

The image features decorative wavy lines in the top corners, composed of multiple thin, parallel lines in a reddish-brown color that create a sense of motion and depth. The main title is centered in a large, bold, sans-serif font.

SPRACHERKENNUNG

Computerlinguistik I - WiSe 21/22

Aesa Budancamanak und Amina Lawal

- Überblick
 - Was ist Spracherkennung?
 - Anwendungen
 - Geschichte
- Funktionsweise
 - Herausforderungen
 - Methoden der Spracherkennung
- Ausblick

SPRACHERKENNUNG - Was ist Spracherkennung?

- Teilgebiet der Angewandten Informatik
- Fähigkeit einer Maschine Sprache zu erkennen und in Text umzuwandeln
- Spracherkennung ist abzugrenzen von der Stimmerkennung
- Je nach Problem können unterschiedliche Ansätze verwendet werden

SPRACHERKENNUNG - Warum Spracherkennung?

- Sprachgesteuerte Software im Auto, z.B. BMW
- Dialogsysteme
- Behindertenunterstützung und Kommunikation
- Bekannte Anwendungen wie Siri, Alexa, Amazon Echo (...)
- Diktiersysteme
- durch Spracherkennung Gerätesteuerung möglich

SPRACHERKENNUNG - Warum Spracherkennung?

- Erleichtert den Alltag, beispielsweise beim Auto fahren, beim Haushalt etc.
- Programme durch Sprache steuerbar
- Zeitersparnis

- **Forschungsbeginn:** 60er - Erkennung einzelner Wörter unter Laborbedingungen
- **1977:** GUS - Beratung und Reservierung von Flugreisen
- **80er:** Unterscheidung von Homophonen
- **1984:** System von IBM erkennt ca. 5.000 englische Einzelwörter
 - lange Dauer + Großrechner
- **1989:** Diktiermaschine Dragon Dictate
 - + 30.000 verschiedene Wortformen
 - + verwendbar auf tragbarem PC
 - kurze Sprechpausen zwischen Worten

SPRACHERKENNUNG - GESCHICHTE

- **1991:** Spracherkennungssystem TANGORA 4 (IBM) erkennt 20.000 bis 30.000 deutsche Wörter
- **1993:** erste Software für Massenmarkt - 1000€
- **1993 - 2000 Projekt Verbmobil:** sprecherunabhängige Übersetzung von Spontansprache (Deutsch, Japanisch und Englisch)
- **1997:** für Heim-PC IBM ViaVoice & Dragon NaturallySpeaking
 - Dragon NaturallySpeaking: meistverbreitete sprecherabhängige Spracherkennungssoftware für Windows
- **2011:** Siri für iPhone 4s - Funktion eines persönlichen Assistenten

