

Sprachsynthese

Text-to-Speech Synthesensysteme im Überblick

Caren Brinckmann
7. Dez 1999

<http://www.coli.uni-sb.de/~cabr/vortraege/tts.pdf>

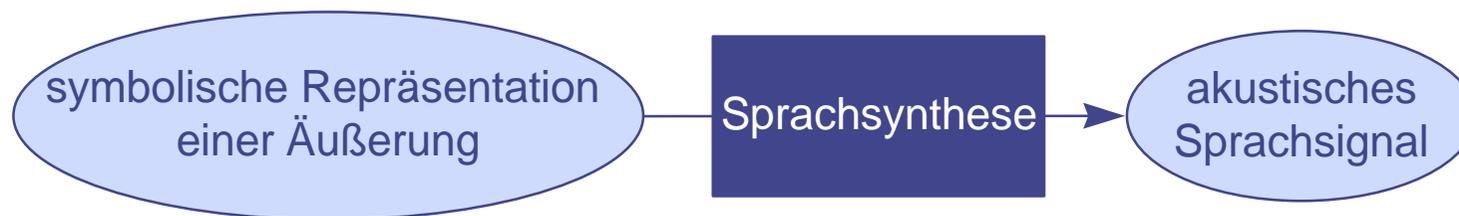


Inhalt

- Was ist Sprachsynthese?
- Wozu das Ganze?
- Ein bisschen Historie: sprechende Maschinen
- Aufbau eines TtS-Synthesystems
- Synthesetechniken (mit Hörbeispielen)

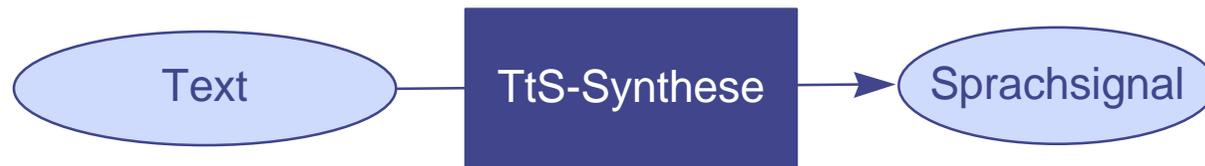
Was ist Sprachsynthese?

“Sprachsynthese beschäftigt sich mit der **maschinellen Transformation** der **symbolischen Repräsentation** einer Äußerung in ein **akustisches Signal**, welches von einem menschlichen Sprecher als der **menschlichen Sprache hinreichend ähnlich** anerkannt wird.” (D. Zboril)



Symbolische Repräsentation?

Text-to-Speech Synthese



‘Wollen wir zusammen essen gehen?’

Concept-to-Speech Synthese



Vorschlag(Sprecher, Hörer, ‘essen gehen’)



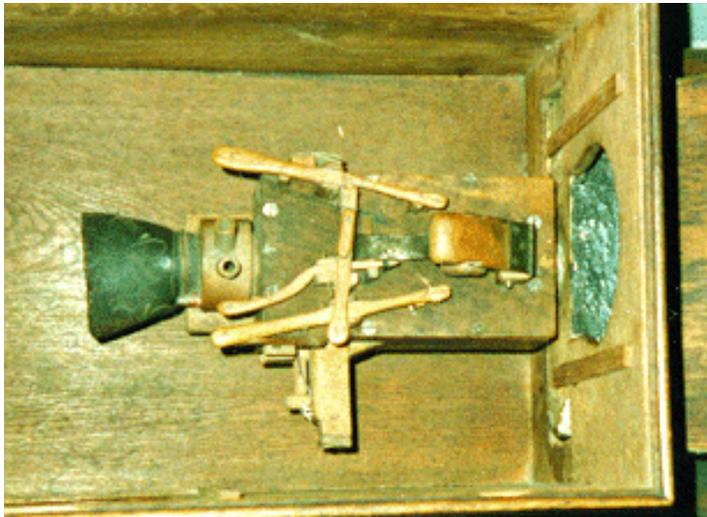
Wozu das Ganze?

