



SPRACHSYNTHESE

EINE PRÄSENTATION VON FELIX FELDMANN UND FABIEN
DALLMANN

GLIEDERUNG


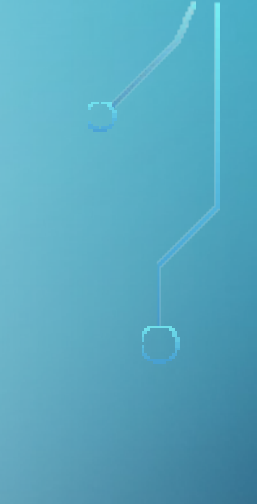
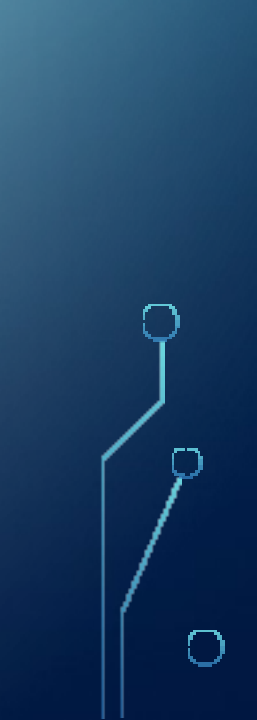
- Was ist Sprachsynthese?
- Geschichtlicher Hintergrund
- Anwendungsbereiche
- Funktionsweise eines TTS-Systems
- Praktische Vorführung eines Systems
- Quellenverzeichnis

WAS IST SPRACHSYNTHESE?

- Künstliche Erzeugung menschlicher Sprache, z.B.:
- Ziele:
 - Klang so natürlich wie möglich
 - Natürliche Satzmelodie / Intonation



• Schwierigkeiten:

- Zahlen- und Buchstabenkombinationen
 - Silbenübergänge
 - Richtige Betonung
 - Semantische Zusammenhänge
 - Sprechpausen
 - Sprechtempo
- 
- 
- 

GESCHICHTLICHER HINTERGRUND

- Erste Versuche in der zweiten Hälfte des 18. Jahrhunderts
- Christian Kratzenstein
- An Orgelpfeifen angeschlossene Resonanzrohre konnten die Laute a, e, i, o, u erzeugen
- Ganze Worte oder Sätze noch nicht möglich

GESCHICHTLICHER HINTERGRUND

- 1791: Wolfgang von Kempelen
- Erfindung der „sprechenden Maschine“
- Nun auch Plosive und Frikative möglich
- Orientierung an menschlichen Sprachorganen
- Blasebalg stellte Lunge dar, die mit dem rechten Arm betätigt wird
- Gegengewicht sorgt für Luftzufuhr („Atmen“)
- Aufschlagendes Rohrblatt simuliert Lippen
- Leerer Gummitrichter simuliert Mund; durch Verändern der Gummiabdichtung werden die Töne beeinflusst
- Weitere Öffnung simuliert Nase (wenn keine Nasale produziert werden, wird sie verschlossen)

GESCHICHTLICHER HINTERGRUND

- 1835: Joseph Faber
 - „Euphonia“: modifizierte Version von Kempelens Erfindung
 - Bedienung erfolgte nicht mehr per Hand, sondern über eine Klaviatur
 - Konnte „God save the Queen“ singen
-
- 1930er
 - „Vocoder“
 - elektronische Töne möglich
 - Bsp.: Das Lied „Mr. Blue Sky“ (wenn nicht bekannt, unbedingt anhören)

GESCHICHTLICHER HINTERGRUND

- 1939:
 - Verbesserte Version der „Vocoder“ – der „Voder“ von Homer Dudley
 - Produktion von Sprachschall auf elektrischem Wege + klarere Intonation möglich

- 1990:
 - „The Talking Machine“ von Martin Riches
 - 23 Rohre – ein Rohr steht für einen Sprechton
 - Ventile, die für Luftstrom sorgen, sind an Computer angeschlossen, über den die Eingaben erfolgen
 - Bsp.: <https://www.youtube.com/watch?v=WCIZcQo9I6Q>

