



(10) **DE 10 2020 119 980 B3** 2021.11.18

(12) **Patentschrift**

(21) Aktenzeichen: **10 2020 119 980.5**
(22) Anmeldetag: **29.07.2020**
(43) Offenlegungstag: –
(45) Veröffentlichungstag
der Patenterteilung: **18.11.2021**

(51) Int Cl.: **G10L 17/06 (2013.01)**
G10L 15/20 (2006.01)
G06F 40/20 (2020.01)
G06F 3/16 (2006.01)

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

(73) Patentinhaber:
**Otto-von-Guericke-Universität Magdeburg,
Körperschaft des öffentlichen Rechts, 39106
Magdeburg, DE**

(56) Ermittelter Stand der Technik:

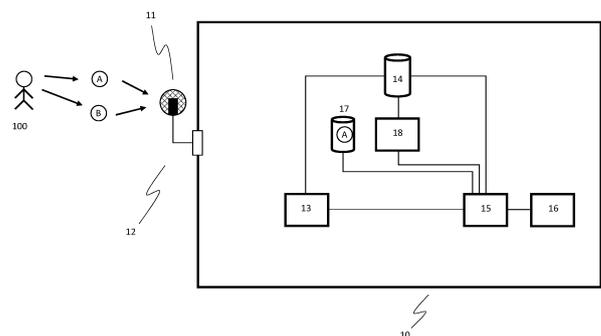
US 2016 / 0 077 792 A1
EP 3 312 831 A1

(72) Erfinder:
**Siegert, Ingo, Prof. Dr., 39114 Magdeburg,
DE; Weißkirchen, Norman, 53804 Much,
DE; Wendemuth, Andreas, Prof. Dr., 38126
Braunschweig, DE**

**SIEGERT, Ingo ; WEISSKIRCHEN, Norman ;
WENDEMUTH, Andreas: Admitting the addressee-
detection faultiness to improve the performance
using a continuous learning framework. In: 8.
Interdisziplinärer Workshop Kognitive Systeme:
Verstehen, Beschreiben und Gestalten Kognitiver
(Technischer) Systeme, Abstractband 2019,
Duisburg, 26.-28. März 2019, S. 38 – 39.**

(54) Bezeichnung: **Sprachassistenzsystem, Verfahren und Computerprogramm zur sprachbasierten Unterstützung**

(57) Zusammenfassung: Die Erfindung betrifft ein Sprachassistenzsystem zur sprachbasierten Unterstützung eines Nutzers, wobei das Sprachassistenzsystem gattungsgemäß eingerichtet ist, eine Sprachakustik zu erkennen und festzustellen, ob die erkannte Sprachakustik an das Sprachassistenzsystem gerichtet war oder nicht, wobei eine Unterstützungsfunktion durch das Sprachassistenzsystem in Abhängigkeit von dem sprachlichen Inhalt der Sprachakustik ausgeführt wird, wenn die Sprachakustik an das Sprachassistenzsystem gerichtet war. Die Erfindung betrifft ebenso ein Verfahren sowie ein Computerprogramm hierzu.



Beschreibung

[0001] Die Erfindung betrifft ein Sprachassistenzsystem zur sprachbasierten Unterstützung eines Nutzers gemäß dem Oberbegriff des Anspruchs 1. Die Erfindung betrifft ebenso ein Verfahren sowie ein Computerprogramm hierzu.

[0002] Es gibt viele verschiedene Formen einer Mensch-Maschine-Schnittstelle. Eine der intuitivsten Formen, eine Mensch-Maschine-Schnittstelle zu etablieren, ist die sprachbasierte Kommunikation zwischen Mensch und Maschine mittels eines Sprachassistenzsystems. Gattungsgemäß werden hierbei permanent die auditiv wahrnehmbaren Schallsignale durch eine auditive Sensorvorrichtung (Aufnahmeeinrichtung) erfasst und eine entsprechende Sprachakustik identifiziert. Mittels Spracherkennungsalgorithmen wird dann der sprachliche Inhalt der Sprachakustik extrahiert und basierend auf dem sprachlichen Inhalt dann eine Unterstützungsfunktion ausgewählt und durch das Assistenzsystem oder einer nachgelagerten (künstlichen) Vorrichtung ausgeführt.

[0003] Derartige Sprachassistenzsysteme bieten somit natürlichen Personen die Möglichkeit, Eingaben an das Assistenzsystem mittels Sprache durchzuführen, wobei das Sprachassistenzsystem den sprachlichen Inhalt der gesprochenen Sprachakustik als Eingabe versteht und entsprechend reagiert. Derartige Reaktionen können beispielsweise wiederum Sprachausgaben sein, wenn die zuvor getätigte Eingabe als Sprachakustik eine Frage an das Assistenzsystem enthielt. Derartige Reaktionen können aber auch das Schalten von Aktoren oder das Durchführen von Berechnungen sein.

[0004] Dabei existieren derzeit zwei große Problemfelder. Zum einen müssen derartige Sprachassistenzsysteme die in der Sprachakustik enthaltenen sprachlichen Inhalte prozesssicher erkennen können. Natürliche Sprache jedoch ist fehlerbehaftet und kann einen großen Interpretationsspielraum aufweisen. Meist nur unter Kenntnis eines konkreten Kontextes lässt sich dann mit hoher Wahrscheinlichkeit feststellen, was tatsächlich gemeint ist. Diesen sprachlichen Kontext maschinell zu erfassen, ist algorithmisch eine große Herausforderung und nicht selten ebenfalls fehlerbehaftet, da zur Ermittlung des sprachlichen Kontextes auch nonverbale Kanäle, wie beispielsweise prosodische Eigenschaften der Sprache, Mimik oder Gesten eine große Rolle spielen. Durch die immer weiter voranschreitende Leistungsfähigkeit moderner Computersysteme sowie den Entwicklungen im Bereich der KI konnte in den letzten Jahren eine signifikante Steigerung der Erkennungsrate von sprachlichen Inhalten einer natürlichen Sprachakustik erreicht werden, wodurch die sprachbasierte Mensch-Maschine-Schnittstelle in

vielen Bereichen des alltäglichen Lebens Einzug erhalten hat.

[0005] Das zweite große Problemfeld besteht darin, dass ein Sprachassistenzsystem von sich aus feststellen muss, ob eine gesprochene Sprachakustik an das Sprachassistenzsystem gerichtet ist oder nicht. Bei einer natürlichen Sprachkommunikation zwischen Menschen erfolgt eine solche Feststellung oftmals nur unter Kenntnis anderer Kommunikationskanäle, wie beispielsweise Blickkontakt, Mimik oder prosodischen Eigenschaften, was einem Sprachassistenzsystem derzeit verwehrt bleibt. Sprachassistenzsysteme müssen demnach einzig und allein basierend auf der empfangenen und erkannten Sprachakustik feststellen, ob die Sprachakustik an das Sprachassistenzsystem gerichtet ist und demzufolge eine Reaktion erwartet wird oder nicht.

[0006] Es ist bekannt, dass Sprachassistenzsysteme gewisse Schlüsselwörter bzw. Kommandowörter erkennen und anhand dieser Erkennung darauf schließen, dass die Sprachakustik bzw. die folgende Sprachakustik an sie gerichtet sind. Gerade im Bereich der Hausautomation finden sich eine Reihe von derartigen Sprachassistenzsystemen, die auf ganz spezielle Kommandowörter reagieren und dann entsprechend die Sprachakustik hinsichtlich der darin enthaltenen natürlichen Sprache analysieren und entsprechende Unterstützungsfunktionen daraufhin ausführen. So kann eine Sprachsteuerung realisiert werden, mit der sich beispielsweise die Beleuchtung oder andere Dinge im Haus steuern lassen.

[0007] Nachteilig hierbei ist jedoch, dass die Erkennung von Kommandowörtern fehleranfällig ist. Denn das Sprachassistenzsystem kann nicht unterscheiden, ob das gesprochene Kommandowort im Kontext der gesamten Sprachakustik dazu dient, das Sprachassistenzsystem zu bedienen oder ob das gesprochene Kommandowort im Kontext Teil einer Konversation zwischen zwei Personen ist. Es kann daher dazu führen, dass das Sprachassistenzsystem fälschlicherweise annimmt, die aufgenommene und erkannte Sprachakustik sei an das System gerichtet, obwohl die erkannte Sprachkommunikation Teil einer zwischenmenschlichen Sprachkommunikation ist und das Sprachassistenzsystem selber gar nicht gemeint ist. Dies ist deshalb problematisch, da das Sprachassistenzsystem grundsätzlich jedes auditiv wahrnehmbare Schallsignal empfängt und verarbeitet, um zur richtigen Zeit das an das Sprachkommunikationssystem gerichtete Kommando Wort auch erkennen zu können.

