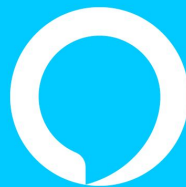


# Sprachassistenten. Wie lernen sie, uns zu antworten?

Am Beispiel Amazon Alexa





# Motivation

**WARUM?**



# Inhalte

1. Was ist Alexa?
2. Marktüberblick
3. Geschichte
4. Nutzerinteraktion
5. Alexa Voice Service
6. Skills & Interaction Model
7. Skill Entwicklung
8. Ausblick



# Hardware Amazon Echo & weitere Hersteller

Amazon Echo Serie & Geräte weiterer Hersteller  
mit Alexa Support (Sonos | JBL)





# Wer/Was ist Alexa wirklich?

*“Eine Stimme, der man Fragen stellen kann  
und darauf Antworten bekommt”*

## Funktionserweiterung klassischer Lautsprecher durch Cloud-Services

Cloud-basierter Service, der

- Spracherkennung
- Machine Learning
- hochwertige Sprachsynthese

kombiniert

Amazon Voice Service  
AVS



# Alexa “Herzstück” - Amazon Voice Service

Dadurch können Geräte mit Alexa-Integration...

- Befehle in Sprachform hören &
- in maschinenlesbare Informationen umwandeln &
- mit relevanten Antworten darauf reagieren



# Marktüberblick

Spracherkennungstechnologie wird von den Usern stark genutzt.

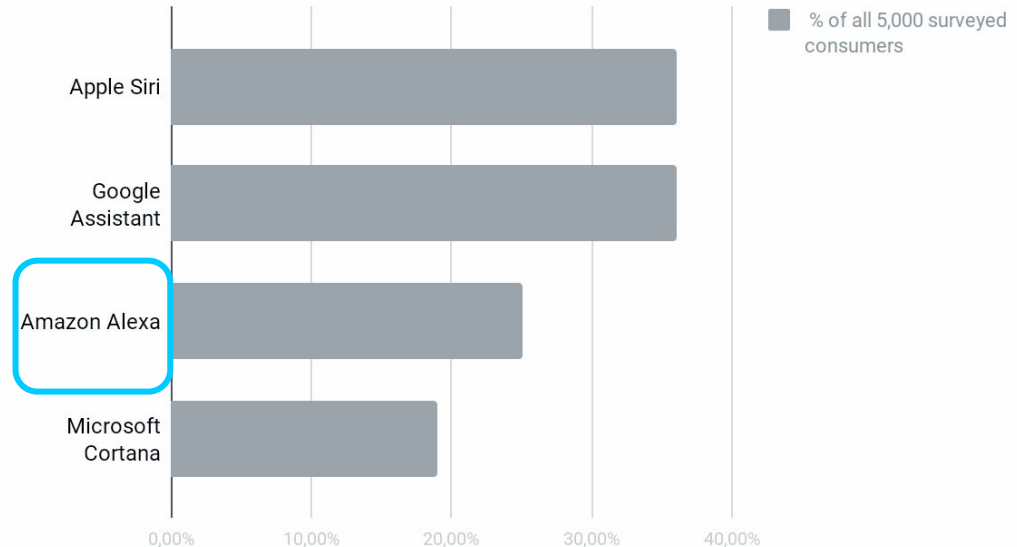
- 325 Millionen Menschen suchen mit **diktierten Suchbegriffen** im Internet
- **20 Prozent aller Google-Suchen** basieren auf Spracherkennung
- **Genauigkeit** der Antworten bei Voice-Google-Searches: 95 Prozent
- US Marktforscher Comscore: im **Jahr 2020** bereits jede zweite Internet-Suche mittels Spracherkennung



# Konkurrenz

- Amazon führend bei Hardware und im Home Segment
- Alexa nur 3. beliebter Sprachassistent  
→ Sprachsteuerung wird am häufigsten auf Smartphones genutzt

Popularity of voice assistants | USA Feb 2019







# Geschichte Alexa

- Markteintritt Echo: 2014
  - “intelligenter Lautsprecher”, mit dem Nutzer nur durch ihre Stimme die Musikwiedergabe steuern können.
  - ursprünglich ein auf Sprachsynthese basierender Text-Reader mit multi-direktionalen Mikros und einem Wi-Fi oder Bluetooth Lautsprecher