- Telekommunikationsdienste
- Mensch-Maschine-Kommunikation
- Hilfe für Behinderte
- Fremdsprachenerwerb
- Forschung

Ein bisschen Historie (1)

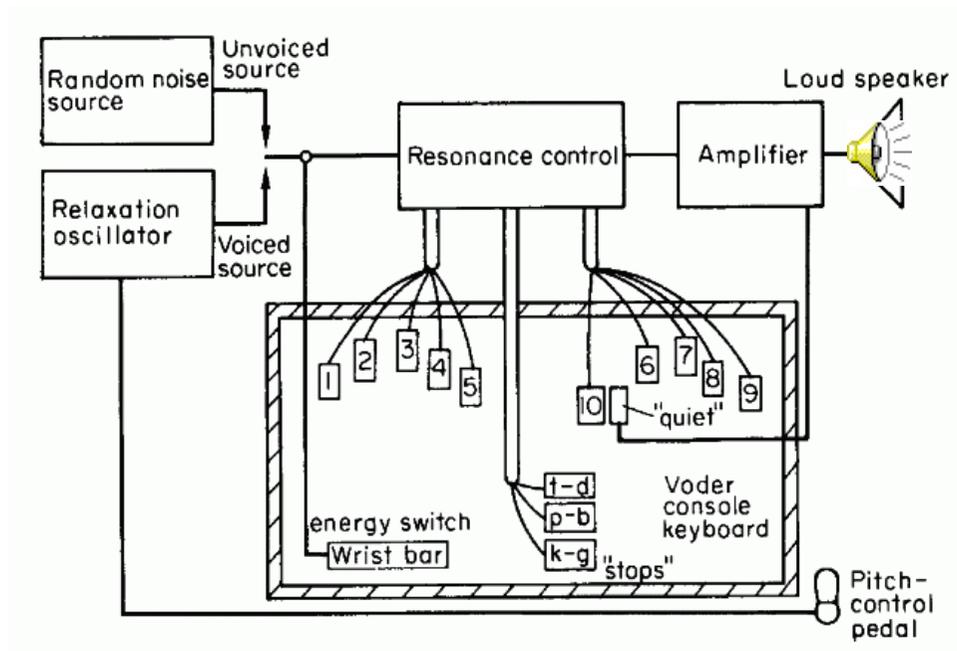
Wolfgang von Kempelens
sprechende Maschine (1791)

- von Hand zu bedienen
- einzelne Sprachlaute, Wörter
und kürzere Sätze



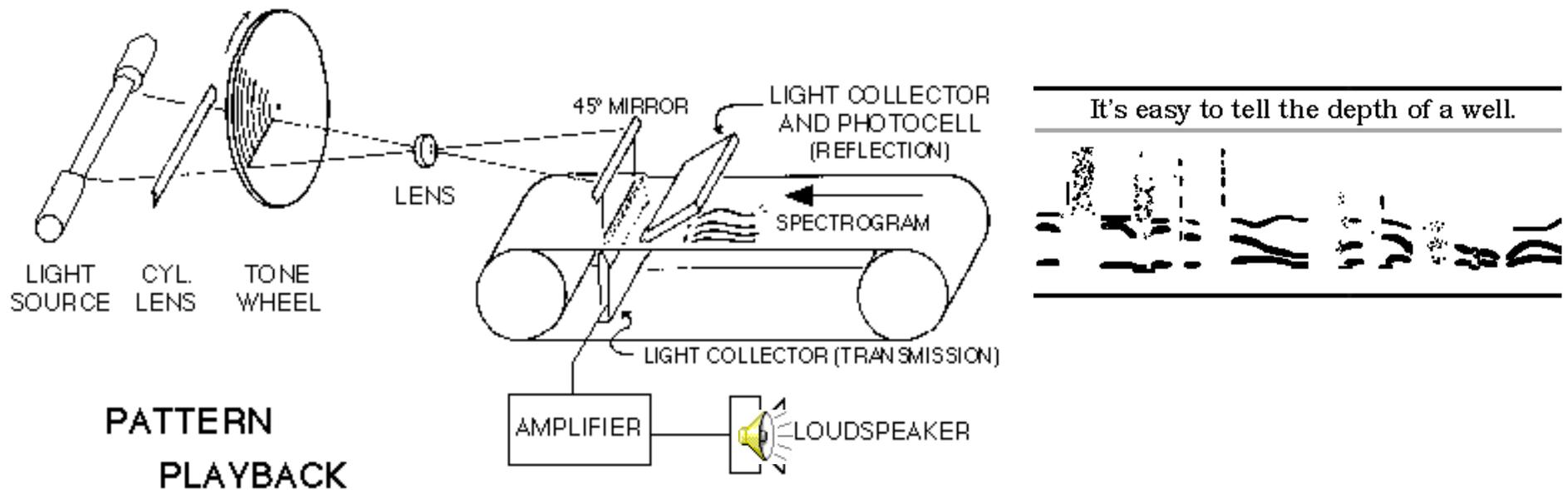
Ein bisschen Historie (2)

Homer Dudleys VODER (1939)

