Einführung in die Computerlinguistik:
Gesprochene Sprache

Dr. Marc Schröder, DFKI
schroed@dfki.de

Kurs-Homepage:
http://www.coli.uni-saarland.de/~hansu/courses/EC07/index.html

Überblick

- Eigenschaften gesprochener Sprache
  - im Vergleich zu geschriebener Sprache
  - Sprachsignal
- Maschinelle Verarbeitung gesprochener Sprache
  - Spracherkennung
  - Sprachsynthese
  - Dialogsteuerung
  - Emotionserkennung
  - Expressive Sprachsynthese
  - Modellierung nonverbaler Kommunikation

Geschriebene Sprache

Mit dem Verweis auf die Tarifautonomie hat Kanzlerin Merkel klargestellt, nicht in den Bahn-Tarifkonflikt eingreifen zu wollen. Damit erteilte sie der Aufforderung von Bahn-Chef Mehdorn eine Absage. Die Lokführergewerkschaft GDL entscheidet heute über Streiks im Güterverkehr.

(tagesschau.de vom 7.11.07)

Gesprochene Sprache: Dialog!

- Theorie:

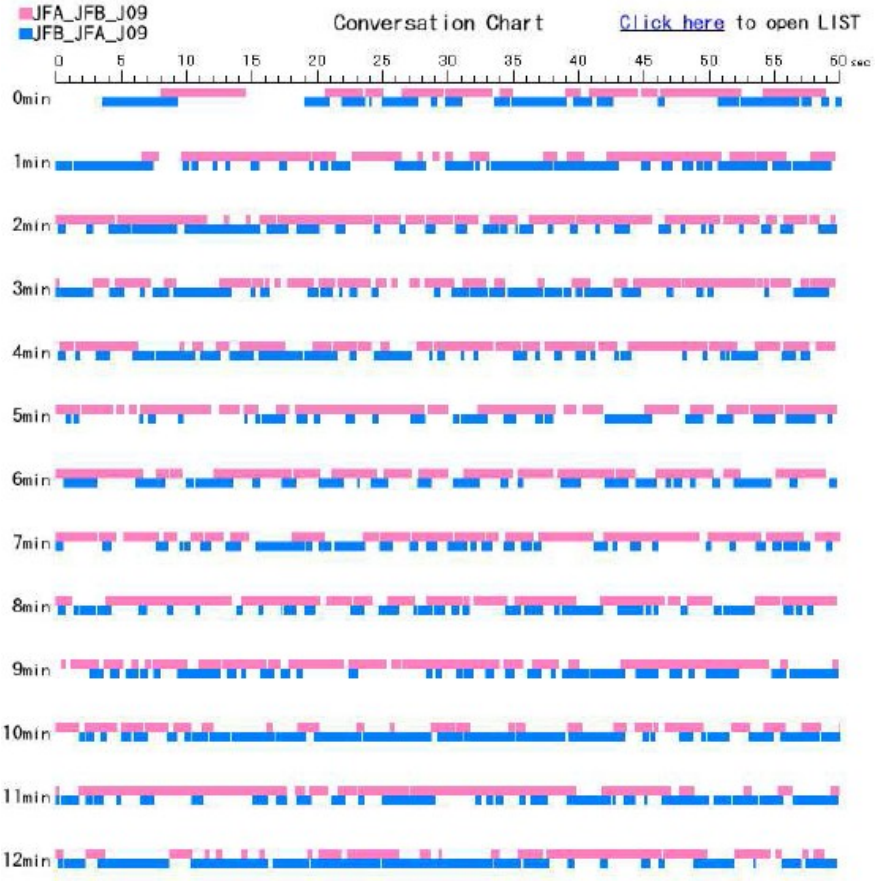
MARGARETE: **Versprich mir, Heinrich!**

FAUST: **Was ich kann!**

MARGARETE: **Nun sag, wie hast du's mit der Religion?**

Gesprochene Sprache: Dialog!

- Praxis:
  - häufige Überlappungen



(aus Campbell, 2007)

Gesprochene Sprache: Dialog!

