Spraakherkenning voor onderzoek in AV-archieven

Twintig jaar ontwikkeling in Nederland

Roeland Ordelman
16 april 2021

AVA_Net

De wens om automatische spraakherkenning in te zetten om audiovisuele archieven beter doorzoekbaar te maken is niet van gisteren. Samen met archieven in Nederland werken al zo'n twintig jaar computer- en geesteswetenschappers hieraan. De recente aandacht voor 'kunstmatige intelligentie' (AI) heeft deze wens weer eens extra op de agenda gezet. Want naast dat spraakherkenning op zichzelf al interessant is, kan het ook als motor dienen voor andere 'intelligente' toepassingen. Hieronder een bondig verslag van hoe het er in Nederland voor staat

Nederlands audiovisueel erfgoed is een schat voor onderzoekers. De 'stoffige' archieven van weleer hebben de afgelopen jaren belangrijke stappen gezet naar 'digitale depots' om erfgoed te bewaren en beter toegankelijk te maken. Naast digitalisering en mogelijkheden om content online beschikbaar te stellen, is het kunnen zoeken in de data een belangrijke voorwaarde: er is metadata nodig om gebruikers te kunnen verwijzen naar items die ze interessant zouden kunnen vinden.
Het creëren van deze metadata was altijd de verantwoordelijkheid van archivarissen of documentalisten. Maar dit veelal handmatige proces staat onder druk door de toestroom van digitale data door digitalisering en de creatie van nieuwe, 'digital born' content. Daar komt bij dat gebruikers behoefte hebben aan tijd-gecodeerde metadata om ook binnen collectie-items te kunnen zoeken. Dit geldt met name in het audiovisuele domein waar je niet zomaar met 'control-F' door tekst kan springen, maar een item zal moeten bekijken en afluisteren om te vinden waar je naar op zoek bent. Doordat dit soort uitgebreide, beschrijvende metadata meestal maar beperkt aanwezig zijn, is de schat aan informatie in de digitale archieven ook maar beperkt toegankelijk. Sterker nog, vaak is de aanwezigheid van pareltjes volledig onbekend.

Kunstmatige intelligentie

Er wordt al decennia gekeken naar 'kunstmatige intelligentie' (artificial intelligence) als hulpmiddel om content automatisch door computers van metadata te voorzien. Een voorbeeld hiervan in het audiovisuele domein is automatische spraakherkenning. Al in 2001 nam het Nederlands Instituut voor Beeld en Geluid samen met Universiteit Twente deel in een Europees onderzoeksproject (ECHO) dat onderzocht of automatische spraakherkenning zou kunnen worden ingezet voor het doorzoekbaar maken van het archief. Hoewel de techniek toen nog niet zo ver was als vandaag de dag, bleek al wel dat het automatisch laten uitschrijven van het gesproken woord in audio en video enorm behulpzaam kan zijn voor het 'ontsluiten' van archiefmateriaal. Want ook al maakt de spraakherkenning fouten en is een automatische transcriptie vaak niet goed leesbaar voor een mens, het gaat erom dat de belangrijke inhoudswoorden goed worden herkend, zodat je met behulp van een zoekmachine via die woorden interessante fragmenten kunt vinden. Met andere woorden: als de computer de data maar kan 'lezen'.

Interesse vanuit onderzoek

Een doelgroep die al vroeg grote interesse had in het gebruik van spraakherkenning waren geesteswetenschappelijke onderzoekers die met 'gesproken getuigenissen' (Eng. 'Oral History') werkten. Een belangrijk aanleiding voor deze interesse was het MALACH project. Doel van dit iconische project dat startte in 2001, was het preserveren van getuigenissen van overlevenden van de holocaust en het breed toegankelijk maken van deze getuigenissen met behulp van 'moderne' digitale zoek- en analysetechnieken. Vanwege de enorme hoeveelheid van circa 50.000 opgenomen interviews waren die automatische technieken voor 'zoeken in gesproken documenten' (Eng. 'Spoken Document Retrieval') onontbeerlijk.