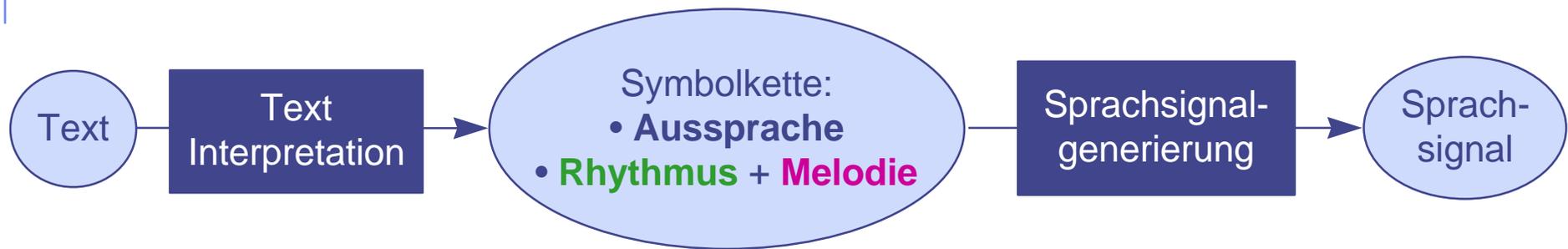


Ein bisschen Historie (3)

Frank Coopers Pattern Playback (1950)



Aufbau eines TtS-Synthesystems

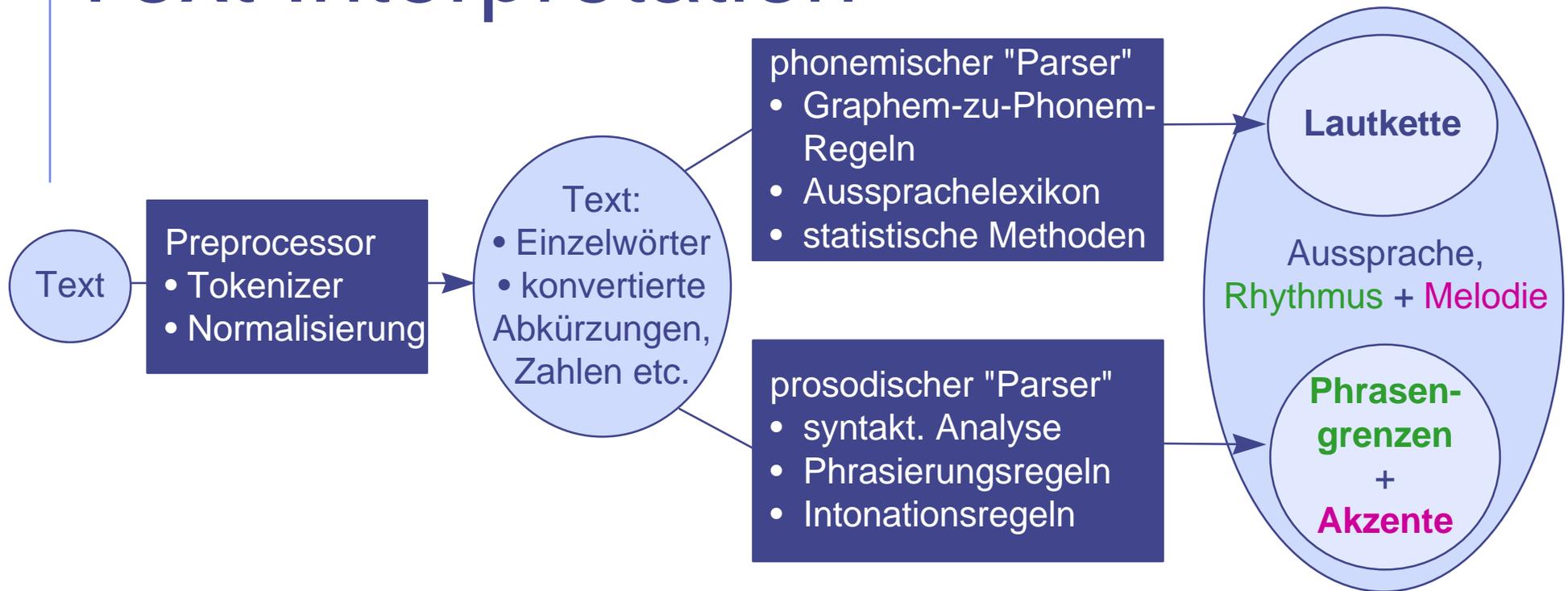


Kommst Du am
9. Dez. nach
Saarbrücken,
oder am 10.?