UNTERSCHIEDUNG IN

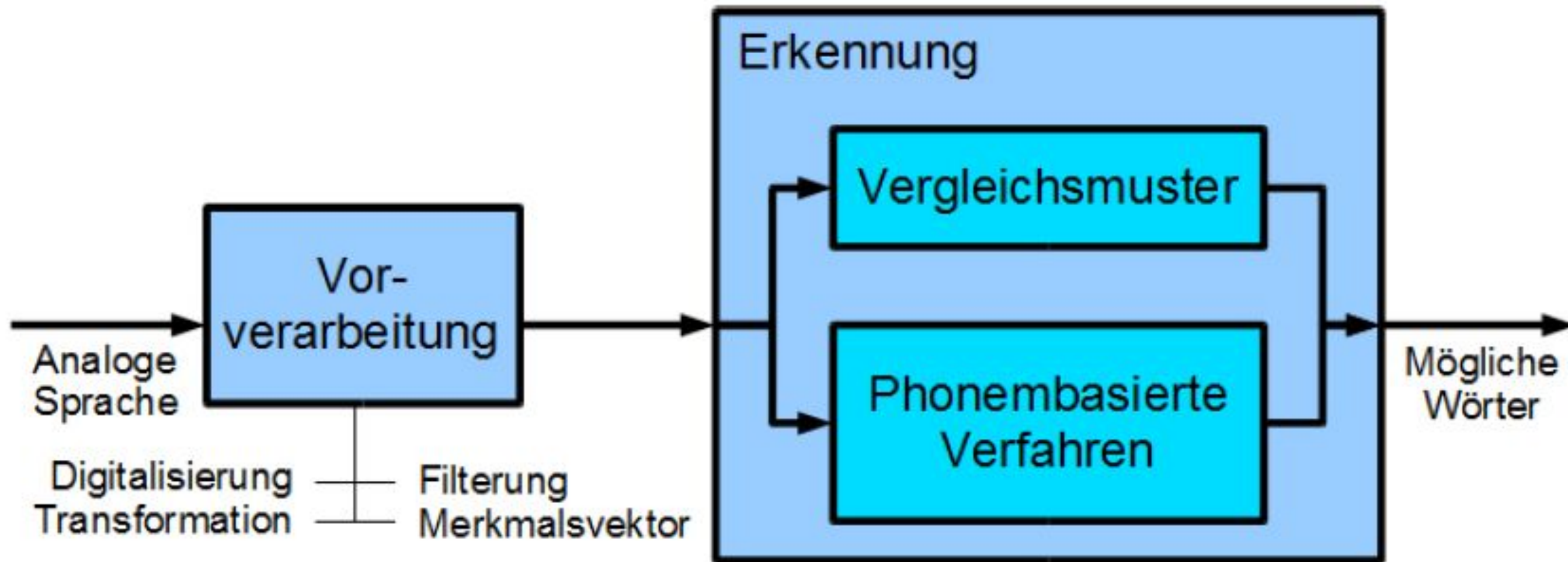
SPRECHERUNABHÄNGIGE

- geringer Wortschatz
- kein Training
- Dialogsysteme wie Fahrplanauskunft/Call-Center; (Einzel-)Worterkennung

SPRECHERABHÄNGIGE

- großer Wortschatz
- Sprecherspezifisches Training vor/während der Verwendung
- Diktiersysteme, persönliche Assistent:innen (Siri), kontinuierliche Sprache

SPRACHERKENNUNG - FUNKTIONSWEISE



HERAUSFORDERUNGEN

KONZEPTION:

- Anzahl der unterstützten Wörter
- Sprechweise
 - Erkennung isolierter Wörter, die von Pausen umgeben sind, ist einfacher als die Erkennung kontinuierlicher Sprache.
 - Erkennung gelesener oder direkt an das Gerät gerichteter Sprache vs. "natürliche" Konversationen

HERAUSFORDERUNGEN

Variationen der Sprachbenutzung
(Störung des Redeflusses, neue Wortschätze)

Intersprechervariabilität
(Verschiedenheit der Sprecher)

Intrasprechervariabilität
(Verschiedenheit des Wortes)