ANWENDUNGSBEREICHE

- Telefonie □ Auskunftssysteme



- Informationssysteme wie Navigation

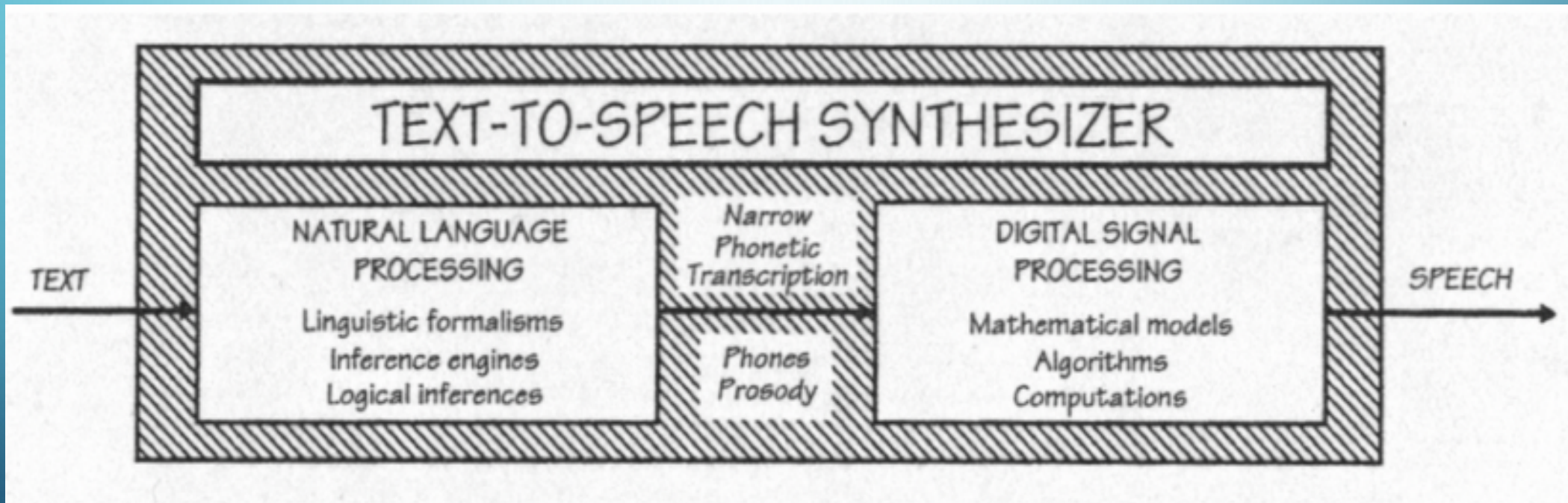


- für Sehbehinderte (Vorlesen der Mails, Websites, ...)

- Künstliche Stimme für Sprechbeh



FUNKTIONSWEISE EINES TTS-SYSTEMS



DIE NLP-KOMPONENTE

- Steht für Natural Language Processing
- Umwandlung von Orthographie in phonetische Zeichen mit ggf. zusätzlichen Informationen über Prosodie z.B Wort-und Satzintonation
- 2 verschiedene Möglichkeiten
 1. Der regelbasierte Ansatz
 2. Der lexikonbasierte Ansatz

DER REGELBASIERTE ANSATZ

- Phonetische Repräsentation durch Regeln zur Graphem-Phonem-Umwandlung
- Liste mit Ausnahmewörter z.B Fremdwörter und Abkürzungen
- Eignet sich besonders für regelmäßige Schriftform-Laut-Entsprechungen wie z.B Spanisch und Französisch

DER LEXIKONBASIERTE ANSATZ

- Phonetische Repräsentation durch ein Lexikon
- Vollformlexika enthalten alle Formen eines Wortes (Speicheraufwendig)
- Alternativ spezielle Wortlisten
- Regeln für Wörter, die nicht im Lexikon enthalten sind
- Eignet sich für Englisch

BEISPIELSYSTEM DER UNI BONN

- Lexikon mit ca. 5000 Einträgen + Flexionstabellen für Deutsch

Gelesen werden kann, falls möglich

S 105 18 176 37 172 56 167

Dauer in Millisekunden

p 90 8 153

Pitchmarker

a: 104 4 150 100 145

s 71 28 145 85 143

DIE DSP-KOMPONENTE

- Steht für Digital Signal Processing
- Erzeugt das akustische Ausgabesignal anhand der überführten phonetischen Zeichen
- 3 verschiedene Realisierungsmöglichkeiten
 1. Artikulatorische Synthese
 2. Formantsynthese
 3. Konkatentative Synthese

ARTIKULATORISCHE SYNTHESE

- Sprachsynthese durch modellieren von artikulatorischen Prozessen
- Sprachsynthesesoftware
- Rechenaufwendig, nicht für die Echtzeitanwendung geeignet

FORMANTSYNTHESE

- Vollständig synthetisch hergestellt
- Anhand von „Quelle-Filter-Modell“
 - Mix aus periodischer Signalquelle (stimmhafte Laute) und Rauschquelle (stimml. Laute)
- Klingt unnatürlich, bietet jedoch ein großes Einsatzgebiet

KONKATENATIVE SYNTHESE

- Sprachsynthese durch das Verketteten von Lautsegmenten
- Segmentdatenbank
- verständliche Sprache

The background is a blue gradient with decorative circuit-like lines in the corners. The lines are light blue and form various geometric shapes, including straight lines, right angles, and circles, resembling a printed circuit board (PCB) layout. The lines are most prominent in the top-left and bottom-left corners, and less so in the top-right and bottom-right corners.

PRAKTISCHE VORFÜHRUNG EINES SYSTEMS

QUELLENVERZEICHNIS

- <https://www2.ling.su.se/staff/hartmut/kempln.htm>
- https://lehre.idh.uni-koeln.de/site/assets/files/2600/sprachsynthese_referat.pdf
- <https://curdt.home.hdm-stuttgart.de/PDF/Sprachsynthese.pdf>
- <https://www.techfak.uni-bielefeld.de/ags/wbski/lehre/digiSA/ECA/Ausarbeitungen/Termin7.pdf>