['kOмst du: ?am \L*+H\ 'nOYnt@n
de'tsEmb6 na:x za:6'brYk@n \H-L%\]
['o:d6 ?am \H+L*\ 'tse:nt@n \L-L%\]



Text Interpretation

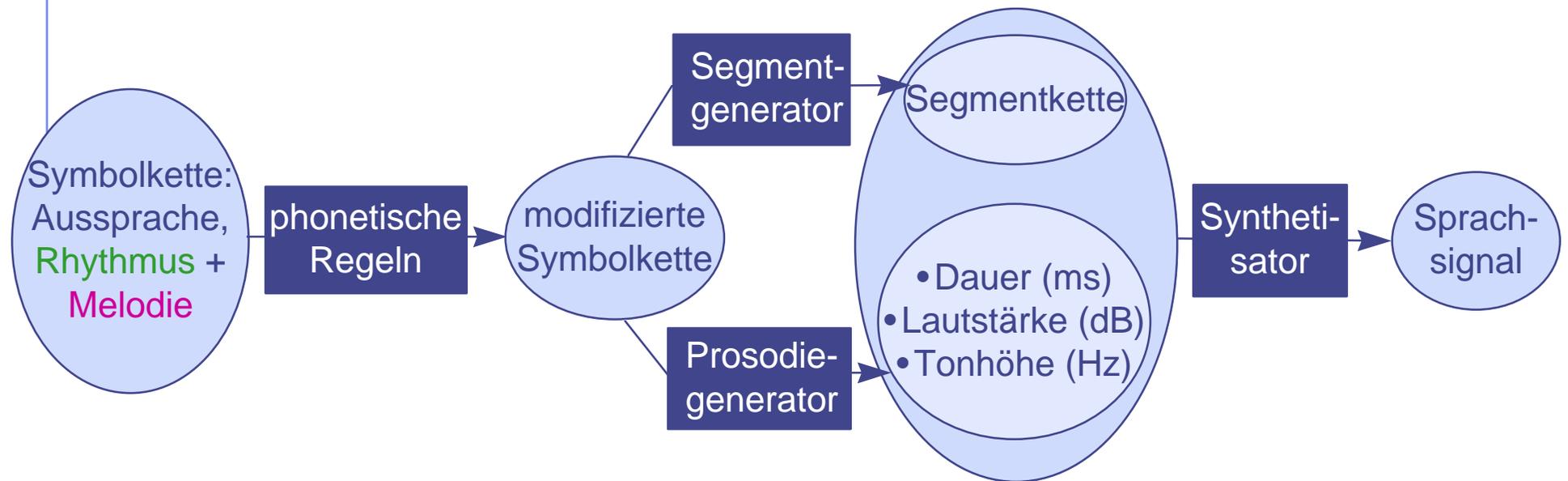


Kommst Du am
9. Dez. nach
Saarbrücken
oder am 10.?

| Kommst | Du | am |
neunten | Dezember | nach |
Saarbrücken | oder |
am | zehnten | ? |

['kOmst du: ?am \L*+H\ 'nOYnt@n
de'tsEmb6 na:x za:6'brYk@n \H-L%\]
['o:d6 ?am \H+L*\ 'tse:nt@n \L-L%\]

Sprachsignalgenerierung



['kOms̩t du:
?am \L*+H\ ...

['kOms̩uam
\L*+H\ ...

[k]O k[O]m ...
30ms 60ms ...
- 230 Hz ...



Synthesetechniken

- artikulatorische Synthese
 - wie produziert der Mensch Sprache?
- Formantsynthese
 - wie hört sich menschliche Sprache an?
- konkatenative Synthese
 - natürliche Sprachschnipsel aneinander gehängt

Artikulatorische Synthese

physikalisches Modell

- von Sprachlauten und ihren Übergängen
- basierend auf
 - Physiologie der Sprachproduktion
 - Physik der Lauterzeugung im Vokaltrakt

→ hauptsächlich zu Forschungszwecken

Bsp.: Uni Köln



Formantsynthese

- akustisch-phonetischer Ansatz
- modelliert die wesentlichen akustischen Merkmale des Sprachsignals

Beispiele:

- Uni Duisburg  Begrüßung
- Infovox (Telia)  Demo

Konkatenative Synthese

natürliche Sprachdaten

- handsegmentiert

- Diphone  BellLabs  Festival  Elan
- Mikrosegmente  Logox
- gemischtes Inventar  Hadifix  RealSpeak (L&H)

- korpusbasiert

- non-uniform unit-selection  CHATR

Ausblick

- regelbasiert (deterministische wenn-dann-Regeln)
vs. korpusbasiert (statistisch)
- Prosodiemodellierung → Natürlichkeit
- Sprechercharakteristika
- Emotionen