Umgebungsgeräusche

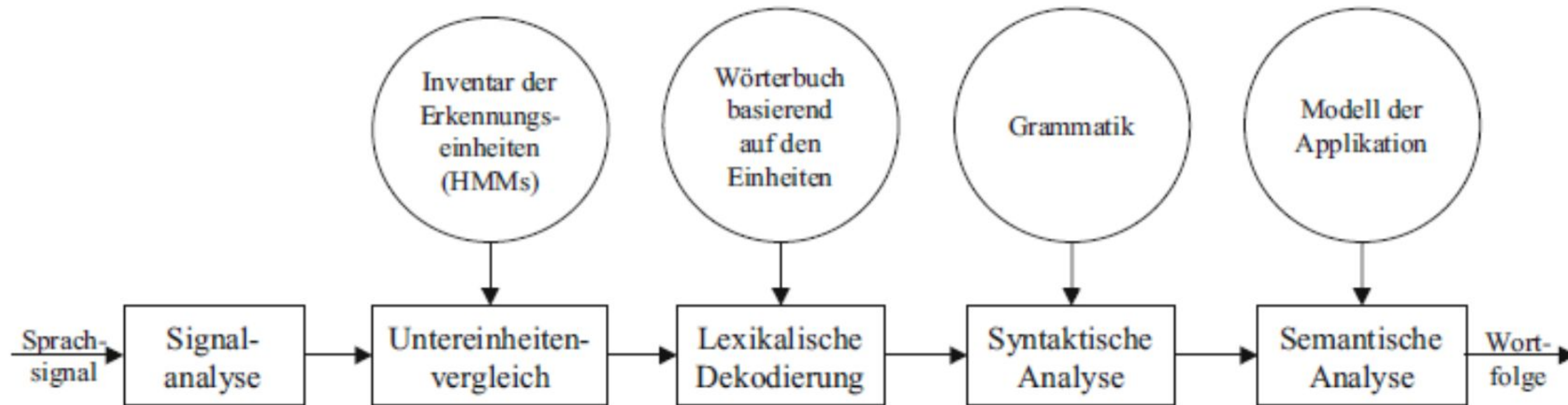
Homonyme

Verschleifungen
(Zusammenziehen der Wörter)

Homophone

Übertragungsschwierigkeiten

HIDDEN MARKOV MODELL(E)



- stochastisches Modell
- Vergleich von bekannten Phonemen mit Eingangssignalen.
 - Das akustische Modell des Phonems wird zerlegt in Anfang, Mittelstücke, Ende.
 - Eingangssignale werden mit den Teilstücken verglichen und mögliche Kombinationen gesucht.

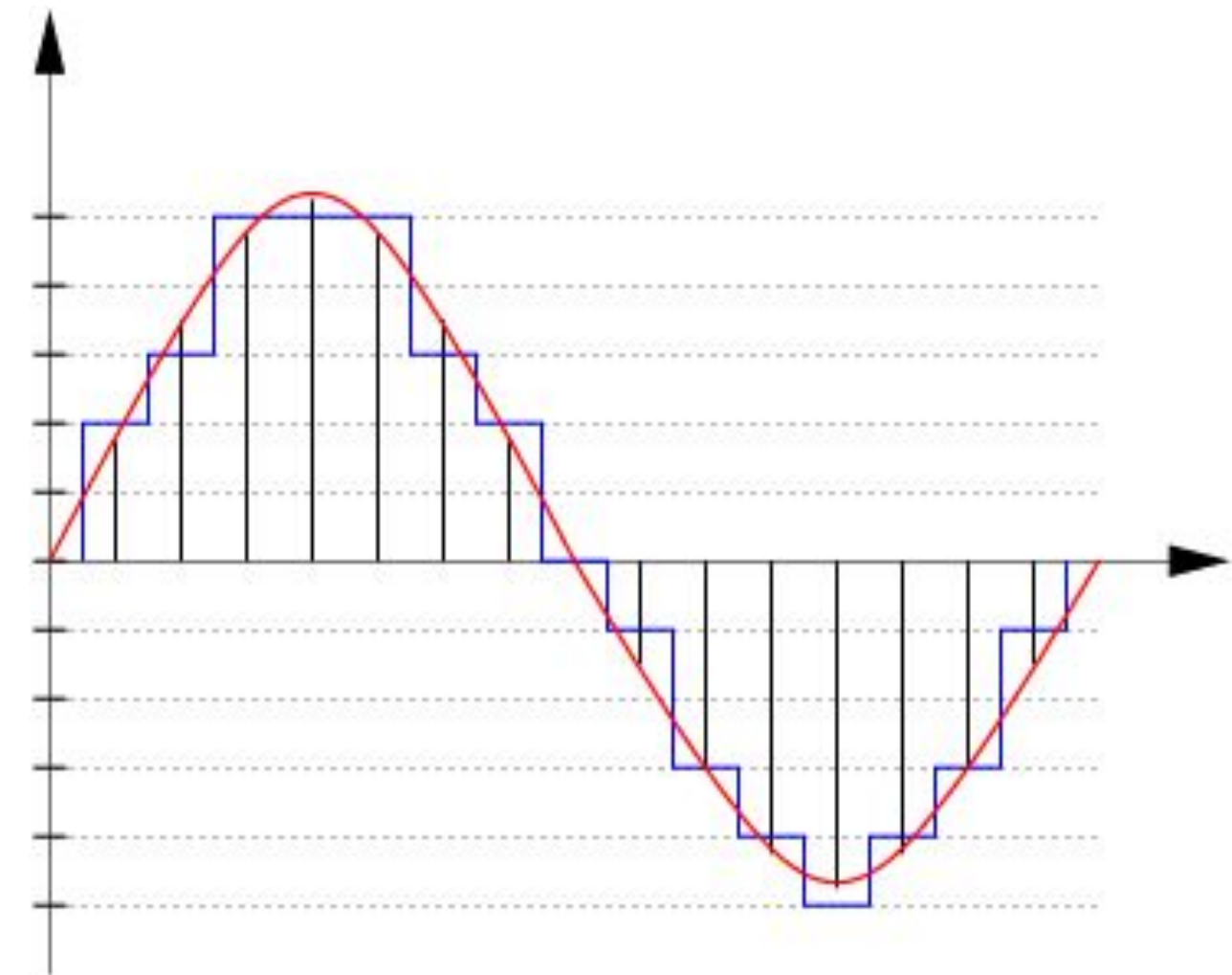
HIDDEN MARKOV MODELL(E)

