

Spraaksynthese



Fasen in tekst-naar-spraak

- Tekstbewerking
- Grafeem-foneemomzetting
- Melodie en ritme
- Synthese

Spraaksynthese

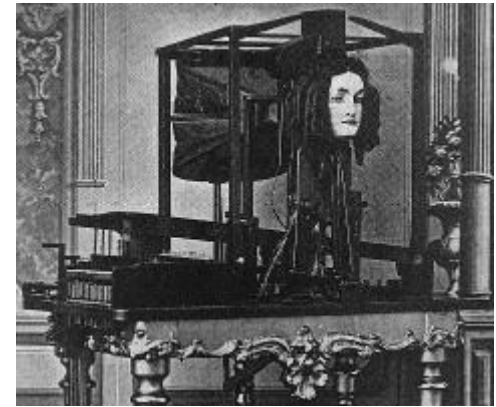
- Het maken en aaneenrijgen van spraakklanken
 - Klank(verloop)
 - Toonhoogte(verloop)
 - Intensiteit(verloop)
 - Duur(verloop)

Methoden

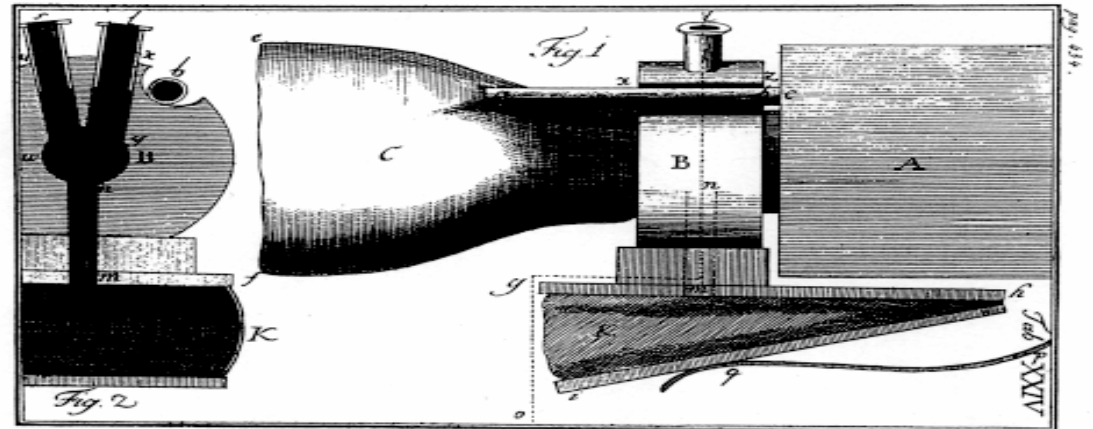
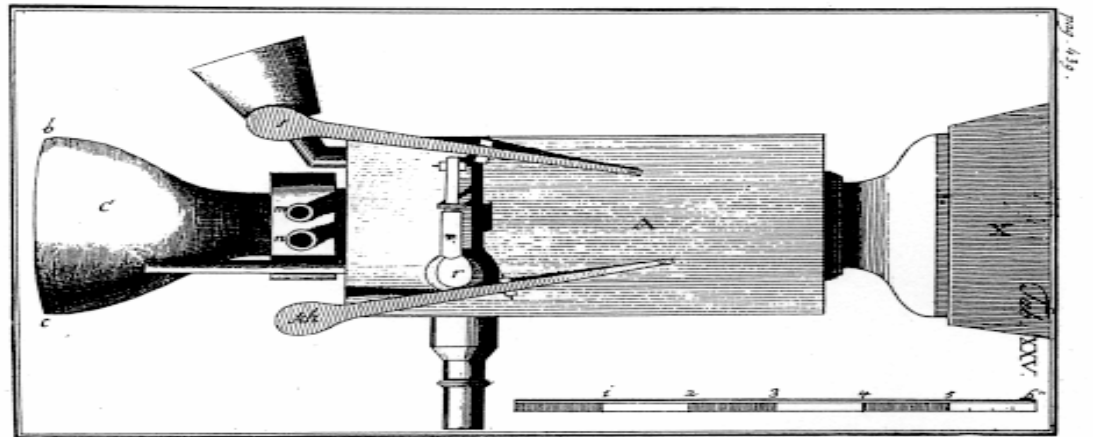
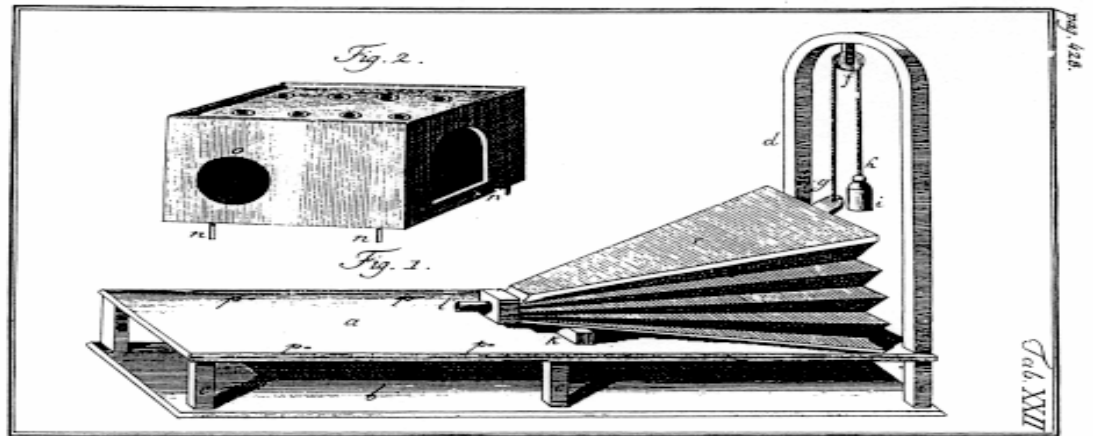
- Mechanisch
- Elektro-akoestisch
- MODEL van de spraakproductie
 - volledig volgens regels
- BOUWSTENEN
 - kant en klare spraakelementen

