung oder Information über ein auf dem Benutzergerät **304** laufendes Betriebssystem. Aktionsauslöser-Generator **360** kann die abgerufene Information als Aktionsauslöser für die mögliche Voice Action nutzen oder kann die abgerufene Information nutzen, um Aktionsauslöserdaten zu erstellen, die in der Lage sind, die Aktionen oder Aktivitäten auszulösen, die mit der möglichen Voice Action auf dem Benutzergerät **304** verknüpft sind. In einigen Implementierungen beinhaltet das Voice-Action-System **300** möglicherweise weder Aktionsauslöser-Generator **360** noch Anwendungsdaten-Datenbank **370**, dergestalt dass eine Absicht als solche an das Benutzergerät **304** übertragen werden kann, und Information in der Absicht das Benutzergerät **304** dazu bringen kann, die mit der Absicht verknüpfte Aktion oder Aktivität auszuführen.

**[0093]** Im Beispiel, dargestellt in **Fig. 3**, kann der Aktionsauslöser-Generator **360** eine Absicht für die Voice Action „play next“ für die Mediaplayer-Anwendung und eine Absicht für die Voice Action „Planets“ zum Öffnen der „Planets“-Anwendung erhalten. Aktionsauslöser-Generator **360** kann Aktionsauslöser für

jede der „play next“- und „Planets“-Voice-Actions erstellen. Um einen Aktionsauslöser für die Voice Action „play next“ zu erstellen, greift der Aktionsauslöser-Generator **360** auf Information in der Anwendungsdaten-Datenbank **370** zu, die die auf dem Benutzergerät **304** installierte Mediaplayer-Anwendung betrifft, kann auf Information zugreifen, die das Benutzergerät **304** oder ein spezifisches Modell des Benutzergeräts **304** betrifft, oder kann auf Information zugreifen, die ein auf dem Benutzergerät **304** laufendes Betriebssystem betrifft. Aktionsauslöser-Generator **360** nutzt die empfangene Absicht und die Information, auf die in der Anwendungsdaten-Datenbank **370** zugegriffen wurde, um einen Aktionsauslöser für die Voice Action „play next“ zu erstellen. Auf ähnliche Weise kann der Aktionsauslöser-Generator **360** einen Aktionsauslöser für die Voice Action „Planets“ erstellen, indem er auf Information im Zusammenhang mit der „Planets“-Anwendung, dem Benutzergerät **304** oder einem Betriebssystem, das auf dem Benutzergerät **304** läuft, zugreift, und kann einen Aktionsauslöser für die Voice Action „Planets“ erstellen, basierend auf der empfangenen und abgerufenen Information.

**[0094]** Wie oben besprochen kann der Matcher **350** in einigen Beispielen mehr als eine mögliche auszulösende Absicht identifizieren, und in diesen Fällen kann es nötig sein, eine bestimmte auszulösende Absicht aus dem Satz von möglichen Absichten auszuwählen. Die Disambiguierungs-Engine **380** kann Operationen ausführen, um aus den möglichen Absichten auszuwählen oder eine Benutzereingabe zu empfangen, die dies tut. Beispielsweise kann die Disambiguierungs-Engine eine Anfrage an das Benutzergerät **304** richten, die den Benutzer **304** auffordert, eine bestimmte Voice Action aus den möglichen Voice Actions, die mit den möglichen Absichten verknüpft sind, auszuwählen. Auf die Anfrage hin kann der Benutzer **305** eine Eingabe am Benutzergerät **304** vornehmen, um eine bestimmte mögliche Voice Action zur Ausführung auszuwählen. Information, die die Benutzerauswahl angibt, kann von der Disambiguierungs-Engine **380** empfangen werden.

**[0095]** Alternativ kann die Disambiguierungs-Engine **380** Zugriff auf einen Benutzeraktivitäts-Verlauf **390** haben, der Information über vorangegangene am Benutzergerät **304** eingegebene Voice Actions oder andere am Benutzergerät **304** vorgenommene Handlungen beinhaltet. Benutzeraktivitäts-Verlauf **390** kann beispielsweise eine Anzahl von Fällen angeben, in denen der Benutzer **305** eine bestimmte mögliche Voice Action ausgelöst hat, die Anzahl von Fällen, in denen der Benutzer **305** spezifische Aktionen auf dem Benutzergerät **304** mit oder ohne Nutzung einer Voice Action ausgeführt hat, oder kann andere Information über die Nutzung des Benutzergeräts **305** durch den Benutzer **304** beinhalten. In einigen Fällen kann der Benutzeraktivitäts-Verlauf **390**

Information beinhalten, die den Nutzungsverlauf von anderen Benutzern als dem Benutzer **305** angibt. Zum Beispiel kann der Benutzeraktivitäts-Verlauf **390** Information beinhalten, die Nutzungsmuster von allen Benutzern, die die Mediaplayer-Anwendung und/oder die „Planets“-Anwendung auf einem Benutzergerät installiert haben.

**[0096]** In einigen Implementierungen kann die Information im Benutzeraktivitäts-Verlauf **390** vom Voice-Action-System **300** genutzt werden, um aus möglichen Absichten auszuwählen, ohne eine Eingabe des Benutzers **305** anzufordern. Zum Beispiel kann die Disambiguierungs-Engine **380** auf Information im Benutzeraktivitäts-Verlauf **390** zugreifen, die angibt, dass der Benutzer **305** zuvor eine bestimmte mögliche Voice Action ausgewählt hat, als er eine ähnliche Spracheingabe vorgenommen hat, und auf Grundlage dieser Information kann die Disambiguierungs-Engine **380** eine bestimmte mögliche auszulösende Absicht auswählen, ohne eine Eingabeaufforderung an den Benutzer **305** zu richten. Auf ähnliche Weise kann die Disambiguierungs-Engine **380** feststellen, dass ein gewisser Anteil von Benutzern, die eine der vom Benutzer **305** empfangenen Spracheingabe ähnliche Spracheingabe vorgenommen haben, gewählt haben, eine bestimmte Voice Action auszulösen, und die Disambiguierungs-Engine **380** kann eine bestimmte mögliche auszulösende Sprachabsicht auswählen, basierend auf dieser Information und ohne eine Eingabeaufforderung an den Benutzer **305** zu richten. In einigen Implementierungen kann jede der möglichen Absichten einen verknüpften Wert besitzen, der die Stärke einer Beziehung zwischen einem bestimmten Auslösersatz und der bestimmten möglichen Absicht angibt. Eine bestimmte mögliche Absicht kann auf Grundlage der den möglichen Absichten zugewiesenen Werte ausgewählt werden, die die Stärke der Beziehung zwischen der möglichen Absicht und einem empfangenen Auslösersatz angeben.

**[0097]** Andere Faktoren können im Benutzeraktivitäten-Verlauf **390** berücksichtigt oder einbezogen werden, der bei der Auswahl einer möglichen Absicht genutzt wird. Voice-Action-System **300** kann Handlungen auf dem Benutzergerät **304** verfolgen, nachdem das Voice-Action-System eine bestimmte mögliche Absicht auslöst um festzustellen, ob die ausgelöste Absicht der Erwartung des Benutzers **305** entspricht. Wenn Benutzer **305** beispielsweise Operationen unternimmt, um eine andere Aktion auf dem Benutzergerät **304** durchzuführen, nachdem das Voice-Action-System **300** eine bestimmte Voice Action auslöst, kann das darauf hindeuten, dass das Voice-Action-System **300** die falsche mögliche Absicht ausgelöst hat, und somit kann das Voice-Action-System **300** Information zum Benutzeraktivitäts-Verlauf **390** hinzufügen, die angibt, dass die empfangene Spracheingabe die ausgelöste Absicht nicht oder sel-

tener auslösen sollte. In anderen Beispielen kann der Benutzeraktivitäts-Verlauf **390** Information enthalten, die angibt, wie häufig der Benutzer **305** oder andere Benutzer eine bestimmte mögliche Absicht bei einem bestimmten Status ihres Benutzergeräts auswählen, z. B. bei einem bestimmten Status, der in der dem Voice-Action-System **300** bereitgestellten Kontextinformation angegeben ist. Auf andere im Benutzeraktivitäts-Verlauf **390** enthaltene Information kann von der Disambiguierungs-Engine **380** zugegriffen werden, und sie kann genutzt werden, um eine bestimmte mögliche Absicht aus einem Satz von möglichen Absichten auszuwählen.

**[0098]** Für das Beispiel, dargestellt in **Fig. 3**, kann die Disambiguierungs-Engine **380** auf Grundlage der Feststellung, dass die möglichen Absichten für den vom Benutzer **305** vorgenommene Spracheingabe die Absicht für die Voice Action „play next“ für die Mediaplayer-Anwendung oder die Absicht für die Voice Action „Planets“ zum Öffnen der „Planets“-Anwendung beinhalten, dem Benutzergerät **305** Information bereitstellen, die eine Benutzerauswahl einer bestimmten möglichen Voice Action anfordert. Wie gezeigt kann eine Anfrage auf dem Display des Benutzergeräts **304** angezeigt werden, die den Benutzer **305** auffordert, entweder eine Voice Action „play next“ oder eine Voice Action zum „Planets öffnen“ auszuwählen. Benutzer **305** nimmt eine Eingabe auf dem Benutzergerät **304** vor, um die Option „nächstes Lied“ auszuwählen. Daten, die die Benutzerauswahl angeben, werden an die Disambiguierungs-Engine **380** übertragen, und auf den Empfang von Daten in, die die Benutzerauswahl angeben, bestimmt die Disambiguierungs-Engine **380**, die Absicht für die Voice Action „play next“ auszulösen. In einigen Beispielen kann die Disambiguierungs-Engine **380** auch Information im Benutzeraktivitäts-Verlauf **390** speichern, die angibt, dass der Benutzer **305** sich dafür entschieden hat, die Absicht für die Voice Action „play next“ und nicht die Absicht für die Voice Action „Planets“ auszulösen. Disambiguierungs-Engine **380** kann sich auf die Information stützen, die im Benutzeraktivitäts-Verlauf **390** gespeichert ist, um automatisch eine mögliche auszulösende Voice Action zu bestimmen, wenn eine ähnliche Spracheingabe später vom Voice-Action-System **300** vom Benutzergerät **304** empfangen wird.

