

Sprachsynthese - ein Überblick

Florian Schiel

SPRACHSYNTHESE

Definition: Produktion von Sprache
mit technischen Mitteln

Gegensatz: Synthese vs. Wiedergabe
(Übergang fließend)

SPRACHSYNTHESE

Einführende Literatur:

Sammy Lemmetti:

„Review of Speech Synthesis Technology“

www.acoustics.hut.fi/~slemmet/dippa/

(Kapitel 5)

Thierry Dutoit:

„A Short Introduction to TTS synthesis“

tcts.fpms.ac.be/synthesis/introtts.html

SPRACHSYNTHESE

(Mechanische Verfahren)

(Elektronische Verfahren)

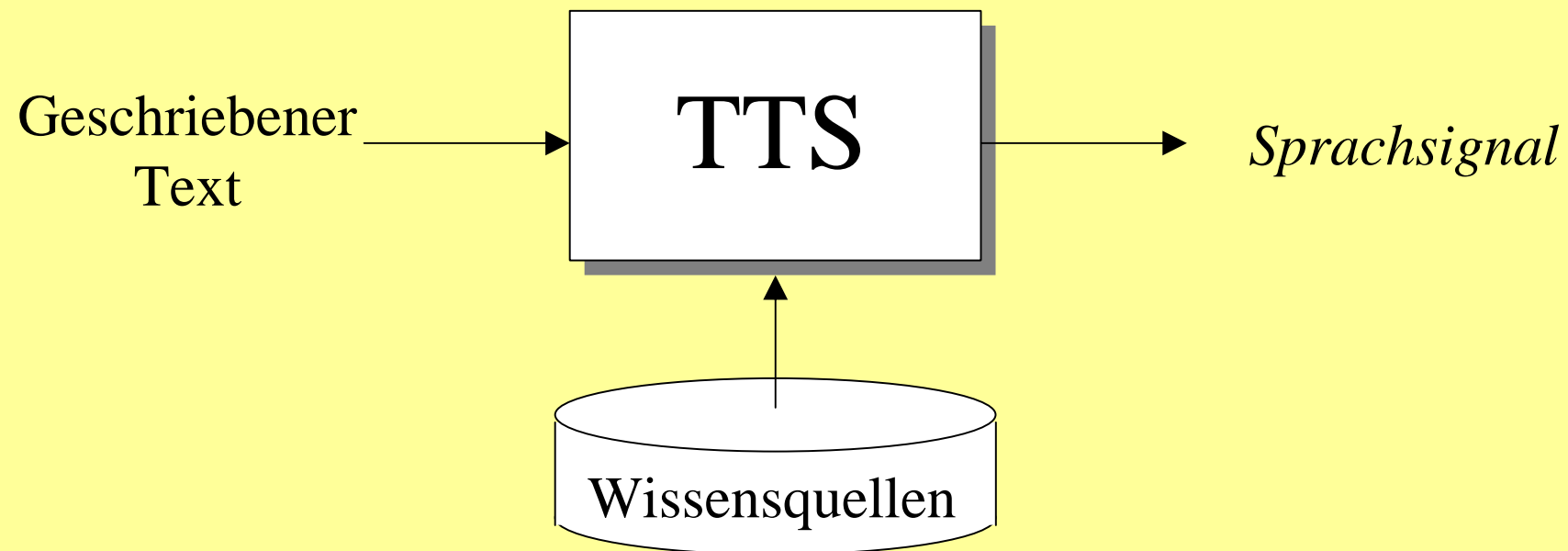
Artikulatorische Synthese

Formant-Synthese, parametric synthesis

Konkatenative Synthese, copy synthesis

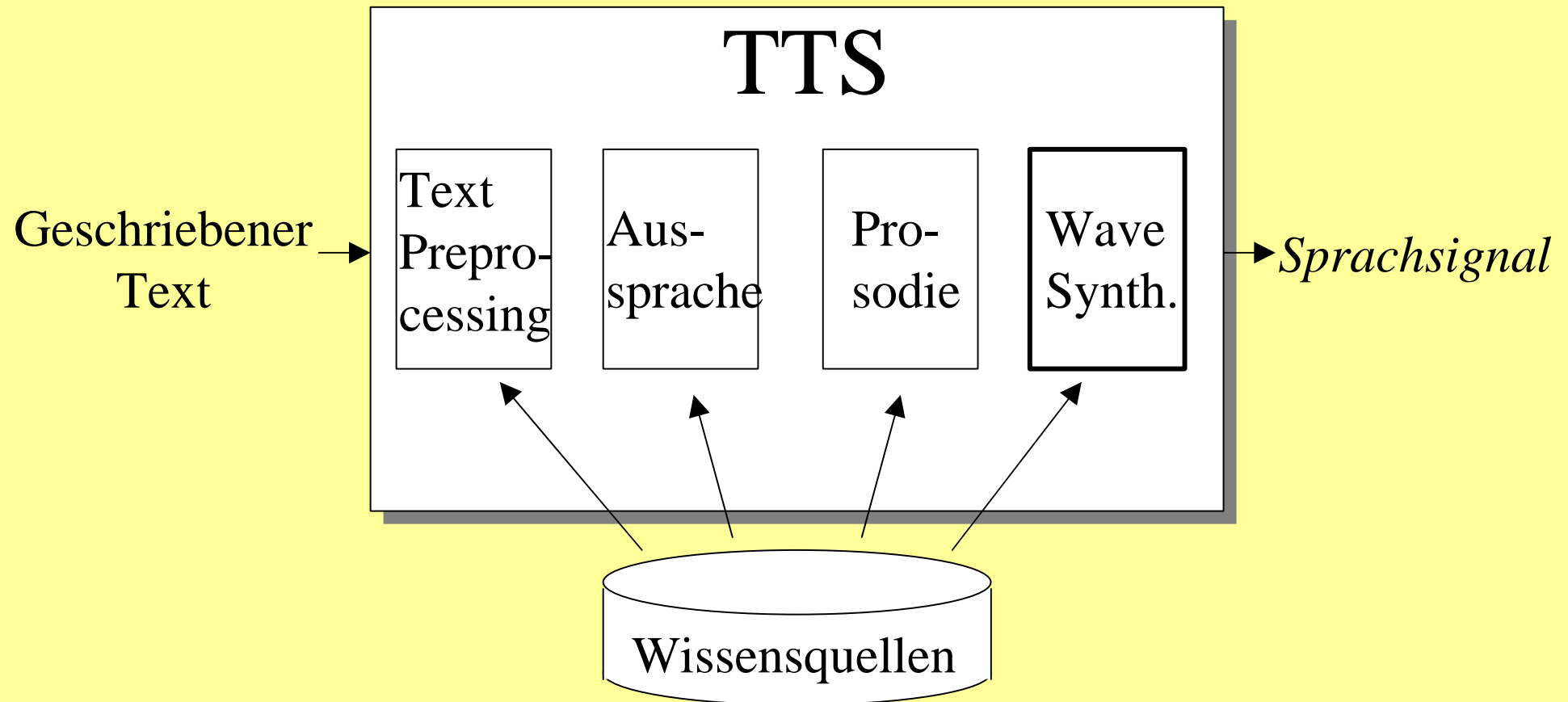
SPRACHSYNTHESE

„Black Box Model“