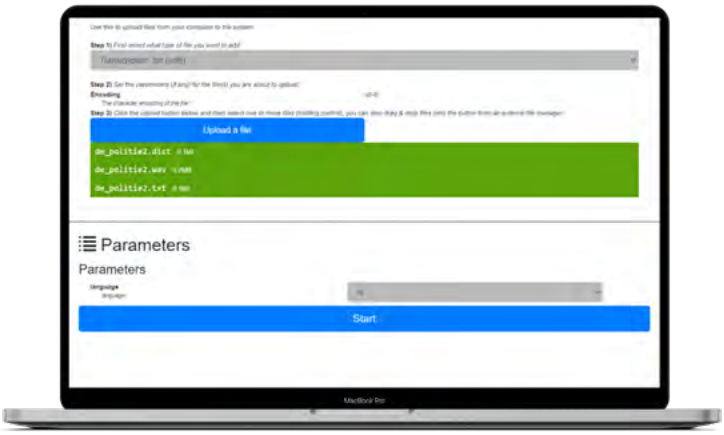
Vanwege de grote verscheidenheid van de spraak in de holocaust getuigenissen, bleek het echter nog een hele toer om de technologie over de volle breedte van de collectie op een acceptabel kwaliteitsniveau te krijgen. Denk aan diverse talen, dialecten en etnolecten die je kon tegenkomen, spraak van soms zeer geëmotioneerde mensen op leeftijd, en specifieke termen, namen en plaatsen gerelateerd aan de oorlog. Delen van de collectie zijn daarom jarenlang gebruikt voor onderzoek in de spraaktechnologie om manieren te vinden om de kwaliteit te verbeteren. In Nederland spitste zich dat toe op het gebruik van Nederlandse spraakherkenning voor Oral History onderzoek in het CHoral[1] project, dat een vervolg heeft gekregen in diverse projecten en samenwerkingsverbanden. Een voorbeeld hiervan is het Buchenwald project waarin in samenwerking met NIOD getuigenissen van overlevenden van kamp Buchenwald online toegankelijk werden gemaakt
(zie Figuur 1).

Figuur 1 - Zoeken in het gesproken woord van getuigenissen van overlevenden van kamp Buchenwald met behulp van spraakherkenning. Via de zoekterm "fiets" springt een gebruiker naar het fragment waarin dhr. De Vries in dit voorbeeld praat over hoe hij op de fiets van Hilversum naar Amsterdam ging om in het sportfondsenbad te zwemmen.

[1] Willemijn Heeren, Laurens van der Werff, Franciska de Jong, Roeland Ordelman, Thijs Verschoor, Arjan van Hessen & Mies Langelaar (2009) Easy Listening: Spoken Document Retrieval in CHoral, Interdisciplinary Science Reviews, 34:2-3, 236-252, DOI: 10.1179/174327909X441135

Transcripties oplijnen

Naast het automatisch uitschrijven van het gesproken woord, kan spraakherkenning ook op een andere manier behulpzaam zijn bij het onderzoek en de ontsluiting van interview data. Het komt regelmatig voor dat onderzoekers de interviews waarmee ze onderzoek doen zelf handmatig transcriberen. Deze 'transcripties' voorzien ze vervolgens weer van commentaar gerelateerd aan het specifieke onderzoek dat ze doen. Onderzoekers voegen aan deze commentaren codes toe die het mogelijk maken om interviewfragmenten met eenzelfde code met elkaar te vergelijken. Het zou heel handig zijn als onderzoekers op basis van die codes door het materiaal zouden kunnen scrollen, maar om dat mogelijk te maken zouden de commentaren ook van tijdcodes moeten worden voorzien. Dit is met de hand onbegonnen werk, maar voor een spraakherkenner relatief eenvoudig. De spraakherkenner 'weet' als het ware al wat er werd gezegd, en hoeft alleen nog maar aan te geven op welke plek in de spraak elk woord werd uitgesproken. Zolang het verschil tussen de transcriptie en wat letterlijk werd gezegd niet te groot wordt, kan de spraakherkenner hier goed mee overweg en tot op de milliseconde nauwkeurige tijdcodes opleveren. Deze spraakherkenningstechniek wordt oplijnen (Eng. 'alignment') genoemd. Radboud Universiteit Nijmegen heeft hier samen met onderzoekers handige tools voor ontwikkeld.