Mechanisch/articulatorisch

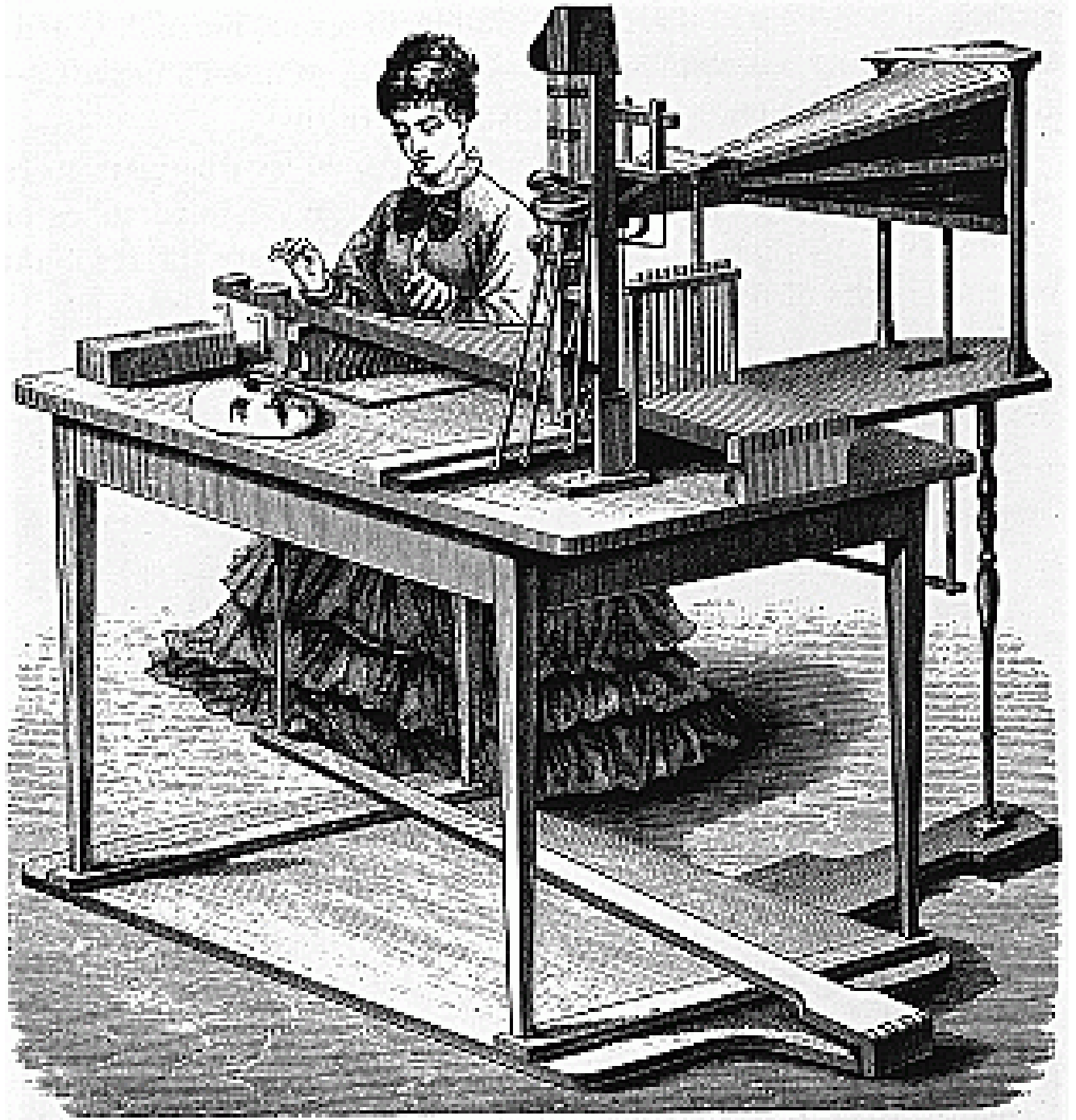
- Historisch
- Stembron
 - trilbron – toonhoogte en luidheid
 - ruisbron
- Stembweg (aanzetstuk)
 - Kaak, tong, lippen, neusholte



