

# Ausarbeitung zur Allgemeinen Arbeitsweise von Voice Assistants

## Komponentenbasierte Softwareentwicklung

Fachhochschule Bielefeld  
Campus Minden  
Studiengang Informatik

---

### Beteiligte Personen:

Name	Email
Jan-Hendrik Sünderkamp	jan-hendrik.suenderkamp@fh-bielefeld.de
Daniel Nagel	daniel.nagel@fh-bielefeld.de
Philipp Clausing	philipp.clausing@fh-bielefeld.de
Tobias Muth	tobias.muth@fh-bielefeld.de
Devin-Alexander Meier	devin-alexander.meier@fh-bielefeld.de
Alexander Jaksties	alexander.jaksties@fh-bielefeld.de
Peter Dick	peter.dick@fh-bielefeld.de

---

20. Dezember 2017

# Inhaltsverzeichnis

1	Was ist ein Voice Assistant?	4
1.1	Digitaler Assistent . . . . .	4
1.2	Spracherkennung . . . . .	4
1.3	Computerlinguistik . . . . .	4
1.4	Sprachsynthese . . . . .	5
1.5	Historie . . . . .	6
2	Generelle Funktionsweise eines Voice Assistant	7
2.1	Voice Assistant . . . . .	7
2.2	Webservice . . . . .	8
2.3	Was ist ein Webservice? . . . . .	8
2.4	Welchen Webservice nutzt Amazons Alexa? . . . . .	8
	Lambda . . . . .	8
	Alexa Voice Service (AVS) . . . . .	8
2.5	Welchen Webservice nutzt Google Assistant? . . . . .	9
3	Alexa	9
3.1	Funktionen/Funktionsweise . . . . .	9
3.2	Hardware . . . . .	10
	Echo Dot (v2) . . . . .	11
	Echo . . . . .	11
	Echo Plus . . . . .	12
3.3	Erweiterbarkeit - Skills . . . . .	12
4	Google Home	12
4.1	Funktionen/Funktionsweise . . . . .	12
4.2	Hardware . . . . .	13
	Google Home . . . . .	14
	Google Home Mini . . . . .	14
	Google Home Max . . . . .	15
4.3	Erweiterbarkeit . . . . .	15
4.4	Actions . . . . .	15

## Vorwort

In dieser Arbeit wird die generelle Funktionsweise von Voice Assistants näher beleuchtet. Dazu wird zunächst geklärt, was unter einem Voice Assistant verstanden wird. Anschließend wird anhand eines vereinfachten Modells, der verallgemeinerte Ablauf eines Befehlsaufrufs dargestellt. An dieser Stelle wird zudem auf die einzelnen Bestandteile des genannten Modells näher eingegangen. Abschließend wird mithilfe von zwei Fallstudien der gesamte Ablauf konkret erläutert.

# 1 Was ist ein Voice Assistant?

Voice Assistants (VAs) sind digitale Assistenten, welche Spracherkennung (SE), Computerlinguistik (CL) und Sprachsynthese (TTS) nutzen, um Nutzern auf Smartphones und Spracherkennungsgeräten verschiedene Dienste zur Verfügung zu stellen.

Um zu klären, was ein Voice Assistant ist, wird auf die obig genannten Begrifflichkeiten im folgenden näher eingegangen. Abschließend wird Anhand von historischen Beispielen die Entwicklung von Voice Assistants dargestellt.

## 1.1 Digitaler Assistent

Digitale Assistenten - auch Intelligente Persönliche Assistenten (IPA) genannt - sind Softwarelösungen mit der Fähigkeit der Spracherkennung und -analyse. Diese Fähigkeit erlaubt ihnen Anwender bei der Suche nach Informationen zu unterstützen und einfache Aufgaben für diese zu übernehmen. Ziel von digitalen Assistenten ist es Nutzern zu ermöglichen Suchanfragen und Kommandos einfacher zu formulieren. Hierzu sind IPAs in der Lage in einen Dialog mit dem Anwender zu treten. Dabei orientieren sich Fragen und Antworten immer mehr an einem normalen Gesprächsfluss.