Figuur 2 - Schermafbeelding van een deel van de alignment service van Radboud Universiteit Nijmegen waar een gebruiker een transcriptie en een audiofile kan uploaden die het systeem vervolgens omzet in een bestand met tijdcodes.

Naast historici en sociale wetenschappers die met interview data werken, ontstond er de laatste jaren ook vanuit andere onderzoeksdomeinen interesse in het gebruik van spraakherkenning. Bijvoorbeeld door mediawetenschappers, om te onderzoeken hoe de verslaglegging van een bepaald onderwerp in nieuws- en actualiteitenrubrieken van de publieke omroep zich heeft ontwikkeld. Of hoe de verslaglegging zich verhoudt tot andere media, zoals kranten (te vinden bij de Koninklijke Bibliotheek) of social media. Maar ook hoe een individueel programma zich door de tijd ontwikkeld heeft. Ook voor (onderzoeks) journalisten zijn dit soort *data analyses* die gebruik maken van automatische spraakherkenning interessant. Tijdens de verkiezingen van 2021 gebruikten journalisten spraakherkenning om te achterhalen welke onderwerpen aan bod kwamen in talkshows tijdens de verkiezingen[2]. Ook werden er woordenwolken gemaakt die mooi lieten zien welke woorden politici het meest gebruikten (zie figuur 3).

- Gertjan Segers

- Jesse Klaver

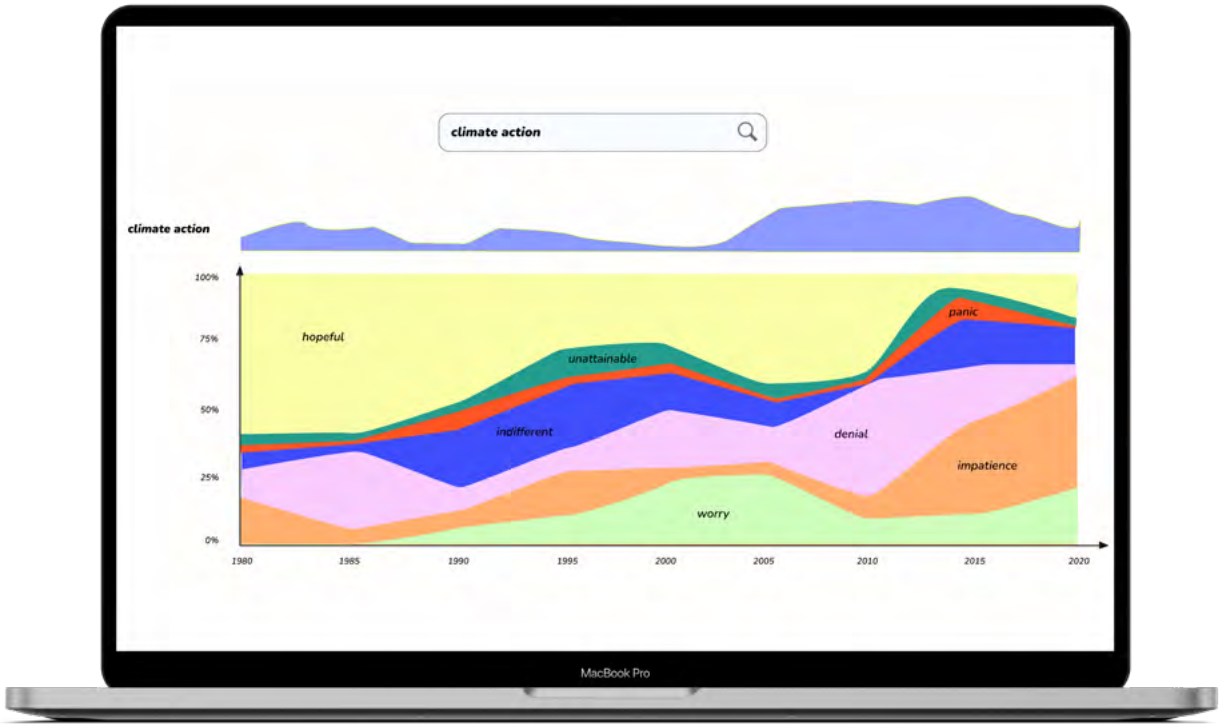
Figuur 3 - Wordclouds van de spraak van Gertjan Segers (boven) en Jesse klaver (onder) in een selectie van publieke radio/TV 1 februari - 14 maart 2021.

