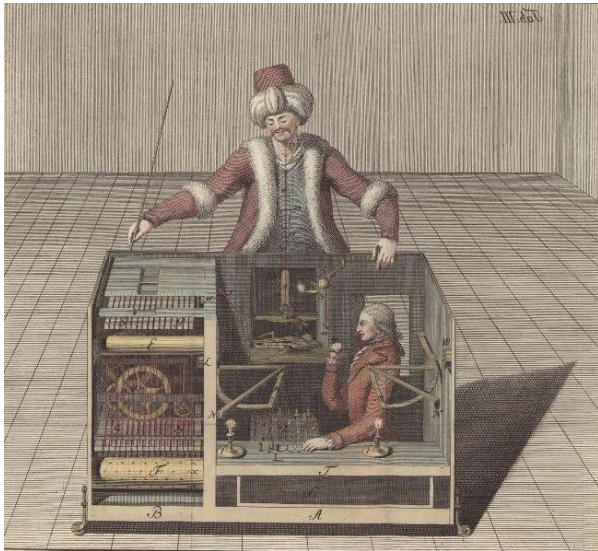


# Taal- en Spraaktechnologie

## 1 Introductie

Al eeuwen dromen mensen over zelfdenkende machines waarmee je “gewoon” kunt praten. Een van de bekendste “denkende machines” was de “Schaakturk” die in 1770 door de Hongaarse edelman

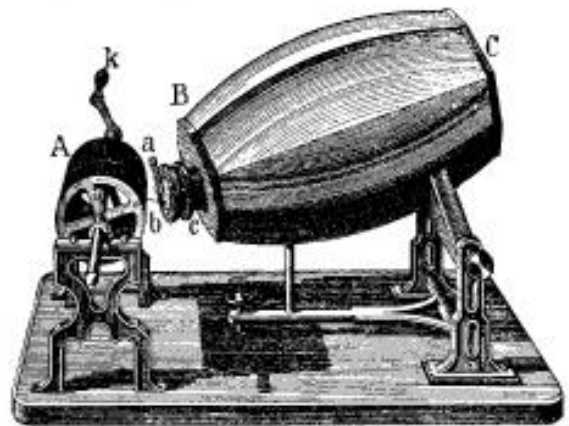


von Kempelen (1734-1804) werd gebouwd op verzoek van keizerin Maria Theresia van Habsburg. Het was een schaakmachine waartegen men kon spelen. Joseph Rahnitz toonde al in 1789 aan dat het bedrog was (er zat een dwerg in die goed kon schaken) maar de meeste mensen bleven geloven dat de machine zelfstandig kon schaken. In 1809 speelde Napoleon Bonaparte nog een partij tegen de machine: d.w.z. tegen de verborgen Turk Allgaier, één van de sterkste schaakspelers van zijn tijd.

Ook sprekende machines bleven de fantasie prikkelen maar pas halverwege de 19<sup>de</sup> eeuw

lukte het [Édouard-Léon Scott de Martinville](#) (1857) om de menselijke stem ("Au Clair de la Lune") vast te leggen op een zogeheten fonautograaf: een apparaat dat de geluidstrillingen op beroet papier kraste. 20 jaar later lukte het Edison om niet alleen de menselijke stem vast te leggen maar ook weer ten gehore te brengen met een zogeheten fonograaf. Deze geluidsopname is verloren gegaan, maar er bestaat nog een soortgelijke [opname](#) uit 1927. Voor een uitgebreid overzicht van de ontwikkelingen van de geluidsdragers kijkt men [hier](#).

Daarna is het snel gegaan en tegenwoordig staat niemand er meer van te kijken dat allerlei apparaten spraak kunnen vastleggen en weer afspelen. Toch zijn dit niet de sprekende machines waarvan men altijd gedroomd heeft. Het is hier de mens die spreekt; de machine legt het alleen maar vast en speelt het af. Voor een sprekende machine is echter meer nodig.



Phonautograph.

*BC*, barrel with opening at *C*; *c*, brass tube with membrane and style at *B*, and movable piece *a*, by which the position of the nodal points can be regulated; *A*, handle to turn cylinder (*A*) covered with lumpblacked paper.