1. Signalanalyse
2. Unit matching (Untereinheitenvergleich)
3. lexikalische Dekodierung
4. syntaktische Analyse
5. Semantische & pragmatische Analyse

HIDDEN MARKOV MODELL(E)

1) SIGNALANALYSE

- Umwandlung:
 - Digitalisierung
 - Filterung von Störgeräuschen
- Vorverarbeitung: spektrale Analyse



1) SIGNALANALYSE

- **Merkmalsextraktion:**

- Merkmale des Signals werden bestimmt, extrahiert und in Vektoren umgewandelt, die zu Vektorsequenzen zusammengefasst werden
- Erstellung einer Repräsentation

Merkmale:

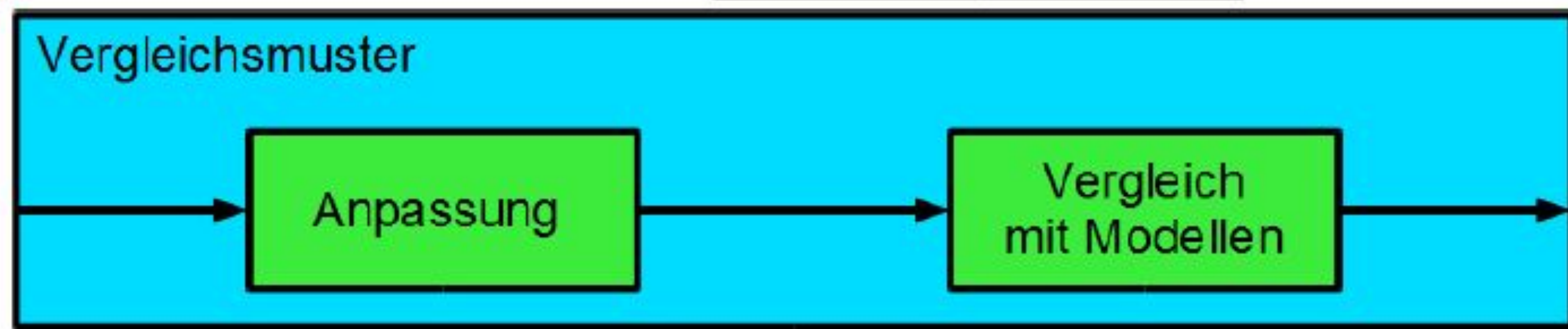
- Seitenverhältnis
- rund oder eckig
- wie viele Ecken
- Ecken abgerundet



2) UNTEREINHEITENVERGLEICH (UNITMATCHING)