**Geschichte Spracherkennung?**



# Anfänge der Spracherkennung

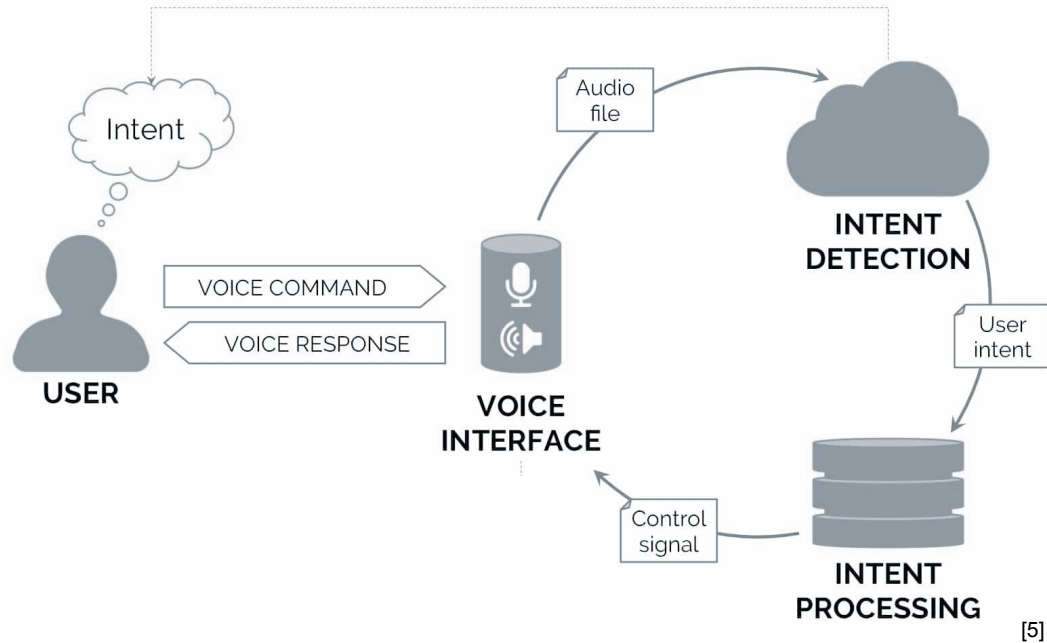
## 1961 - IBM "Shoebbox"

- Eingabegerät: Mikrofon
  - Umwandlung Geräusch in elektronische Impulse
  - Relay-System als Ausgangssignal-Empfänger
- erkannte und reagierte auf  
16 gesprochene Wörter  
(inkl. Zahlen von 0 bis 9)
- leichte arithmetische Operationen



