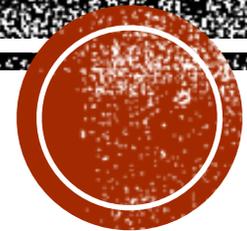


# AUTOMATISCHE SPRACHERKENNUNG

Computerlinguistik im WS 2020/21

Cindy Hartmann

Alessandro Marini



# SPRACHERKENNUNG

- Was ist automatische Spracherkennung(ASR)?
- Engl. Speech Recognition
- Unter automatischer Spracherkennung versteht man das maschinelle Erkennen gesprochener Sprache. Sie ist ein Teilgebiet der angewandten Informatik, Ingenieurwissenschaften und Computerlinguistik
- Kommunikation zwischen Mensch und Maschine



# SPRACHERKENNUNG

- Sprecherabhängige Spracherkennung
  - > Trainingsphase
  - > Anpassung auf jeweiligen Benutzer
  - > Keine wechselnden Benutzer möglich
  
- Sprecherunabhängige Spracherkennung
  - > Keine Trainingsphase
  - > Geringer Wortschatz
  - > Mehrere Benutzer möglich



# GESCHICHTE

- Forschung begann in 1960er Jahren aber weitestgehend erfolglos
- 1984 wurde erstes Spracherkennungssystem von IBM vorgestellt mit ~ 5.000 englischen Wörtern, Erkennungsvorgänge dauerten einige Minute
- 1991 kam erstes Spracherkennungssystem das bis zu 30.000 deutschen Wörter erkannte
- 1993 erstes System für den breiten Markt, Kostenpunkt um die 1000 Dollar
- 2004 gab IBM Teile der Spracherkennungsanwendung als Open Source frei
- 2007 gründete sich die Firma Siri Inc.
- 2010 von Apple gekauft -> 2011 Siri als Spracherkennungssoftware im iPhone 4s vorgestellt



# HERAUSFORDERUNG



# HERAUSFORDERUNGEN

- Umgebungsgeräusche
  - Bürolärm
  - Andere Personen
  - Fahrgeräusche (u.v.a.m.)
- Homonyme – Mehrdeutigkeit der Sprache
  - *Bank* (Möbel, Geldinstitut, Sportgerät)
  - *Brücke* (Bauwerk, Zahnersatz)
- Homophone – Wörter die identisch klingen, aber eine andere Bedeutung haben
  - Meer ≠ mehr
  - Graf ≠ Graph ( u. v. a. m.)



# ANWENDUNGSGEBIETE



# ANWENDUNGSGEBIETE

- Diktierprogramme (z.B. IBM ViaVoice , Dragon Systems Naturally Speaking)
- Telefonauskunftssysteme/ Kundenservice
- Fremdsprachensoftware
- Behindertenunterstützung und kommunikation
- Smartphones (z.B Sprache-zu-Text)
- Smarthome Systeme (Alexa, Bosch Smart Home u.a.)



# BENUTZBARKEIT VON SPACHGESTEUERTEN ASSISTENTEN



# STÄRKEN

- **Aufgaben beschleunigen:** Voice Control kann einfache Aufgaben abkürzen, z.B. Termine und Erinnerungen im Smartphone anlegen, den Wetterbericht abrufen etc.
- **Aufgaben-Schritte überspringen:** Mit Sprachsteuerung kann beispielsweise ein Termin mit nur einem Kommando erstellt werden
- **Die Hände sind frei:** Solange sich das Gerät in der Nähe befindet, kann es rein mit der Stimme gesteuert werden.
- **Gut für klare Kommandos:** Aufgaben, die sich klar und deutlich ausdrücken lassen, werden von der Spracherkennung meist richtig gedeutet.



# SCHWÄCHEN

- **Komplexe Kommandos:** Bei komplexeren Kommandos oder Fragen funktioniert Sprachsteuerung in der Mehrzahl der Fälle noch nicht.
- **Störgeräusche:** Mit Umgebungslärm, undeutlicher Aussprache, Akzenten und Dialekten tun sich sprachgesteuerte Assistenten schwer. Auch das wird sich in Zukunft sicherlich verbessern.
- **Datenschutz:** Die Datenschutzlage ist in manchen Fällen noch bedenklich, vor allem, wenn Geräte ununterbrochen „zuhören“.
- **Gewohnheitssache:** Für viele Menschen ist es ungewohnt mit einem Computer zu sprechen und sie nutzen aus purer Gewohnheit die bekannten Interaktionsformen, wie Touch oder Point&Click.
- **Intransparente Möglichkeiten:** Im Gegensatz zu grafischen User-Interfaces sind den Nutzern von Sprachsteuerung die möglichen Kommandos völlig unbewusst.