- Faust etwas realistischer:

hör mal heinrich ich wollte dich
(sieht, nickt)

schon lange mal was emm was
fragen also mit mit der
jaa frag ruhig

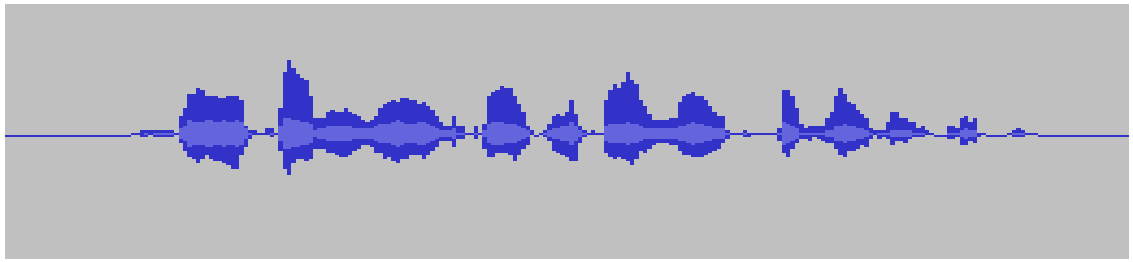
religion wie ist das eigentlich
(runzelt die Stirn)

so für dich
hmmm (lächelt)

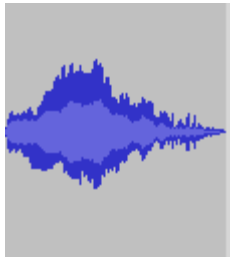
Gesprochene Sprache

- Eigenschaften
  - Dialog: man spricht **mit jemandem**
  - Wir sprechen keine Satzzeichen
  - Wir sprechen nicht nur mit Worten, sondern auch mit Intonation, mit Blicken, mit Gesten: "**multimodal**"
  - Gesprochene Äußerungen sind oft nicht wohlgeformte Sätze: Häsitationen, Wiederholungen
  - Dialog ist kein Ping-Pong-Spiel: viele Überlappungen
  - Hörer gibt Sprecher Rückmeldung: "feedback"

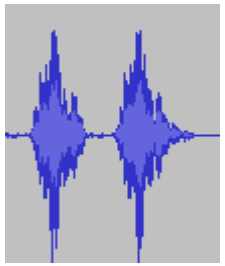
Sprachsignal



Faye Dunaway war zweimal verheiratet.