[0008] Andere Umstände hingegen, wie beispielsweise eine qualitativ schlechte Sprachqualität bzw. Aufnahmequalität, können hingegen dazu führen, dass das Sprachassistenzsystem trotz eines gesprochenen Kommandowortes die aufgenommene

Sprachakustik nicht als an sich gerichtet erkennt und somit keine Funktion ausführt.

[0009] Ein gattungsgemäßes Sprachassistenzsystem ist aus der US 2016 / 0 077 792 A1 bekannt. Aus der EP 3 312 831 A1 ist die Erkennung einer Aktivierungsphrase mit geringer Leistung bekannt. Weitere Verfahren zur Spracherkennung sind bekannt aus: SIEGERT, Ingo ; WEISSKIRCHEN, Norman ; WENDEMUTH, Andreas: Admitting the addresser-detection faultiness to improve the performance using a continuous learning framework. In: 8. Interdisziplinärer Workshop Kognitive Systeme: Verstehen, Beschreiben und Gestalten Kognitiver (Technischer) Systeme, Abstractband 2019, Duisburg, 26.-28. März 2019, S. 38-39.

[0010] Es ist daher Aufgabe der vorliegenden Erfindung ein verbessertes Sprachassistenzsystem sowie Verfahren zur sprachbasierten Unterstützung anzugeben, mit dem sich die Erkennungsrate bezüglich des Ansprechverhaltens von Sprachassistenzsystemen erhöhen lässt.

[0011] Die Aufgabe wird mit dem Sprachassistenzsystem gemäß Anspruch 1 sowie dem Verfahren gemäß Anspruch 6 erfindungsgemäß gelöst. Vorteilhafte Ausgestaltungen der Erfindung finden sich in den entsprechenden Unteransprüchen.

[0012] Gemäß Anspruch 1 wird ein Sprachassistenzsystem zur sprachbasierten Unterstützung eines Nutzers vorgeschlagen, dass zunächst eine Aufnahmeeinrichtung hat, um auditiv wahrnehmbare Schallsignale aufnehmen zu können. Eine derartige Aufnahmeeinrichtung kann beispielsweise einen Schallsensor (Mikrofon) aufweisen, der ein auditiv wahrnehmbares Schallsignale empfängt und in ein digitales Signal umwandelt.

[0013] Das Sprachassistenzsystem weist gattungsgemäß des Weiteren eine Erkennungseinrichtung auf, die beispielsweise mittels einer mikroprozessorgesteuerten Recheneinheit in den aufgenommenen auditiv wahrnehmbaren Schallsignalen eine Sprachakustik erkennt und daraus sprachbasierte Informationen extrahiert. Die Erkennungseinrichtung ist demzufolge so ausgebildet, dass sie eine Sprachanalyse durchführt und so beispielsweise anhand von Modellen aus der Sprachakustik die darin enthaltenen sprachbasierten Informationen extrahiert. Derartige sprachbasierte Informationen können neben den sprachlichen Inhalten (d. h. den gesprochenen Wörtern der in der Sprachakustik enthaltenen Sprachkommunikation) auch weitere sprachbasierte Informationen enthalten, wie beispielsweise prosodische Eigenschaften der Sprachakustik. Unter den prosodischen Eigenschaften werden dabei insbesondere ein jene Eigenschaften der Sprache im Sinne der vorliegenden Erfindung verstanden, die sich

nicht auf die sprachlichen Inhalte, d. h. die gesprochenen Wörter, beziehen. Dies umfasst insbesondere Akzente, Tonsprache, Intonation, Quantität, Tempo, Rhythmus, Pausen beim Sprechen und dergleichen. Sprachbasierte Informationen, die aus den auditiv wahrnehmbaren Schallsignalen einer Sprachakustik extrahiert wurden, können somit insbesondere sprachliche Inhalte und/oder prosodische Eigenschaften der Sprachakustik bzw. der gesprochenen Inhalte enthalten.

[0014] Das Sprachassistenzsystem weist gattungsgemäß des Weiteren eine Aktivierungseinrichtung auf, die beispielsweise ebenfalls mittels einer mikroprozessorgesteuerten Recheneinheit basierend auf einem Sprachmodell in Abhängigkeit von den extrahierten sprachbasierten Informationen erkennt, ob die erkannte Sprachakustik an das Sprachassistenzsystem gerichtet ist oder nicht. Die Aktivierungseinrichtung kann somit eine Aktivierungsabsicht in Abhängigkeit von den extrahierten sprachbasierten Informationen erkennen. Eine Aktivierungsabsicht meint hierbei, dass der Nutzer das Sprachassistenzsystem bewusst anspricht, um eine Sprachkommunikation aufzubauen.

[0015] Basierend auf dieser Erkennung, ob die Sprachakustik bzw. Sprachkommunikation an das Sprachassistenzsystem gerichtet war oder nicht, wird das Sprachassistenzsystem zur sprachbasierten Unterstützung aktiviert, nicht aktiviert oder wiederum deaktiviert. Wurde die Sprachakustik als an das Sprachassistenzsystem gerichtet erkannt, so wird demzufolge mittels der Aktivierungseinrichtung das Sprachassistenzsystem aktiviert bzw. in einem aktivierten Zustand versetzt. Wurde die Sprachakustik hingegen als an das Sprachassistenzsystem nicht gerichtet erkannt, so wird demzufolge mittels der Aktivierungseinrichtung das Sprachassistenzsystem nicht aktiviert oder gegebenenfalls deaktiviert (sofern es zuvor aktiviert war) bzw. in einem nicht aktivierten Zustand versetzt. Die Erkennungseinrichtung und die Aktivierungseinrichtung können dabei baulich eine Einheit darstellen.

[0016] Schließlich weist das Sprachassistenzsystem gattungsgemäß eine Assistenzeinrichtung auf, die eingerichtet ist, in Abhängigkeit von extrahierten sprachbasierten Informationen eine Unterstützungsfunktion auszuwählen und zur sprachbasierten Unterstützung des Nutzers auszuführen, sofern das Sprachassistenzsystem zuvor durch die Aktivierungseinrichtung aktiviert wurde.

[0017] Dies bedeutet, dass beispielsweise in einer Sprachakustik durch die Aktivierungseinrichtung ein Kommandowort erkannt wurde, was dazu führt, dass die Aktivierungseinrichtung das Sprachassistenzsystem in einem aktivierten Zustand versetzt und die weiterhin in der Sprachakustik enthaltenen sprach-

basierten Informationen zur Auswahl der Unterstützungsfunktion verwendet. Dies ist in der Regel dann der Fall, wenn die Sprachakustik neben dem Kommandowort auch die Anweisung an das Sprachassistenzsystem enthält. Neben dem Kommandowort zur Erkennung einer Aktivierungsabsicht ist es alternativ oder zusätzlich aber auch möglich, dass basierend auf prosodischen Eigenschaften, die in den extrahierten sprachbasierten Informationen enthalten sind, eine Aktivierungsabsicht erkannt wird. Es wird somit anhand von prosodischen Eigenschaften festgestellt, ob die erfasste Sprachakustik an das Sprachassistenzsystem gerichtet ist oder nicht.

[0018] Das Sprachassistenzsystem kann alternativ oder zusätzlich darüber hinaus auch eingerichtet sein, in einer ersten Sprachakustik das entsprechende Kommandowort zu erkennen, woraufhin das Sprachassistenzsystem durch die Aktivierungseinrichtung in den aktivierten Zustand versetzt wird, wobei in einer danach empfangenen und erkannten zweiten Sprachakustik dann die entsprechenden Anweisungen an das Sprachassistenzsystem enthalten sind, die dann zur Auswahl und Umsetzung der Unterstützungsfunktion verwendet werden. In diesem Fall sind das Kommandowort und die Anweisungen an das Sprachassistenzsystem in jeweils voneinander verschiedenen Sprachakustiken enthalten. Nach der ersten Sprachakustik, die von dem Sprachassistenzsystem empfangen wurde, kann das Sprachassistenzsystem gegebenenfalls eine Bestätigung in sprachlicher Form an den Nutzer ausgeben, um den Nutzer darüber zu informieren, dass das Sprachassistenzsystem nunmehr auf die an das Sprachassistenzsystem gerichtet und Anweisungen wartet. Das Sprachassistenzsystem befindet sich demnach nach der ersten Sprachakustik für einen gewissen Zeitraum in einem aktivierten Zustand und erwartet demnach eine zweite Sprachakustik.