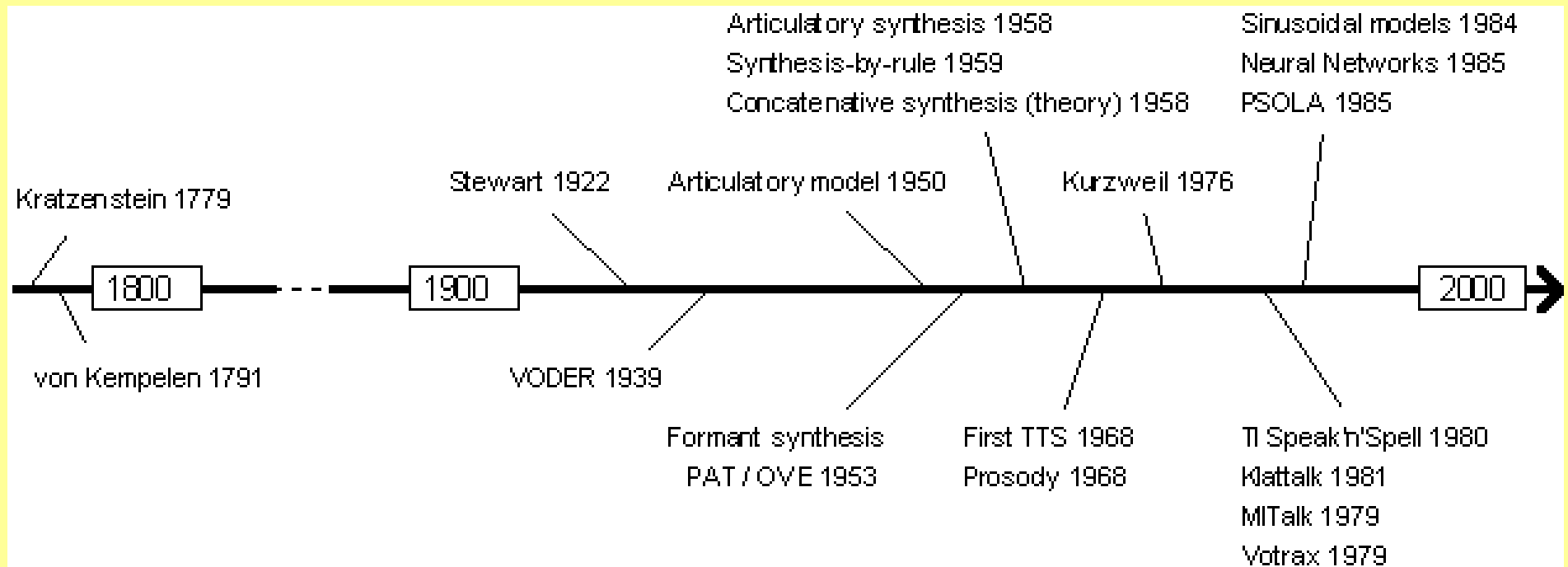
SPRACHSYNTHESE

Verfeinertes ‚Black Box Model‘



SPRACHSYNTHESE

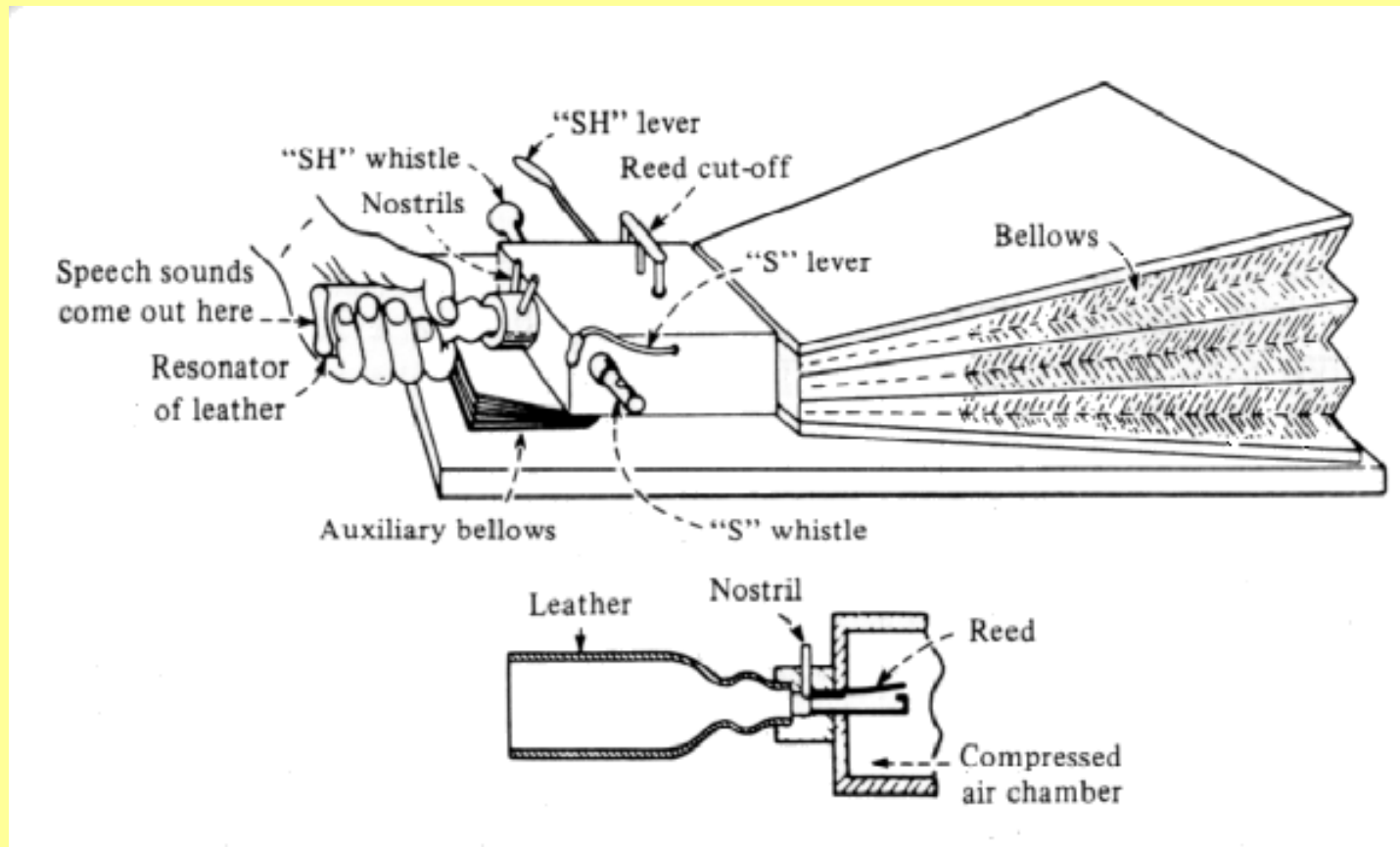
Historischer Überblick



Aus: S. Lemmetti, Kap. 2

SPRACHSYNTHESE

Mechanische Verfahren: van Kempelen (1791)