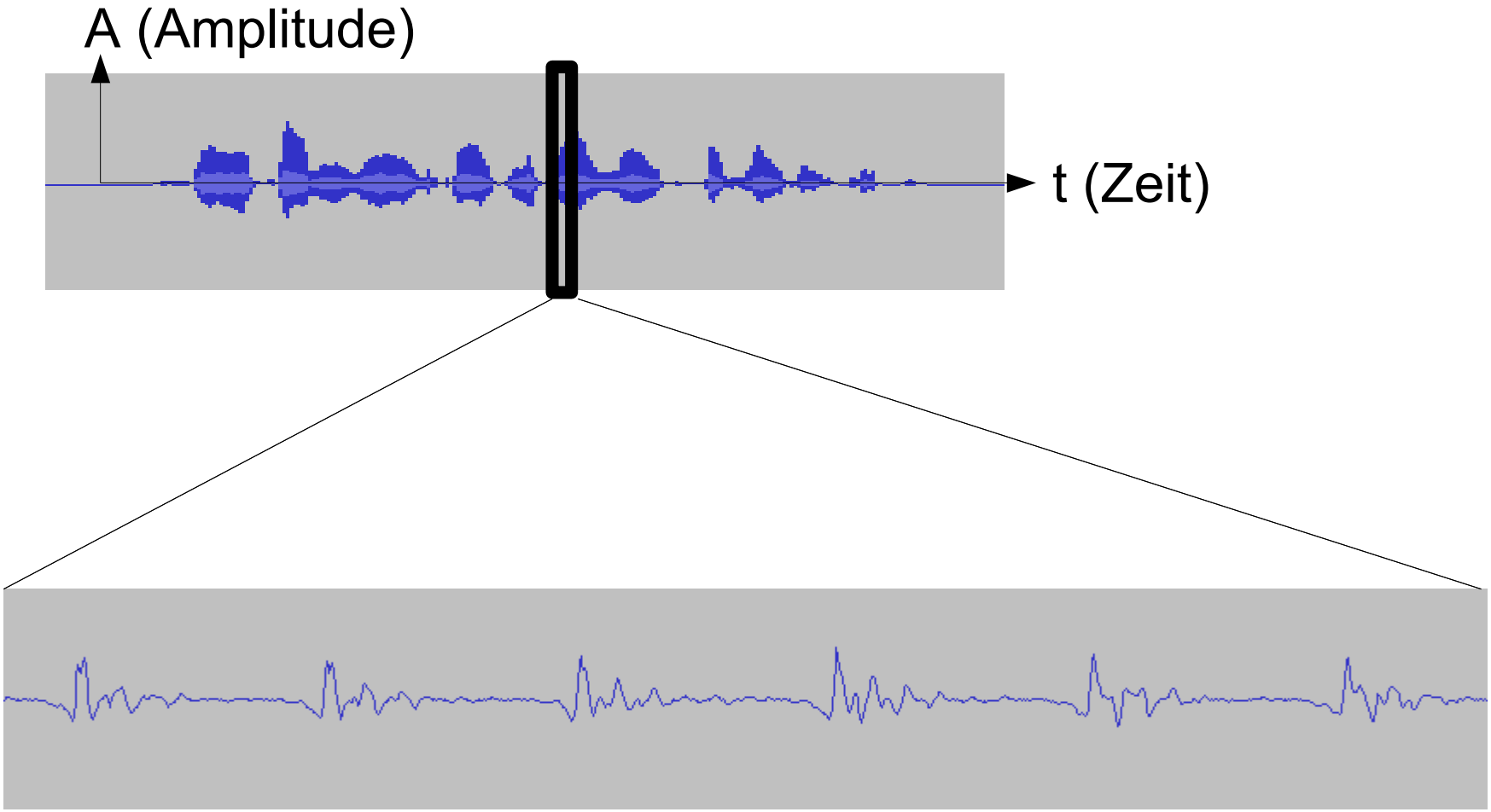


??

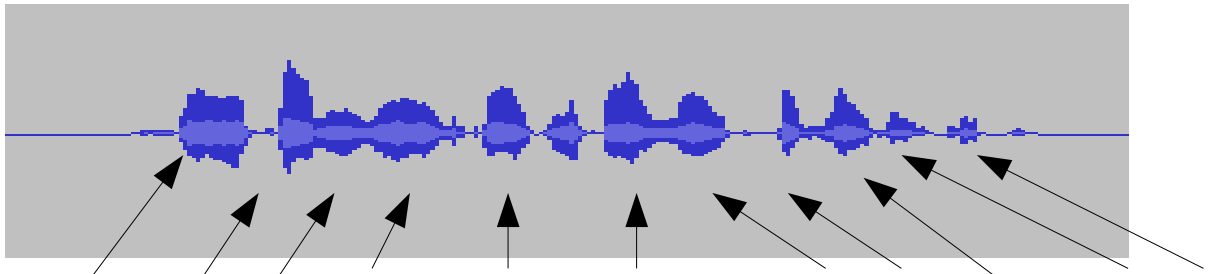


??

Sprachsignal



Sprachsignal und Lautschrift



[fɛɪ da nə wɛɪ waɐ tsvaɪ mal fɐ haɪʁa tət]

Lautschrift im IPA (International Phonetic Alphabet)

- Lautschrift lässt sich mit dem Sprachsignal alignieren – näher an der akustischen Wirklichkeit als Buchstaben
  - keine Wortgrenzen im akustischen Signal!
  - aber Silben relativ klar erkennbar

Zusammenfassung der Ausgangssituation

- Spontansprache dient zur Kommunikation
- Dialog wird multimodal von Sprecher und Hörer gemeinsam getragen
- Sprachsignal ist kontinuierlich
  - Silben sind erkennbar
- Lautschrift näher am Sprachsignal als Buchstaben

Die Herausforderung

- Was müssen Computer können, damit sie mit gesprochener Sprache genau so gut umgehen können wie der Mensch?