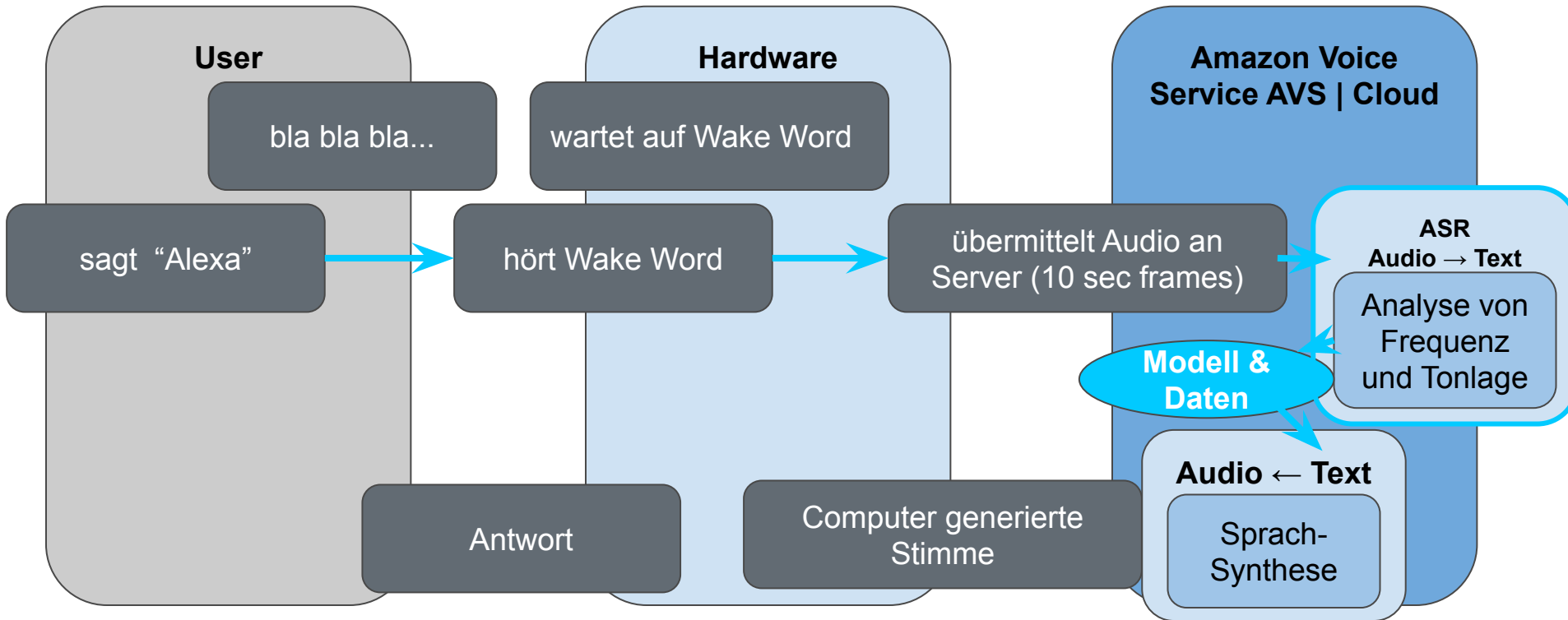


# Konversationen mit Alexa? Interaktion





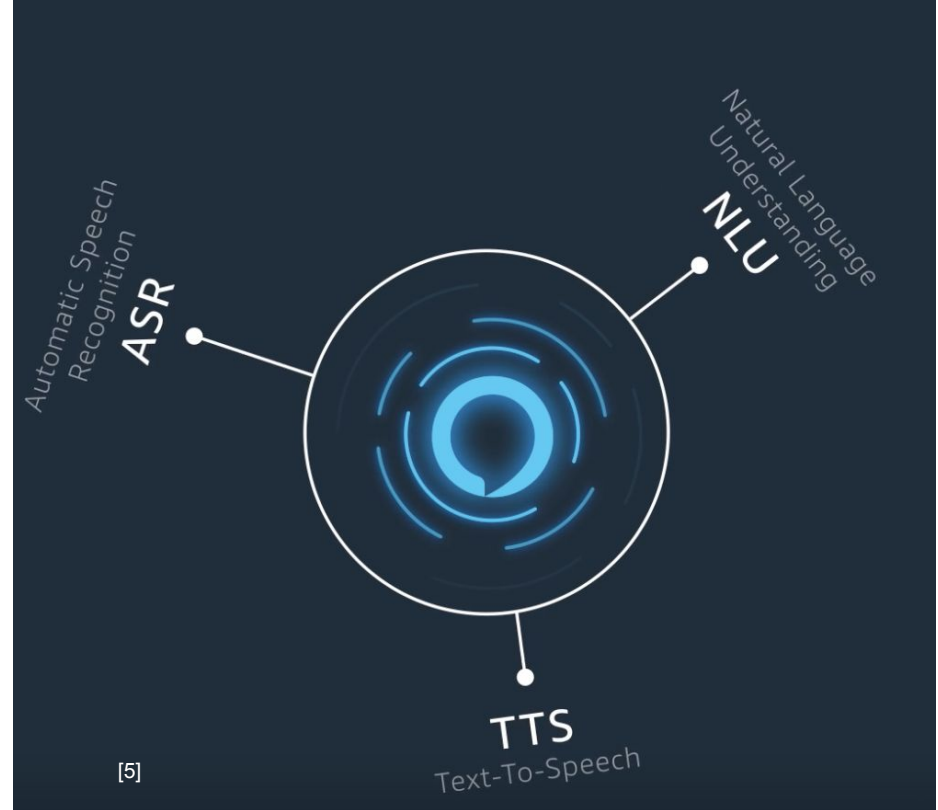
# Konversationen mit Alexa? Interaktion





# Amazon Voice Service Kit

- Skills-Entwicklung, ohne Kenntnis von Machine Learning Technologien und Spracherkennung
  - Amazon
    - übernimmt Spracherkennung
    - stellt Framework zur Verfügung
      - Anlernen des Machine Learning Models mit eigenen Trainingsdaten
- somit kann Alexa auf requests mit individuellen Antworten reagieren





# ASR Automatic Speech Recognition

= Technologie, die gesprochenes Wort in Text umwandelt

- **filtert** Sprache aus den Umgebungsgeräuschen heraus und erkennt dies als Wörter
- vor ASR war Sprache “nur” als eine Aufnahme von Tonlagen im Computer abgebildet.
- ASR entdeckt **Muster** in den Tonwellen und vergleicht diese mit den Geräuschen in einer best. Sprache → Identifikation der gesprochenen Wörter.