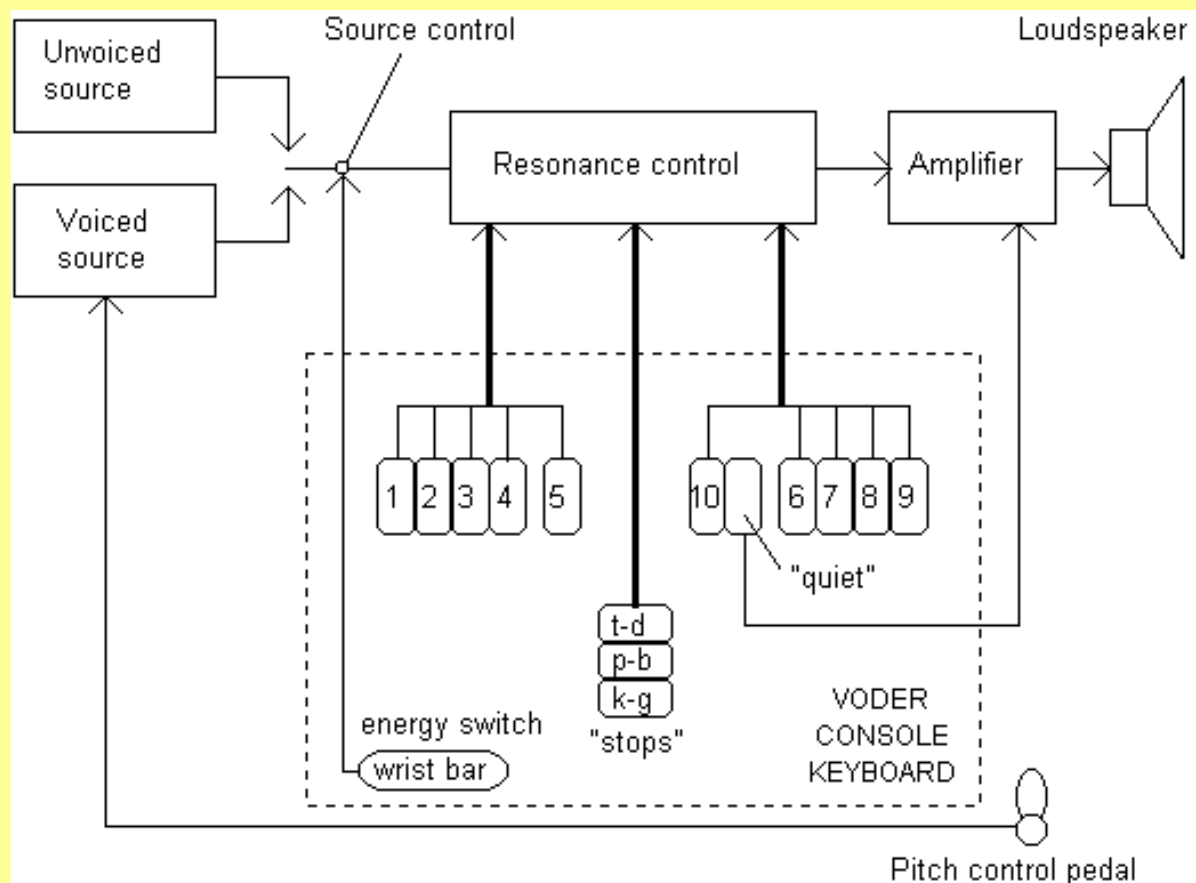


Aus: Flanagan 1972

SPRACHSYNTHESE

Elektronische Verfahren:

Homer Dudley,
1939: VODER



SPRACHSYNTHESE

Elektronische Verfahren:

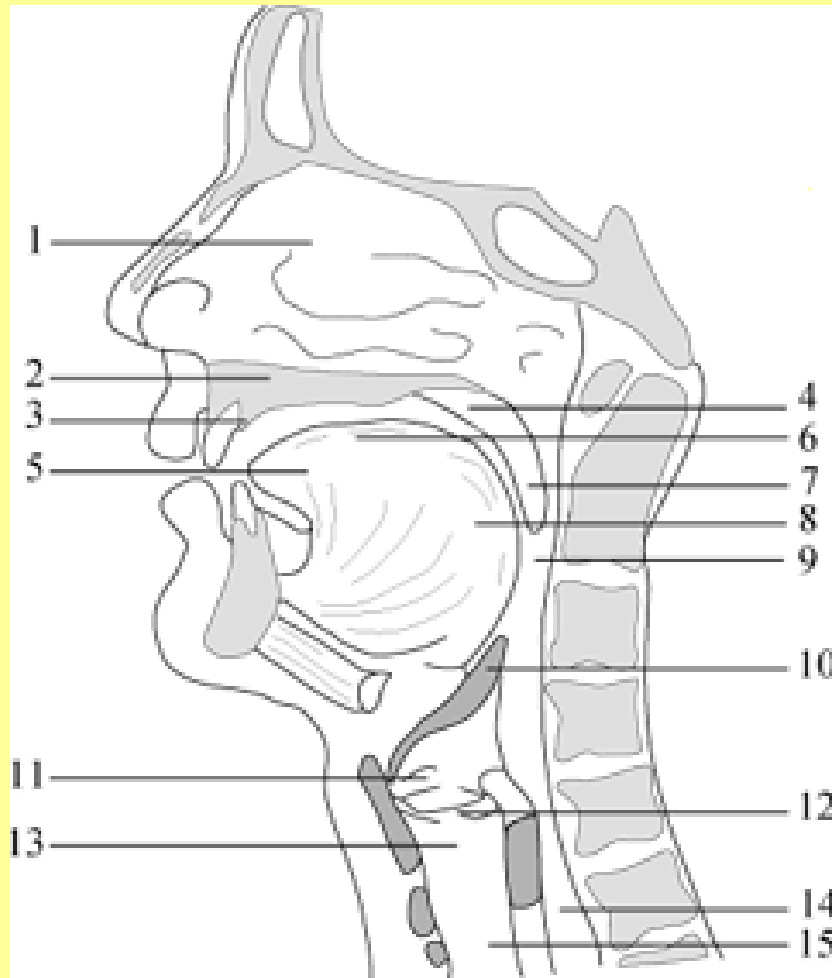
Homer Dudley

1939: VODER



SPRACHSYNTHESE

Artikulatorische Synthese: Modellierung des Verhaltens der Sprechorgane und dessen Auswirkung auf die Akustik

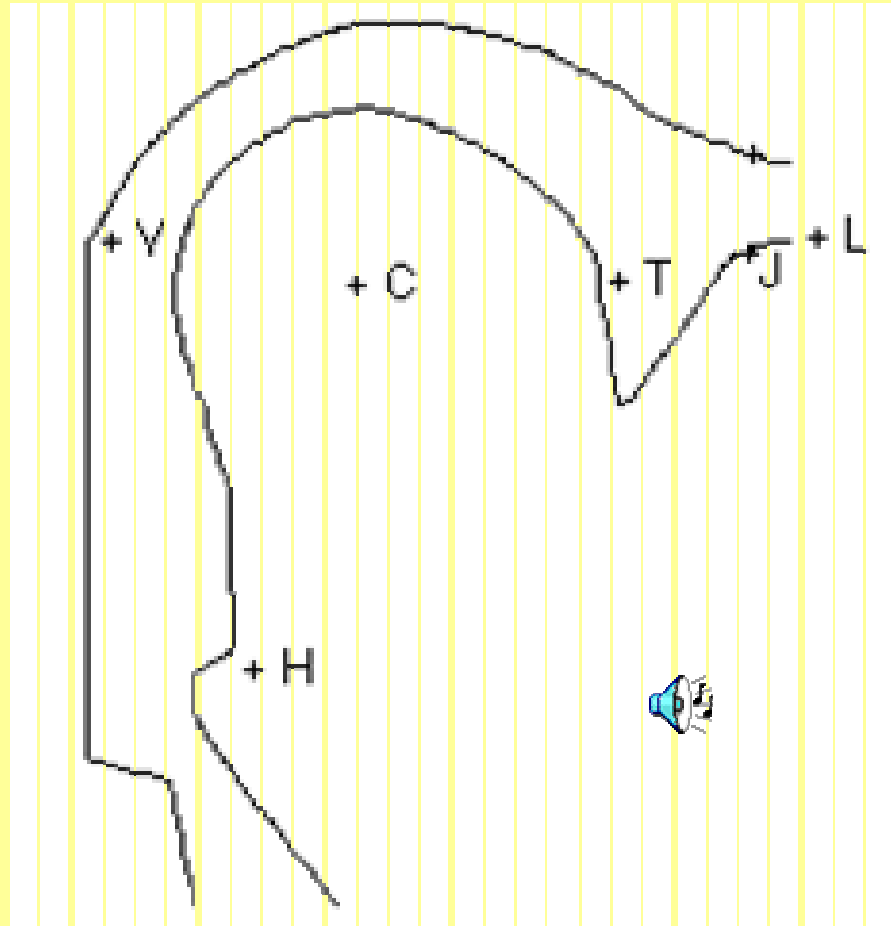


⑩ Model der schwingenden Stimmlippen

⑩ Model des Vokaltrakts (mit Nase und Lippen)

SPRACHSYNTHESE

Artikulatorische Synthese: Beispiel ASY



Probleme:

- ⑩ sehr rechenintensiv
- ⑩ wenig Messdaten über 3D des Vokaltrakts
- ⑩ Akustische Modellierung komplex; daher oft nur mit Näherungen
- ⑩ in der Praxis bis jetzt nicht eingesetzt

Aus: Haskins Web-Page

SPRACHSYNTHESE

Format-Synthese, Parametrische Verfahren:

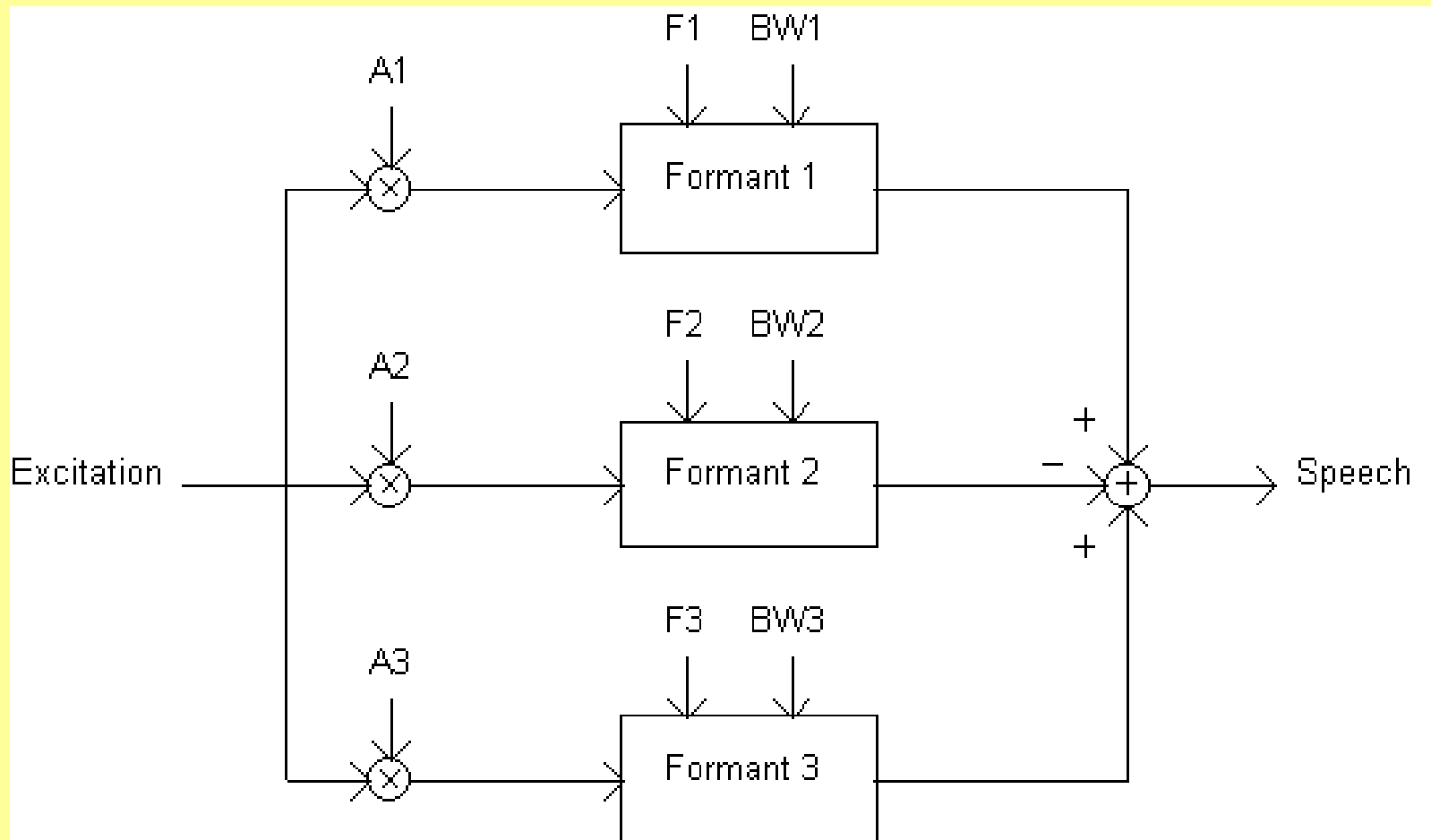
Modellierung nach dem ‚Quelle-Filter-Modell‘ (Fant 1970) durch Parameter die jeweils Quelle und Filter zu jedem Zeitpunkt beschreiben.