Von Kempelen



Faber's Euphonia



Ontwikkeling van elektro-akoestiek

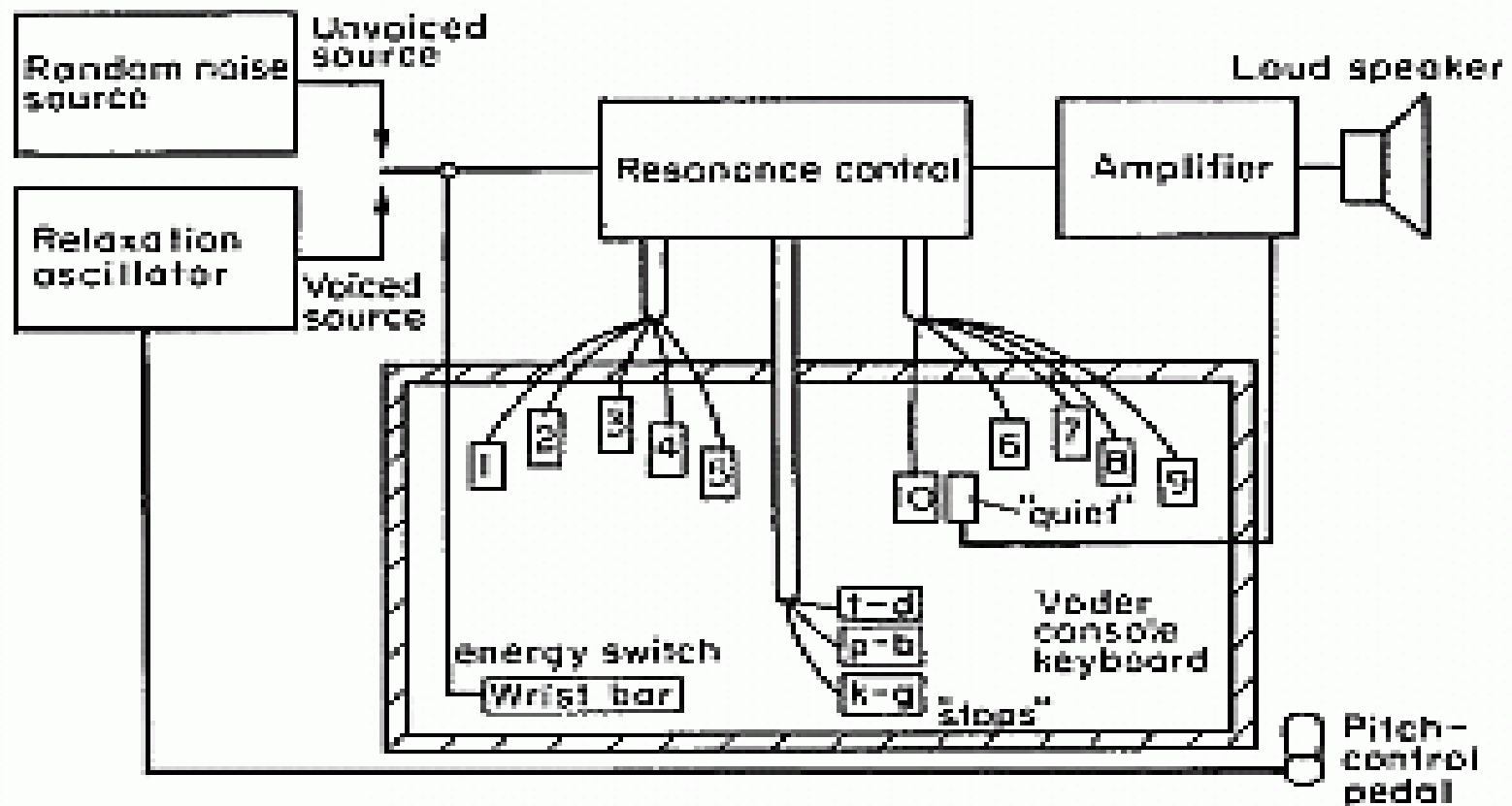
- Elektrisch analogon van akoestische processen
 - Analoog
 - Digitaal
- Klankbeschrijving
 - Articulatorische basis niet noodzakelijk
 - Bron + complex akoestisch filter