[2] https://mediasuitedatastories.clariah.nl/elections-2021-first-results/

Spraakherkenning als basis

Bovenstaande voorbeelden geven aan hoe spraakherkenning van nut kan zijn voor audiovisuele archieven. Niet alleen om te zoeken naar woorden als 'fiets' of 'vluchtelingencrisis', maar ook als basis voor verdere of diepere analyse. Die analyse kan bestaan uit eenvoudig woordjes tellen en het resultaat hiervan op een mooie manier laten zien, maar kan ook complexere vormen aannemen.
Om verbindingen te leggen *tussen* verschillende bronnen, binnen een audiovisueel archief, of tussen meerdere archieven (denk aan het voorbeeld van televisie en kranten hierboven), kan een thesaurus van persoons- of plaatsnamen als verbindende factor worden gebruikt. Het herkennen van personen of plaatsen in de automatische transcripties (Eng. 'Named Entity Extraction') is voor dit aan elkaar relateren van databronnen (Eng. 'Linked Data') een belangrijk hulpmiddel. Andere voorbeelden van technieken waarvoor spraaktranscripties kunnen worden gebruikt zijn het automatisch segmenteren op basis van (de verandering in) onderwerpen waarover gesproken wordt, of het inschatten van het sentiment (Bijv. positief, neutraal of negatief) van een bron.

Niet alleen voor onderzoekers en journalisten zijn technieken om te zoeken en relaties te leggen interessant. Ook voor het 'algemeen publiek' kunnen dit soort technieken helpen bij het toegankelijk maken van collecties. We weten na decennia van onderzoek best veel over het gedrag van dit type gebruikers als het gaat om toegang tot audiovisuele content. Bijvoorbeeld dat gebruikers niet altijd precies weten waar ze naar op zoek zijn. Dat ze vaak helemaal niet zo'n beeld hebben bij wat er in AV-collecties van erfgoedinstellingen zoal te vinden is. En ook, dat ze snel afgeleid of weer vertrokken zijn. Als ze dan gaan zoeken is het fijn als ze vanuit de zoekresultaten direct naar een fragment in de data kunnen springen (Eng. 'jump-in point') die relateert aan de zoekvraag. En dat ze kunnen grasduinen, 'exploratief' kunnen zoeken zoals dat heet, om al browsend op persoonlijke pareltjes in de collectie te stuiten. Het *stimuleren van serendipiteit* wordt dat wel genoemd, het toevallig ergens tegenaan lopen. Namen of plaatsen als startpunt nemen, ligt hierbij voor de hand. Want gebruikers die geen expliciete informatiebehoefte hebben over een onderwerp, typen vaak als eerste hun eigen naam in, of die van bekende personen of plaatst waar ze vandaan komen of geboren zijn.