## 2 Luisterende & Sprekende machines

Onder luisterende-en-sprekende machines verstaan we apparaten die de spraak tot zich nemen, er vervolgens “iets” mee doen en (eventueel) een gesproken antwoord teruggeven. Sinds de komst van de elektronische computers in de Tweede Wereldoorlog, zijn wetenschappers bezig computers echt te laten luisteren en spreken. In eerste instantie werd geprobeerd de computers de mens te laten nadoen maar al snel werd duidelijk dat we eigenlijk niet precies weten hoe de mens kan luisteren en spreken. We weten er tegenwoordig wel veel van maar beslist nog niet alles. Daarom werd er al snel teruggegrepen op allerlei wiskundige algoritmes die dan niet precies werken zoals de mens, maar waarmee het wel mogelijk was om de computers te laten luisteren, praten en, het allerbelangrijkste, te laten begrijpen.

### Technologie

Het gaat te ver om diep op de techniek in te gaan, maar we kunnen er wel iets algemeen over zeggen.

### TTS

Text-to-Speech (TTS) is de techniek waarmee de computer tekst kan uitspreken. De tekst kan door mensen worden geschreven (bv. een krantenartikel) of door computers m.b.v. databases en Nederlandse spellingregels in elkaar gezet worden. Het mooiste resultaat wordt verkregen door alle veel voorkomende woorden op te nemen (bv. *het, de, een, zijn, worden, ben, jij, vandaag, lopen, etc.*). Ook veelvoorkomende delen van woorden (zoals bv. *ver-, ge-, -heid, -lijk, -ing*) worden opgenomen. Voor de overgebleven klanken worden dan nog kleinere stukjes spraak gebruikt. Een slim algoritme zet dan de zinnen in elkaar door al deze opgeslagen stukjes spraak op de juiste manier aan elkaar te plakken. Deze methode is al oud maar werd niet vaak gebruikt doordat computergeheugen verschrikkelijk duur was. Doordat dat tegenwoordig niet meer zo is, kan deze methode in allerlei apparaten gebruikt worden. Een mooi voorbeeld is de nieuwe TomTom die behalve de opgenomen commando's (*na 100m links af*) tegenwoordig ook straatnamen en andere niet vooraf opgenomen woorden kan uitspreken.



TTS klinkt tegenwoordig zo goed dat men bij korte zinnen nauwelijks meer hoort dat het om een computerspraak gaat. Bij langere fragmenten hoort men dat wel omdat computerspraak op den duur een beetje saai gaat klinken. De reden is dat de computer nog niet begrijpt waar in een verhaal de klemtoon gelegd moet worden. Neem een verhaal als Roodkapje. Om dat spannend voor te lezen, moet je met je stem gaan spelen: laag en langzaam, hoog of snel, pauzeren voor een “spannend” woord etc. Technisch kan de computer dit wel, maar we weten nog niet hoe we dat automatisch uit een zin moeten bepalen.

### ASR

Automatic Speech Recognition (ASR) is precies het omgekeerde van TTS. De mens spreekt en de computer zet de spraak om in:

spraakklanken → letters → woorden → zinnen → uitingen

Anders dan bij het spreken, hebben we hier met een enorme variatie aan menselijke spraak te maken. Mensen spreken in de regel nu eenmaal niet zo netjes als bv. Sacha de Boer op het 8-uur

journaal. Ook hier is het onmogelijk om diep op de technologie in te gaan, maar grosso modo komt spraakherkenning neer op het volgende:

Het spraaksignaal wordt met een microfoon opgenomen en door de computer omgezet in een digitaal signaal. Vervolgens wordt van kleine stukjes signaal (25 milliseconde lang) bepaald op welke bekende menselijke spraakklank het stukje het meest lijkt. Het resultaat is een hele rij klanken die op een slimme manier in mogelijke woorden moet worden omgezet.

Dit lijkt eenvoudig maar dat is het niet. Een van de problemen is dat mensen woorden zo uitspreken dat de luisteraar ze verstaat. Zo wordt een woord als verzekeringsmaatschappij bijna nooit netjes uitgesproken. Meestal klinkt het als: *fe-see-krings-maat-sge-pij*. De computer moet dus de juiste woorden “raden” ondanks de ontbrekende klanken.

## Statistiek

Bij het bepalen welke woorden waarschijnlijk werden uitgesproken, wordt gebruik gemaakt van statistiek. Hoe waarschijnlijk is het dat woord B wordt uitgesproken na woord A? De kans hierop wordt berekend door uit miljoenen woorden de kans te berekenen op 1) het voorkomen van woord B en 2) de kans op woord B gegeven woord A. Het lijkt daarmee een beetje op de invuloefeningen die we wel van de lagere school kennen en waar gevraagd werd:

Maak de volgende zin af:

Jantje zag eens .....