[0019] Darüber hinaus kann das Sprachassistenzsystem alternativ oder zusätzlich auch eingerichtet sein, mittels der Aktivierungseinrichtung anhand von prosodischen Eigenschaften der erkannten Sprachakustik zu erkennen, ob die erkannte Sprachakustik an das Sprachassistenzsystem gerichtet ist oder nicht.

[0020] Ein solches gattungsgemäßes Sprachassistenzsystem wird nun dahingehend erfindungsgemäß weitergebildet, dass in einem digitalen Datenspeicher eine erste Sprachakustik und/oder die aus der ersten Sprachakustik extrahierten sprachbasierten Informationen hinterlegt werden und dass eine Adaptionseinrichtung vorgesehen ist, die eingerichtet ist, das Sprachmodell in Abhängigkeit von der zuvor in dem Datenspeicher hinterlegten ersten Sprachakustik und/oder den aus der ersten Sprachakustik extrahierten sprachbasierten Informationen anzupassen, wenn bei der ersten Sprachakustik eine Unsicherheit und/oder Fehlerhaftigkeit darüber bestand, ob

die erste Sprachakustik an das Sprachassistenzsystem gerichtet war oder nicht, und diese Unsicherheit und/oder Fehlerhaftigkeit durch eine auf die erste Sprachakustik zeitlich nachfolgende erkannte oder durch erkanntes Ausbleiben einer entsprechenden zweiten Sprachakustik ausgeräumt wird.

[0021] Es wird somit die erste Sprachakustik bzw. die aus der ersten Sprachakustik extrahierten sprachbasierten Informationen dazu verwendet, um das Sprachmodell anzupassen, wenn eine entsprechende Unsicherheit und/oder Fehlerhaftigkeit darüber erkannt wurde, ob die erste Sprachakustik an das Sprachassistenzsystem gerichtet war oder nicht. Diese Unsicherheit und/oder Fehlerhaftigkeit wird dabei durch eine auf die erste Sprachakustik folgende zweite Sprachakustik oder dem Ausbleiben einer solchen zweiten Sprachakustik erkannt.

[0022] Eine Unsicherheit darüber, ob eine erste Sprachakustik an das Sprachassistenzsystem gerichtet war oder nicht, tritt in der Regel immer dann auf, wenn die Aktivierungseinrichtung nicht mit hinreichender Sicherheit erkennen kann, ob die Sprachakustik nun anders Sprachassistenzsystem gerichtet war oder nicht. Diese hinreichende Sicherheit wird beispielsweise über einen Schwellwert (beispielsweise einen prozentualen Schwellwert oder einem Schwellwert zwischen 0 und 1) abgebildet. Das System kann dabei die Erkennung der Aktivierungsabsicht in mindestens drei Bereiche unterteilen. Der erste Bereich wird immer dann verwendet, wenn keine Unsicherheit über eine Aktivierungsabsicht besteht. Der zweite Bereich wird immer dann verwendet, wenn das System unsicher darüber ist, ob eine Aktivierungsabsicht vom Nutzer vorlag oder nicht (es liegt eine Unsicherheit vor). Der dritte Bereich wird immer dann verwendet, wenn das System mit Sicherheit erkennen kann, dass eine Aktivierungsabsicht nicht vorliegt. Unsicherheit ist Systemsicht.

[0023] So kann ein erster Schwellwert angegeben, dass die Sprachakustik mit hinreichender Sicherheit an das Sprachassistenzsystem gerichtet ist. Ein zweiter Schwellwert kann gegenüber dessen angegeben, dass die Sprachakustik mit hinreichender Sicherheit nicht an das Sprachassistenzsystem gerichtet ist. Wird weder der erste noch der zweite Schwellwert überschritten, kann das Sprachassistenzsystem nicht mit hinreichender Sicherheit annehmen, dass die Sprachakustik an das Sprachassistenzsystem gerichtet ist oder eben gerade nicht an das Sprachassistenzsystem gerichtet ist. In den meisten Fällen erfolgt bei einer Unsicherheit in der Regel keine Aktivierung.

[0024] Ein solcher Schwellenwert bzw. die mehreren Schwellenwerte können dabei Bestandteil des Sprachmodells sein, auf deren Basis entschieden wird, ob die erkannte Sprachakustik an das

Sprachassistenzsystem gerichtet ist oder nicht. So kann die Anpassung des Sprachmodells beispielsweise dadurch erfolgen, dass die Werte der Schwellenwerte des Sprachmodells entsprechend verändert werden, um so eine verbesserte Entscheidungsgrundlage zu generieren.

[0025] Eine Fehlerhaftigkeit besteht dann, wenn die Aktivierungseinrichtung das Sprachassistenzsystem aktiviert hat, obwohl die Sprachakustik nicht an das Sprachassistenzsystem gerichtet war oder wenn die Aktivierungseinrichtung das Sprachassistenzsystem nicht aktiviert hat oder deaktiviert hat, obwohl die Sprachakustik definitiv an das Sprachassistenzsystem gerichtet war. Dies kann beispielsweise durch den Inhalt der zweiten Sprachakustik erkannt werden oder dadurch, dass eine auf die erste Sprachakustik folgende zweite Sprachakustik ausbleibt und somit dieses Ausbleiben einer zweiten Sprachakustik erkannt wird. Fehlerhaftigkeit ist Nutzersicht.

[0026] Die Tatsache, ob eine Unsicherheit oder Fehlerhaftigkeit vorlag, kann meist durch wenigstens eine zweite Sprachakustik erkannt und gegebenenfalls ausgeräumt werden. In diesem Fall werden jedoch die zuvor in dem digitalen Datenspeicher hinterlegten extrahierten sprachbasierten Informationen der ersten Sprachakustik dazu verwendet, das Sprachmodell zur Erkennung, ob die Sprachakustik an das Sprachassistenzsystem gerichtet ist oder nicht, dahingehend anzupassen, dass es zukünftig derartige Unsicherheiten und/oder Fehlerhaftigkeit und bei der Erkennung der Ansprache des Assistenzsystems vermeidet bzw. die Unsicherheitsrate und/oder Fehlerrate verringert.

[0027] Es hat sich gezeigt, dass eine solche nachträgliche Anpassung des Sprachmodells bei einer bestehenden Unsicherheit oder Fehlerhaftigkeit, ob das Sprachassistenzsystem durch eine Sprachakustik angesprochen wurden ist oder nicht, basierend auf derjenigen Sprachakustik, die fehlerhaft oder und sicher erkannt wurde, zu einem verbesserten Aktivierungsverhalten des Sprachassistenzsystems führt, sodass die Akzeptanz derartiger Sprachassistenzsystem deutlich gesteigert werden kann.

[0028] Gemäß der Erfindung ist vorgesehen, dass das Sprachassistenzsystem eingerichtet ist, mittels der Adaptionseinrichtung das Sprachmodell in Abhängigkeit von der zuvor hinterlegten ersten Sprachakustik und/oder den aus der ersten Sprachakustik extrahierten sprachbasierten Informationen anzupassen, wenn innerhalb eines gewissen Zeitraumes nach der ersten Sprachakustik die zweite Sprachakustik erkannt oder innerhalb des gewissen Zeitraumes das Ausbleiben der zweiten Sprachakustik erkannt wird.

[0029] Demnach sollte die zweite Sprachakustik innerhalb eines gewissen Zeitraumes nach dem Empfang bzw. dem Erkennen der ersten Sprachakustik empfangen bzw. erkannt werden, um aufgrund des zeitlichen Bezuges einen kontextuellen Zusammenhang beider Sprachakustiken herzustellen. Die erste Sprachakustik bzw. die aus der ersten Sprachakustik extrahierten sprachbasierten Informationen, wie beispielsweise prosodische Eigenschaften oder sprachbasierte Informationen, werden dann zur Anpassung des Sprachmodells verwendet, dann diesem Fall davon ausgegangen werden kann, dass die erste Sprachakustik und die zweite Sprachakustik in einem kontextuellen Zusammenhang stehen und somit zusammengehören.