Vocoder

- Voice coder / decoder
- Codering van spraak, elke 10 ms
 - Toonhoogte
 - Luidheid
 - Spectrumcodering
 - Bandfilterenergie

Voder, 1939

demonstratie



Voder

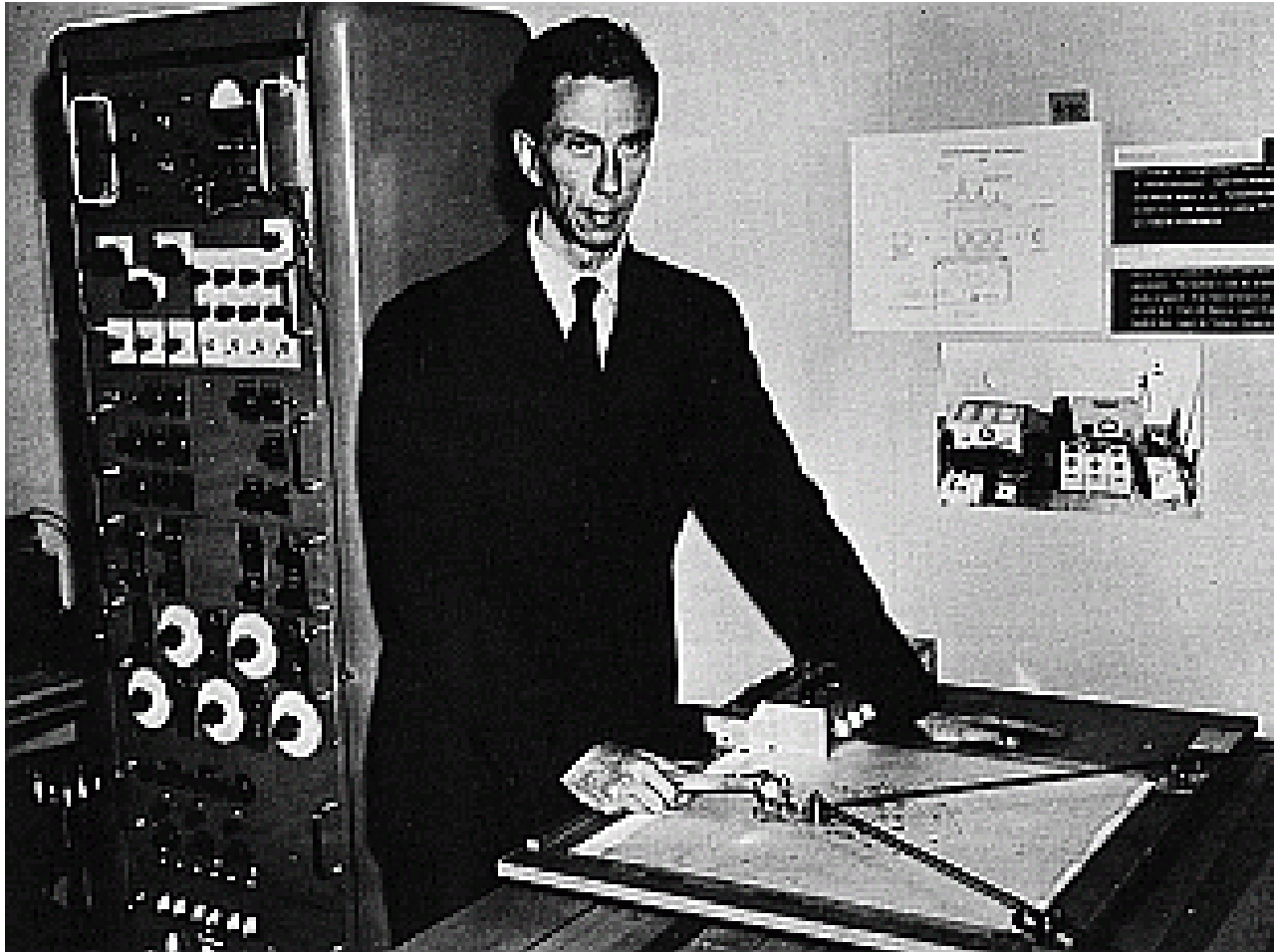


Formant synthese

- Codering van spraak, elke 10 ms
 - Toonhoogte
 - Luidheid
 - Spectrumcodering
 - Resonanties: formanten
 - Uit vorm van aanzetstuk berekenen (articulatorische synthese)
- Maar hoe kom je aan de juiste waarden?

Formantsynthesiser (Fant, 1953)

demonstratie



Klatt-synthesiser (DEC-talk, 1987)

demonstratie

- Geheel digitaal
- Geheel modelmatig (formantsynthesiser)
alle parameters regelbaar (ca 30)
 - Stembron
 - Formanten
- Allerlei stemmen mogelijk
- Lange tijd succesvol

stemmen

Articulatorische synthese

demonstratie (1968)

- Geef vorm van aanzetstuk
- Bereken dwarsdoorsneden van strot tot lippen
- Bereken bijbehorende akoestische resonanties
- Synthetiseer met filtermodel

Nadelen regelsynthese

- Zeer bewerkelijk om goede regels te maken
- Vooral de **overgang** tussen spraakklanken is moeilijk te beregelen
- Contekstafhankelijkheden
- Door modelbeperkingen klinkt spraak minder natuurlijk

Oplossing

- Gebruik bouwstenen (spraakfragmenten) die aanelkaar worden geregen
- Flexibiliteit versus grootte bouwsteen

Eenvoudigste oplossing

- Woorden en zinnen als bouwstenen gebruiken
- Opnemen met direct de juiste intonatie
- NS informatiesystemen op stations
- Vliegveld informatiesystemen
- Alleen bij beperkte vocabulaire, en veel standaard zinnen
- Taalkundig en fonetisch weinig interessant

Kleinere bouwstenen

- Kant en klare stukjes vooraf opgenomen spraak die aan elkaar geregen worden

Voordeel

- Klankovergangen zitten in bouwsteen

Nadeel

- Beregelen toonhoogte blijft nodig
- Overgangen tussen bouwstenen
- Nieuwe stem: nieuwe bouwstenen

Populaire bouwstenen

demonstratie

- Difonen
 - Overgang tussen twee spraakklanken, van midden spraakklank tot midden volgende spraakklank

Hoeveel nodig?

