Universität Hamburg, Fachbereich Informatik
Arbeitsbereich
Natürlichsprachliche Systeme

Herausforderung: **Gesprochene Sprache**

Die größte Herausforderung bei der Verarbeitung von gesprochener Sprache ist die **große Variabilität des Sprachsignals**:

- dasselbe Wort wird nie zweimal wirklich *identisch* ausgesprochen
- Buchstaben werden je nach Kontext ganz unterschiedlich gesprochen („der na**ss**e Ha**s**e **s**chläft nacht**s**“)
- unser Gefühl täuscht uns oft: zum Beispiel wird im Deutschen das „r“ eigentlich nie gesprochen
- auch die Bedeutung von Wörtern ist je nach Kontext verschieden

„Nur weil Menschen es können, heißt es leider nicht, dass es einfach ist!“

Wäre es nicht besser, wenn Sprache viel einfacher wäre?

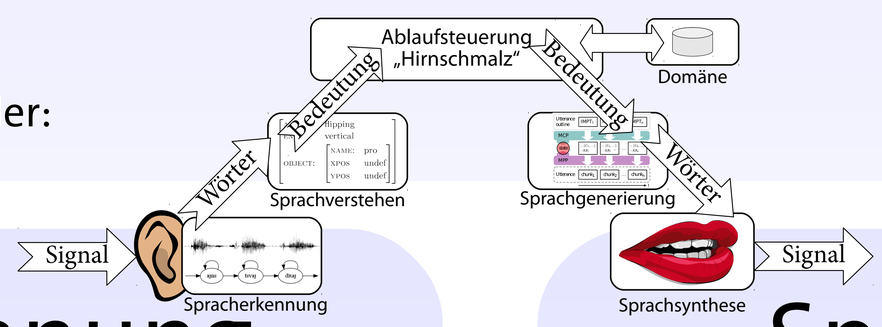
- dann wäre sie sicher leichter zu verstehen (gerade für Computer), aber
- unsere Sprache ist sehr *ausdrucksstark*: was immer wir sagen wollen, wir können es mit Sprache tun!

Aufbau von sprachverarbeitenden Systemen

Sprache ist vielschichtig und wird von unterschiedlichen Teildisziplinen der Linguistik erforscht:

- Phonetik → Sprachlaute
- Morphologie → Worte und Wortbildung
- Syntax → Kombination von Wörtern zu Sätzen
- Semantik → Bedeutung von ganzen Äußerungen

Die Vielschichtigkeit spiegelt sich in sprachverarbeitenden Systemen wider:



vielschichtiges Problem → komplexe Systeme

Spracherkennung

Ziel: im Audiosignal die gesprochenen Wörter erkennen

Formal: Gegeben einen Sinneseindruck **O** gilt es, die Wortfolge **W** zu finden, die am wahrscheinlichsten zu diesem Sinneseindruck passt: $\hat{\mathbf{W}} = \arg\max_{\mathbf{W}} P(\mathbf{W}|\mathbf{O})$.

Es können nur bekannte Wörter berücksichtigt werden, d.h. „neue“ Wörter sind in der Regel nicht erlaubt.

Spracherkennungsqualität ist *nie perfekt*, aber heute für viele Anwendungen „gut genug“ (Siri, ...)

- Menschen gleichen Fehler durch Integration von Weltwissen aus, was Computer nur schlecht können

Sprachsynthese

Ziel: aus Wortfolge „lebendige“ Sprache erzeugen, dabei muss die *natürliche Variabilität* hinzugefügt werden

Problem: Schriftsprache ist oft mehrdeutig, korrekte Aussprache erschließt sich erst aus dem Kontext

Die Satzmelodie entspricht der Informationsstruktur. Gutes vorlesen setzt daher Verständnis voraus!

“I didn't say we should kill him.”
- *someone else said* we should kill him
- I am denying that I said we should kill him
- I wrote it down or implied it, but I didn't *say* it
- I said *someone else should* do the job
- I said that we absolutely *must* kill him
- *scaring* him a little would've been enough
- we got the *wrong guy*!

Spontane Sprache zeichnet sich durch Füllwörter, Lachen, und Feedback aus – dies ist für Sprachsynthesesysteme noch schwieriger als reines Vorlesen

Unsere Forschung: Schritthaltende Verarbeitung

Menschen fallen sich ins Wort, korrigieren sich, helfen sich, kurz: **Menschen verstehen während der Verarbeitung**.
Die meisten Systeme interpretieren Äußerungen erst als ganzen Satz, und sprechen selbst immer ganze Sätze.

Probleme:
- keine Systemreaktionen während der Nutzer spricht
- keine Anpassung von laufenden Sprachausgaben

→ keine natürliche Interaktion im Dialog

Lösungen:
- sofortige Weitergabe von vorläufigen Ergebnissen
- spätere Korrektur und Ergänzung der Hypothesen

→ ermöglicht natürlicheres Verhalten