**[0099]** In einer alternativen Implementierung des Beispiels, dargestellt in **Fig. 3**, kann die Disambiguierungs-Engine **380** eine bestimmte auszulösende mögliche Absicht auf Grundlage des Zugriffs auf den Benutzeraktivitäts-Verlauf **390** auswählen. Benutzeraktivitäts-Verlauf **390** kann beispielsweise angeben, dass der Benutzer **305** des Benutzergeräts **304** zuvor eine ähnliche Spracheingabe vorgenommen hat und sich dazu entschieden hat, die Absicht für die Voice Action „nächstes Lied abspielen“ auszulösen. Basierend darauf, dass der Benutzeraktivitäts-Verlauf **390**

angibt, dass die Absicht für die Voice Action „play next“ zuvor ausgelöst wurde, nachdem die ähnliche Spracheingabe empfangen wurde, kann die Disambiguierungs-Engine **380** bestimmen, die Absicht für die Voice Action „play next“ auszulösen, ohne zusätzliche Eingaben vom Benutzer **305** anzufordern.

**[0100]** Nachdem eine bestimmte mögliche Absicht ausgewählt wurde, und, wenn nötig, ein Aktionsauslöser für die ausgewählte mögliche Voice Action erstellt wurde, überträgt das Voice-Action-System **300** Information an das Benutzergerät **304**, die dazu führt, dass die ausgewählte Voice Action ausgeführt wird. Zum Beispiel kann die Disambiguierungs-Engine **380** oder eine andere Komponente des Voice-Action-Systems **300** die Absicht oder die aktionsauslösenden Daten für die gewählte Voice Action an das Benutzergerät **304** übertragen, um das Benutzergerät **304** dazu zu bringen, die ausgewählte Voice Action auszuführen.

**[0101]** Im Beispiel, dargestellt in **Fig. 3**, kann die Disambiguierungs-Engine **380** beispielsweise den Aktionsauslöser an das Benutzergerät **304** übertragen, nachdem der Benutzer **305** die „play next“-Option ausgewählt hat, um die Voice Action „play next“ auszuwählen, die die Mediaplayer-Anwendung dazu bringt, zu einem nächsten Lied zu springen, und nachdem ein Aktionsauslöser erstellt wurde, der Daten zur Steuerung der Mediaplayer-Anwendung zum Springen zu einem nächsten Lied enthält. Benutzergerät **304** kann den Aktionsauslöser empfangen, und der Aktionsauslöser kann das Benutzergerät **304** dazu bringen, die Voice Action „play next“ auszuführen, indem er die Mediaplayer-Anwendung dazu bringt, zu einem nächsten Lied zu springen.

**[0102]** In einigen Implementierungen kann eine Benachrichtigung, die eine ausgelöste Voice Action angibt, zur Anzeige auf dem Benutzergerät **304** bereitgestellt werden. Voice-Action-System **300** oder ein Betriebssystem der Benutzergeräts **304** können beispielsweise veranlassen, dass eine Benachrichtigung auf dem Benutzergerät **304** angezeigt wird, die die ausgelöste Voice Action identifiziert. In einigen Fällen kann die Benachrichtigung, die die ausgelöste Voice Action identifiziert, dergestalt bereitgestellt werden, dass es für den Benutzer **305** ersichtlich ist, dass die mit der ausgelösten Voice Action verknüpfte Anwendung nicht für die Erkennung der Voice Action oder die Auslösung der Voice Action verantwortlich ist. Eine auf dem Benutzergerät **304** bereitgestellte Benachrichtigung kann beispielsweise eine Stimme nutzen, die als mit dem Voice-Action-System **300**, dem Benutzergerät **304** oder einem Betriebssystem des Benutzergeräts **304** verknüpft erkennbar ist, und nicht mit der Mediaplayer-Anwendung, die auf dem Benutzergerät **304** läuft. In einem anderen Beispiel kann eine Benachrichtigung auf eine Art angezeigt werden, dass das System oder die Anwendung, die

die Spracheingabe des Benutzers **305** verarbeitet von der Anwendung unterscheidet, die die ausgelöste Voice Action ausführt. Eine Benachrichtigung kann beispielsweise angeben: „Sie haben die Voice Action ‘play next’ ausgelöst. Ich werde jetzt die Mediaplayer-Anwendung veranlassen, zu einem nächsten Lied zu springen“, wodurch dem Benutzer **305** angezeigt wird, dass die Mediaplayer-Anwendung nicht für die Verarbeitung der Spracheingabe verantwortlich war.

**[0103]** **Fig. 4** zeigt ein Flussdiagramm für einen beispielhaften Prozess **400**, der von einem Voice-Action-System ausgeführt wird. In einigen Beispielen kann der Prozess **400** aus **Fig. 4** ausgeführt werden vom Voice-Action-System **100** aus **Fig. 1**, oder von Komponenten des Voice-Action-Systems **100** aus **Fig. 1**. Prozess **400** beschreibt einen Prozess durch den ein Anwendungsentwickler eine neue Voice Action für eine Anwendung anlegen kann, die auf einem oder mehreren Benutzergeräten installiert wurde, und durch das ein Benutzer eines Benutzergeräts, auf dem die Anwendung installiert ist, die neue Voice Action auslösen kann, um die Ausführung von Operationen zu veranlassen, die mit der von der Anwendung auf dem Benutzergerät auszuführenden neuen Voice Action verknüpft sind.

**[0104]** Zunächst werden Daten von einem Voice-Action-System empfangen, die eine neue Voice Action für eine Software-Anwendung spezifizieren, die vom Voice-Action-System verschieden ist, wobei die Daten einen oder mehrere Operationen zur Ausführung der neuen Voice Action und einen oder mehrere Auslöserbegriffe zur Auslösung der neuen Voice Action umfassen (**402**). Zum Beispiel kann ein mit der Software-Anwendung verknüpfter Anwendungsentwickler **101a–101n** unter Nutzung eines Terminals **102a–102n** Daten in das Voice-Action-System **100** eingeben, die eine neue Voice Action definieren.

**[0105]** Die Daten, die die neue Voice Action definieren, können Daten beinhalten, die eine oder mehrere Operationen zur Ausführung der neuen Voice Action und einen oder mehrere Auslöserbegriffe zur Auslösung der neuen Voice Action spezifizieren. Zum Beispiel können Daten, die eine neue Voice Action definieren den Auslöserbegriff „play next“ spezifizieren, und können spezifizieren, dass die neue Voice Action Operationen ausführen soll, die eine Mediaplayer-Anwendung dazu bringen, zu einem nächsten Lied zu springen. Die Daten können auch zusätzliche Information im Zusammenhang mit der neuen Voice Action spezifizieren, so wie die Anwendung, die die neue Voice Action betrifft, einen Kontext, in dem die neue Voice Action zur Nutzung durch einen Benutzer aktiviert sein soll, oder andere Information. Zum Beispiel können die Daten, die die Voice Action „play next“ definieren, auch spezifizieren, dass die Voice Action die Mediaplayer-Anwendung betrifft, und kann spezifizieren, dass die Voice Action aktiviert sein soll,

wenn die Mediaplayer-Anwendung im Vordergrund läuft und sich im Audioplayer-Modus befindet.

**[0106]** Eine kontextuelle Voice-Action-Absicht für die Software-Anwendung wird vom Voice-Action-System auf Grundlage von mindestens den empfangenen Daten erstellt, wobei die kontextuelle Voice-Action-Absicht Daten umfasst, die, wenn sie von der Software-Anwendung empfangen werden, anfordern, dass die Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt (**404**). Zum Beispiel das Voice-Action-System **100** aus **Fig. 1** kann die Daten empfangen, die die neue Voice Action spezifiziert, und kann auf Grundlage der empfangenen Daten eine kontextuelle Voice-Action-Absicht für die neue Voice Action erstellen. Voice-Action-System **100** kann die kontextuelle Voice-Action-Absicht in der Anwendungs-Voice-Action-Datenbank **110** speichern. In einigen Implementierungen kann das Voice-Action-System **100** die Absicht unter Nutzung eines Prozesses erstellen, der ähnlich zu dem beschriebenen ist in **Fig. 2**, durch die Validierung der Daten und der Einfügung der empfangenen Grammatik, um die kontextuelle Voice-Action-Absicht zu erstellen. In einigen Implementierungen kann die erstellte kontextuelle Voice-Action-Absicht Information speichern, die die Software-Anwendung steuern kann, um die eine oder die mehreren Operationen zur Ausführung der neuen Voice Action ausführen kann.

**[0107]** Die kontextuelle Voice-Action-Absicht ist mit dem einen oder den mehreren Auslöserbegriffen für die neue Voice Action durch das Voice-Action-System verknüpft (**406**). Nachdem beispielsweise die kontextuelle Voice-Action-Absicht für die neue Voice Action vom Voice-Action-System **100** erstellt wurde, kann das Voice-Action-System **100** die kontextuelle Voice-Action-Absicht für die neue Voice Action mit dem einen oder den mehreren Auslöserbegriffen verknüpfen, die in den empfangenen Daten spezifiziert werden. Die kontextuelle Voice-Action-Absicht mit dem/den verknüpften Auslöserbegriff(en) kann in der Anwendungs-Voice-Action-Datenbank **110** gespeichert werden. In einigen Implementierungen kann die Verknüpfung des einen oder der mehreren Auslöserbegriffe mit der kontextuellen Voice-Action-Absicht beinhalten, den einen oder die mehreren Auslöserbegriffe in der kontextuellen Voice-Action-Absicht zu spezifizieren. In anderen Implementierungen kann die Verknüpfung dadurch hergestellt werden, dass anderweitig der eine oder die mehreren Auslöserbegriffe mit der kontextuellen Voice-Action-Absicht verbunden werden, dergestalt dass der Empfang eines oder mehrerer Auslöserbegriffe zur Identifizierung der kontextuellen Voice-Action-Absicht führt.