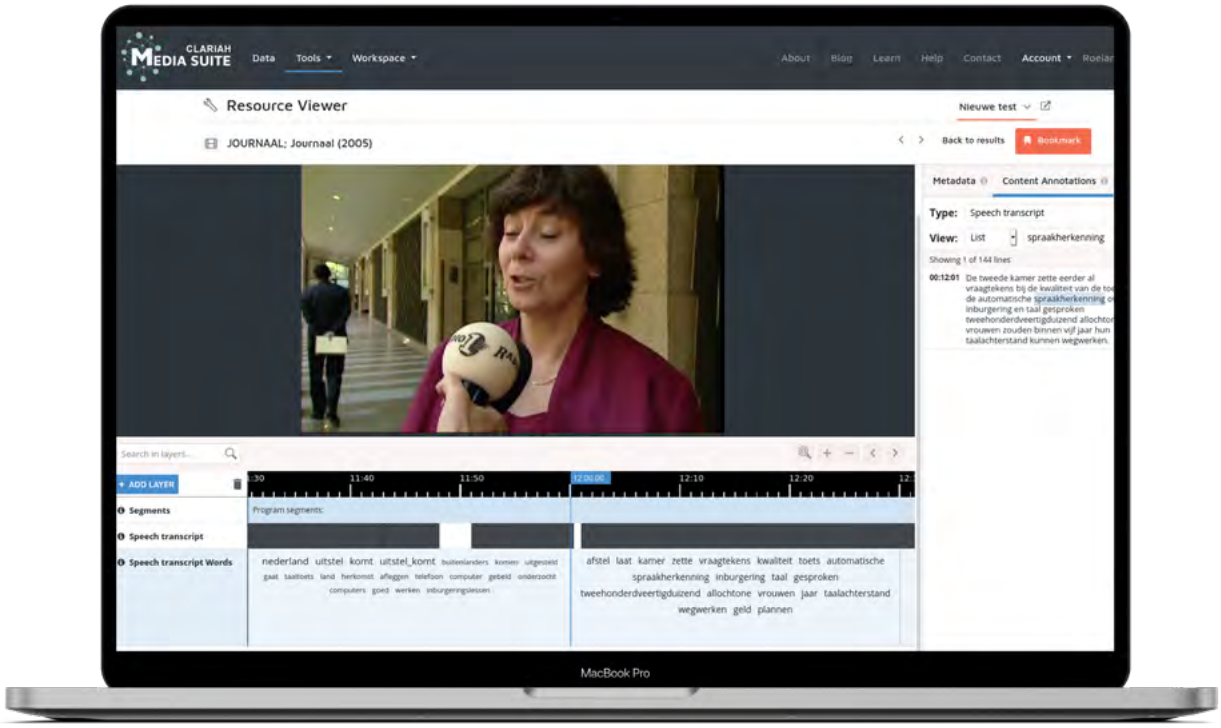


Figuur 4 - Impressie van sentiment analyse op basis van spraak voor het onderwerp "klimaat actie" door de tijd heen waaraan ontwikkeld wordt in het Newsslider project, een samenwerking tussen Beeld en Geluid, IDfuse, Cream on Chrome and Fontys Hogeschool (Stimuleringsfonds Creatieve Industrie).

Gebruiken van spraak-herkenning: digitale infrastructuur

Automatische spraakherkenning biedt dus diverse mogelijkheden om audiovisuele bronnen toegankelijk te maken. Hoe kun je als archief deze techniek praktisch inzetten? Het korte antwoord is dat allereerst de spraakherkenning bij de bronbestanden van een archief moet kunnen om de audio op te halen, te verwerken, en de resultaten op te leveren aan een bestaand archivering- of metadateringsproces. Vervolgens kunnen de spraaktranscripts als metadata door een aanwezige zoekmachine worden geindexeerd zodat gebruikers er in kunnen zoeken. Wil je gebruikers direct naar een relevant fragment binnen een item kunnen sturen, in plaats van naar het hele AV-document, dan moet de zoekmachine er rekening mee houden dat de metadata *tijd-gecodeerd* is. Daarbij moet de software die gebruikt wordt om bestanden af te spelen (Eng. 'play-out') kunnen omgaan met verwijzingen *naar een specifieke plek* in het bestand.

Achter het korte antwoord zit in de praktijk vaak een wereld van nieuwe vragen. Het beschikbaar maken van bronbestanden aan een spraakherkenningsproces kan al een lastige horde zijn. Bijvoorbeeld omdat het vanwege ingewikkelde digitale werkprocessen (Eng. 'workflows') niet zo eenvoudig is om bronbestanden uit het archief te trekken om aan een spraakherkenner te voeren. Ook privacy-gevoeligheid of auteursrechten kunnen de beschikbaarstelling aan een spraakherkenner in de weg staan, vooral omdat spraakherkenning doorgaans als online dienst wordt geleverd. De bestanden moeten daarvoor 'naar buiten' of nog spannender, naar 'de cloud'. Denk bijvoorbeeld aan de spraakherkenningsdiensten die partijen zoals Google[3] en Microsoft[4] bieden. Vanuit archiefperspectief volkomen terecht, noopt dit tot voorzichtigheid en aarzeling om groen licht te geven voor het versturen van bronbestanden. Tenslotte spelen ook vragen met betrekking tot de spraakherkenning zelf, met name: wat kost het en is de kwaliteit goed genoeg?



Figuur 5 - Schermafbeelding van de CLARIAH Media Suite waar zoeken op "spraakherkenning" een item uit het NOS Journaal uit 2005 oplevert waarin je direct kunt springen naar Rita Verdonk die spreekt over de kwaliteit van de inburgeringstoets die gebruik maakt van automatische spraakherkenning.