Erstellung hypothetischer Wortfolgen, die dem System bekannt sind

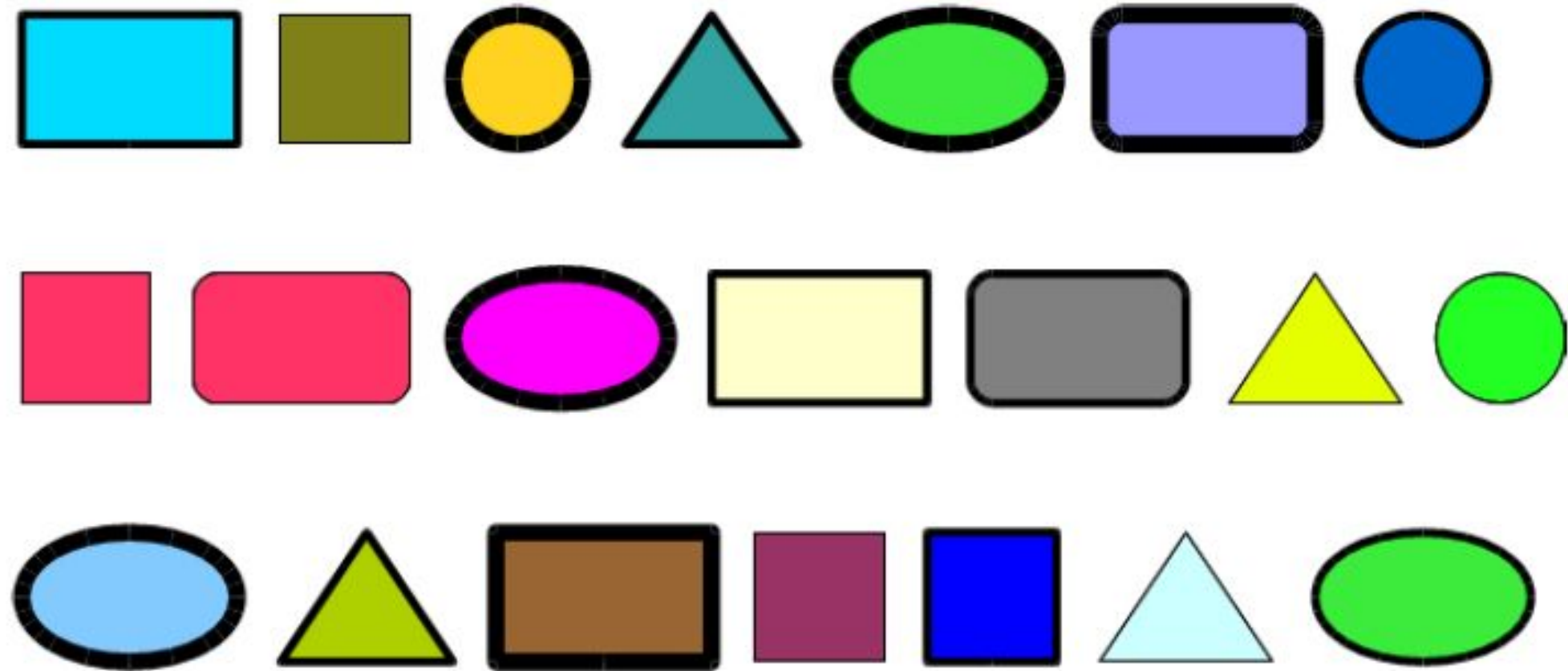
- umgewandeltes und extrahiertes Eingangssignal wird so lange angepasst bis es dem Modell entspricht und vergleichbar ist