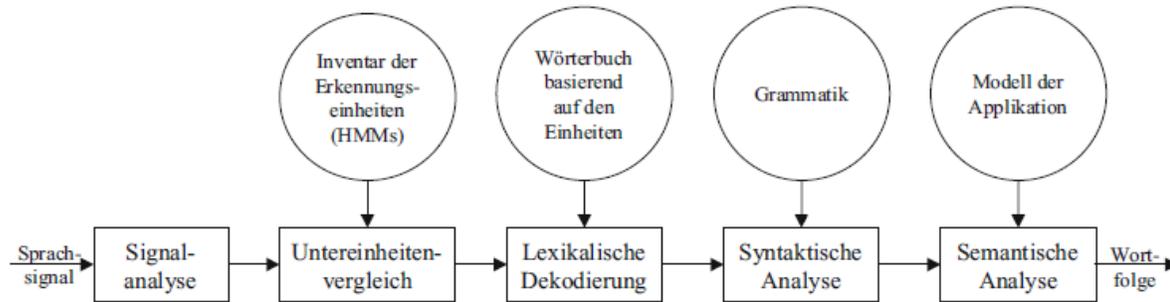
# FUNKTIONSWEISE

Spracherkennung isoliert in einer Audio-Spur enthaltene Informationen

- Linguistischer Inhalt
- Sprecher-Spezifikationen
- Raumcharakteristika und Umgebungsgeräusche



# FUNKTIONSWEISE



Der wohl bekannteste Ansatz:  
HMM (Hidden Markov Model)

Aufteilung in 5 Schritte

- Signalanalyse
- Untereinheitenvergleich (oder unit matching)
- Lexikalische Dekodierung
- Syntaktische Analyse
- Semantische und pragmatische Analyse



# FUNKTIONSWEISE

## Signalanalyse

- Umwandlung bzw. Digitalisierung des Gesagten
- Erstellung einer kompakten parametrischen Repräsentation durch:
  - Vorverarbeitung: spektrale Analyse (Lautstärkenverlauf)
  - Merkmalsextraktion: Transformation in Merkmalsvektoren



# FUNKTIONSWEISE

Untereinheitenvergleich (oder unit matching)

- Merkmalsvektoren werden in Laute/ Silben/ Wörter gewandelt
- Prozess wechselt physikalische Darstellung in eine linguistische
- Für jeden Vektor wird jede Möglichkeit in Betracht gezogen
- Lexikalische und syntaktische Einschränkungen folgen



# FUNKTIONSWEISE

## Lexikalische Dekodierung

- Vektorsequenzen werden mit Lexikoninhalten abgeglichen
- Lexikoninhalte entsprechen der Merkmalsgröße

## Syntaktische Analyse

- Sprachenspezifische syntaktische Regeln
- Modellpfade die aus dem vorherigen Schritt entstanden sind wenn Sie in einer syntaktisch geeigneten Reihenfolge stehen



# FUNKTIONSWEISE

## Semantische und pragmatische Analyse

- Dieser Schritt wird häufig als Sprachverstehen bezeichnet
- Prozess erfolgt ganz am Schluss, ohne Rückkopplung zu vorherigen Schritten



# ANSÄTZE

Erinnerung an die Liste von Folie 9:

- Linguistischer Inhalt
- Sprecher-Spezifikationen
- Raumcharakteristika und Umgebungsgeräusche
- HMM konzentriert sich vor allem auf die linguistischen Inhalte
- Alternativ kann Spracherkennung als Sprechererkennung genutzt werden
- Ganz Theoretisch kann auch Raumcharakteristika oder Umgebungsgeräusche extrahiert werden



# ANSÄTZE

- Für linguistische Spracherkennung gibt es statt HMM's auch das Prinzip der „neuronalen Netze“ (oder sogar Hybride aus beiden)
- „neuronale Netze“ versuchen durch mit Vektoren verknüpften Neuronen das menschliche Gehirn „nachzubauen“
  - Diese Systeme sind auf Einzelwörterkennungen spezialisiert



# QUELLEN

- <https://www.ei.ruhr-uni-bochum.de/media/ei/lehmaterialien/spracherkennung/ab137310d5bb3a2cf37dc230b67cb6edf9ac8b86/Skript-2018.pdf>
- <https://userlutions.com/blog/usability-insights/sprachsteuerung-virtuelle-assistenten-usability/>
- [https://www.dragon-speaking.de/loesung\\_handicap.php](https://www.dragon-speaking.de/loesung_handicap.php)
- <https://www.janko.at/Raetsel/Gedichte/Homonyme.htm>
- <https://de.wikipedia.org/wiki/Spracherkennung>