## 1.2 Spracherkennung

Spracherkennung ist ein Teilgebiet der angewandten Informatik und Computerlinguistik. Das Ziel der Spracherkennung ist es Automaten die gesprochene Sprache zugänglich zu machen. Insbesondere helfen die entwickelten Verfahren Computern bei der automatischen Datenerfassung. Momentan wird zwischen zwei Arten der Spracherkennung unterschieden: Der sprecherunabhängigen und der sprecherabhängigen Spracherkennung. Während bei der sprecherunabhängigen Erkennung jeder Nutzer direkt erkannt wird, ist bei der sprecherabhängigen Erkennung ein kurzes Training des Systems nötig, um die Nutzererfahrung zu optimieren. Während erstere nur über einen Wortschatz von einigen Tausend Wörtern verfügen, erreichen letztere ein Vokabular von über 300.000 Worten.

## 1.3 Computerlinguistik

Computerlinguistik untersucht, wie natürliche Sprache von Computern algorithmisch Verarbeitet werden kann. Folgendes Zitat bietet eine gute Definition für den Begriff Computerlinguistik:

Computerlinguistik erforscht die maschinelle Verarbeitung natürlicher Sprachen. Sie erarbeitet die theoretischen Grundlagen der Darstellung, Erkennung und Erzeugung gesprochener und geschriebener Sprache durch Maschinen.  
- Universität München

Die Erfassung von Sprache ist Computern auf zwei Arten möglich. Sprache wird entweder als Schallinformation - akustisch - oder in Buchstabenketten - textuell - registriert.

Häufig wird bei der Verarbeitung das Saarbrücker Pipelinemodell genutzt, auf welches nun weiter eingegangen wird.

1. Spracherkennung  
Umwandlung der Schallinformation in Text, falls nötig.
2. Tokenisierung  
Segmentierung der Buchstabenkette in Wörter und Sätze.
3. Morphologische Analyse  
Extraktion grammatischer Informationen und Rückführung der Wörter auf Grundform.
4. Syntaktische Analyse  
Analyse der einzelnen Wörter eines Satzes auf ihre strukturelle Funktion.
5. Semantische Analyse  
Zuordnung der Bedeutung einzelner Sätze. Kann viele Einzelschritte enthalten.
6. Dialog- und Diskursanalyse  
In Beziehung setzen aufeinander folgender Sätze.

Zu beachten ist, dass durch maschinelles Lernen einige dieser Schritte ignoriert werden können. Dies lässt sich darauf zurückführen, dass auf jeder Analyseebene statische Regelmäßigkeiten existieren, welche zur Modellierung zu Hilfe genommen werden können. Viele Modelle maschineller Übersetzung beschränken sich darauf Korrespondenzmuster auf Wortebene auszunutzen ohne Syntax und Semantik stark zu beachten.

Momentan existieren drei Hauptprobleme bei der Sprachverarbeitung:

1. Manche Sätze lassen sich auf mehrere Weisen deuten. Dementsprechend kann die Auflösung syntaktischer Mehrdeutigen zusätzliche semantische Informationen erfordern, allerdings mindestens ein statisches Vorwissen über gemeinsames Auftreten von Wörtern.
2. Da die gleiche Wortform über unterschiedliche Bedeutungen verfügen kann - je nach Kontext -, gestaltet sich die Bestimmung der Semantik als schwierig.
3. Einige Sätze sind nicht wörtlich gemeint. Die Absicht hinter dem gesagten muss erkannt werden, um eine korrekte Interpretation zu erhalten.

## 1.4 Sprachsynthese

Sprachsynthese bezeichnet die künstliche Erzeugung der menschlichen Sprechstimme. Ein Text-to-Speech (TTS) System ermöglicht dies.

Bei der Erzeugung von Sprache unterscheidet man zwischen zwei verschiedenen Methoden. Einerseits kann durch den Zugriff auf Sprachaufnahmen eine Stimme modelliert werden. Diesen Vorgang bezeichnet man als Signalmodellierung. Andererseits kann die Stimme auch vollständig digital erzeugt werden. Dies wird als Formatsynthese bezeichnet.