**[0108]** Das Voice-Action-System ist dazu konfiguriert, eine Angabe einer Benutzeräußerung, die von

einem Gerät empfangen wurde, auf dem die Software-Anwendung installiert ist, zu empfangen (**408**). Voice-Action-System **100** kann beispielsweise dazu konfiguriert sein, eine Angabe zu erhalten, dass eine Äußerung vom Benutzergerät **104** empfangen wurde, wobei die Äußerung vom Benutzergerät **104** als Spracheingabe des Benutzers **105** erhalten wird. Auf Benutzergerät **104** kann die mit der neuen Voice Action verknüpfte Software-Anwendung bereits installiert sein, beispielsweise die Mediaplayer-Anwendung, besprochen in Zusammenhang mit **Fig. 1**.

**[0109]** Das Voice-Action-System ist dazu konfiguriert festzustellen, dass eine Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit der kontextuellen Voice-Action-Absicht verknüpft sind (**410**). Voice-Action-System **100** kann beispielsweise dazu konfiguriert sein, Audiodaten zu empfangen, die der Benutzeräußerung entsprechen, und kann unter Nutzung der Spracherkennungs-Engine **130** des Voice-Action-Systems **100** eine Transkription der Benutzeräußerung erstellen. Die von der Spracherkennungs-Engine **130** erstellte Transkription der Benutzeräußerung kann mit mehreren kontextuellen Voice-Action-Absichten verknüpften Auslöserbegriffen verglichen werden, inklusive der kontextuellen Voice-Action-Absicht für die neue Voice Action und kontextueller Voice-Action-Absichten für andere Voice Actions.

**[0110]** In einigen Fällen kann der Vergleich durchgeführt werden vom Voice-Action-Selektor **150** aus **Fig. 1**, wobei der Voice-Action-Selektor **150** die Transkription der Benutzeräußerung mit Auslöserbegriffen für kontextuelle Voice-Action-Absichten vergleichen kann, die in der Anwendungs-Voice-Action-Datenbank **110** oder der OS-Voice-Action-Datenbank **130** gespeichert sind. Auf Grundlage des Vergleichs kann der Voice-Action-Selektor **150** feststellen, dass die Transkription der Benutzeräußerung mit einem oder mehreren von dem einen oder den mehreren Auslöserbegriffen übereinstimmt, die mit der kontextuellen Voice-Action-Absicht der neuen Voice Action verknüpft sind. Beispielsweise kann Spracherkennungs-Engine **130** eine Transkription für die vom Benutzergerät **104** empfangene Benutzeräußerung erstellen, die als „play next“ angezeigt wird. Voice-Action-Selektor **150** kann die Voice Action „play next“ mit Auslöserbegriffen vergleichen, die mit möglichen kontextuellen Voice-Action-Absichten verknüpft sind, um festzustellen, dass die Transkription mit dem Auslösersatz „play next“ übereinstimmt, der mit der kontextuellen Voice-Action-Absicht für die neue Voice Action „play next“ verknüpft ist.

**[0111]** Das Voice-Action-System ist dazu konfiguriert, die kontextuelle Voice-Action-Absicht dem bestimmten Gerät bereitzustellen und dadurch anzufordern, dass die auf dem bestimmten Gerät installierte Software-Anwendung die eine oder die meh-

reren Operationen der neuen Voice Action ausführt (412). Beispielsweise kann auf die Feststellung hin, dass die Transkription der Benutzeräußerung, die von der Spracherkennungs-Engine 130 des Voice-Action-Systems 100 erstellt wurde, mit dem einen oder den mehreren Auslöserbegriffen übereinstimmt, die mit der kontextuellen Voice-Action-Absicht für die neue Voice Action verknüpft sind, der Voice-Action-Selektor 150 dem Benutzergerät 104 die kontextuelle Voice-Action-Absicht für die neue Voice Action bereitstellen. Die kontextuelle Voice-Action-Absicht für die neue Voice Action kann effektiv anfordern, dass die Software-Anwendung, die auf dem Benutzergerät 104 installiert ist, die eine oder die mehreren Operationen der neuen Voice Action ausführt. Beispielsweise kann die kontextuelle Voice-Action-Absicht Daten beinhalten, die die eine oder die mehreren Operationen auslösen, die mit der neuen Voice Action verknüpft sind, sobald das Benutzergerät 104 die kontextuelle Voice-Action-Absicht empfängt.

**[0112]** Es wurde eine Reihe von Implementierungen beschrieben. Trotzdem versteht es sich, dass verschiedene Modifikationen durchgeführt werden können, ohne vom Geist und Umfang der Offenbarung abzuweichen. Zum Beispiel können verschiedene Formen der vorstehend dargestellten Abläufe verwendet werden, wobei Schritte neu geordnet, hinzugefügt oder entfernt werden. Dementsprechend liegen andere Implementierungen im Geltungsbereich der folgenden Ansprüche.

**[0113]** In Fällen, in denen die Systeme und/oder Verfahren, die hier besprochen werden, persönliche Informationen über Benutzer sammeln oder persönliche Informationen nutzen können, kann für die Benutzer eine Möglichkeit der Kontrolle bereitgestellt werden, ob Programme oder Funktionen persönliche Informationen sammeln, z. B. Informationen über das soziale Netzwerk eines Benutzers, soziale Handlungen oder Aktivitäten, Beruf, Präferenzen oder den derzeitigen Standort, oder um zu kontrollieren, ob und/oder wie das System und/oder die Verfahren für den Benutzer relevantere Operationen ausführen können. Zusätzlich können gewisse Daten auf einem oder mehreren Wegen anonymisiert werden, bevor sie gespeichert oder verwendet werden, so dass personenbezogene Informationen entfernt werden. Eine Benutzeridentität kann beispielsweise so anonymisiert werden, dass keine persönlichen identifizierbaren Informationen für den Benutzer bestimmt werden können, oder ein Standort des Benutzers kann verallgemeinert werden, wobei Standortinformationen entnommen werden, wie beispielsweise eine Stadt, Postleitzahl oder ein Bundesland, sodass ein bestimmter Standort eines Benutzers nicht festgestellt werden kann. So hat der Benutzer Kontrolle darüber, wie Informationen über ihn oder eingeholt und verwendet werden.

**[0114]** Während die vorgenannten Ausführungsformen hauptsächlich im Hinblick auf die Entwicklung von Voice Actions zur Nutzung mit auf Benutzergeräten installierten Anwendungen beschrieben wurden, können die beschriebenen Funktionen auch im Zusammenhang mit Maschinen, anderen Geräten, Robotern oder anderen Systemen genutzt werden. Zum Beispiel das Voice-Action-System 100 aus Fig. 1 kann genutzt werden, um Voice Actions zur Interaktion mit Maschinen zu entwickeln und zu implementieren, wobei die Maschinen ein verknüpftes Computersystem besitzen; es kann außerdem genutzt werden, um Voice Actions zur Interaktion mit Robotern oder Systemen mit Roboter-Komponenten zu entwickeln und zu implementieren; es kann genutzt werden, um Voice Actions zur Interaktion mit Haushaltsgeräten, Entertainment-Systemen oder anderen Geräten zu entwickeln und zu implementieren; oder es kann genutzt werden, um Voice Actions zur Interaktion mit einem Fahrzeug oder einem anderen Transportsystem zu entwickeln und zu implementieren.

**[0115]** Ausführungsformen und alle funktionsfähigen in dieser Beschreibung beschriebenen Operationen können in einer digitalen elektronischen Schaltung, in physisch greifbarer Computer-Software oder Firmware, in Computer-Hardware, darunter auch in aus dieser Beschreibung hervorgehenden Strukturen und deren strukturellen Äquivalenten oder in Kombinationen einer oder mehrerer derselben implementiert werden. Ausführungsformen können als eines oder mehrere Computerprogrammprodukte, d. h. als eines oder mehrere Module von Computerprogrammanweisungen, die auf einem computerlesbaren Medium für die Ausführung durch oder für die Kontrolle der Operation der datenverarbeitenden Einrichtung verschlüsselt sind, implementiert werden. Das maschinenlesbare Speichermedium kann eine maschinenlesbare Speichervorrichtung, ein maschinenlesbares Speichersubstrat, eine Speichervorrichtung, eine Stoffzusammensetzung, die ein maschinenlesbares verbreitetes Signal bewirkt, oder eine Kombination aus einem oder mehreren davon sein. Der Begriff „datenverarbeitende Einrichtung“ umfasst jedwede Einrichtung, Vorrichtung oder Maschine zur Verarbeitung von Daten, einschließlich beispielsweise eines programmierbaren Prozessors, eines Computers bzw. mehrerer Prozessoren oder Computer. Der Apparat kann neben der Hardware auch einen Code beinhalten, der eine Ausführungsumgebung für das betreffende Computerprogramm erstellt, z. B. einen Code, der Prozessor-Firmware, einen Protokollstapel, ein Datenbank-Managementsystem, ein Betriebssystem oder eine Kombination einer oder mehrerer der genannten darstellt. Ein propagiertes Signal ist ein künstlich erzeugtes Signal, z. B. ein maschinell erzeugtes elektrisches, optisches oder elektromagnetisches Signal, das erzeugt wird, um Informationen für die Übertragung an eine geeignete Empfangereinrichtung zu kodieren.