[3] https://cloud.google.com/speech-to-text

[4] https://azure.microsoft.com/nl-nl/services/cognitive-services/speech-to-text/

CLARIAH infrastructuur

Zoals hierboven reeds aangestipt, was er vanuit het onderzoeksveld al jaren veel vraag naar spraakherkenning en de toepassing ervan in de praktijk van audiovisuele archieven waar de voor onderzoekers interessante data veelal ligt opgeslagen. Een aantal jaar geleden is daarom in het CLARIAH[5] project gestart met het ontwikkelen van een spraakherkenningsdienst voor onderzoeks- en erfgoedinstellingen. CLARIAH bouwt aan een landelijke digitale onderzoeks *infrastructuur*. Kort gezegd moet deze infrastructuur het voor onderzoekers eenvoudiger maken om vanuit hun bureaustoel met een paar drukken op de knop online databestanden te raadplegen en te analyseren. Met het beschikbaar stellen van spraakherkenning binnen deze infrastructuur kunnen AV-collecties door onderzoekers veel beter worden benut.
De CLARIAH Media Suite[6] is de portal binnen de CLARIAH infrastructuur die zich expliciet richt op onderzoek met mediabronnen, met name audiovisuele media. Via de Media Suite kunnen onderzoekers de collecties van Beeld en Geluid en partners (denk bijvoorbeeld aan Tweede Kamer data) raadplegen. Ook worden stapsgewijs collecties van bijvoorbeeld Eye Filminstituut, DANS, en Meertens Instituut beschikbaar gemaakt. Het ultieme doel is om via de Media Suite zoveel mogelijk audiovisuele collecties aan te bieden: van nationale archieven tot regionale en lokale archieven en zelfs de micro-archieven die 'in de kast' liggen bij kleine instellingen, projectorganisaties of personen.

Om deze collecties van spraakherkenning te kunnen voorzien, is de afgelopen jaren gewerkt aan een dienst die binnen CLARIAH kan worden gebruikt voor het efficiënt verwerken van (soms grootschalige) AV-bronnen. Deze dienst haalt audiobestanden op en laat daar vervolgens meerdere spraakherkenners tegelijk op los. Dit gebeurt doorgaans op een relatief krachtige computer die supersnel kan rekenen. Om te beginnen is de dienst getest op een groot deel van de collectie van Beeld en Geluid, honderdduizenden uren aan data, gebruik makend van een rekencluster van SURF[7],de organisatie die voor onderzoek en onderwijs digitale diensten beschikbaar stelt. Maar inmiddels draait de dienst ook op een lokale server bij Beeld en Geluid. Dat is wat handzamer en betekent dat de dienst ook op kleinere collecties kan worden ingezet en ook bij andere CLARIAH partners met AV-collecties kan draaien. Met het inzetten van de spraakherkenning als CLARIAH dienst starten we dit jaar de eerste testen.

Inclusieve spraakherkenning

De spraakherkenner die bij Beeld en Geluid en CLARIAH gebruikt wordt is gebaseerd op de 'open-source' spraakherkenner Kaldi[8]. Deze maakt gebruik van de nieuwste technieken op het gebied van machine learning ('deep learning'). Wetenschappers aan de Universiteit Twente en Radboud Universiteit hebben met behulp van een grote hoeveelheid Nederlandse spraak- en tekstdata, Nederlandse modellen getraind. Deze modellen zijn 'open-source' beschikbaar gesteld onder de vlag Kaldi-NL. Om het gebruik van Nederlandse spraakherkenning te bevorderen hebben spraakonderzoekers van Nederlandse universiteiten en kennisinstellingen zich verzameld onder de vlag van de Stichting Open Spraaktechnologie[9]. Deze stichting stelt Kaldi-NL beschikbaar en denkt mee over de verdere ontwikkeling van spraakherkenning in samenwerking met universiteiten en diverse private partijen.