Natuurlijk zal iedereen hier onmiddellijk “pruimen hangen” antwoorden omdat dit een soort standaardzin is geworden. Maar ook woorden als *appels, schilderijen, auto's, meisjes, of iPhone's* zouden in principe ingevuld kunnen worden. Eigenlijk kan ieder zelfstandig naamwoord maar de kans op *zwembaden of hooibergen* is waarschijnlijk niet erg hoog.

Kortom, de computer leert door heel veel Nederlandstalige teksten in te voeren welke woorden en woordvolgordes in het Nederlands waarschijnlijk zijn.

## AI

De laatste stap is de inzet van AI (Artificial Intelligence ofwel kunstmatige intelligentie). Hierbij wordt niet alleen gekeken wat waarschijnlijk is, maar ook wat logisch is, gegeven de context van een gesprek. Als je bv. de telefoon vraagt je vrouw te bellen na de vergadering vanmiddag dan moet de computer een aantal zaken weten:

- Wie is jouw vrouw?
- Wat is haar telefoonnummer?
- Hoe laat is de middagvergadering afgelopen?
- Welke vergadering wordt er bedoeld als één vergadering van 10:00 tot 12:15 gepland staat en de ander van 14:00 tot 16:15?

## 3 Toepassingen

Door de toenemende rekenkracht, de steeds kleiner wordende apparaatjes, de steeds betere batterijen en het overall aanwezige (draadloze) internet zijn steeds meer mensen “always connected”. Doordat het kan willen we de informatie en services dan ook onmiddellijk kunnen benaderen.



Je loopt door een straat en ziet een huis dat te koop staat. Wat zou dat kosten en hoeveel kamers heeft het? Je pakt de telefoon en zegt: “geef mij de informatie over dit te koop staande huis”. De telefoon weet waar je bent, en dus ook welk huis je waarschijnlijk bedoelt. Hij haalt de informatie van de Funda website en toont die en/of leest het voor.

Dit soort applicaties is sinds een halfjaar voor het Engels, Frans en Duits beschikbaar via het door Apple gekochte bedrijf SIRI. De applicatie wordt gezien als het meeste innovatieve onderdeel van de nieuwste iPhone en zal naar verwachting snel voor andere toestellen en andere talen beschikbaar komen.

Taal- en Spraaktechnologie wordt, behalve voor dit

soort Persoonlijke Assistentes, veelvuldig in bijna alle onderdelen van het dagelijks leven toegepast.

- Het aangeven/corrigeren van schrijf- en stijlfouten in je documenten
- Het inspreken van een rapport
- Het benaderen van bedrijven voor informatie via het web (tekst) en telefoon (spraak)
- Het zoeken in spraak (TV-programma's, interviews, persoonlijke herinneringen, toespraken in het parlement, etc.)
- Het ondersteunen van griffiers bij het maken van een verslag
- Het mondeling overdragen van relevante politie-informatie (van agent op straat → politie-database op het hoofdbureau)
- Het schatten van de kans dat iemand aan een bepaalde stoornis lijdt aan de hand van geschreven en/of gesproken egodocumenten of afgenomen interviews
- Het ondertitelen van voorgelezen boeken voor kinderen en/of mensen met leesproblemen

En dit is nog niet alles! In februari 2011 heeft IBM's nieuwste, met taaltechnologie volgestopte computer WATSON op overtuigende wijze de Amerikaanse quiz Jeopardy gewonnen waarbij de vorige wereldkampioenen verpletterend verslagen werden.



## 4 Conclusie

De enorme hoeveelheid informatie, de wens om altijd direct toegang tot alles te hebben, de mogelijkheid om je altijd verbonden computer in je binnenzak mee te nemen, leidt ertoe dat we behoefte hebben aan “menselijke” ondersteuning in ons dagelijks leven. Taal- en Spraaktechnologie, Statistiek en Kunstmatige Intelligentie, zoals hierboven kort beschreven, zijn hier uitstekend voor geschikt en zullen dus steeds vaker en intensiever gebruikt gaan worden.

Het aardige van deze technieken is dat, naarmate ze beter werken, ze steeds onzichtbaarder worden. Nog even en we weten niet beter en dan dat we gewoon kunnen praten met de machines om ons heen. En daarmee is de lang gekoesterde wens van de mensheid om met levenloze objecten op een intelligente wijze te kunnen praten eindelijk in vervulling gegaan!