Das größte Hindernis hierbei ist in beiden Fällen die Erzeugung einer Stimme mit natürlicher Sprachmelodie.

## 1.5 Historie

Den Grundstein für Voice Assistants legte Bell Laboratories 1952 mit dem ersten SE-Gerät namens AUDREY. AUDREY erkannte einzelne gesprochene Ziffern. Allerdings musste zwischen den gesprochenen Ziffern eine deutliche und klare Pause existieren. Da die Technologie noch nicht sehr ausgereift war und umständlich zu benutzen war, war sie kein kommerzieller Erfolg.

Im Jahr 1961 brachte IBM mit der IBM-Shoebbox das erste kommerzielle SE-Gerät auf den Markt. Dieses erkannte 16 unterschiedliche Wörter, welche die Ziffern von 0-9 sowie die Befehle minus, plus, subtotal, total, false und of umfassten. Dadurch war das Gerät in der Lage einfache mathematische Operationen durchzuführen.

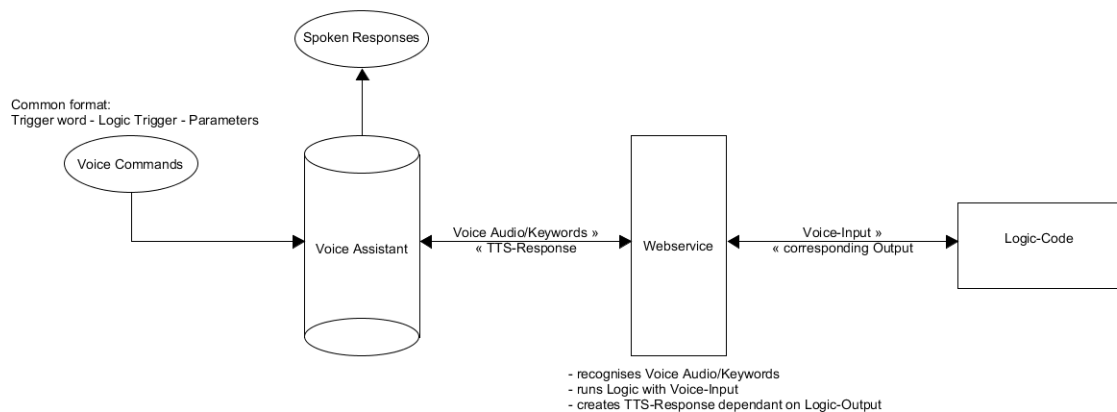
Die DARPA - Defense Advanced Research Projects Agency - entwickelte 1970 mit HARPY ein SE-Gerät welches bis zu 1000 Wörter erfassen konnte. 10 Jahre später erkannte HARPY durch Nutzung des Hidden Markov Models sogar ganze Sätze. Das Hidden Markov Model wird genutzt um die Wahrscheinlichkeit, dass ein bestimmtes Wort auf anderes folgt, zu bestimmen.

Der kommerzielle Durchbruch von VAs erfolgte am 4. Oktober 2011 mit der Veröffentlichung des iPhone 4S. Dieses verfügte über einen Voice Assistant namens Siri, welcher erstmals die Datenübertragung an einen Server zur Verarbeitung der Eingaben nutzte. Dadurch konnte der Funktionsumfang drastisch erhöht werden.

Heutzutage erfreuen sich eigenständige Voice Assistants immer größerer Beliebtheit. Nicht nur lassen sich diese Geräte überall im Haus aufstellen, sie verfügen auch über einen extrem erweiterbaren Funktionsumfang. Besonders bekannt sind die beiden Dienste Alexa des Unternehmens Amazon.com, Inc., welcher am 23. Juni 2015 erschienen ist, und Google Home des Unternehmens Alphabet Inc., welcher am 4. November 2016 herausgebracht wurde.