**[0116]** Ein Computerprogramm (auch als Programm, Software, Software-Anwendung, Skript oder Code bezeichnet) kann in einer beliebigen Form von Programmiersprache geschrieben sein, einschließlich kompilierter oder interpretierter Sprachen, und das Computerprogramm kann in jeder beliebigen Form eingesetzt sein, darunter als unabhängiges Programm oder als ein Modul, Komponente, Subroutine, oder andere Einheit, die zur Benutzung in einer Rechenumgebung geeignet ist. Ein Computerprogramm entspricht nicht unbedingt einer Datei in einem Dateisystem. Ein Programm kann in einem Teil einer Datei gespeichert sein, die andere Programme oder Daten enthält (z. B. eines oder mehrere Scripts, die in einem Dokument in Markup-Sprache gespeichert sind), in einer einzelnen Datei speziell für das betreffende Programm oder in mehreren koordinierten Dateien (z. B. Dateien, die ein oder mehrere Module, Unterprogramme oder Teile von Code speichern). Ein Computerprogramm kann auf einem Computer oder auf mehreren Computern bereitgestellt und ausgeführt werden, die sich an einem Standort befinden oder über mehrere Standorte verteilt und durch ein Kommunikationsnetzwerk miteinander verbunden sind.

**[0117]** Die in dieser Beschreibung beschriebenen Prozesse und Logikabläufe können von einem oder mehreren programmierbaren Computern ausgeführt werden, die ein oder mehrere Computerprogramme ausführen, die durch den auf Eingabedaten basierenden Betrieb und das Erzeugen von Ausgabedaten bestimmte Funktionen aktivieren. Die Prozesse und die logischen Abläufe können auch durch logische Sonderzweckschaltungen durchgeführt werden, und der Apparat kann als Sonderzweckschaltungen implementiert werden, z. B. ein FPGA (Field Programmable Gate Array) oder eine ASIC (anwendungsspezifische integrierte Schaltung).

**[0118]** Prozessoren, die für die Ausführung eines Computerprogramms geeignet sind, umfassen beispielsweise sowohl allgemeine als auch Spezialmikroprozessoren sowie alle Arten von einem oder mehreren Prozessoren von jeglicher Art von digitalen Computern. Ein Prozessor nimmt im Allgemeinen Anweisungen und Daten von einem Festspeicher oder einem Arbeitsspeicher oder von beiden entgegen.

**[0119]** Die wesentlichen Elemente eines Computers sind ein Prozessor für das Ausführen von Befehlen und ein oder mehrere Speichergeräte für das Speichern von Befehlen und Daten. Ganz allgemein gehören zu einem Computer auch ein oder mehr Massenspeichergeräte für das Speichern von Daten, z. B. Magnet-, magnetooptische oder optische Disketten, um Daten entgegenzunehmen und/oder zu übertragen. Ein Computer muss jedoch nicht über solche Geräte verfügen. Außerdem kann ein Computer in ein anderes Gerät eingebettet sein, z. B. in ein Tablet, in

ein Mobiltelefon, einen Organizer (PDA), einen mobilen Audioplayer, einen GPS-Empfänger (Global Positioning System), um nur einige zu nennen. Computerlesbare Medien, die für das Speichern von Computerprogrammbefehlen und -daten geeignet sind, schließen alle Formen von nichtflüchtigen Speichern, Medien und Speichergeräten ein, einschließlich beispielsweise Halbleiter-Speichergeräte, z. B. EPROM, EEPROM und Flash-Speicher; Magnetplatten, z. B. interne Festplatten oder Wechselplatten; magneto-optische Platten; und CD-ROM- und DVD-ROM-Platten. Der Prozessor und der Speicher können mit Speziallogikschaltungen ergänzt werden oder darin integriert sein.

**[0120]** Um eine Interaktion mit einem Benutzer bereitzustellen, können Ausführungsformen auf einem Computer mit einer Displayvorrichtung, wie z. B. ein CTR-(Kathodenstrahlröhren-) oder LCD-(Flüssigkristalldisplay-)Monitor zum Anzeigen von Informationen für den Benutzer und einer Tastatur und einem Zeigegerät, z. B. einer Maus oder einem Trackball, über die der Nutzer Eingaben in den Computer bereitstellen kann, erfolgen. Andere Arten von Geräten können auch verwendet werden, um eine Interaktion mit einem Benutzer bereitzustellen; zum Beispiel kann eine dem Benutzer bereitgestellte Rückmeldung irgendeine Form von Sinnesrückmeldung sein, z. B. visuelle Rückmeldung, auditive Rückmeldung oder Tastrückmeldung; und eine Eingabe vom Benutzer kann in einer beliebigen Form empfangen werden, einschließlich akustischer, Sprach- oder Tasteingaben.

**[0121]** Ausführungsformen können in einem Rechensystem implementiert werden, das Folgendes umfasst: eine Backendkomponente, z. B. als Datenserver; eine Middlewarekomponente, z. B. einen Applikationsserver; eine Frontendkomponente, z. B. einen Clientcomputer mit grafischer Benutzeroberfläche oder einen Webbrowser, mit dem ein Benutzer mit einer Implementierung oder einer beliebigen Kombination einer oder mehrerer dieser Backend-, Middleware- oder Frontendkomponenten interagieren kann. Die Komponenten des Systems können durch eine beliebige Form oder ein beliebiges Medium von digitaler Datenkommunikation, z. B. ein Kommunikationsnetzwerk, miteinander verbunden sein. Beispiele für Kommunikationsnetzwerke beinhalten ein lokales Netzwerk („LAN“) und ein Fernnetz („WAN“), z. B. das Internet.

**[0122]** Das Computersystem kann Clients und Server umfassen. Client und Server sind generell voneinander entfernt und interagieren in der Regel über ein Kommunikationsnetzwerk. Die Beziehung von Client und Server ergibt sich durch Computerprogramme, die auf den jeweiligen Computern ausgeführt werden und eine Client-Server-Beziehung zueinander haben.



**[0123]** Während diese Beschreibung viele Einzelheiten enthält, sollten diese nicht als Beschränkungen des Umfangs der Offenbarung oder von dem, was beansprucht werden kann aufgefasst werden, sondern vielmehr als Beschreibungen von bestimmten Merkmalen bestimmter Ausführungsformen. Bestimmte Merkmale, die innerhalb dieser Beschreibung im Zusammenhang mit separaten Ausführungsformen beschrieben werden, können auch in Kombination in einer einzelnen Ausführungsform implementiert werden. Umgekehrt können verschiedene Merkmale, die im Zusammenhang mit einer einzelnen Ausführungsform beschrieben werden, auch in mehreren Ausführungsformen separat oder in einer geeigneten Teilkombination implementiert werden. Außerdem können, auch wenn die Merkmale weiter oben als in bestimmten Kombinationen wirkend beschrieben und sogar zunächst als solche beansprucht werden, in einigen Fällen ein oder mehrere Merkmale einer beanspruchten Kombination aus der Kombination herausgeschnitten und die beanspruchte Kombination auf eine Teilkombination oder eine Variante einer Teilkombination gerichtet werden.

**[0124]** Gleichermaßen sollte, obwohl die Vorgänge in den Zeichnungen in einer bestimmten Reihenfolge dargestellt sind, dies nicht so verstanden werden, dass die besagten Vorgänge in der dargestellten Reihenfolge oder in fortlaufender Reihenfolge durchgeführt werden müssen, oder dass alle veranschaulichten Vorgänge durchgeführt werden, um die erwünschten Ergebnisse zu erzielen. Unter bestimmten Umständen können Multitasking und Parallelverarbeitung von Vorteil sein. Darüber hinaus sollte die Trennung verschiedener Systemkomponenten in den oben beschriebenen Ausführungsformen nicht in allen Ausführungsformen als erforderlich aufgefasst werden, und es versteht sich, dass die beschriebenen Programmkomponenten und Systeme im Allgemeinen zusammen in ein einziges Softwareprodukt integriert oder zu mehreren Softwareprodukten verkapselt werden können.

**[0125]** In allen Fällen, in denen eine HTML-Datei erwähnt wird, kann diese durch andere Dateitypen oder -formate ersetzt werden. Eine HTML-Datei kann beispielsweise durch eine XML- oder JSON-Datei, eine Klartextdatei oder andere Dateitypen ersetzt werden. Darüber hinaus können wo eine Tabelle oder Hashtabelle erwähnt wird andere Dateistrukturen (wie Spreadsheets, relationale Datenbanken oder strukturierte Dateien) genutzt werden.

**[0126]** Weitere Implementierungen sind in den folgenden Beispielen zusammengefasst:

Beispiel 1: Computerimplementiertes System, umfassend:

das Empfangen von Daten durch ein Voice-Action-System, die eine neue Voice Action für eine Software-Anwendung spezifizieren, die vom besagten Voice-

Action-System verschieden ist, wobei die Daten eine oder mehrere Operationen zur Ausführung der neuen Voice Action und einen oder mehrere Auslöserbegriffe zur Auslösung der neuen Voice Action umfassen. das Erstellen einer kontextuellen Voice-Action-Absicht für die Software-Anwendung durch das Voice-Action-System auf Grundlage von mindestens den empfangenen Daten, wobei die kontextuelle Voice-Action-Absicht Daten umfasst, die, wenn sie von der Software-Anwendung empfangen werden, anfordern, dass die Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt

das Verknüpfen, durch das Voice-Action-System, der kontextuellen Voice-Action-Absicht mit dem einen oder den mehreren Auslöserbegriffen für die neue Voice Action;

wobei das Voice-Action-System dazu konfiguriert ist: eine Angabe einer Benutzeräußerung, die von einem Gerät erhalten wurde, auf dem die Software-Anwendung installiert ist, zu empfangen, festzustellen, dass eine Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit der kontextuellen Voice-Action-Absicht verknüpft sind; und auf die Feststellung hin dem Gerät die kontextuelle Voice-Action-Absicht bereitzustellen und dadurch anzufordern, dass die auf dem Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt.

**[0127]** Beispiel 2: Computerimplementiertes System nach Beispiel 1, wobei die empfangenen Daten einen Kontext spezifizieren, wobei der Kontext einen Status eines Geräts oder der Software-Anwendung spezifiziert, wenn die neue Voice Action aktiviert ist.