Difoonsynthese

Fluency

Als ingrediënten opgeven:

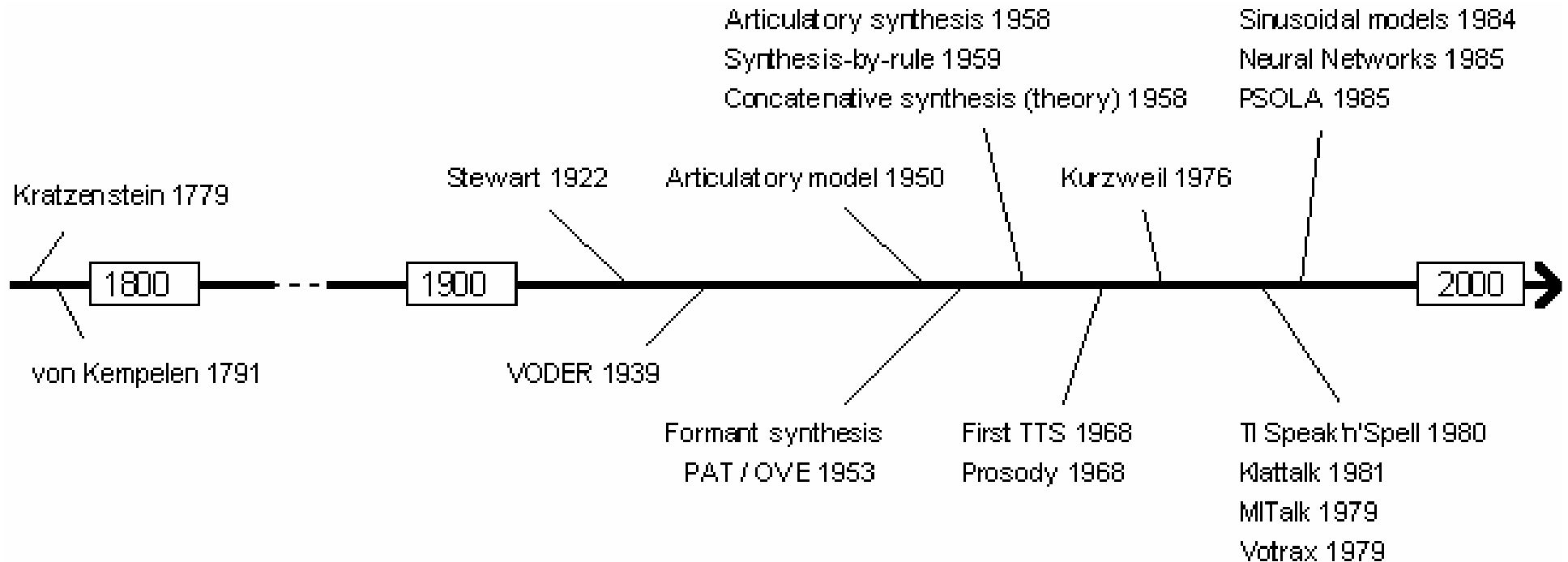
- Fonemen
- Toonhoogte
- Duren

Moeilijkheden

- Overgangen tussen difonen
 - Spectrale verschillen tussen einde ene difoon en begin van volgende difoon
- Voldoende difonen?
- Prosodie (Toonhoogte en luidheid)
 - Bouwsteen heeft opgenomen toonhoogteverloop en duur, die moeten veranderbaar zijn
 - Prosodieberegeling blijft lastig
- Elke nieuwe stem eist nieuwe difoonset

Spraaksynthese tijdlijn tot 2000

[demonstraties](#)



na 2000: Unit-selection synthesis

AT&T Acapela

- Neem zeer veel spraak op van een spreker (vele uren)
- Segmenteer dat op foneem-, syllabe-, woordniveau
- Zoek precies die bouwstenen die passen voor de te synthetiseren zin

Unit-selection kenmerken

- Geen toonhoogte / duur manipulaties
 - Zeer grote natuurlijkheid
- Lang niet altijd zijn precies de noodzakelijke segmenten beschikbaar
 - Compromis: kies dan kleinere units
 - maar wel steeds groter aansluitingsproblemen

Audiovisuele spraaksynthese

- Modelleren van gezichtsuitdrukking tijdens het spreken -> talking agents
- Visuele **versus** auditieve informatie
[McGurk effect](#)

Opdracht

- Spraaksynthese verkenningen
 - Maak verslag en stuur dat op voor 1 juni
 - Via Ephorus, inlevercode TST2