[0030] Dabei kann es auch vorgesehen sein, dass innerhalb des gewissen Zeitraumes keine weitere zweite Sprachakustik erkannt wird (das Ausbleiben einer folgenden zweiten Sprachakustik wird erkannt), sodass davon auszugehen ist, dass die Erkennung der ersten Sprachakustik fehlerhaft war. Dies ist meist dann der Fall, wenn die erste Sprachakustik fehlerhafterweise als an das Sprachassistenzsystem gerichtet erkannt wurde, was von dem Sprecher jedoch nicht beabsichtigt war. Der Sprecher bleibt nun für den gewissen Zeitraum stumm und erzeugt keine weitere Sprachakustik, sodass das Sprachassistenzsystem nunmehr annimmt, dass die ursprüngliche erste Sprachakustik nicht an das Sprachassistenzsystem gerichtet war. In diesem Fall wird ebenfalls die erste Sprachakustik bzw. die aus der ersten Sprachakustik extrahierten sprachbasierten Informationen zur Anpassung des Sprachmodells verwendet.

[0031] Gemäß der Erfindung ist vorgesehen, dass das Sprachassistenzsystem eingerichtet ist, den gewissen Zeitraum in Abhängigkeit von mindestens einem akustischen Qualitätskriterium der ersten Sprachakustik festzulegen oder zu variieren. Solche akustischen Qualitätskriterien können beispielsweise das Signal-Rausch-Verhältnis bzw. der Signal-Rausch-Abstand, schon erfolgte Fehlaktivierungen, Anzahl der Sprecher, etc. sein.

[0032] Dabei ist es denkbar, dass an diesen gewissen Zeitraum, der beispielsweise wenige Sekunden betragen kann (vorzugsweise weniger als 10 Sekunden, sondern vorzugsweise weniger als 5 Sekunden), auch die Speicherdauer der ersten Sprachakustik bzw. der aus der ersten Sprachakustik extrahierten sprachbasierten Informationen gekoppelt ist, sodass nach einer gewissen Zeit diese Informationen aus dem digitalen Datenspeicher wieder gelöscht werden.

[0033] Gemäß einer Ausführungsform ist vorgesehen, dass die Erkennungseinrichtung eingerichtet ist, prosodische Eigenschaften der ersten Sprachakustik als Teil der sprachbasierten Informationen aus

der ersten Sprachakustik zu extrahieren, die Aktivierungseinrichtung eingerichtet ist, basierend auf dem Sprachmodell in Abhängigkeit von den in den extrahierten sprachbasierten Informationen enthaltenen prosodischen Eigenschaften der ersten Sprachakustik zu erkennen, ob die erkannte Sprachakustik an das Sprachassistenzsystem gerichtet ist oder nicht, und die Adaptionseinrichtung eingerichtet ist, das Sprachmodell hinsichtlich der prosodischen Erkennung in Abhängigkeit von der zuvor in dem Datenspeicher hinterlegten ersten Sprachakustik, den aus der ersten Sprachakustik extrahierten sprachbasierten Informationen und/oder den in den extrahierten sprachbasierten Informationen enthaltenen prosodischen Eigenschaften der ersten Sprachakustik anzupassen.

[0034] In dieser Ausführungsform wird das Sprachmodell hinsichtlich der prosodischen Erkennung, ob die Sprachakustik an das Sprachassistenzsystem gerichtet ist oder nicht, angepasst, sodass zukünftig die Erkennungsrate verbessert und die Unsicherheitsrate bzw. Fehlerrate verringert werden kann.

[0035] Gemäß einer Ausführungsform ist vorgesehen, dass die Erkennungseinrichtung eingerichtet ist, sprachliche Inhalte der ersten Sprachakustik als Teil der sprachbasierten Informationen aus der ersten Sprachakustik zu extrahieren, die Aktivierungseinrichtung eingerichtet ist, basierend auf dem Sprachmodell in Abhängigkeit von den in den extrahierten sprachbasierten Informationen enthaltenen sprachlichen Inhalten zu erkennen, ob die erkannte Sprachakustik an das Sprachassistenzsystem gerichtet ist oder nicht, und die Adaptionseinrichtung eingerichtet ist, das Sprachmodell hinsichtlich der Erkennung von sprachlichen Inhalten in Abhängigkeit von der zuvor in dem Datenspeicher hinterlegten ersten Sprachakustik, den aus der ersten Sprachakustik extrahierten sprachbasierten Informationen und/oder den in den extrahierten sprachbasierten Informationen enthaltenen sprachlichen Inhalten der ersten Sprachakustik anzupassen.

[0036] In dieser Ausführungsform wird das Sprachmodell hinsichtlich der sprachbasierten Informationen angepasst, sodass beispielsweise Aktivierungsabsichten (bspw. Kommandowörter, prosodische Eigenschaften der Sprachakustik) zukünftig besser erkannt werden können.

[0037] Gemäß einer Ausführungsform ist vorgesehen, dass die Aktivierungseinrichtung eingerichtet ist, bei einer Unsicherheit darüber, ob die erkannte Sprachakustik als an das Sprachassistenzsystem gerichtet oder nicht erkannt werden soll (d.h. ob eine Aktivierungsabsicht vorliegt oder nicht), eine optische, haptische, olfaktorische und/oder akustische Systemrückfrage zu generieren und mittels einer Ausgabe-einrichtung des Sprachassistenzsystems auszuge-

ben. Die Systemrückfrage kann eine akustische Systemrückfrage, bspw. eine Sprachausgabe in Form einer Sprachakustik sein, wobei die Sprachausgabe eine Nachfrage in Bezug auf die erkannte Sprachakustik an den Nutzer enthält.

[0038] In dieser Ausführungsform wird das Sprachassistenzsystem bei einer erkannten Unsicherheit zunächst eine Systemrückfrage generieren, die eine optische, haptische, olfaktorische und/oder akustische Nachfrage in Bezug auf die erkannte erste Sprachakustik enthält. Das Assistenzsystem fragt demnach den sprechenden Nutzer an, was gemeint ist bzw. ob die Sprachakustik an das Assistenzsystem tatsächlich gerichtet war oder nicht.

[0039] Gemäß einer Ausführungsform ist vorgesehen, dass das Sprachmodell ein auf einem maschinellen Lernsystem basierendes gelerntes Sprachmodell ist.

[0040] Die Aufgabe wird im Übrigen auch mit dem Verfahren zur sprachbasierten Unterstützung eines Nutzers gemäß Anspruch 6 gelöst. Dies beinhaltet, dass mittels einer Aufnahmeeinrichtung auditiv wahrnehmbare Schallsignale aufgenommen werden, mittels einer Erkennungseinrichtung in den aufgenommenen, auditiv wahrnehmbaren Schallsignalen eine Sprachakustik erkannt und, sofern eine Sprachakustik erkannt wurde, sprachbasierte Informationen aus der erkannten Sprachakustik extrahiert werden, mittels einer Aktivierungseinrichtung basierend auf einem Sprachmodell in Abhängigkeit von den extrahierten sprachbasierten Informationen erkannt wird, ob die erkannte Sprachakustik an das Sprachassistenzsystem gerichtet ist oder nicht, wobei das Sprachassistenzsystem zur sprachbasierten Unterstützung aktivieren wird, wenn die Sprachakustik als an das Sprachassistenzsystem gerichtet erkannt wurde, und/oder nicht aktiviert wird, wenn die Sprachakustik nicht als an das Sprachassistenzsystem gerichtet erkannt wurde, und indem mittels einer Assistenzeinrichtung in Abhängigkeit von extrahierten sprachbasierten Informationen eine Unterstützungsfunktion ausgewählt und zur sprachbasierten Unterstützung des Nutzers ausgeführt wird, wenn das Sprachassistenzsystem zuvor aktiviert wurde, wobei dass in einem digitalen Datenspeicher eine erste Sprachakustik und/oder die aus der ersten Sprachakustik extrahierten sprachbasierten Informationen hinterlegt werden, und dass mittels einer Adaptionseinrichtung das Sprachmodell in Abhängigkeit von der zuvor in dem Datenspeicher hinterlegten ersten Sprachakustik und/oder den aus der ersten Sprachakustik extrahierten sprachbasierten Informationen angepasst wird, wenn bei der ersten Sprachakustik eine Unsicherheit und/oder Fehlerhaftigkeit darüber bestand, ob die erste Sprachakustik an das Sprachassistenzsystem gerichtet war oder nicht, und diese Unsicherheit und/oder Fehlerhaftig-

keit durch eine auf die erste Sprachakustik zeitlich nachfolgende erkannte zweite Sprachakustik oder durch erkanntes Ausbleiben einer entsprechenden zweite Sprachakustik ausgeräumt wird.

[0041] Vorteilhafte Ausgestaltungen des Verfahrens finden sich in den Unteransprüchen.