Eén van de onderwerpen waar de stichting zich hard voor maakt is de *kwaliteit* van de spraakherkenning, die doorgaans wordt uitgedrukt in 'Word Error Rate' (WER): het aantal fout herkende woorden gedeeld door het totaal aantal woorden. Fout herkend kan betekenen dat een verkeerd woord werd herkend (substitutie), foutief ingevoegd (insertie) of juist foutief weggelaten (deletie). Kenmerkend voor AV-collecties is dat de data heel divers is: de kwaliteit van de brondata wisselt (denk aan oude opnamen), er komen heel veel typen spraak voorbij (verschillende sprekers, leeftijden, dialecten) in diverse contexten (interview, redevoering, vergadering) en opgenomen onder verschillende omstandigheden (in een studio, een kerk, op straat, in een zoom meeting). Om onderzoekers spraakherkenning te kunnen laten verbeteren, is het belangrijk om te weten wanneer spraakherkenning het

[5] https://clariah.nl/

[6] https://mediasuite.clariah.nl/

[7] https://www.surf.nl/over-surf

[8] https://kaldi-asr.org/

[9] https://openspraaktechnologie.org/

goed doet en wanneer niet. We weten bijvoorbeeld al wel dat spraakherkenners het moeilijk hebben met spraak van kinderen, dialectspraak of spraak van Nederlanders met een migratieachtergrond. Omdat in AV-collecties dit type spraak ook voorkomt en ook allerlei publieke diensten worden ontwikkeld waar spraakherkenning een rol in speelt (denk aan Google Home) is het belangrijk dat onderzoekers werken aan wat wel inclusieve *spraakherkenning* wordt genoemd.

Als we het over AV-archieven hebben geldt die 'inclusiviteit' ook voor domeinen. Hierboven kwam al even ter sprake dat in Oral History interviews rond de Tweede Wereldoorlog veel woorden worden gebruikt die specifiek met de oorlog te maken hebben. De spraakherkenner moet deze woorden *kénnen* om ze te kunnen *hérkennen*. Dit speelt ook bij bijvoorbeeld stadsarchieven waar personen, straatnamen of gebouwen uit de stad voorkomen in de spraak, maar mogelijk niet bekend zijn bij de spraakherkenner. Een ander voorbeeld is collecties waar veel medische termen in voorkomen. In het Homo Medicinalis (HOMED) project[11] onderzoeken we hoe we spraakherkenning zo eenvoudig mogelijk kunnen aanpassen aan dit soort specifieke domeinen in CLARIAH en daarbuiten, zodat spraakherkenning in de volle breedte van de bijzondere collecties in archieven succesvol kan worden ingezet.

Omdat er naast Kaldi-NL andere spraakherkenners voor het Nederlands bestaan, zou het ook mooi zijn om verschillende spraakherkenners te kunnen vergelijken zodat partijen kunnen kiezen welke spraakherkenner het beste past bij hun wensen. Om de kwaliteit van spraakherkenning op een gestructureerde manier te kunnen testen is de Stichting Open Spraaktechnologie met een aantal mediapartijen waaronder Beeld en Geluid, NPO, en RTL een project gestart rond *benchmarking*. Het doel is om het voor partijen eenvoudiger te maken om aan de hand van voorbeelden uit hun eigen datacollectie een beeld te krijgen of de kwaliteit van spraakherkenning of van verschillende spraakherkenners voldoende is voor hun doeleinden. Een partij die geïnteresseerd is in automatische ondertiteling stelt uiteraard andere eisen aan de kwaliteit dan een partij die vooral wil kunnen zoeken.

Figuur 6 - In het project Kinderen in Gesprek met Media'[10] mogelijk gemaakt door SIDNfonds en CLICKNL wordt onderzocht hoe kinderen op een verantwoorde manier met een robot kunnen communiceren, bijvoorbeeld om interessante content in het archief te ontdekken door een gesprek aan te gaan met de robot.

Categorie	WER
Nieuws	18.15
Opinie	26.13
Radio	31.03
Sport	34.36
Samenleving	36.74
Kennis	42.36
Expressie	44.61
Amusement	46.88
Kinderprogrammering	62.15

Figuur 7 - Score van de KALDI-NL open-source spraakherkenner op een aantal media categorieën in termen van 'Word Error Rate' (WER) geëvalueerd in het benchmarking project. Hoewel de categorieën niet altijd wat zeggen over het type spraak en geven de scores wel een globaal beeld van de kwaliteitsverschillen voor verschillende typen data.