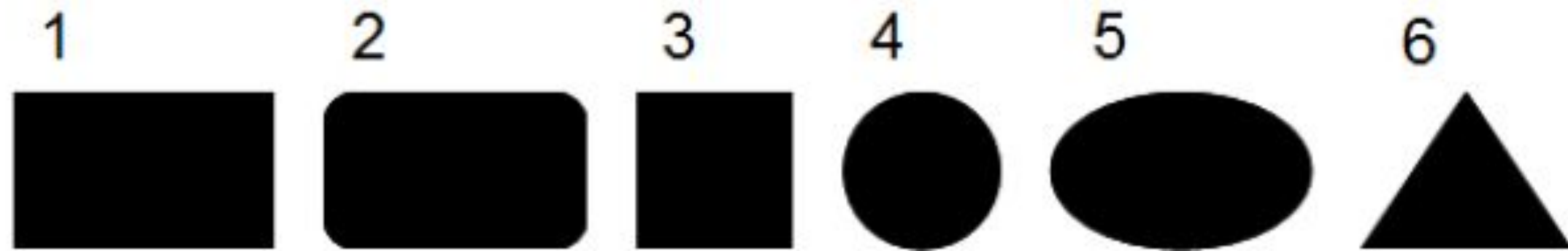
- Verwendung von z.B. Phoneme, Silben oder ganzen Wörtern als atomare linguistische Bausteine (units), die im Sprachsignal identifiziert werden sollen