Automatic Speech  
Recognition  
**ASR**





# Vorteile der ASR Sprachsteuerung

## 1. Feels Fast

- Text To Speech Transformation geschieht in Millisekunden (nicht jedoch das Antrainieren des ML Modelles)

## 2. Make Educated Decisions

- Homonyme werden dank dahinter liegenden Statistiken richtig erkannt.

## 3. It Helps Voice Get Smarter

- ASR bildet die Basis für Natural Language Understanding

Automatic Speech  
Recognition  
**ASR**





# NLU Natural Language Understanding

ASR ist die grundlegende Technologie, auf der NLU aufbaut.

## Ziel NLU

Ein möglichst natürliches Sprache-Verstehen durch...

- das Generieren eines Kontextes aus einzelnen Wörtern
- und somit Vermeiden von Missverständnissen durch
  - Homonyme
  - Betonung
  - Satzzeichen

NLU ist ein permanenter Optimierungsprozess (durch Training des Machine Learning Modells)





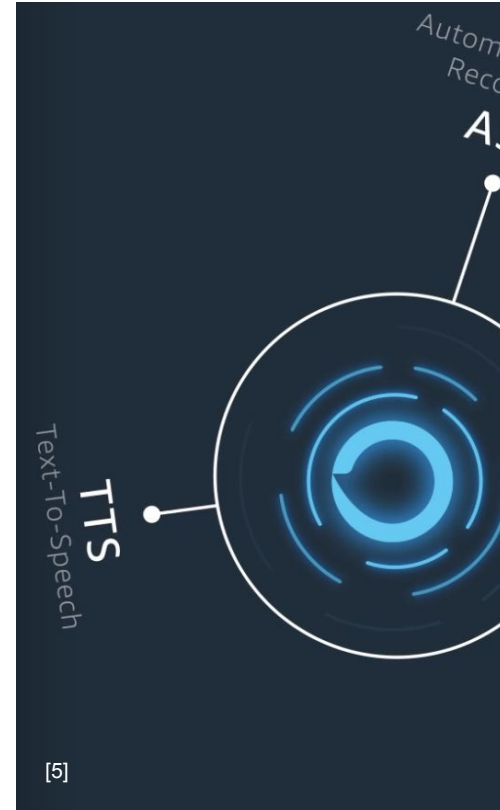


# TTS Text-To-Speech

= komplexe Technologie der Sprachsynthese, um die Textantwort in Sprache ausgeben zu können.

Zu berücksichtigen:

- Sprachkenntnis wird benötigt
  - Aussprache einzelner Laute
  - Pausen bei Satzzeichen
  - etc.





# Alexa Skills

= Third Party Erweiterungen der Alexa Funktionalitäten

Dadurch ist es möglich, die Spracherkennung zu erweitern und individuelle Antworten auf spezielle Spracheingaben zu entwickeln.

Vorgefertigte Skills standardmäßig verfügbar:

- Zeitungs-Skill
- Wetter-Skill



# Alexa Skill entwickeln. Aber wie?





# Interaction Model

Das Interaction Model gibt vor, welche Requests (Anfragen) der Skill bewältigen kann und welche Befehle den Skill aktivieren.

Bestandteile:

- **Name** für den Skill → “invocation name”  
Mit der Invocation springt der User in die Custom Anwendung
- Die Anfragen (“**requests**”), die der User dem Skill stellen kann.  
Werden “**intents**” genannt.  
z.B.: “Bier bestellen” → OrderBeer
- Die **Wörter**, die der User verwenden kann, um die Anfrage zu formulieren.  
→ *Das ist das eigentliche Interaction Model,*  
*da es bestimmt, wie der User interagieren kann.*  
z.B.: “Bestelle Bier”

# Custom Skill entwickeln



"90% des Designs sollte stehen, bevor eine einzige Zeile Code geschrieben wird."