## 2 Generelle Funktionsweise eines Voice Assistant

In diesem Kapitel soll die generelle Funktionsweise eines Voice Assistants anhand eines Schaubilds dargestellt werden. Anschließend wird auf die einzelnen Bestandteile kurz eingegangen.



**Abbildung 1:** Komponentendiagramm der generellen Funktionsweise eines Voice Assistants

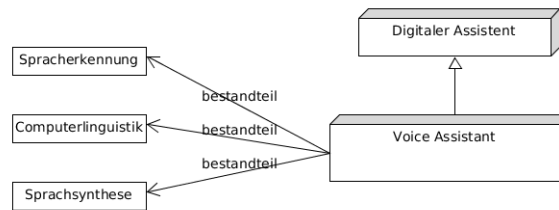
Wie in Abbildung 1 beschrieben ist, muss der VA zunächst durch ein Trigger word aktiviert werden. Darauf folgt die so genannte Utterance. Diese setzt sich aus Logic Trigger und gegebenenfalls notwendigen Parametern zusammen. Diese wird mithilfe von SE auf dem Gerät erkannt und digitalisiert. Anschließend sendet der VA die Utterance an einen Webservice, der die weitere Verarbeitung übernimmt. So können die Voice Assistants schlank gehalten werden, da ein Großteil der Rechenarbeit auf dem Webservice erfolgt.

Der Webservice erkennt nun anhand des Logic Triggers den konkreten Anwendungsfall und ermittelt, falls nötig, die gegebenen Parameter. Der konkrete Anwendungsfall wird als Intent bezeichnet. Die Anfrage wird daraufhin vom Webservice bearbeitet. Hierzu wird der auf einem Webserver abgelegten Logic-Code verwendet. Für jeden definierten Intent enthält der Logic-Code eine passende Funktion. Dies kann von einer einfachen Rechnung, über Suchanfragen im Internet bis hin zur Wiedergabe von akustischen Medien auf dem VA reichen.

Der Logic-Code generiert eine zur Anfrage passenden Antwort, welche an den Webservice weitergereicht wird. Der Webservice erzeugt daraus eine TTS-Response, welche an den Voice Assistant gesendet wird. Letzterer ist nun in der Lage dem Anwender eine gesprochene Antwort auf seine Anfrage zu geben.

### 2.1 Voice Assistant

Im folgenden Schaubild wird veranschaulicht, was in Kapitel 1 textuell beschrieben wird.



**Abbildung 2:** Konzeptionelles Modell eines Voice Assistant

## 2.2 Webservice

### 2.3 Was ist ein Webservice?

Ein Webservice ermöglicht eine Maschine-zu-Maschine Kommunikation mittels HTTP oder HTTPS über das Internet. Auf einem entfernten Computer werden Funktionen ausgeführt. Der Webservice wird mittels URI (Uniform Resource Identifier) eindeutig identifiziert. Die Kommunikation kann HTTP, XML oder JSON basiert sein.

### 2.4 Welchen Webservice nutzt Amazons Alexa?

Alexa arbeitet mit dem Amazon Web Service (AWS). AWS ist ein Tochterunternehmen von Amazon. AWS ist ein Cloud Computing Anbieter. Ein Nutzer von AWS bekommt einen virtuellen Server, welcher auf einem echten Server innerhalb einer Serverfarm läuft. Für diesen virtuellen Server bietet AWS diverse Dienste an. Die für den Voice Assistant benötigten sind dabei folgende:

#### Lambda

Der AWS Dienst Lambda führt beliebigen Code auf dem virtuellen Server aus. Meistens wird Lambda dazu verwendet eine Anfrage zu verarbeiten. In dieser Anfrage wird eine Lambda Funktion aufgerufen, ein von einem Programmierer entwickeltes Programm. Die Lambda Funktion generiert eine Antwort, meist im JSON-Format. Zurzeit unterstützte Programmiersprachen mit den entsprechenden Bibliotheken sind: Java, Node.js, Python und C#. In Zusammenarbeit mit einem Voice Assistenten wird dem AWS Dienst Lambda eine Anfrage im JSON-Format übermittelt, in dieser steht die Lambda Funktion welche ausgeführt werden soll mit eventuellen Parametern. Der hinterlegte Programmcode (Logic Code) wird ausgeführt. Die Antwort im JSON-Format sendet AWS zurück an den Voice Assistant. Dieser wandelt die Antwort mittels TTS in Sprache um und antwortet dem Nutzer auf seine Anfrage.