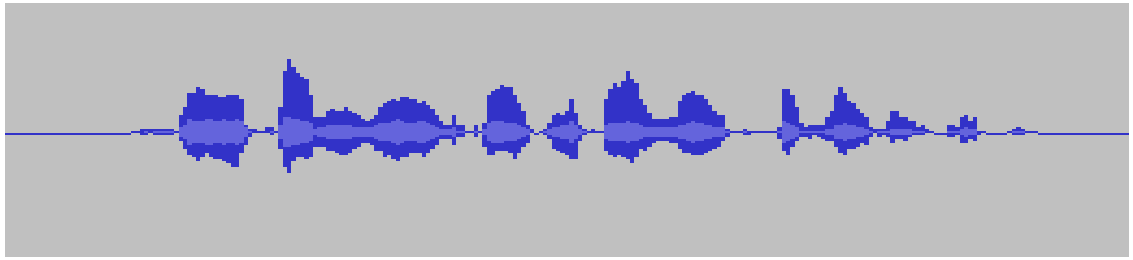
Laut Prof. R. Moore (U. Sheffield) ist gesprochene Sprache: **“the most sophisticated behaviour of the most complex organism in the known universe!”**

Maschinelle Verarbeitung gesprochener Sprache

- Sprache verstehen: Spracherkennung
- Sprache erzeugen: Sprachsynthese
- passende Antwort finden: Dialogsteuerung

- Sprecherzustand analysieren: Emotionserkennung
- Ausdrucksstarke Sprache erzeugen: Expressive Sprachsynthese
- Rückmeldung ohne Worte: nonverbales Hörerverhalten

Spracherkennung



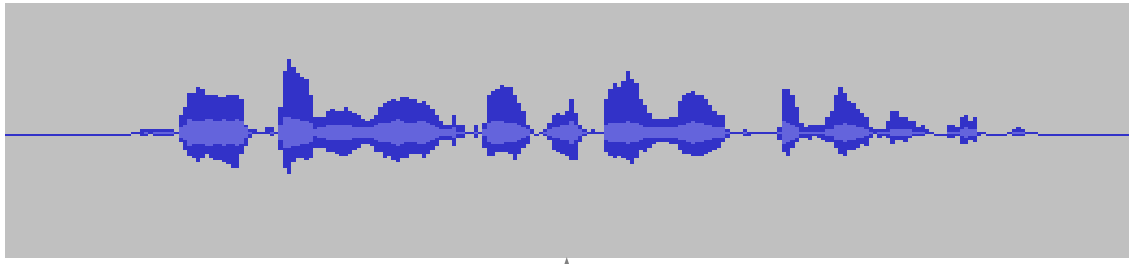
akustische Modelle

[fɛɪ da nə wɛɪ waɐ tsvaɪ mal fɐ haɪʁa tət]

Sprachmodell (language model)

Faye Dunaway war zweimal verheiratet.

Spracherkennung: akustische Modelle

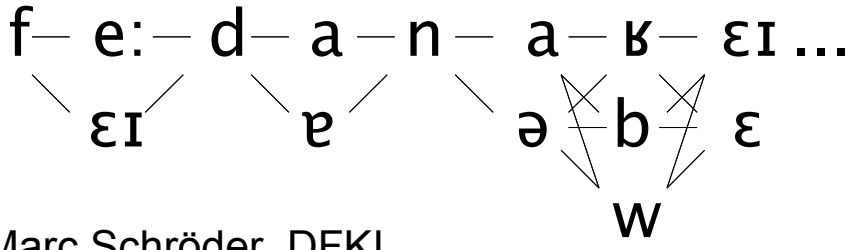


Anwendung auf das Sprachsignal:
ist dieser Teil des Sprachsignals eher
ein [a], ein [f], ein [h] etc.?

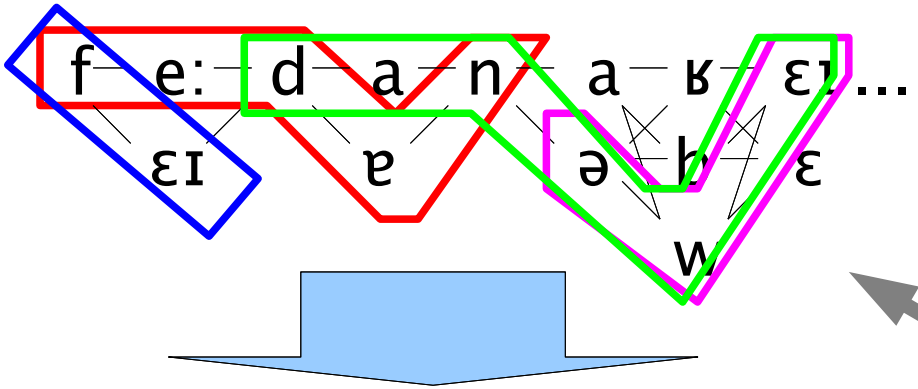
statistische Modelle für alle Laute

akustische Messungen / Durchschnittswerte und Verteilung für jeden Laut

Lauthypothesengraph:



Spracherkennung: Sprachmodell

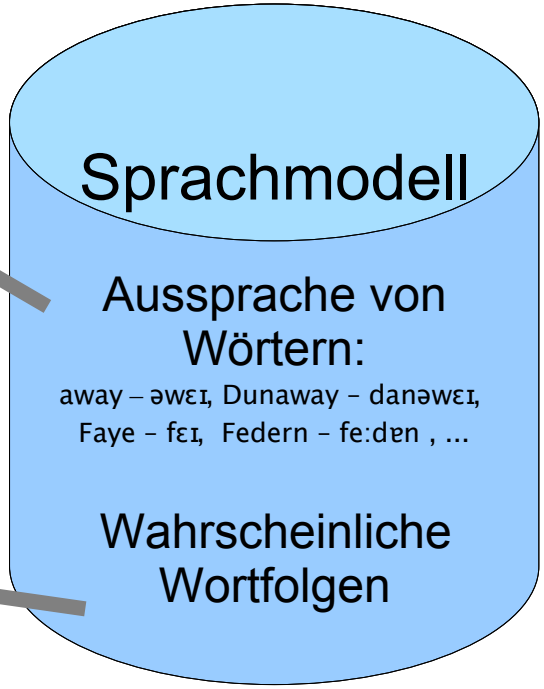


Worthypothesengraph:

Federn — away — war — zwei

Faye — Dunaway — drei

...



Faye Dunaway war zweimal ...

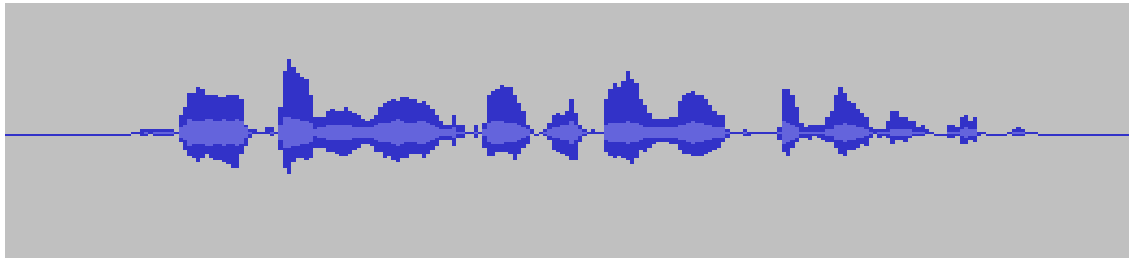
Sprachsynthese

Faye Dunaway war zweimal verheiratet.

Textanalyse

[fɛɪ da nə wɛɪ waɐ tsvaɪ mal fɐ haɪ ʁa tət]

Klangerzeugung



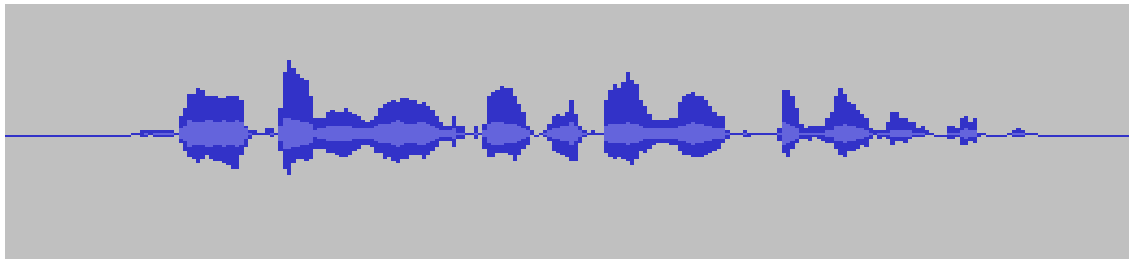
Sprachsynthese

Faye Dunaway war zweimal verheiratet.