[0042] Die Erfindung wird anhand der beigefügten Figuren beispielhaft näher erläutert. Es zeigt:

Fig. 1 schematische Darstellung des erfindungsgemäßen Sprachassistenzsystems.

[0043] **Fig. 1** zeigt in einer schematisch vereinfachten Darstellung ein Sprachassistenzsystem **10**, das mittels einer Aufnahmeeinrichtung **11** zum Aufnehmen von auditiv wahrnehmbaren Schallsignalen ausgebildet ist. Eine solche Aufnahmeeinrichtung **11** kann dabei beispielsweise ein Mikrofon **12** aufweisen, mit der die Schallsignale erfasst und anschließend mittels einer Wandlereinheit in ein digitales Signal umgewandelt werden.

[0044] Eine Erkennungseinrichtung **13** erhält nun diese in digitaler Form vorliegenden Schallsignale als Eingang und ist eingerichtet, basierend auf den auditiv wahrnehmbaren Schallsignalen eine darin enthaltene Sprachakustik **A, B** zu erkennen. Die Erkennungseinrichtung **13** ist somit ausgebildet, in einem ersten Schritt zu identifizieren, ob in dem aufgenommenen Schallsignal und Sprachakustik **A, B** enthalten ist oder nicht.

[0045] Hat die Erkennungseinrichtung **13** erkannt, dass in den auditiv wahrnehmbaren Schallsignalen eine Sprachakustik **A, B** enthalten ist, so ist die Erkennungseinrichtung **13** des Weiteren ausgebildet, sprachbasierte Informationen aus der erkannten Sprachakustik **A, B** in an sich bekannter Weise zu extrahieren. Derartige sprachbasierte Informationen können dabei prosodische Eigenschaften der Sprachakustik **A, B** enthalten. Die sprachbasierten Informationen enthalten darüber hinaus aber auch sprachliche Inhalte der Sprachakustik **A, B**, d. h. jene Wörter und Sätze, die von einem Nutzer **100** zuvor gesprochen wurden.

[0046] Im Ausführungsbeispiel der **Fig. 1** hat ein Nutzer **100** eine erste Sprachakustik **A** und eine zweite Sprachakustik **B** erzeugt, indem der Nutzer **100** bestimmte Wörter oder Sätze ausgesprochen hat. Die erste Sprachakustik **A** wurde dabei zu einem ersten Zeitpunkt t ausgesprochen, während die zweite Sprachakustik **B** zu einem darauffolgenden zweiten Zeitpunkt $t+1$ ausgesprochen wurde. Zwischen dem ersten Zeitpunkt und dem zweiten Zeitpunkt liegt dabei ein gewisser Zeitraum, der einen kontextuellen Zusammenhang beider Sprachakustiken **A** und **B** nahelegt.

[0047] Die Erkennungseinrichtung **13** hat dabei Zugriff auf einem Datenspeicher, auf dem ein Sprachmodell **14** hinterlegt ist. Basierend auf diesem Sprachmodell **14** ist die Erkennungseinrichtung **13** in der Lage, zu erkennen, ob in den auditiv wahrnehmbaren Schallsignalen eine Sprachakustik **A, B** erkannt wurde und darüber hinaus in der Lage, die sprachbasierten Informationen aus der erkannten Sprachakustik **A, B** zu extrahieren.

[0048] Die sprachbasierten Informationen werden dann an eine Aktivierungseinrichtung **15** übermittelt, die ebenfalls Zugriff auf das Sprachmodell **14** hat und basierend darauf eingerichtet ist zu erkennen, ob die erkannte Sprachakustik **A, B** an das Sprachassistenzsystem **10** gerichtet ist oder nicht.

[0049] So kann die erste Sprachakustik **A** dergestalt sein, dass in der Sprachakustik **A** ein Kommandowort und/oder eine prosodische Eigenschaft enthalten ist, wodurch angezeigt wird, dass die erste Sprachakustik **A** oder die zweite Sprachakustik **B** an das Sprachassistenzsystem **10** gerichtet ist (erste Sprachakustik **A** und zweite Sprachakustik **B** können zeitlich in einem Zusammenhang gesprochen sein oder zeitlich nacheinander versetzt, um eine Reaktion des Systems zwischen der ersten Sprachakustik **A** und zweiten Sprachakustik **B** abzuwarten). Durch die Erkennungseinrichtung **13** wurde dabei erkannt, dass in der Sprachakustik **A** ein Kommandowort und/oder eine prosodische Eigenschaft enthalten sind sowie ggf. welches Wort bzw. welcher Satz enthalten ist bzw. was durch die prosodische Eigenschaft ausgedrückt werden soll, was in den sprachbasierten Informationen hinterlegt wird. Somit können durch die Erkennungseinrichtung **13** weiterhin auch prosodische Eigenschaften aus der Sprachakustik ermittelt und somit Teil der sprachbasierten Informationen sein, die durch die Erkennungseinrichtung **13** generiert werden.

[0050] In der Aktivierungseinrichtung **15** wird nun basierend auf dem Sprachmodell **14** festgestellt, ob in den sprachbasierten Informationen ein Kommandowort und/oder eine prosodische Eigenschaft enthalten ist, was darauf hindeutet, dass die erste Sprachakustik **A** oder die darauf folgende zweite Sprachakustik **B** an das Sprachassistenzsystem **10** gerichtet ist. Die Aktivierungseinrichtung **15** kann basierend auf dem Sprachmodell **14** auch feststellen, ob anhand der prosodischen Eigenschaften der Sprachakustik **A**, die in den sprachbasierten Informationen hinterlegt sind, eine Sprachkommunikation mit dem Sprachassistenzsystem **10** aufgebaut werden soll. Dabei ist es denkbar, dass nur anhand von prosodischen Eigenschaften der ersten Sprachakustik **A** eine Aktivierungsabsicht festgestellt wird.

[0051] Wird durch die Aktivierungseinrichtung **15** erkannt, dass die Sprachakustik **A** dergestalt ist,

dass eine Sprachkommunikation mit dem Sprachassistenzsystem **10** aufgebaut werden soll somit die Sprachakustik **A** und/oder **B** an das Sprachassistenzsystem **10** gerichtet ist, so aktiviert die Aktivierungseinrichtung **15** das Sprachassistenzsystem **10**, wodurch die Weiterverarbeitung durch eine Assistenzeinrichtung **16** erfolgt.

[0052] Die Assistenzeinrichtung **16** wählt dabei in Abhängigkeit von sprachbasierten Informationen eine Unterstützungsfunktion aus und führt diese aus, wenn das Sprachassistenzsystem **10** zuvor aktiviert wurde. Wurde das Sprachassistenzsystem **10** zuvor durch die Aktivierungseinrichtung **15** nicht aktiviert oder bestand eine Unsicherheit darüber, ob die Sprachkommunikation an das Sprachassistenzsystem **10** gerichtet war, so passiert nichts. Dabei kann bei einer Unsicherheit über die Aktivierungsabsicht des Nutzers von System vorab eine Systemrückfrage generiert werden, um beim Nutzer nachzufragen, ob eine Aktivierung gewünscht ist oder nicht, d.h. ob zumindest die erste Sprachakustik eine Aktivierungsabsicht beinhaltet oder nicht.

[0053] Die zuvor empfangene erste Sprachakustik **A** wird dabei in einem Zwischenspeicher **17** für mindestens einen Zeitraum zwischen gespeichert, der geeignet ist, eine weitere Sprachakustik **B** zu empfangen, die kontextuellen zu der ersten Sprachakustik **A** gehört.

[0054] Wurde nach dem Empfang einer zweiten Sprachakustik **B**, die ebenfalls durch die Erkennungseinrichtung **13** und die Aktivierungseinrichtung **15** gelaufen ist, festgestellt, dass eine Unsicherheit und/oder Fehlerhaftigkeit darüber bestand, ob die erste Sprachakustik **A** an das Sprachassistenzsystem gerichtet war oder nicht, so wird mithilfe einer Adaptionseinrichtung **18** das Sprachmodell **14** basierend auf der zwischengespeicherten ersten Sprachakustik **A** angepasst und optimiert, um zukünftig derartige Unsicherheiten und/oder Fehlerhaftigkeiten zu vermeiden. Durch die Anpassung des Sprachmodells wird dabei nicht nur die Aktivierungseinrichtung **15** verbessert, sondern auch die Arbeitsweise der Erkennungseinrichtung **13**. Die Anpassung des Sprachmodells kann dabei derart erfolgen, dass Parameter eines gelernten Modells angepasst werden, oder dass Schwellwerte zur Entscheidung, ob die Sprachakustik an das Sprachassistenzsystem **10** gerichtet ist oder nicht, angepasst werden.