#### Alexa Voice Service (AVS)

AVS versteht natürliche Sprache und kann diese in Text umwandeln. Umgekehrt kann AVS Text wieder in Sprache umwandeln. Die aktuellste Version des AVS wird von Alexa auf das Voice Assistant Gerät heruntergeladen. Der Voice Assistant wird dann lokal die Sprache in Text umwandeln und an den AWS senden. Da AVS ein Teil der Amazon Web Services ist wird die Spracherkennung mittels Machine Learning stetig verbessert,



d.h. viele Geräte senden erfolgreiche oder fehlgeschlagene Anfragen an AWS, um die Spracherkennung zu verbessern. Durch das Alexa Skill Kit kann ein Entwickler dem AVS zeigen welche Reaktion der Entwickler bei einem bestimmten Stichwort oder Satz erwartet. Dabei muss nahezu jede mögliche Variation des Satzes bedacht sein, damit entsprechend darauf reagiert werden kann.

## 2.5 Welchen Webservice nutzt Google Assistant?

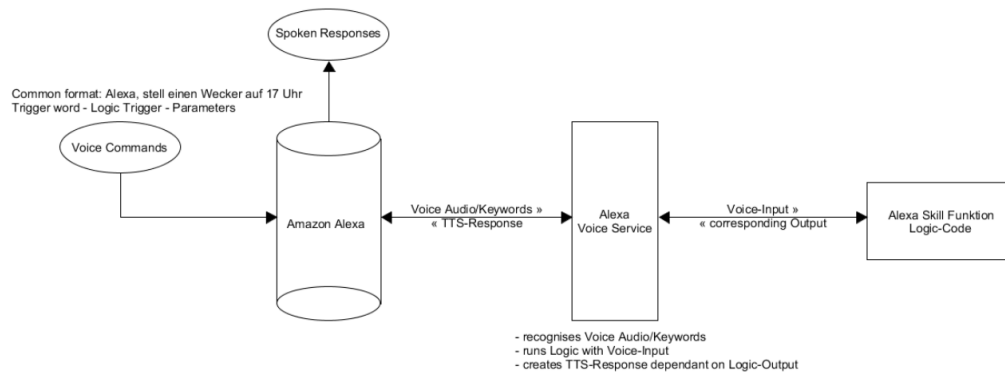
Die Google Home Geräte integrieren den Google Assistant. Dieser kann mittels Dialogflow (früher API.ai) erweitert werden, wie vorher bereits mit mobilen Geräten oder dem Google Chrome Browser möglich. Dialogflow wird durch eine Node.js Node auf einem eigenen Server oder z.B. einem Server des Google Cloud Projects ausgeführt. Der Entwickler hat einen Agent (Anwendung) entwickelt, welcher auf die Anfrage des Nutzers reagiert.

## 3 Alexa

Alexa ist der Name eines von der Firma Amazon entwickelten Voice Assistenten. Die verschiedenen Gerätevarianten sind jeweils mit mindestens einem Lautsprecher und sieben Mikrofonen ausgestattet. In Deutschland werden diese Geräte seit dem 26. Oktober 2016 angeboten. Durch die Mikrofone nimmt das Gerät die Geräusche in der Umgebung auf und versucht daraus die menschliche Sprache herauszufiltern. Dadurch bieten die Geräte die Funktionen eines Digitalen-Assistenten mittels Spracherkennung an, auch genannt Voice Assistant. Alexa reagiert nicht dauerhaft auf gesagte Textpassagen sondern erst sobald ein aus einer beschränkten Auswahl gewähltes Wort genannt wird. Hier bietet Amazons Alexa folgende Aktivierungswörter an: Alexa, Echo, Amazon oder Computer. Sollte eines der Aktivierungswörter genannt werden beginnt ein LED-Indikator blau zu leuchten. Alles nun gesagte wird zur Erkennung an die Alexa Voice Services gesendet, wodurch auch sensible Daten über das Internet übertragen werden. Lediglich das erste Wort wird lokal verarbeitet, auch bietet das Gerät einen fest verlöteten Schalter um das Mikrofon zu deaktivieren. Dies wird über einen roten LED-Indikator angezeigt.