SPRACHERKENNUNG - METHODEN

Phoneme



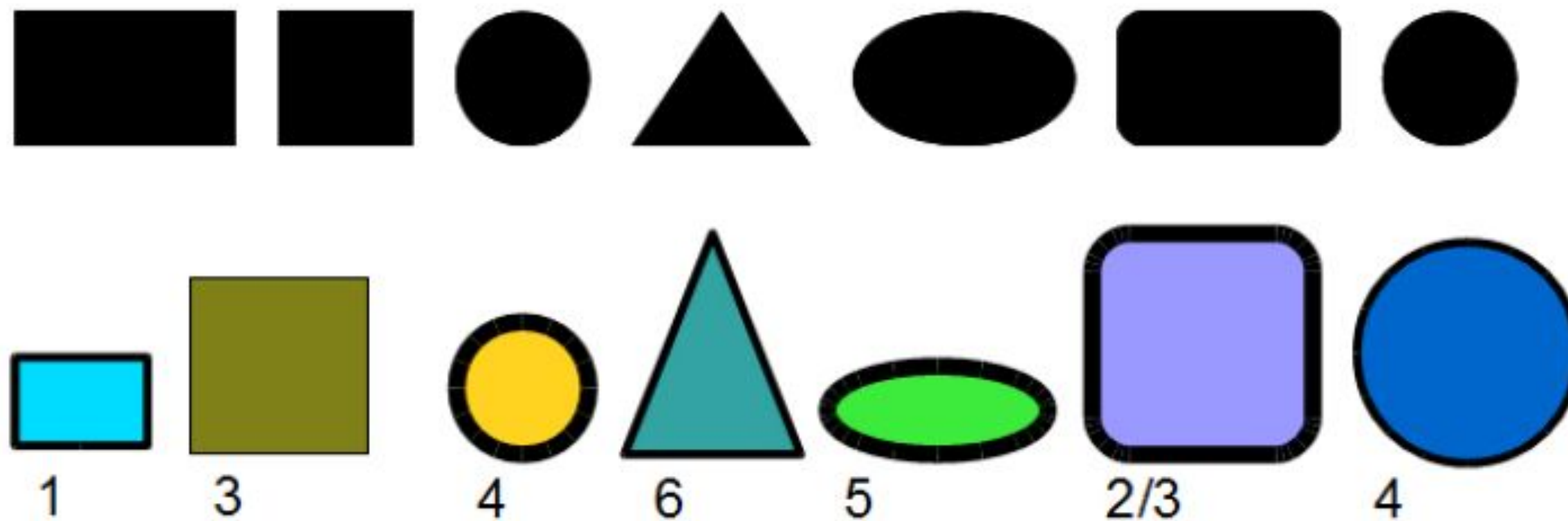
Vergleichsmodelle



3) LEXIKALISCHE DEKODIERUNG

Vergleich der Kombinationen mit den Wörtern des Wörterbuchs

Passende Wörter im Wörterbuch

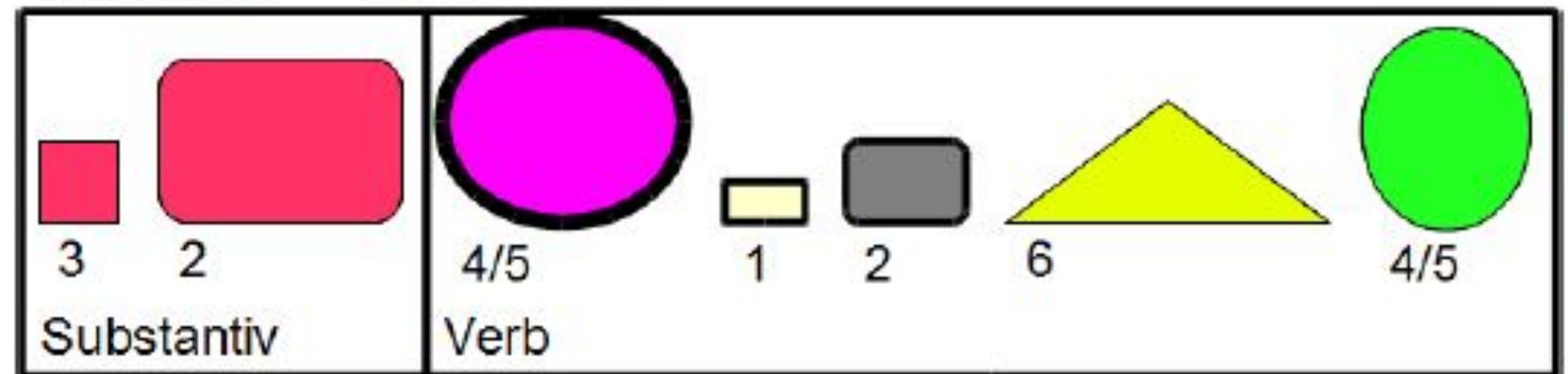


Lexikoneinträge enthalten zu jedem Worteintrag auch die Sequenz von Untereinheiten, aus denen sich das Wort zusammensetzt

4) SYNTAKTISCHE ANALYSE

Grammatikprüfung

- jedem Wort wird eine Funktion zugeordnet
- Reihenfolge anhand syntaktischer Regeln, die in einer Grammatik festgelegt sind
- Reduktion der Einheitensequenzen



1. Verb



2. Substantiv



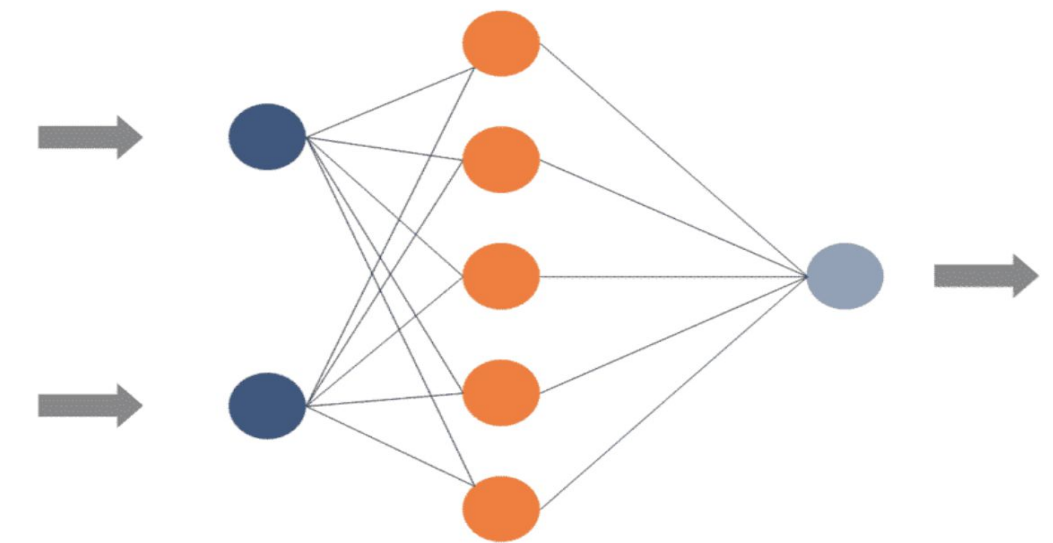
5) SEMANTISCHE & PRAGMATISCH ANALYSE

- Sprachverstehen
- Reduktion der Sequenzen auf grammatisch wohlgeformte Wortfolgen
- Vorgang wird so oft wie nötig wiederholt, bis die Wortfolgen linguistisch sinnvoll und dem kommunikativen Kontext angemessen sind