[0055] Bei der Anpassung des Sprachmodells **14** können dabei verschiedene Fälle unterschieden werden.

1. Fall: Fehlerhafte Aktivierung

[0056] Im ersten Fall, der hier betrachtet wird, ist die erste Sprachakustik **A** nicht an das Sprachassistenz-

system gerichtet und beispielsweise Teil einer zwischenmenschlichen Sprachkommunikation. Die Aktivierungseinrichtung **15** erkennt jedoch fälschlicherweise, dass die Sprachakustik **A** an das Sprachassistenzsystem gerichtet ist, wodurch eine Aktivierung des Sprachassistenzsystems **10** erfolgt.

[0057] In der zweiten Sprachakustik **B**, die auf die erste Sprachakustik **A** innerhalb eines gewissen Zeitraums folgt, hat der Nutzer sprachlich signalisiert, dass die vorherige erste Sprachakustik **A** nicht an das Sprachassistenzsystem **10** gerichtet war. Dies kann beispielsweise durch Schlüsselwörter wie „Stopp“ oder „Abbruch“ erfolgen, die in der zweiten Sprachakustik **B** enthalten sind.

[0058] Die Aktivierungseinrichtung **15** erkennt basierend auf der zweiten Sprachakustik **B** nun, dass die vorherige Aktivierung des Sprachassistenzsystems **10** basierend auf der ersten Sprachakustik **A** fehlerhaft war und deaktiviert Sprachassistenzsystem **10** entsprechend. Außerdem wird mithilfe der Adaptionseinrichtung **18** und der zuvor gespeicherten Sprachakustik **A** bzw. der aus der Sprachakustik **A** extrahierten sprachbasierten Informationen das Sprachmodell **14**, auf deren Basis sowohl die Erkennung der sprachbasierten Informationen als auch die Aktivierung des Sprachassistenzsystems **10** erfolgt, angepasst.

[0059] Das Erkennen, dass die Aktivierung basierend auf der ersten Sprachakustik **A** fehlerhaft war, kann auch dadurch erfolgen, dass innerhalb des zuvor erwähnten gewissen Zeitraumes gerade keine weitere zweite Sprachakustik **B** empfangen wurde, wodurch das Sprachassistenzsystem ebenfalls erkennt, dass die erste Sprachakustik an nicht einer Sprachassistenzsystem gerichtet war (warten auf einen „time out“).

2. Fall: Fehlerhafte Nicht-Aktivierung

[0060] Im zweiten Fall ist in der Sprachakustik **A** eine sprachliche oder prosodische Information enthalten, die signalisieren soll, dass die Sprachakustik **A** oder eine nachfolgende Sprachakustik **B** an das Sprachassistenzsystem **10** gerichtet ist. Dies kann beispielsweise dadurch erfolgen, dass der Sprachakustik **A** ein Kommandowort wie beispielsweise „Hallo“ oder dergleichen enthalten ist, was durch die Erkennungseinrichtung **13** identifiziert und durch die Aktivierungseinrichtung **15** dahingehend interpretiert werden soll, dass die Sprachakustik **A** oder die in einem gewissen Zeitraum nachfolgenden Sprachakustiken **B** an das Sprachassistenzsystem **10** gerichtet ist.

[0061] Im Fall der fehlerhaften Nicht-Aktivierung wird jedoch durch die Aktivierungseinrichtung **15** fälschlicherweise das Sprachassistenzsystem nicht aktiviert,

obwohl die Sprachakustik **A** als an das Sprachassistenzsystem **10** gerichtet vorgesehen ist. Mit anderen Worten, das Sprachassistenzsystem **10** wird nicht aktiviert, obwohl der zur 100 das Sprachassistenzsystem **10** anspricht und eine Sprachkommunikation mit diesem aufbauen wollte bzw. möchte.

[0062] Diese Nicht-Aktivierung basierend auf der ersten Sprachakustik **A** wird dann durch eine weitere nachfolgende Sprachakustik **B** erkannt. Dies kann dadurch erfolgen, dass nunmehr in der zweiten Sprachakustik **B** festgestellt wird, dass das Sprachassistenzsystem **10** aktiviert werden soll, wobei die zweite Sprachakustik **B** innerhalb eines gewissen Zeitraumes nach der ersten Sprachakustik **A** erkannt wurde. Durch den engen zeitlichen Zusammenhang der ersten Sprachakustik **A** und der zweiten Sprachakustik **B** wird ein kontextueller Zusammenhang hergestellt und angenommen, dass die erste Sprachakustik **A** bereits eine Aktivierung des Sprachassistenzsystems **10** beinhaltet hat.

[0063] Nachdem nun durch die nachfolgende zweite Sprachakustik **B** festgestellt wurde, dass bereits die erste Sprachakustik **A** zu einer Aktivierung des Sprachassistenzsystems **10** führen sollte, wird mithilfe der zwischengespeicherten ersten Sprachakustik Art nunmehr das Sprachmodell **14** angepasst.

3. Fall: Unsichere Aktivierung oder Nicht-Aktivierung

[0064] Im dritten Fall kann das Sprachassistenzsystem **10**, genauer gesagt die Aktivierungseinrichtung **15**, nicht mit hinreichender Sicherheit feststellen, ob basierend auf der ersten Sprachakustik **A** eine Aktivierung oder Nicht-Aktivierung erfolgen soll. Eine solche Unsicherheit entsteht immer dann, wenn die für eine hinreichende sichere Feststellung einer Aktivierung oder Nicht-Aktivierung vorgesehenen Schwellwerte nicht überschritten wurden und sich beispielsweise innerhalb eines und Sicherheitsbereiches befinden.

[0065] Wird diese Unsicherheit durch eine nachfolgende zweite Sprachakustik **B** ausgeräumt bzw. wird die vorherige unsichere Entscheidung über die Aktivierung oder Nicht-Aktivierung durch die zweite Sprachakustik **B** bestätigt, so wird ebenfalls das Sprachmodell **14** durch die Adaptionseinrichtung **18** entsprechend angepasst.

[0066] So ist es beispielsweise denkbar, dass in der ersten Sprachakustik **A** eine Information enthalten ist, die kennzeichnet, dass die Sprachakustik **A** bzw. die nachfolgende Sprachakustik **B** an das Sprachassistenzsystem **10** gerichtet ist. Allerdings kann nicht mit hinreichender Sicherheit darauf geschlossen werden. Trotz der verbleibenden Unsicherheit wird basierend auf den Aktivierungsinformationen darauf geschlossen, dass das Sprachassistenzsystem **10** zu

aktivieren ist. Gegebenenfalls kann das Sprachassistenzsystem **10** auch so eingerichtet sein, dass hier durch eine Sprachausgabe an den Nutzer **100** nachgefragt wird, ob die Sprachakustik **A** tatsächlich an das Sprachassistenzsystem **10** gerichtet war.

[0067] Wird durch eine im zeitlichen Zusammenhang stehende zweite Sprachakustik **B** nunmehr eindeutig festgestellt, dass die erste Sprachakustik **A** bereits an das Sprachassistenzsystem **10** gerichtet war, was beispielsweise dadurch erkannt werden kann, dass in der Sprachakustik **B** zweifelsfrei eine Anweisung an das Sprachassistenzsystem **10** enthalten ist, so wird das Sprachmodell **14** entsprechend zur Ausräumung von Unsicherheiten basierend auf der ersten Sprachakustik **A** oder den daraus extrahierten sprachbasierten Informationen angepasst.

[0068] Denkbar ist aber auch der umgekehrte Fall, dass eine Unsicherheit darüber besteht, ob in der ersten Sprachakustik **A** eine Nicht-Aktivierung enthalten ist, dass durch eine nachfolgende zweite Sprachakustik **B** ausgeräumt werden kann.