### 3.1 Funktionen/Funktionsweise

Mit Beginn der Einrichtung, beherrscht der VA bereits grundlegende Dinge wie: Wecker, Erinnerungen, Wetter abfrage, Internetsuche, Musik abspielen etc. . Diese sind immer gleich aufgebaut. Mit dem mitteilen des "WakeWords"(z.B. Alexa oder Computer) wird der VA aktiviert und dann kann mittels des anschließend gesprochenem Textes die entsprechende Aktion ausgeführt werden. Mit Alexa, stell meinen Wecker auf 6:30 Uhr spricht man den Wecker-Skill an. Es gibt mittlerweile über 10.000 Skills welche auch von vielen Drittanbietern entwickelt werden. Außerdem kann man mit Alexa auch über den Online-Store von Amazon einkaufen.



**Abbildung 3:** Amazon Voice-Assistant Architektur

Oben wird die Struktur des VA als Diagramm gezeigt. Hier sieht man wie das Endgerät mit dem Alexa Voice Service kommuniziert und dieser dann alles entsprechend zusammenfügt und an die eigentliche Funktion überträgt. Diese verarbeitet dann die Abfrage und sendet eine Antwort zurück welche mittels TTS zu einer Stimme formatiert wird.

### 3.2 Hardware

Nachfolgend wird eine kleine Auswahl an Alexa-Geräten gezeigt. Insgesamt gibt es allerdings deutlich mehr.

Amazon Alexa Geräte:

- Echo Spot (Alexa Lautsprecher mit kleinem Display)
- Echo Buttons (SSmart Home"Buttons)
- Echo Connect (Verbindet Alexa mit dem Telefon für Anrufe)
- Echo Show (7"Display - Anzeige von Videos und anderen Infos)
- Echo, Echo Dot und Echo Plus
- Alexa Sprachferbedienung (FireTV)

## Echo Dot (v2)

- 59,99€
- 7 Mikrofone - 1 Lautsprecher
- Hardwaretasten für die Lautstärke
- Arbeitet als Smart Home Hub
- ein Mute-Button für das Mikrofon auf dem Gerät



Abbildung 4: Echo Dot

## Echo

- 99,99€
- Ähnlich wie Echo Dot
- Größere und bessere Lautsprecher
- 63mm-Woofer
- 16mm-Hochtonlautsprecher

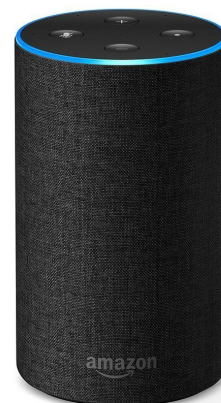


Abbildung 5: Echo

## Echo Plus

- 149,99€
- Ähnlich wie Echo Dot und Echo
- integrierten ZigBee-Smart Home Hub
- Größere und bessere Lautsprecher
- 63mm-Woofer
- 20mm-Hochtonlautsprecher



**Abbildung 6:** Echo Plus

### 3.3 Erweiterbarkeit - Skills

Der Alexa Skills Service bietet die Möglichkeit eigene Skills zu entwickeln und über den Store den anderen Nutzer bereitzustellen. Die Skills bedienen sich dabei einer Skills API, für welche es mittlerweile eine breite Unterstützung für viele verschiedene Programmier- und Skriptsprachen gibt. Jeder Skill verfolgt dabei ein festes Muster welches sich aus einem Dialogmuster, also einer Datei welche aussagt wie mit dem Nutzer kommuniziert wird und einer Funktion zusammensetzt. Diese Funktion wird in der Regel durch einen Webservice bereitgestellt, hierfür bietet sich der von Amazon bereitgestellte Service AWS Lambda an, da dieser bereits eine gute Integration in die Entwicklung darstellt.