Textanalyse

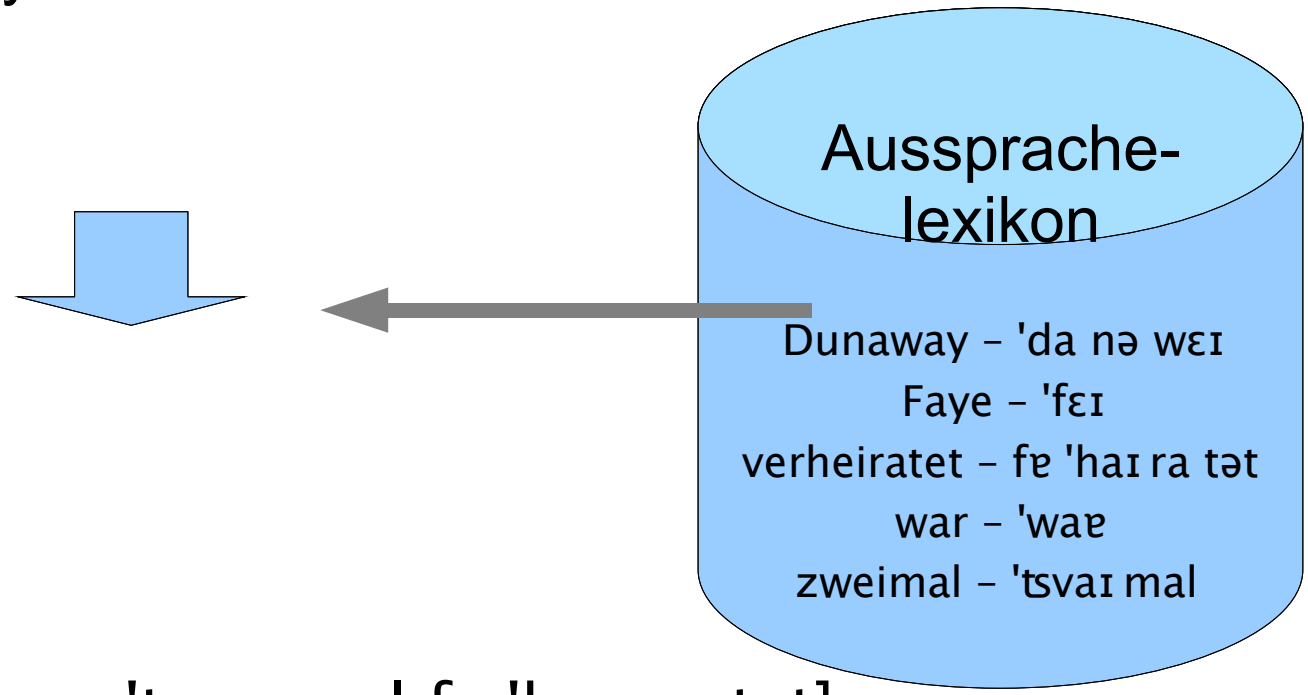
['fɛɪ 'da nə wɛɪ 'waɐ 'tsvaɪ mal fɐ 'haɪʁa tət] + Intonation!

Klangerzeugung



Sprachsynthese: Textanalyse

Faye Dunaway war zweimal verheiratet.



['fɛɪ 'da nə wɛɪ 'waɐ 'ʦvaɪ mal fɐ 'haɪ ʁa tət]

Sprachsynthese: Textanalyse

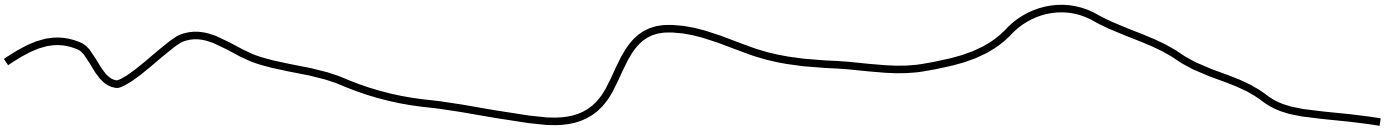
Faye Dunaway war zweimal verheiratet.

Eigenname Eigenname Hilfsverb Adverb Partizip

Wortart-Bestimmung

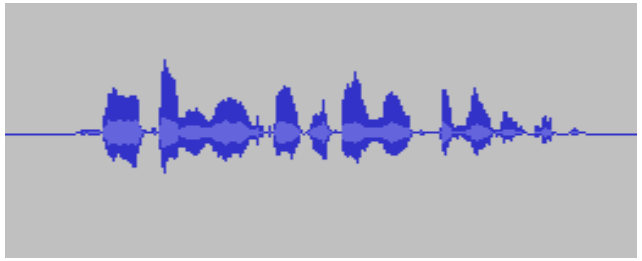
Intonation: Akzente auf die Inhaltswörter!