1. Voice User Interface designen
  - Abläufe und mögliche Interaktionen vorab planen
2. Skill in der Amazon Developer Console aufsetzen
3. Interaction Model bauen anhand des Voice User Interface Design (Pkt. 1)
4. Coding...
  - Entwickeln eine API, welche auf Intents reagiert &
  - Textantworten retour gibt
  - TTS spricht die Antworten dann aus
5. Veröffentlichungsprozess:
  - Skill Beta-testen
  - finaler Check durch Amazon
  - Veröffentlichung



# Alexa Skill Blueprints

Amazon verfolgt den Ansatz der Plattform Ökonomie.

- strebt danach, möglichst viele Skills verfügbar zu haben
- steigert die Relevanz von Alexa

## **Daher**

- unterstützen Skill-Entwicklung für “Jedermann/Frau” - auch ohne Code-Kenntnisse.

Blueprints ermöglichen es, mit Vorlagen Skills in weniger als 10 Minuten zu erstellen.



# Alexa Skill Blueprints

HEAR for yourself...

- Gast Guide
- Meine Antworten

# Fragen?







# Sources

[https://www.codecademy.com/courses/learn-alexa/lessons/intro-to-alexa/exercises/intro-to-alexa-skills?action=resume\\_content\\_item](https://www.codecademy.com/courses/learn-alexa/lessons/intro-to-alexa/exercises/intro-to-alexa-skills?action=resume_content_item)  
<https://www.bigdata-insider.de/was-ist-alexa-a-581289/>  
<https://blog.wiwo.de/look-at-it/2018/07/26/ok-google-hey-siri-alexa-106-spannende-zahlen-fakten-rund-um-spracherkennung/>  
<https://www.digitaltrends.com/home/what-is-amazons-alexa-and-what-can-it-do/>  
<https://developer.amazon.com/de/alexa-voice-service>  
<https://onlim.com/eine-einfuehrung-in-sprachassistenten-mit-amazon-alexa/>  
<https://blog.wiwo.de/look-at-it/2018/07/26/ok-google-hey-siri-alexa-106-spannende-zahlen-fakten-rund-um-spracherkennung/>  
[https://www.ibm.com/ibm/history/exhibits/specialprod1/specialprod1\\_7.html](https://www.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html)  
<https://www.cnet.com/news/alexa-vs-google-assistant-vs-siri-the-state-of-voice-after-google-io-and-wwdc-2019/>  
<https://uk.pcmag.com/smart-home/92422/the-best-smart-speakers>  
<https://www.digitaltrends.com/home/siri-google-asistant-most-used-voice-assistants-alexa/>  
<https://developer.amazon.com/de/docs/ask-overviews/understanding-the-different-types-of-skills.html>  
<https://developer.amazon.com/de/docs/devconsole/create-a-skill-and-choose-the-interaction-model.html>  
<https://developer.amazon.com/de/docs/devconsole/build-your-skill.html>  
<https://developer.amazon.com/de/docs/smapi/interaction-model-schema.html>  
<https://developer.amazon.com/de/docs/ask-overviews/build-skills-with-the-alexa-skills-kit.html>  
<https://vui.agency/de/services/voice-design/>  
<https://medium.com/crowdbotics/how-to-build-a-custom-amazon-alexa-skill-step-by-step-my-favorite-chess-player-dcc0edae53fb>  
[https://www.codecademy.com/courses/learn-alexa/lessons/intro-to-alexa/exercises/test-service-simulator?action=resume\\_content\\_item](https://www.codecademy.com/courses/learn-alexa/lessons/intro-to-alexa/exercises/test-service-simulator?action=resume_content_item)



# Images

- [1] <https://www.amazon.de/Sonos-One-Speaker-Sprachsteuerung-AirPlay/dp/B075XVVFHJ>
- [2] <http://www.affilistarter.de/wp-content/uploads/2019/06/JBL-Live-500BT-mit-Alexa-und-Google.jpg>
- [3] <https://www.amazon.com/Echo-Buttons-Pack-gaming-companion/dp/B072C4KCQH>
- [4] <https://www.ibm.com/ibm/history/exhibits/specialprod1/images/overlay/1404011.jpg>
- [5] <https://vui.agency/de/services/voice-design/>
- [6] <https://youtu.be/oTaFX3ZQInA?t=19>
- [7] <https://www.macexpertguide.com/wp-content/uploads/2018/09/Alexa-Skills-for-Emergencies.png>