Weitere Modelle: Künstliche Neuronale Netze

Entwicklung der KNN

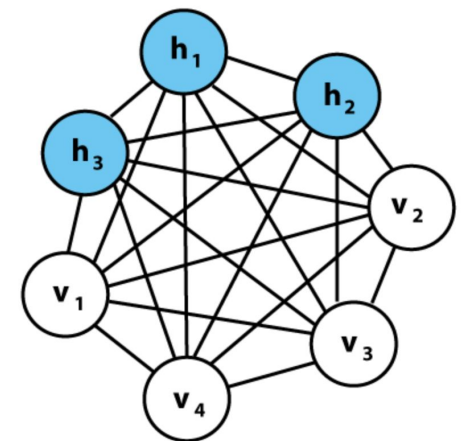
- Neuronale Netze wurden bereits in den 80ern eingesetzt, hat aber keine vergleichbaren Ergebnisse gezeigt wie HMM
- Durch den Einsatz von deep neural networks wurde die Erkennungsrate verbessert
- Zwischen der Ein und Ausgabeschicht liegt eine hohe Anzahl an versteckten Schichten
- Moderner Ansatz um Phoneme zu erkennen
- Durch DNN ist es weiterhin aufwendig



Weitere Modelle: Künstliche Neuronale Netze

Entwicklung der KNN

- **DBN (Deep Belief Network)**
- Jede Schicht isoliert trainiert
- Durch Einsatz von GPU Berechnungen möglich und Verfahren wird beschleunigt
- Möglichkeit des Einsatzes von DBN für Spracherkennungssysteme einzusetzen
- In der Realität geschieht das allerdings in Verbindung mit HMM als „hybrides Verfahren“
- Mit der Kombination ist eine Verbesserung der Erkennungsrate um bis zu 9,2% möglich



Ausblick

- Zukunftsaussichten vor allem in der sprachbasierten **Mensch-Maschine Kommunikation**
- große Fortschritte im Bereich der Spracherkennung in den letzten Jahren
- Alle großen Technologieanbieter arbeiten an der Weiterentwicklung von Spracherkennung und Sprachsteuerung
- In verschiedenen **Bereichen einsetzbar** bspw. Medizin, Haushalt, Telekommunikation, Automobilbranche, Forschung etc.

Anwendungen an denen gearbeitet werden:

- Software, die die Dialekte der gesprochenen Sprache erkennt (nur ein Indiz)
- Spracherkennung zur Nutzung fürs Transkribieren (Niedrige Fehlerrate)
- **Zukunftsprojekte** wie Apples “ios in the car”
- Anwendungen und Innovationen in der Spracherkennung birgen auch Datenschutzrechtliche Fragen, da die Möglichkeit besteht abgehört zu werden (z.B Vorfälle bei Amazon)

Quellen

- Carstensen, Kai Uwe et al. 2010. *Computerlinguistik und Sprachtechnologie - Einführung*. Heidelberg: Spektrum Akademischer Verlag
- https://www.h-brs.de/files/20171215_fbinf_mclab_ss15_spracherkennung_flock_spieker_sa_mk.pdf
- <https://www.computerweekly.com/de/definition/Spracherkennung>
- <https://www.dfcsystems.de/blog/trends/spracherkennung/>
- https://www.parlament.gv.at/ZUSD/FTA/066_sprachassistenten.pdf
- <https://medium.com/swlh/what-are-rbms-deep-belief-networks-and-why-are-they-important-to-deep-learning-491c7de8937a>
- <https://www.tu-chemnitz.de/informatik/KI/edu/prosem/ss2011/tischer/spracherkennung.pdf>
- <https://www.tu-chemnitz.de/urz/archiv/kursunterlagen/multimedia/rsrc/quantisierung.png>