Ein Skill besteht aus Utterances, Intents und einer Funktion. Utterances (auch Aussagen) werden einem Intent zugeordnet, der AVS nutzt diese um dem gesagten ein Intent zuzuweisen damit der korrekt Code ausgeführt wird. Ein paar Beispiele:

- stell den Wecker
- wie spät ist es ?
- wann klingelt der Wecker ?

Diese zugeordneten Intents werden dann über einen eindeutigen Namen mit dem Code verknüpft. Im Code kann dann die entsprechende Antwort in Form eines Strings zurückgegeben werden.

## 4 Google Home

Google Home ist ein von Google entwickelter Smart-Home-Lautsprecher, der mit integriertem Mikrofon ausgestattet ist und über einen persönlichen Voice-Assistenten verfügt.

Das Gerät wurde am 4. November 2016 in den Vereinigten Staaten veröffentlicht. In Deutschland erfolgte die Markteinführung am 8. August 2017 mit einer deutscher Stimme für den Voice-Assistenten. Es gibt momentan drei verschiedene Google Home Geräte (Home, Home Mini, Home Max), die sich in erster Linie in ihrer Größe und der Qualität ihrer Lautsprecher unterscheiden.

#### 4.1 Funktionen/Funktionsweise

Google Home kommt mit einer Reihe von Funktionen sowohl von Google als auch von Drittanbietern, die es dem Benutzer beispielsweise erlauben Musik zu hören, die Wiedergabe von Videos oder Fotos auf Smart-Home Geräten zu steuern oder aktuelle Nachrichten zu erhalten. Mehrere Google Home Geräte können in unterschiedlichen Räumen zum synchronisierten Abspielen von Musik verwendet werden. Ein Update im April 2017 ermöglicht es den Geräten bis zu sechs Benutzer anhand ihrer Stimme zu erkennen und individuell zu reagieren.

Um dem Gerät Befehle zu erteilen muss es zunächst mit den Weckworten *"Hey Google"* oder *"Ok Google"* aufgeweckt werden. Um Drittanbietersoftware über die sogenannten Actions on Google zu nutzen ist keine Installation notwendig.

Der integrierte Google Assistant, den es auch außerhalb von Google Home Geräten (z.B. auf Android Smartphones) gibt, bietet dem Benutzer sein eigenes persönliches Google um ihm beim Suchen, Organisieren oder Interagieren zu helfen.

Actions on Google erweitert den Google Assistant um sogenannte Actions, die es Entwicklern ermöglicht eigene Funktionen hinzuzufügen. Anders als bei traditionellen mobilen oder Desktopapplikationen interagiert der Nutzer mit den Apps für den Assistenten über verbale Konversation (natürlich klingender Austausch zwischen Nutzer und App).

Ein typischer Ablauf eines solchen Austausches sieht wie folgt aus. Wenn ein Nutzer den Assistenten auffordert eine bestimmte Aktion auszuführen, sendet der Assistent eine Anfrage an Actions on Google um eine passende App zu finden, die der Intention des Nutzers entspricht. Diese wird dann an den Assistenten weitergereicht und der Nutzer wird gefragt, ob er diese App ausführen möchte. Wenn der Nutzer dies bestätigt, wird die App vorgestellt und die Konversation an diese weitergereicht (siehe Abbildung 1).