Bezugszeichenliste

10	Sprachassistenzsystem
11	Aufnahmeeinrichtung
12	Mikrofon
13	Erkennungseinrichtung
14	Sprachmodell
15	Aktivierungseinrichtung
16	Assistenzeinrichtung
17	digitaler Datenspeicher
18	Adaptionseinrichtung
100	Nutzer
A	erste Sprachakustik
B	zweite Sprachakustik

Patentansprüche

1. Sprachassistenzsystem (10) zur sprachbasierten Unterstützung eines Nutzers (100) mit
 - einer Aufnahmeeinrichtung (11) zum Aufnehmen von auditiv wahrnehmbaren Schallsignalen,
 - einer Erkennungseinrichtung (13), die eingerichtet ist, in den aufgenommenen, auditiv wahrnehmbaren Schallsignalen eine Sprachakustik zu erkennen und, sofern eine Sprachakustik erkannt wurde, sprachbasierte Informationen aus der erkannten Sprachakustik zu extrahieren,
 - einer Aktivierungseinrichtung (15), die eingerichtet ist, basierend auf einem Sprachmodell (14) in Abhängigkeit von den extrahierten sprachbasierten Informationen zu erkennen, ob die erkannte Sprachakus-

tik an das Sprachassistenzsystem (10) gerichtet ist oder nicht, und das Sprachassistenzsystem (10) zur sprachbasierten Unterstützung zu aktivieren, wenn die Sprachakustik als an das Sprachassistenzsystem (10) gerichtet erkannt wurde, und/oder nicht zu aktivieren, wenn die Sprachakustik nicht als an das Sprachassistenzsystem (10) gerichtet erkannt wurde, und

- einer Assistenzeinrichtung (16), die eingerichtet ist, in Abhängigkeit von extrahierten sprachbasierten Informationen eine Unterstützungsfunktion auszuwählen und zur sprachbasierten Unterstützung des Nutzers (100) auszuführen, wenn das Sprachassistenzsystem (10) zuvor aktiviert wurde,

- wobei das Sprachassistenzsystem (10) eingerichtet ist, in einem digitalen Datenspeicher (17) eine erste Sprachakustik (A) und/oder die aus der ersten Sprachakustik (A) extrahierten sprachbasierten Informationen zu hinterlegen, und

- das Sprachassistenzsystem (10) eine Adaptionseinrichtung (18) hat, die eingerichtet ist, das Sprachmodell (14) in Abhängigkeit von der zuvor in dem Datenspeicher hinterlegten ersten Sprachakustik (A) und/oder den aus der ersten Sprachakustik (A) extrahierten sprachbasierten Informationen anzupassen, wenn bei der ersten Sprachakustik (A) eine Unsicherheit und/oder Fehlerhaftigkeit darüber bestand, ob die erste Sprachakustik (A) an das Sprachassistenzsystem (10) gerichtet war oder nicht, und diese Unsicherheit und/oder Fehlerhaftigkeit durch eine auf die erste Sprachakustik (A) zeitlich nachfolgende erkannte oder durch erkanntes Ausbleiben einer entsprechenden zweiten Sprachakustik (B) ausgeräumt wird, wobei das Sprachassistenzsystem (10) eingerichtet ist, mittels der Adaptionseinrichtung (18) das Sprachmodell (14) in Abhängigkeit von der zuvor hinterlegten ersten Sprachakustik (A) und/oder den aus der ersten Sprachakustik (A) extrahierten sprachbasierten Informationen anzupassen, wenn innerhalb eines gewissen Zeitraumes nach der ersten Sprachakustik (A) die zweite Sprachakustik (B) erkannt oder innerhalb des gewissen Zeitraumes das Ausbleiben der zweiten Sprachakustik (B) erkannt wird, **dadurch gekennzeichnet**, dass das Sprachassistenzsystem (10) eingerichtet ist, den gewissen Zeitraum in Abhängigkeit von mindestens einem akustischen Qualitätskriterium der ersten Sprachakustik (A) festzulegen oder zu variieren.

2. Sprachassistenzsystem (10) nach Anspruch 1, **dadurch gekennzeichnet**, dass

- die Erkennungseinrichtung (13) eingerichtet ist, prosodische Eigenschaften der ersten Sprachakustik (A) als Teil der sprachbasierten Informationen aus der ersten Sprachakustik (A) zu extrahieren,

- die Aktivierungseinrichtung (15) eingerichtet ist, basierend auf dem Sprachmodell (14) in Abhängigkeit von den in den extrahierten sprachbasierten Informationen enthaltenen prosodischen Eigenschaften der ersten Sprachakustik (A) zu erkennen, ob die erkannt-

te Sprachakustik an das Sprachassistenzsystem (10) gerichtet ist oder nicht, und

- die Adaptionseinrichtung (18) eingerichtet ist, das Sprachmodell (14) hinsichtlich der prosodischen Erkennung in Abhängigkeit von der zuvor in dem Datenspeicher hinterlegten ersten Sprachakustik (A), den aus der ersten Sprachakustik (A) extrahierten sprachbasierten Informationen und/oder den in den extrahierten sprachbasierten Informationen enthaltenen prosodischen Eigenschaften der ersten Sprachakustik anzupassen.

3. Sprachassistenzsystem (10) nach einem der vorhergehenden Ansprüche,

dadurch gekennzeichnet, dass

- die Erkennungseinrichtung (13) eingerichtet ist, sprachliche Inhalte der ersten Sprachakustik (A) als Teil der sprachbasierten Informationen aus der ersten Sprachakustik (A) zu extrahieren,

- die Aktivierungseinrichtung (15) eingerichtet ist, basierend auf dem Sprachmodell (14) in Abhängigkeit von den in den extrahierten sprachbasierten Informationen enthaltenen sprachlichen Inhalten zu erkennen, ob die erkannte Sprachakustik an das Sprachassistenzsystem (10) gerichtet ist oder nicht, und

- die Adaptionseinrichtung (18) eingerichtet ist, das Sprachmodell (14) hinsichtlich der Erkennung von sprachlichen Inhalten in Abhängigkeit von der zuvor in dem Datenspeicher hinterlegten ersten Sprachakustik (A), den aus der ersten Sprachakustik (A) extrahierten sprachbasierten Informationen und/oder den in den extrahierten sprachbasierten Informationen enthaltenen sprachlichen Inhalten der ersten Sprachakustik (A) anzupassen.

4. Sprachassistenzsystem (10) nach einem der vorhergehenden Ansprüche,

dadurch gekennzeichnet, dass die Aktivierungseinrichtung (15) eingerichtet ist, bei einer Unsicherheit darüber, ob die erkannte erste Sprachakustik (A) als an das Sprachassistenzsystem (10) gerichtet oder nicht erkannt werden soll, eine optische, haptische, olfaktorische und/oder akustische System rückfrage zu generieren und mittels einer Ausgabereinrichtung des Sprachassistenzsystems (10) auszugeben.

5. Sprachassistenzsystem (10) nach einem der vorhergehenden Ansprüche,

dadurch gekennzeichnet, dass das Sprachmodell (14) ein auf einem maschinellen Lernsystem basierendes gelerntes Sprachmodell (14) ist.

6. Verfahren zur sprachbasierten Unterstützung eines Nutzers (100) mittels eines Sprachassistenzsystems (10), indem mittels einer Aufnahmeeinrichtung (11) auditiv wahrnehmbare Schallsignale aufgenommen werden, mittels einer Erkennungseinrichtung (13) in den aufgenommenen, auditiv wahrnehmbaren Schallsignalen eine Sprachakustik erkannt und, sofern eine Sprachakustik erkannt wurde, sprachba-

sierte Informationen aus der erkannten Sprachakustik extrahiert werden, mittels einer Aktivierungseinrichtung (15) basierend auf einem Sprachmodell (14) in Abhängigkeit von den extrahierten sprachbasierten Informationen erkannt wird, ob die erkannte Sprachakustik an das Sprachassistenzsystem (10) gerichtet ist oder nicht, wobei das Sprachassistenzsystem (10) zur sprachbasierten Unterstützung aktiviert wird, wenn die Sprachakustik als an das Sprachassistenzsystem (10) gerichtet erkannt wurde, und/oder nicht aktiviert wird, wenn die Sprachakustik nicht als an das Sprachassistenzsystem (10) gerichtet erkannt wurde, und indem mittels einer Assistenzeinrichtung (16) in Abhängigkeit von extrahierten sprachbasierten Informationen eine Unterstützungsfunktion ausgewählt und zur sprachbasierten Unterstützung des Nutzers (100) ausgeführt wird, wenn das Sprachassistenzsystem (10) zuvor aktiviert wurde, wobei in einem digitalen Datenspeicher (17) eine erste Sprachakustik (A) und/oder die aus der ersten Sprachakustik (A) extrahierten sprachbasierten Informationen hinterlegt werden, wobei mittels einer Adaptionseinrichtung (18) das Sprachmodell (14) in Abhängigkeit von der zuvor in dem Datenspeicher hinterlegten ersten Sprachakustik (A) und/oder den aus der ersten Sprachakustik (A) extrahierten sprachbasierten Informationen angepasst wird, wenn bei der ersten Sprachakustik (A)