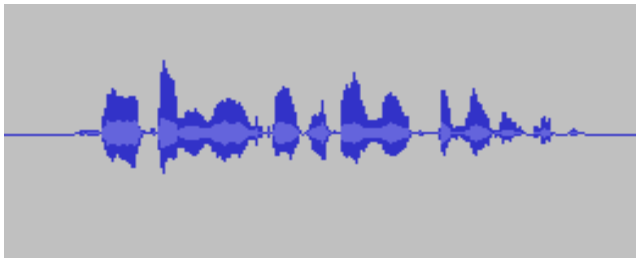
['fɛɪ 'da nə wɛɪ 'waɐ 'tsvaɪ mal fɐ 'haɪʁa tət]



Sprachsynthese: Klangerzeugung

- Konkatenative Sprachsynthese
  - klebe geeignete Sprachschnipsel aneinander
- Parametrische Sprachsynthese
  - berechne akustische Parameter und realisiere sie mit einem Vocoder





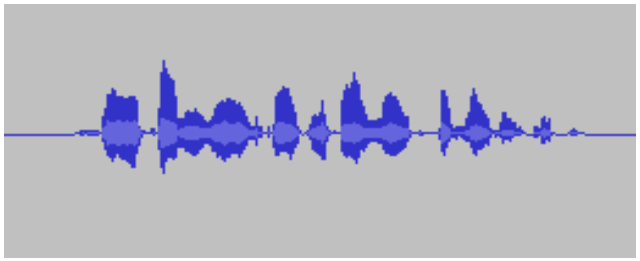
Dialogsteuerung

- Semantische Modellierung einer Domäne
  - z.B. Kauf Bahnticket:
    - wann, von wo, nach wo, zu welchem Tarif, etc.
  - im Dialog wird ein Skript abgearbeitet, bis die “Slots” gefüllt sind:
    - “Wann möchten Sie fahren?” etc.
  - Nachfragen bei Verständnisproblemen
  - Bestätigung des Verstandenen
    - (“Sie möchten also am 14.11. um 15:30 von Saarbrücken nach München mit vier Personen”)
- Nachteil: Mensch muss sich an die Struktur der Maschine anpassen

Maschinen und non-verbale Kommunikation

- State-of-the art im Mensch-Maschine Dialog:
  - Semantische Analyse basierend auf rein linguistischer Information
    - Wortbedeutung
    - Satzstruktur
    - Slot-filling
  - Dialog als “Ping-pong Spiel” organisiert
  - anstrengend, unnatürlich
- Non-verbale Kommunikation modellieren!

Emotionserkennung an Hand der Stimme



“freudig” /
“traurig” /
“ärgerlich” /
...

- Erkenner wird trainiert auf emotionalen Sprachdaten
  - schwer zu beschaffen, denn Aufnahmen mit Schauspielern eignen sich nicht für die Erkennung spontaner Emotionen (übertrieben vs. subtil)
  - typische Erkennungsraten auf Spontansprache: ca. 60% bei einem 4-Klassen-Problem (z.B. freudig/traurig/ärgerlich/neutral)

Expressive Sprachsynthese

- Klang der synthetischen Stimme an auszudrückende Emotion anpassen
  - Konkatenative Synthese: geeignete Aufnahmen machen -> klingt natürlich, aber nur im aufgenommenen Stil
    - Beispiel Fußballansage: 
  - Parametrische Synthese: akustische Parameter auf emotionalen Daten trainieren, dann miteinander “mischen”:

	trained from corpus:		trained from corpus:	
Miyanaga et al. (2004)	**neutral**	0.5 joyful interpolated	**joyful**	1.5 joyful interpolated

Non-verbales Hörerverhalten

- Menschliche Hörer signalisieren dem Sprecher, ob sie noch zuhören, ob sie der gleichen Meinung sind oder zweifeln, etc.
- Erste Experimente, um dies in Mensch-Maschine-Dialoge einzubauen
  - Hörer-Rückmeldung: 
  - Hörtests: was ist angemessen?
- Neues Projekt wird System bauen

Zusammenfassung

- Gesprochene Sprache stellt andere Herausforderungen als geschriebene Sprache
- Verbaler und non-verbaler Anteil
  - beide sind nötig für angemessene Verarbeitung gesprochener Sprache