**[0128]** Beispiel 3: Computerimplementiertes System nach Beispiel 2, wobei der Kontext spezifiziert, dass die Software-Anwendung im Vordergrund eines Geräts läuft, auf dem die Software-Anwendung installiert ist.

**[0129]** Beispiel 4: Computerimplementiertes System aus Beispiel 2, wobei der Kontext spezifiziert, dass die Software-Anwendung im Hintergrund eines Geräts läuft, auf dem die Software-Anwendung installiert ist.

**[0130]** Beispiel 5: Computerimplementiertes System nach Beispiel 2, wobei der Kontext spezifiziert, dass die Software-Anwendung eine spezifische Aktivität ausführt.

**[0131]** Beispiel 6: Computerimplementiertes System nach Beispiel 2, wobei der Kontext spezifiziert, dass eine von der Software-Anwendung ausgeführte Aktivität spezifische Aktivität sich in einem bestimmten Aktivitätszustand befindet.

**[0132]** Beispiel Nr. 7: Computerimplementiertes System nach einem der Beispiele 2 bis 6, umfassend: das Verknüpfen der kontextuellen Voice-Action-Absicht mit dem Kontext für die neue Voice-Action; das Empfangen von Kontextinformation durch das Voice-Action-System, die einen Status eines bestimmten Geräts, auf dem die Software-Anwendung installiert ist, oder der auf dem bestimmten Gerät installierten Software-Anwendung angibt; das Feststellen, dass die Kontextinformation den Kontext für die neue Voice Action erfüllt; und das Bereitstellen, in Reaktion auf das Feststellen, dass eine Transkription einer Benutzeräußerung, die vom bestimmten Gerät erhalten wird, mit dem einen oder den mehreren Auslöserbegriffen übereinstimmt, die mit der kontextuellen Voice-Action-Absicht verknüpft sind, und dass die Kontextinformation den mit der kontextuellen Voice-Action-Absicht verknüpften Kontext erfüllt, durch das Voice-Action-System für das bestimmte Gerät, der kontextuellen Voice-Action-Absicht, und dadurch die Anfrage, dass die auf dem bestimmten Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt.

**[0133]** Beispiel Nr. 8: Computerimplementiertes System nach Beispiel 7, wobei das Empfangen der Kontextinformation, die den Status des bestimmten Geräts oder der auf dem bestimmten Gerät installierten Software-Anwendung angibt, Folgendes umfasst: das Bereitstellen, durch das Voice-Action-System für das bestimmte Gerät, einer Anfrage für bestimmte Kontextinformation; und das Empfangen der bestimmten Kontextinformation auf die Anfrage hin.

**[0134]** Beispiel Nr. 9: Computerimplementiertes System nach einem der Beispiele 7 oder 8, umfassend: das Feststellen, dass die Kontextinformation einen Kontext für eine zweite Voice Action erfüllt, und dass die Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit einer Absicht für die zweite Voice Action verknüpft sind, wobei die Absicht für die zweite Voice Action eine oder mehrere Operationen zur Ausführung der zweiten Voice Action spezifiziert. das Auswählen, in Reaktion auf das Feststellen, einer Voice Action aus der neuen Voice Action und der zweiten Voice Action; und das Bereitstellen, durch das Voice-Action-System für das bestimmte Gerät, der Absicht, die mit der ausgewählten Voice Action verknüpft ist, wodurch die Anfrage ergeht, dass die auf dem bestimmten Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der ausgewählten Voice Action ausführt.

**[0135]** Beispiel Nr. 10: Computerimplementiertes System nach Beispiel 9, wobei das Auswählen der

ausgewählten Voice Action aus der neuen Voice Action und der zweiten Voice Action die Auswahl der ausgewählten Voice Action auf den Empfang von Daten hin umfasst, die eine Benutzerauswahl einer aus der neuen Voice Action oder der zweiten Voice Action angeben.

**[0136]** Beispiel 11: Computerimplementiertes System nach Beispiel 9 oder 10, wobei die Auswahl der ausgewählten Voice Action aus der neuen Voice Action und der zweiten Voice Action Folgendes umfasst: das Zuweisen eines jeweiligen Werts für die neue Voice Action und die zweite Voice Action; und das Auswählen der ausgewählten Voice Action basierend auf mindestens dem jeweils zugewiesenen Wert für die neue Voice Action und die zweite Voice Action.

**[0137]** Beispiel Nr. 12: Computerimplementiertes System nach einem der Beispiele 9 bis 11, wobei das Auswählen der ausgewählten Voice Action aus der neuen Voice Action und der zweiten Voice Action das Auswählen der ausgewählten Voice Action auf die Feststellung hin umfasst, dass die mit der ausgewählten Voice Action verknüpfte Software-Anwendung im Vordergrund läuft.

**[0138]** Beispiel Nr. 13: Computerimplementiertes System nach einem der Beispiele 1 bis 12, wobei die Erstellung der kontextuellen Voice-Action-Absicht für die Software-Anwendung umfasst festzustellen, dass die eine oder die mehreren Operationen der neuen Voice Action von der Software-Anwendung ausführbar sind.

**[0139]** Beispiel Nr. 14: Computerimplementiertes System nach einem der Beispiele 1 bis 13, umfassend: das Feststellen, dass eine Transkription einer Benutzeräußerung, die von einem bestimmten Gerät erhalten wird, auf dem die Software-Anwendung installiert ist, dem einen oder den mehreren Auslöserbegriffen ähnelt, die mit der kontextuellen Voice-Action-Absicht verknüpft sind; das Bereitstellen, in Reaktion auf das Feststellen, durch das Voice-Action-System für das bestimmte Gerät, von Daten, die eine Anfrage einer Benutzer-eingabe angeben, die bestätigt, ob die Benutzeräußerung den einen oder die mehreren Auslöserbegriffe angegeben hat oder mit der Absicht erfolgt ist, die Software-Anwendung zu veranlassen, die neue Voice Action auszuführen; das Empfangen, in Reaktion auf die Anfrage, durch das Voice-Action System und vom bestimmten Gerät, von Daten, die eine Bestätigung angeben; und das Bereitstellen, in Reaktion auf das Empfangen der Daten, die die Bestätigung angeben, der kontextuellen Voice-Action-Absicht an das bestimmte Gerät durch das Voice-Action-System, wodurch die Anfrage ergeht, dass die auf dem bestimmten Gerät instal-

lierte Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt.

**[0140]** Beispiel Nr. 15: Computerimplementiertes System nach einem der Beispiele 1 bis 14, umfassend:

das Empfangen, durch das Voice-Action-System von einem mit der Software-Anwendung verknüpften Entwickler, einer Anfrage, die neue Voice Action für die Software-Anwendung auszurollen; und  
das Bereitstellen, in Reaktion auf die Anfrage, der neuen Voice Action für die Software-Anwendung in Reaktion auf die Anfrage, wobei das Ausrollen der neuen Voice Action für die Software-Anwendung die neue Voice Action für die Software-Anwendung freischaltet.

**[0141]** Beispiel Nr. 16: Computerimplementiertes System nach einem der Beispiele 1 bis 15, umfassend:

das Empfang, durch das Voice-Action-System von einem mit der Software-Anwendung verknüpften Entwickler, einer Anfrage, das Ausrollen der neuen Voice Action für die Software-Anwendung rückgängig zu machen; und  
das Rückgängigmachen, in Reaktion auf die Anfrage, des Ausrollens der neuen Voice Action für die Software-Anwendung auf die Anfrage hin, wobei das Rückgängigmachen des Ausrollens der neuen Voice Action für die Software-Anwendung die neue Voice Action für die Software-Anwendung deaktiviert.

**[0142]** Beispiel Nr. 17: Computerimplementiertes System nach einem der Beispiele 1 bis 16, umfassend:

das Empfang, durch das Voice-Action-System von einem mit der Software-Anwendung verknüpften Entwickler, einer Anfrage, die Erprobung der neuen Voice Action freizugeben, wobei die Anfrage ein oder mehrere Geräte spezifiziert, für die die neue Voice Action freigegeben werden soll; und  
das Freigeben, in Reaktion auf die Anfrage, der neuen Voice Action für das eine oder die mehreren spezifizierten Geräte, wobei die neue Voice Action für Geräte, die nicht in den spezifizierten Geräten enthalten sind, deaktiviert ist.

**[0143]** Beispiel Nr. 18: Computerimplementiertes System nach einem der Beispiele 1 bis 17, umfassend:

das Empfang, durch das Voice-Action-System, einer Angabe einer Benutzeräußerung, die von einem bestimmten Gerät erhalten wird, auf dem die Software-Anwendung installiert ist.  
das Feststellen, durch das Voice-Action-System, dass eine Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit der kontextuellen Voice-Action-Absicht verknüpft sind; und

das Bereitstellen, in Reaktion auf das Feststellen, der kontextuellen Voice-Action-Absicht an das bestimmte Gerät durch das Voice-Action-System, wodurch die Anfrage ergeht, dass die auf dem bestimmten Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt.

**[0144]** Beispiel Nr. 19: Nichtflüchtiges computerlesbares Speichergerät, das Software speichert, die durch einen oder mehrere Computer ausführbare Anweisungen umfasst, die bei einer solchen Ausführung den einen oder die mehreren Computer zum Ausführen von Tätigkeiten veranlassen, umfassend die Schritte der Beispiele 1 bis 18.