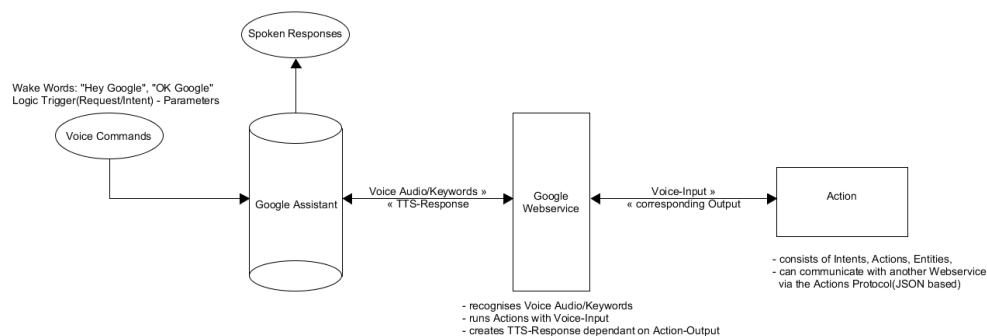


Abbildung 7: Google Voice-Assistant Architektur

Der Google Assistant hat Zugriff auf Google Maps und kann daher Fragen zu Position örtlicher Geschäfte, sowie deren Öffnungszeiten beantworten. Generell steht dem Assistenten die Google Suche zur Verfügung um enzyklopedisches Wissen oder einfache Übersetzungen wiederzugeben.

## 4.2 Hardware

Es folgt ein kurzer Überblick über die verschiedenen Google Home Geräte, die momentan im Handel erhältlich sind.

### Google Home

- 149€
- farbige Status-LEDs an der Oberfläche
- kapazitiver Berührungssensor um Musik zu starten und stoppen oder die Lautstärke anzupassen
- ein Mute-Button für das Mikrofon an der Rückseite



Abbildung 8: Google Home

### Google Home Mini

- 59€
- gleiche Funktionalität
- Berührungssensor zum Anpassen der Lautstärke
- weiße Statuslichter scheinen durch den Stoff auf der Oberseite
- ein Mute-Schalter für das Mikrofon an der Rückseite



Abbildung 9: Google Home Mini

## Google Home Max

- 400\$
- Stereo Lautsprecher (einschließlich zweier Hochtöner und Subwoofer)
- magnetisch befestigbarer Ständer für vertikale Ausrichtung
- beinhaltet *Smart Sound*, ein Soundsystem das *machine learning* nutzt um sich automatisch an die Umgebung anzupassen (Position und Geräuschquellen)



Abbildung 10: Google Home Max

### 4.3 Erweiterbarkeit

Um eine App für den Google Assistant zu entwickeln bietet Google verschiedene Optionen. Zum einen gibt es Templates, die einem einen Großteil der Programmierarbeit abnehmen und bereits vorgefertigte Konversationen enthalten. Allerdings gibt es bisher nicht viele Templates aus denen man wählen kann.

Für einfache Apps mit wenig Inputmöglichkeiten bietet sich die *Actions SDK* an. Sie bietet keine NLU (natural language understanding), es kann allerdings eine eigene NLU eingesetzt werden. Weiterhin müssen *action packages* mit einem Text Editor geschrieben und per Kommando-Zeilen Tool an das jeweilige Google Developer Projekt gesendet werden.

*Dialogflow* ist eine Web IDE, die die Funktionalität der *Actions SDK* einschließt. *Dialogflow* beinhaltet eine *NLU engine* die natürliche, menschliche Sprache parsen kann.

### 4.4 Actions

Im Folgenden wird auf den Aufbau der Actions in Dialogflow eingegangen.

Intents sind ein Hauptbestandteil der Actions. Sie werden mithilfe von *machine learning* von Dialogflow aus *Utterances* (Äußerungen des Nutzers) bestimmt. Die Intents können mehrere Parameter enthalten. Diese können von der *NLU engine* als Entities erkannt werden.

Entities sind Tabellen in denen Einträge von Objekten eines Typs vordefiniert werden können. Zur Unterstützung der Spracherkennung können Synonyme für die Objekte eingetragen werden. Sie dienen zur Erkennung von relevanten Daten in den Intents.

Als Antwort auf ein Intent kann eine Text Response zurückgegeben werden. Unter dem Menüpunkt Fullfilment kann allerdings auch auf einen Webservice verlinkt werden, an den geparsete Intents über das *Actions Protocol* (JSON basiert) versendet werden. Der Webservice kann daraufhin den Inhalt auswerten und eine Antwort zurücksenden, die dann von der Action zurückgegeben werden kann.