F0, OQ, VO (Quellenparameter)

F1, F2, F3 jeweils Frequenz, Amplitude und Bandbreite (Filterparameter)

SPRACHSYNTHESE

Format-Synthese, parametric synthesis:



SPRACHSYNTHESE

Format-Synthese, parametric synthesis:

- + volle Kontrolle
- + jede Stellung der Sprechorgane
- Quelle/Filter-Modell ist nur Näherung
- Modellierung der Anregung
- Modellierung von mehrfachen Quellen
(z.B. stimmhafte Frikative) schwierig

SPRACHSYNTHESE


Konkatenative Synthese:

Synthese durch Verknüpfung von Sprachsignalstücken, Stimmqualität nicht modelliert

- ⑩ **General** concatenative synthesis
- ⑩ Corpus based / **unit selection** synthesis
- ⑩ **Phrase splicing** / limited domain synthesis

SPRACHSYNTHESE

Konkatenative Synthese: phrase splicing

- ⑩ Kleines Inventar, üblicherweise Wörter
- ⑩ Sehr begrenzte Anwendung:
Sprechende Uhr, Navigationssystem, Mailbox
- ⑩ Problem: Inventar so auswählen/manipulieren,
dass beliebige Verkettungen immer natürlich
klingen
- ⑩ Beispiele 

SPRACHSYNTHESE


Konkatenative Synthese: **unit selection**

- ⑩ Großes Inventar aus natürlich-sprachlichem Korpus: Halbphone, Phone, Diphone, Halbsilben
- ⑩ Segmentierung und prosodische Labellung der Segmente
- ⑩ Auswahl der prosodisch ‚richtigen‘ Segmente
- ⑩ Beispiele: CHATR (N. Campbell), AT&T



SPRACHSYNTHESE

Konkatenative Synthese: general concatenative s.

⑩ Mittelgroßes Inventar aus kontrolliert gesprochenem Korpus: meistens Diphone 

⑩ Segmentierung, jede Einheit kommt nur einmal vor (i.G. zur ‚unit selection‘)

⑩ Prosodische Manipulation der Einheiten, TD-PSOLA, MBROLA bei der Verkettung

⑩ Beispiele: IMS-Festival, HADIFIX 

KONKATENATIVE SPRACHSYNTHESE

	Phrase Slicing	Unit Selection	General
Einheiten	Wörter, Phasen	Phone, Diphone, Halbsilben, Wörter	Diphone, Halbsilben
Anzahl d. Einheiten	< 200	> 10000	2500 - 3500
Abdeckung	feste Satzstruktur, enge Domäne	hängt von der Größe des Korpus ab	volle Abdeckung
Sprachaufnahme	Kontrollierte Sätze	Natürliche Sprache	Logatome
Nachbearbeitung	Kontrolle der Schnittstellen, ev. prosodische Manipulation	Segmentierung, prosodische Labellung	Segmentierung
Manipulation bei Verkettung	keine	keine	F0, Energie, Dauer
Natürlichkeit	++	+	-

SPRACHSYNTHESE-KORPUS BITS

Logatom Spr. 2002 Headset



Logatom Spr. 2002 Raum



Logatom Spr. 2002 Laryngo



Logatom Spr. 2001 Headset



Logatom Spr. 2003 Headset



Logatom Spr. 2004 Headset



Einfache Verkettung Spr. 2002 Headset



Einfache Verkettung Spr. 2002 Raum



Einfache Verkettung Spr. 2001 Headset



www.ims.uni-stuttgart.de/~moehler/synthspeech/examples.html