**[0145]** Verfahren, Systeme und Apparate zum Empfangen, durch ein Voice-Action-System, von Daten, die eine neue Voice Action für eine Anwendung spezifizieren, die vom Voice-Action-System verschieden ist, wurden beschrieben. Eine Voice-Action-Absicht für die Anwendung wird auf der Grundlage von zumindest den Daten erstellt, wobei die Voice-Action-Absicht Daten umfasst, die, wenn sie von der Anwendung empfangen werden, anfordern, dass die Anwendung eine oder mehrere Operationen ausführt, die für die neue Voice Action spezifiziert sind. Die Voice-Action-Absicht ist mit für die neue Voice Action spezifizierten Auslöserbegriffen verknüpft. Das Voice-Action-System ist dazu konfiguriert, eine Angabe einer Benutzeräußerung zu empfangen, die von einem Gerät erhalten wird, auf dem die Anwendung installiert ist, und stellt fest, dass eine Transkription der Benutzeräußerung den Auslöserbegriffen entspricht, die mit der Voice-Action-Absicht verknüpft sind. Auf die Feststellung hin stellt das Voice-Action-System dem Gerät die Voice-Action-Absicht bereit.

**[0146]** Somit wurden bestimmte Ausführungsformen beschrieben. Weitere Ausführungsformen liegen innerhalb des Schutzzumfangs der folgenden Patentansprüche. Die in den Ansprüchen ausgeführten Vorgänge können beispielsweise in einer anderen Reihenfolge ausgeführt werden und erzielen immer noch gewünschte Ergebnisse.

### Schutzansprüche

#### 1. Computersystem, umfassend:

das Empfang von Daten durch ein Voice-Action-System, die eine neue Voice Action für eine Software-Anwendung spezifizieren, die vom besagten Voice-Action-System verschieden ist, wobei die Daten eine oder mehrere Operationen zum Ausführen der neuen Voice Action und einen oder mehrere Auslöserbegriffe zur Auslösung der neuen Voice Action umfassen.  
das Erstellen einer kontextuellen Voice-Action-Absicht für die Software-Anwendung durch das Voice-Action-System auf Grundlage von mindestens den empfangenen Daten, wobei die kontextuelle Voice-

Action-Absicht Daten umfasst, die, wenn sie von der Software-Anwendung empfangen werden, anfordern, dass die Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt;

Verknüpfung, durch das Voice-Action-System, der kontextuellen Voice-Action-Absicht mit dem einen oder den mehreren Auslöserbegriffen für die neue Voice Action;

wobei das Voice-Action-System dazu konfiguriert ist: eine Angabe einer Benutzeräußerung, die von einem Gerät erhalten wurde, auf dem die Software-Anwendung installiert ist, zu empfangen, festzustellen, dass eine Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit der kontextuellen Voice-Action-Absicht verknüpft sind; und dem Gerät die kontextuelle Voice-Action-Absicht bereitzustellen, in Reaktion auf das Feststellen, und dadurch anzufordern, dass die auf dem Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt.

2. Computersystem nach Anspruch 1, wobei die empfangenen Daten einen Kontext spezifizieren, wobei der Kontext einen Status eines Geräts oder der Software-Anwendung spezifiziert, wenn die neue Voice Action aktiviert ist.

3. Computersystem nach Anspruch 2, wobei der Kontext spezifiziert, dass die Software-Anwendung im Vordergrund eines Geräts läuft, auf dem die Software-Anwendung installiert ist.

4. Computersystem nach Anspruch 2, wobei der Kontext spezifiziert, dass die Software-Anwendung im Hintergrund eines Geräts läuft, auf dem die Software-Anwendung installiert ist.

5. Computersystem nach Anspruch 2, wobei der Kontext spezifiziert, dass die Software-Anwendung eine spezifische Aktivität ausführt.

6. Computersystem nach Anspruch 2, wobei der Kontext spezifiziert, dass eine von der Software-Anwendung ausgeführte Aktivität spezifische Aktivität sich in einem bestimmten Aktivitätszustand befindet.

7. Computersystem nach Anspruch 2, umfassend: das Verknüpfen der kontextuellen Voice-Action-Absicht mit dem Kontext für die neue Voice-Action; das Empfangen von Kontextinformation durch das Voice-Action-System, die einen Status eines bestimmten Geräts, auf dem die Software-Anwendung installiert ist, oder der auf dem bestimmten Gerät installierten Software-Anwendung angibt; das Feststellen, dass die Kontextinformation den Kontext für die neue Voice Action erfüllt; und das Bereitstellen, in Reaktion auf das Feststellen, dass eine Transkription einer Benutzeräußerung, die

vom bestimmten Gerät erhalten wird, mit dem einen oder den mehreren Auslöserbegriffen übereinstimmt, die mit der kontextuellen Voice-Action-Absicht verknüpft sind, und dass die Kontextinformation den mit der kontextuellen Voice-Action-Absicht verknüpften Kontext erfüllt, durch das Voice-Action-System für das bestimmte Gerät, der kontextuellen Voice-Action-Absicht, und dadurch die Anforderung, dass die auf dem bestimmten Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt.

8. Computersystem nach Anspruch 7, wobei das Empfangen der Kontextinformation, die den Status des bestimmten Geräts oder der auf dem bestimmten Gerät installierten Software-Anwendung angibt, Folgendes umfasst:

das Bereitstellen, durch das Voice-Action-System für das bestimmte Gerät, einer Anforderung für bestimmte Kontextinformation; und das Empfangen der bestimmten Kontextinformation auf die Anforderung hin.

9. Computersystem nach Anspruch 7, umfassend: das Feststellen, dass die Kontextinformation einen Kontext für eine zweite Voice Action erfüllt, und dass die Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit einer Absicht für die zweite Voice Action verknüpft sind, wobei die Absicht für die zweite Voice Action eine oder mehrere Operationen zur Ausführung der zweiten Voice Action spezifiziert.

das Auswählen, in Reaktion auf das Feststellen, einer Voice Action aus der neuen Voice Action und der zweiten Voice Action; und

das Bereitstellen, durch das Voice-Action-System für das bestimmte Gerät, der Absicht, die mit der ausgewählten Voice Action verknüpft ist, wodurch die Anforderung ergeht, dass die auf dem bestimmten Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der ausgewählten Voice Action ausführt.

10. Computersystem nach Anspruch 9, wobei das Auswählen der ausgewählten Voice Action aus der neuen Voice Action und der zweiten Voice Action das Auswählen der ausgewählten Voice Action auf den Empfang von Daten hin umfasst, die eine Benutzer-Auswahl einer aus der neuen Voice Action oder der zweiten Voice Action angeben.

11. Computersystem nach Anspruch 9, wobei das Auswählen der ausgewählten Voice Action aus der neuen Voice Action und der zweiten Voice Action Folgendes umfasst:

das Zuweisen eines jeweiligen Werts für die neue Voice Action und die zweite Voice Action; und das Auswählen der ausgewählten Voice Action basierend auf mindestens dem jeweils zugewiesenen

Wert für die neue Voice Action und die zweite Voice Action.

12. Computersystem nach Anspruch 9, wobei das Auswählen der ausgewählten Voice Action aus der neuen Voice Action und der zweiten Voice Action das Auswählen der ausgewählten Voice Action in Reaktion auf das Feststellen umfasst, dass die mit der ausgewählten Voice Action verknüpfte Software-Anwendung im Vordergrund läuft.

13. Computersystem nach Anspruch 1, wobei das Erstellen der kontextuellen Voice-Action-Absicht für die Software-Anwendung das Feststellen umfasst, dass die eine oder die mehreren Operationen der neuen Voice Action von der Software-Anwendung ausführbar sind.

14. Computersystem nach Anspruch 1, umfassend:  
das Feststellen, dass eine Transkription einer Benutzeräußerung, die von einem bestimmten Gerät erhalten wird, auf dem die Software-Anwendung installiert ist, dem einen oder den mehreren Auslöserbegriffen ähnelt, die mit der kontextuellen Voice-Action-Absicht verknüpft sind;  
das Bereitstellen, in Reaktion auf das Feststellen, durch das Voice-Action-System für das bestimmte Gerät, von Daten, die eine Anfrage einer Benutzereingabe angeben, die bestätigt, ob die Benutzeräußerung den einen oder die mehreren Auslöserbegriffe angegeben hat oder mit der Absicht erfolgt ist, die Software-Anwendung zu veranlassen, die neue Voice Action auszuführen;  
das Empfangen in Reaktion auf die Anfrage, durch das Voice-Action System und vom bestimmten Gerät, von Daten, die eine Bestätigung angeben; und  
das Bereitstellen, in Reaktion auf das Empfangen der Daten, die die Bestätigung angeben, der kontextuellen Voice-Action-Absicht an das bestimmte Gerät durch das Voice-Action-System, wodurch die Anfrage ergeht, dass die auf dem bestimmten Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt.

15. Computersystem nach Anspruch 1, umfassend:  
das Empfangen, durch das Voice-Action-System von einem mit der Software-Anwendung verknüpften Entwickler, einer Anfrage, die neue Voice Action für die Software-Anwendung auszurollen; und  
das Bereitstellen, in Reaktion auf die Anfrage, der neuen Voice Action für die Software-Anwendung auf die Anfrage hin, wobei das Bereitstellen der neuen Voice Action für die Software-Anwendung die neue Voice Action für die Software-Anwendung freischaltet.

16. Computersystem nach Anspruch 1, umfassend:

das Empfangen, durch das Voice-Action-System von einem mit der Software-Anwendung verknüpften Entwickler, einer Anfrage, das Ausrollen der neuen Voice Action für die Software-Anwendung rückgängig zu machen; und

auf die Anfrage hin, Rückgängigmachen des Ausrollens der neuen Voice Action für die Software-Anwendung auf die Anfrage hin, wobei das Rückgängigmachen des Ausrollens der neuen Voice Action für die Software-Anwendung die neue Voice Action für die Software-Anwendung deaktiviert.

17. Computersystem nach Anspruch 1, umfassend:

das Empfangen, durch das Voice-Action-System von einem mit der Software-Anwendung verknüpften Entwickler, einer Anfrage, die Erprobung der neuen Voice Action freizugeben, wobei die Anfrage ein oder mehrere Geräte spezifiziert, für die die neue Voice Action freigegeben werden soll; und

das Freigeben, in Reaktion auf die Anfrage, der neuen Voice Action für das eine oder die mehreren spezifizierten Geräte, wobei die neue Voice Action für Geräte, die nicht in den spezifizierten Geräten enthalten sind, deaktiviert ist.