eine Unsicherheit und/oder Fehlerhaftigkeit darüber bestand, ob die erste Sprachakustik (A) an das Sprachassistenzsystem (10) gerichtet war oder nicht, und diese Unsicherheit und/oder Fehlerhaftigkeit durch eine auf die erste Sprachakustik (A) zeitlich nachfolgende erkannte zweite Sprachakustik (B) oder durch erkanntes Ausbleiben einer entsprechenden zweiten Sprachakustik (B) ausgeräumt wird, wobei mittels der Adaptionseinrichtung (18) das Sprachmodell (14) in Abhängigkeit von der zuvor hinterlegten ersten Sprachakustik (A) und/oder den aus der ersten Sprachakustik (A) extrahierten sprachbasierten Informationen angepasst wird, wenn die zweite Sprachakustik (B) innerhalb eines gewissen Zeitraumes nach der ersten Sprachakustik (A) erkannt oder innerhalb des gewissen Zeitraumes das Ausbleiben der zweiten Sprachakustik (B) erkannt wird, **dadurch gekennzeichnet**, dass der gewisse Zeitraum in Abhängigkeit von mindestens einem akustischen Qualitätskriterium der ersten Sprachakustik (A) festgelegt oder variiert wird.

7. Verfahren nach Anspruch 6, **dadurch gekennzeichnet**, dass

- mittels der Erkennungseinrichtung (13) prosodische Eigenschaften der ersten Sprachakustik (A) als Teil der sprachbasierten Informationen aus der ersten Sprachakustik (A) extrahiert werden,
- mittels der Aktivierungseinrichtung (15) basierend auf dem Sprachmodell (14) in Abhängigkeit von den in den extrahierten sprachbasierten Informationen enthaltenen prosodischen Eigenschaften der ersten

Sprachakustik (A) erkannt wird, ob die erkannte erste Sprachakustik (A) an das Sprachassistenzsystem (10) gerichtet ist oder nicht, und

- mittels der Adaptionseinrichtung (18) das Sprachmodell (14) hinsichtlich der prosodischen Erkennung in Abhängigkeit von der zuvor in dem Datenspeicher hinterlegten ersten Sprachakustik (A), den aus der ersten Sprachakustik (A) extrahierten sprachbasierten Informationen und/oder den in den extrahierten sprachbasierten Informationen enthaltenen prosodischen Eigenschaften der ersten Sprachakustik (A) angepasst wird.

8. Verfahren nach einem der Ansprüche 6 bis 7, **dadurch gekennzeichnet**, dass

- mittels der Erkennungseinrichtung (13) sprachliche Inhalte der ersten Sprachakustik (A) als Teil der sprachbasierten Informationen aus der ersten Sprachakustik (A) extrahiert werden,

- mittels der Aktivierungseinrichtung (15) basierend auf dem Sprachmodell (14) in Abhängigkeit von den in den extrahierten sprachbasierten Informationen enthaltenen sprachlichen Inhalten der ersten Sprachakustik (A) erkannt wird, ob die erkannte erste Sprachakustik (A) an das Sprachassistenzsystem (10) gerichtet ist oder nicht, und

- mittels der Adaptionseinrichtung (18) das Sprachmodell (14) hinsichtlich der Erkennung von sprachlichen Inhalten in Abhängigkeit von der zuvor in dem Datenspeicher hinterlegten ersten Sprachakustik (A), den aus der ersten Sprachakustik (A) extrahierten sprachbasierten Informationen und/oder den in den extrahierten sprachbasierten Informationen enthaltenen sprachlichen Inhalten der ersten Sprachakustik (A) angepasst wird.

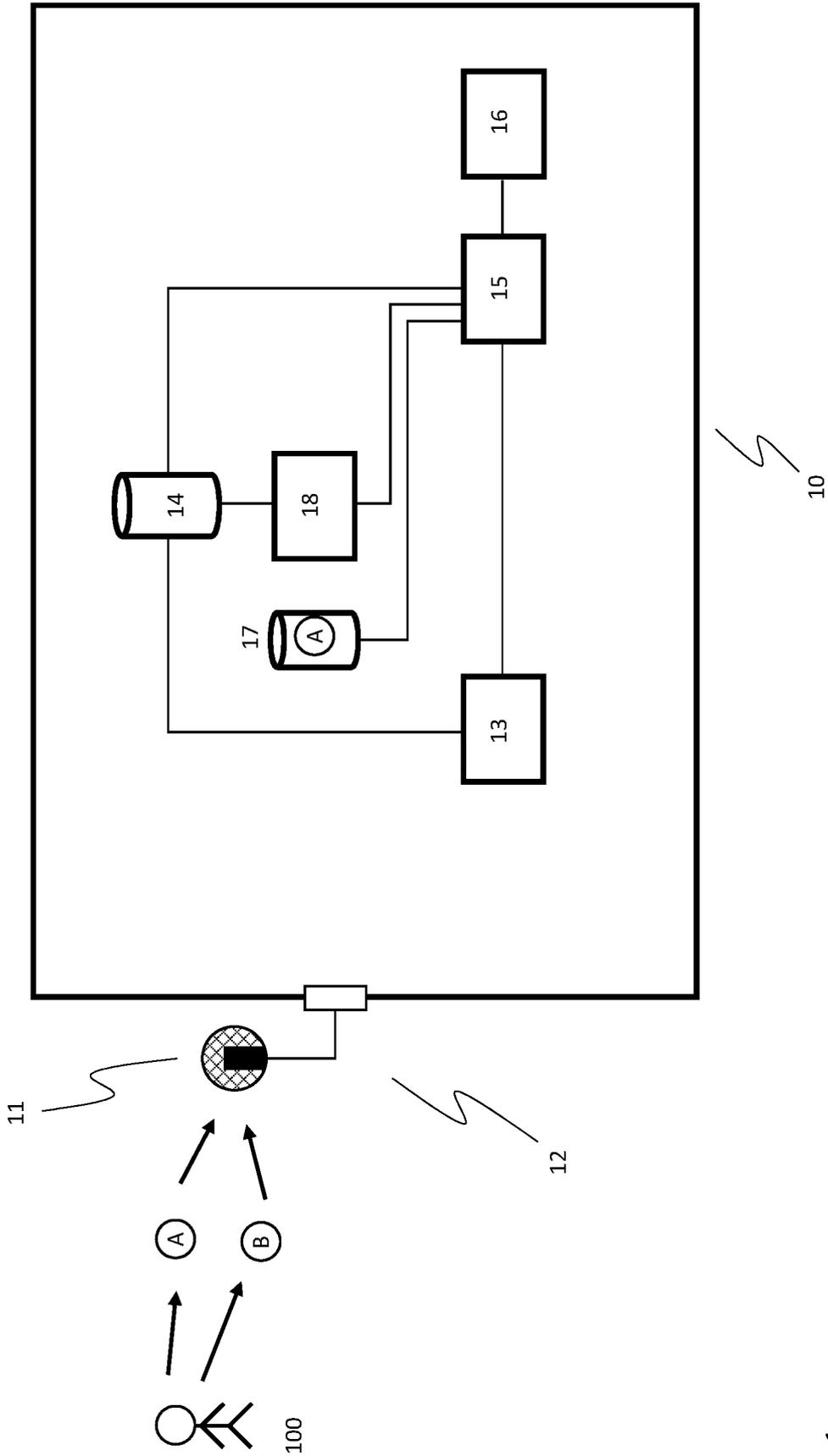
9. Verfahren nach einem der Ansprüche 6 bis 8, **dadurch gekennzeichnet**, dass mittels der Aktivierungseinrichtung (15) bei einer Unsicherheit darüber,

ob die erkannte Sprachakustik als an das Sprachassistenzsystem (10) gerichtet oder nicht erkannt werden soll, eine Sprachausgabe in Form einer Sprachakustik generiert und mittels einer Ausgabereinrichtung des Sprachassistenzsystems (10) ausgegeben wird, wobei die Sprachausgabe eine Nachfrage in Bezug auf die erkannte Sprachakustik an den Nutzer (100) enthält.

10. Computerprogramm mit Programmcodemitteln, eingerichtet zur Durchführung des Verfahrens nach einem der Ansprüche 6 bis 9, wenn das Computerprogramm auf einer Datenverarbeitungsanlage ausgeführt wird.

Es folgt eine Seite Zeichnungen

Anhängende Zeichnungen



Figur 1