[10] https://beeldengeluid.nl/kennis/projecten/kinderen-in-gesprek-met-media

[11] http://homed.ruhosting.nl/

Vorderingen

We zijn in Nederland dus al best lang bezig met het onderwerp spraakherkenning in audiovisuele archieven. Enerzijds zien we daarbij dat het in gebruik nemen van nieuwe technologieën in het archiefdomein een ingewikkeld proces is. Het vergt niet alleen de inrichting van nieuwe digitale processen die als 'complex' worden ervaren, maar ook een mindset: de wil om mee te bewegen met veranderende behoeften en wensen van gebruikers van archieven, en met de veranderende maatschappelijke en technologische context. Technologie niet zien als gevaar of iets engs dat je misschien liever uitbesteedt, maar als kans om op andere manieren met AV-bronnen bezig te gaan.

Anderzijds zien we ook dat er aan een aantal basisvoorwaarden moet worden voldaan om het gebruik van spraakherkenning bij audiovisuele archieven te helpen realiseren. Daar zijn de afgelopen jaren belangrijke stappen in gemaakt: de kwaliteit is sterk verbeterd, het is makkelijker geworden om zelf spraakherkenning te gebruiken en er is ondertussen ruime ervaring met het gebruik van spraakherkenning om archieven doorzoekbaar te maken. De recente aandacht voor kunstmatige intelligentie zal ook helpen om samen met computerwetenschappers in Nederland onderzoek te doen naar de verdere verbetering van spraakherkenning. Dus archieven, laat die stem maar horen!

Resources

- Open source spraakherkenning met o.a. Kaldi NL
- CLARIAH Media Suite
- Language & Speech Tools Radboud Universiteit

Projecten

- CLARIAH
- HOMED
- Kinderen in Gesprek met Media
- Newsslider

SIDNfonds

Roeland Ordelman schreef zijn proefschrift aan de Universiteit Twente (2003) over de inzet van spraakherkenning voor het ontsluiten van audiovisuele archieven. Sindsdien staat zijn werk in het teken van de inzet van informatietechnologie om audiovisuele archieven beter toegankelijk te maken, waarbij hij onderzoek en het gebruik in de praktijk combineert bij respectievelijk Universiteit Twente en het Nederlands Instituut voor Beeld en Geluid. Bij Beeld en Geluid coördineert hij de innovatie-Labs waar de CLARIAH Media Suite onderdeel van uitmaakt. Daarnaast is hij CTO van CLARIAH en voorzitter van de Stichting Open Spraaktechnologie.

CLARIAH is een meerjarig programma dat startte in 2014 en wordt gefinancierd door NWO. Het heeft als doel om wetenschappers in de Geestes- en Sociale Wetenschappen te ondersteunen bij de voortgaande digitalisering van onderzoek door een infrastructuur te ontwikkelen die het eenvoudiger moet maken om online onderzoek te doen met databronnen die verspreid zijn over diverse grote en kleine instellingen in Nederland. Naast toegang tot bronnen biedt de infrastructuur ook allerhande gereedschappen om dit onderzoek uit te voeren. De CLARIAH Media Suite is één van de online onderzoeksomgevingen in de infrastructuur die zich specifiek richt op (audiovisuele) media.

De **Stichting Open Spraaktechnologie** stelt zich ten doel om hoogwaardige Nederlandse spraaktechnologie ter beschikking te stellen aan de Nederlandse onderzoeksgemeenschap, non-profit instellingen en geïnteresseerde derde partijen ten behoeve van het stimuleren van onderzoek en innovatie rond spraaktechnologie en het gebruik ervan. Denk aan het beheren van ontwikkelde Nederlandse, open-source spraaktechnologie, het stimuleren van doorontwikkeling van de technologie, en het faciliteren van laagdrempelige gebruik via diensten voor niet-commercieel gebruik bij publieke instellingen zoals archieven, publieke omroepen, overheidsinstellingen, universiteiten en hogescholen, en andere instellingen zonder winstoogmerk. Deelnemende instituten: Universiteit Twente, Radboud Universiteit Nijmegen, Universiteit Groningen, Beeld en Geluid, Meertens Instituut.