18. Computersystem nach Anspruch 1, umfassend:

das Empfangen, durch das Voice-Action-System, einer Angabe einer Benutzeräußerung, die von einem bestimmten Gerät erhalten wird, auf dem die Software-Anwendung installiert ist.

das Feststellen, durch das Voice-Action-System, dass eine Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit der kontextuellen Voice-Action-Absicht verknüpft sind; und

das Bereitstellen, in Reaktion auf das Feststellen, der kontextuellen Voice-Action-Absicht an das bestimmte Gerät durch das Voice-Action-System, wodurch die Anfrage ergeht, dass die auf dem bestimmten Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt.

19. System, umfassend:

ein oder mehrere Computer, und eine oder mehrere Speichereinrichtungen, welche Anweisungen speichern, die betreibbar sind, wenn sie von einem oder mehreren Computern ausgeführt werden, um den ein oder die mehreren Computer dazu zu veranlassen, Operationen auszuführen, welche das Folgende umfassen:

das Empfangen von Daten durch ein Voice-Action-System, die eine neue Voice Action für eine Software-Anwendung spezifizieren, die vom besagten Voice-Action-System verschieden ist, wobei die Daten eine oder mehrere Operationen zur Ausführung der neuen Voice Action und einen oder mehrere Auslöserbegriffe zur Auslösung der neuen Voice Action umfassen.

das Erstellen einer kontextuellen Voice-Action-Absicht für die Software-Anwendung durch das Voice-Action-System auf Grundlage von mindestens den empfangenen Daten, wobei die kontextuelle Voice-Action-Absicht Daten umfasst, die, wenn sie von der Software-Anwendung empfangen werden, anfordern, dass die Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt;

das Verknüpfen, durch das Voice-Action-System, der kontextuellen Voice-Action-Absicht mit dem einen oder den mehreren Auslöserbegriffen für die neue Voice Action;

wobei das Voice-Action-System dazu konfiguriert ist: das Empfangen einer Angabe einer Benutzeräußerung, die von einem Gerät erhalten wurde, auf dem die Software-Anwendung installiert ist,

das Feststellen, dass eine Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit der kontextuellen Voice-Action-Absicht verknüpft sind; und

das Bereitstellen, in Reaktion auf das Feststellen, der kontextuellen Voice-Action-Absicht für das Gerät und dadurch anzufordern, dass die auf dem Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt.

eine Angabe einer Benutzeräußerung, die von einem Gerät erhalten wurde, auf dem die Software-Anwendung installiert ist, zu empfangen;

das Feststellen, dass eine Transkription der Benutzeräußerung dem einen oder den mehreren Auslöserbegriffen entspricht, die mit der kontextuellen Voice-Action-Absicht verknüpft sind; und

das Bereitstellen, in Reaktion auf das Feststellen, der kontextuellen Voice-Action-Absicht dem Gerät und dadurch anzufordern, dass die auf dem Gerät installierte Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt.

Es folgen 4 Seiten Zeichnungen

20. System nach Anspruch 19, wobei die empfangenen Daten einen Kontext spezifizieren, wobei der Kontext einen Status eines Geräts oder der Software-Anwendung spezifiziert, wenn die neue Voice Action aktiviert ist.

21. Nichtflüchtiges computerlesbares Speichergerät, welches Software speichert, die durch einen oder mehrere Computer ausführbare Anweisungen umfasst, die bei einer solchen Ausführung den einen oder die mehreren Computer zum Ausführen von Tätigkeiten veranlassen, umfassend:

das Empfangen von Daten durch ein Voice-Action-System, die eine neue Voice Action für eine Software-Anwendung spezifizieren, die vom besagten Voice-Action-System verschieden ist, wobei die Daten eine oder mehrere Operationen zur Ausführung der neuen Voice Action und einen oder mehrere Auslöserbegriffe zur Auslösung der neuen Voice Action umfassen.

das Erstellen einer kontextuellen Voice-Action-Absicht für die Software-Anwendung durch das Voice-Action-System auf Grundlage von mindestens den empfangenen Daten, wobei die kontextuelle Voice-Action-Absicht Daten umfasst, die, wenn sie von der Software-Anwendung empfangen werden, anfordern, dass die Software-Anwendung die eine oder die mehreren Operationen der neuen Voice Action ausführt;

Verknüpfung, durch das Voice-Action-System, der kontextuellen Voice-Action-Absicht mit dem einen oder den mehreren Auslöserbegriffen für die neue Voice Action;

wobei das Voice-Action-System dazu konfiguriert ist:

## Anhängende Zeichnungen

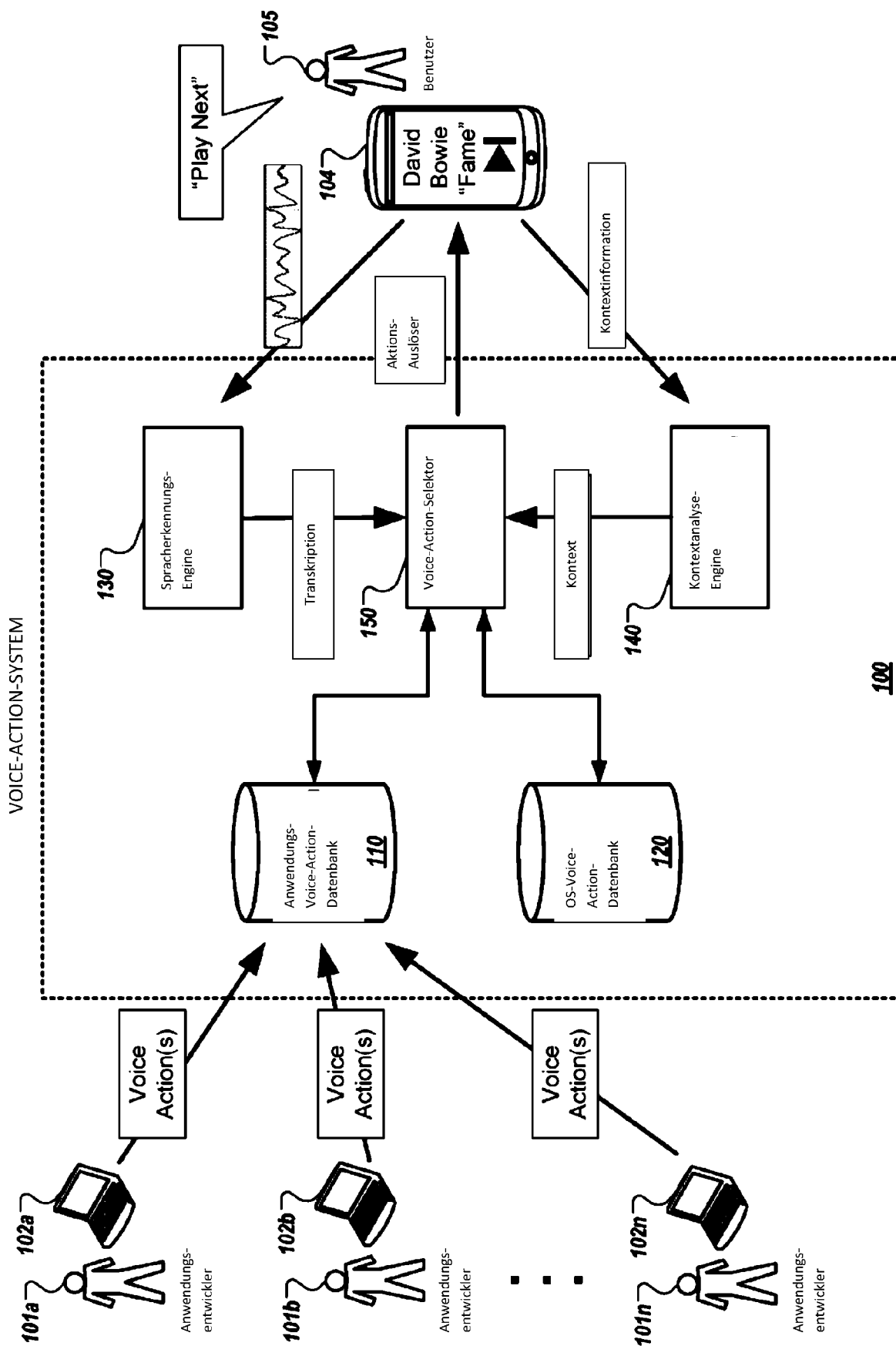


FIG. 1

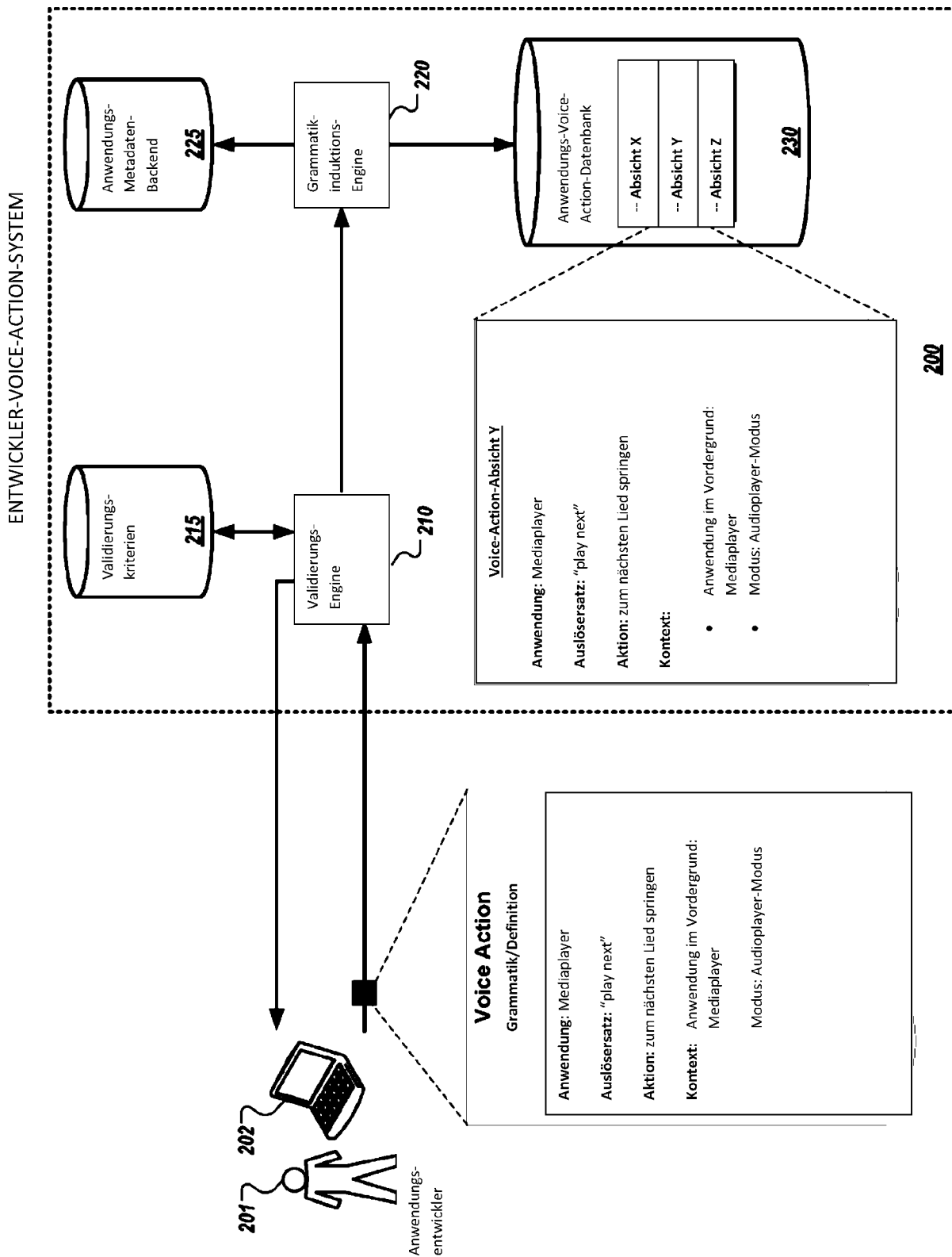


FIG. 2



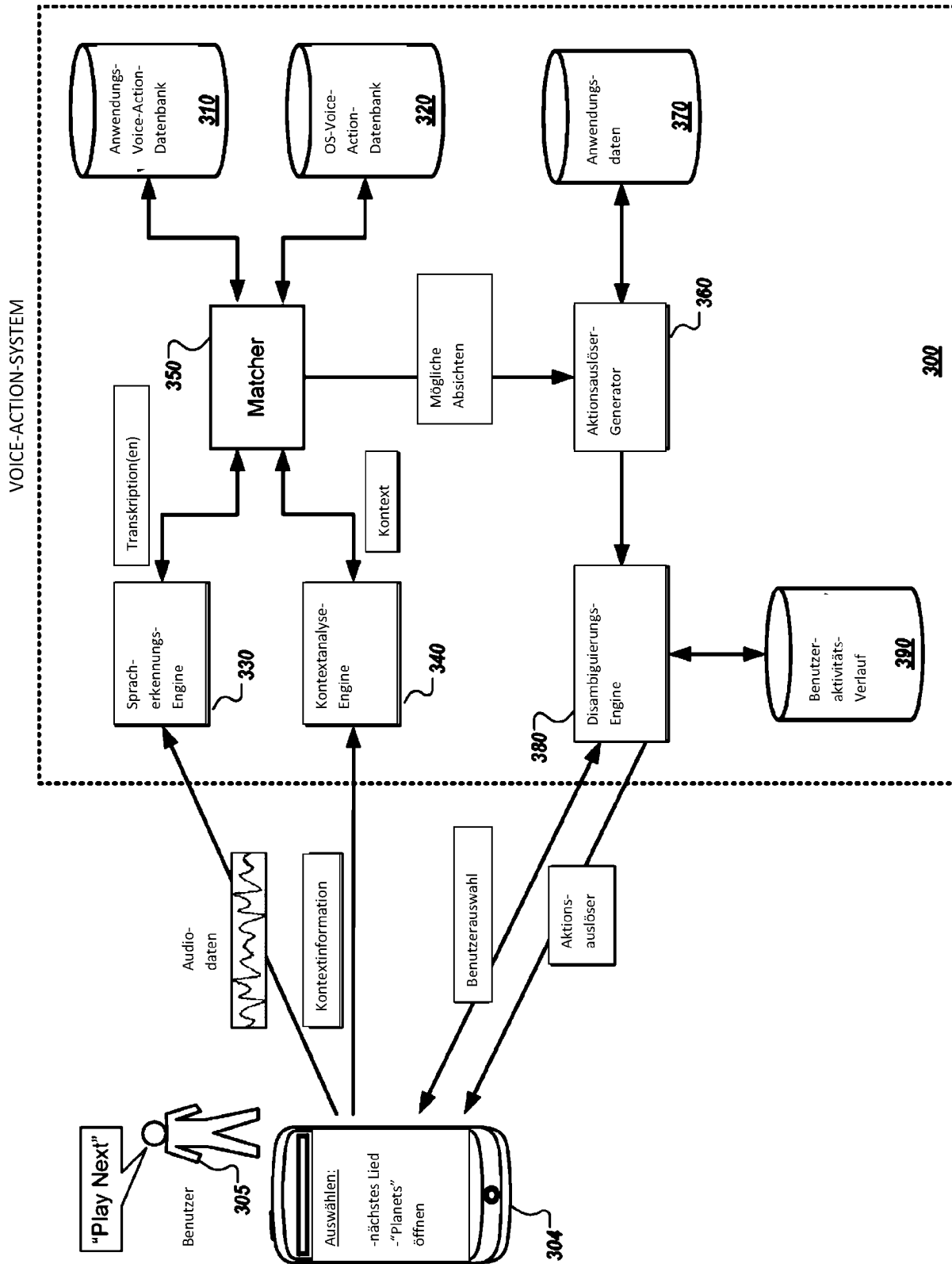


FIG. 3

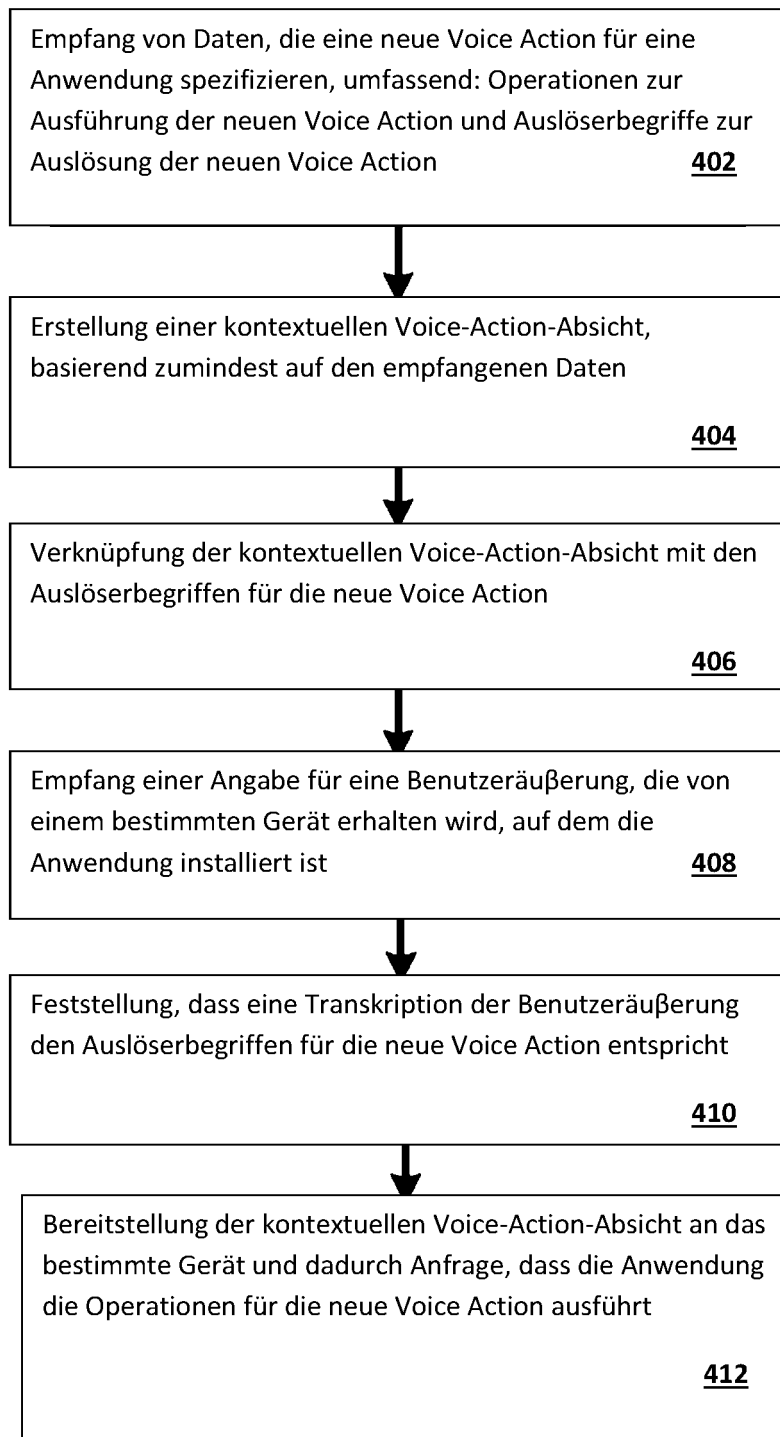
400  
↘

FIG. 4