



Cognitieve achteruitgang detecteren met behulp van spraakherkenning via sensoren

In de enquête SHARE meten interviewers onder andere cognitieve achteruitgang van 50-plussers in pan-Europese landen. Cognitieve achteruitgang kan bijvoorbeeld een indicatie zijn van beginnende dementie. In het kader van de Sensor Data Challenge hebben wij onderzoek gedaan naar alternatieven om kwaliteit van spraak op een objectieve manier te meten door middel van sensoren.

RIANNE DRIJVER, SIGRID VAN HOEK, JONAS KLINGWORT & ROB WILLEMS

Op 22 en 23 april 2021 heeft het Centraal Bureau voor de Statistiek (CBS) samen met De Haagse Hogeschool, het Rijksinstituut voor Volksgezondheid en Milieu (RIVM) en de Universiteit Utrecht voor de derde keer de Sensor Data Challenge georganiseerd. Tijdens deze 24 uur durende hackathon werkten studenten en professionals – met een achtergrond in statistiek, data science, sensor studies,

engineering en IT – in acht kleine groepen van 3 tot 5 personen aan verschillende challenges met behulp van sensoren. De challenges gingen over specifieke problemen op het gebied van gezondheid, veiligheid en tijdsgebruik.

Vanwege de coronamaatregelen werd de hackathon online gehouden. De vorige twee edities van de Sensor Data Challenge waren op locatie, te weten bij het CBS en

op de Haagse Hogeschool. Teams waren daardoor flexibeler in de keuze voor sensoren vanwege de mogelijkheid om hardware te lenen. Deze keer kregen de deelnemers vooraf een sensorkit van Arduino. Deze kit maakt het mogelijk om automatisch verschillende variabelen zoals lichaamstemperatuur en vochtigheid te meten en te analyseren.

Sensoren worden steeds lichter en goedkoper. Door continue innovaties in Sensor Technology worden de metingen ook van steeds betere kwaliteit. Bovendien hebben waarnemingen met behulp van sensoren door het objectieve karakter de potentie uitkomsten van een vragenlijst aan te vullen of zelfs gedeeltelijk te vervangen. Vandaar dat onderzoekinstellingen waaronder het CBS al enige jaren onderzoek doen naar de toepasbaarheid van deze sensoren. Zo onderzoekt het CBS bijvoorbeeld de mogelijkheid om statistieken te verbeteren met behulp van Weigh-in-Motion sensoren, gegevens van verkeerslussen and Automatic Identification System (AIS) (zie [1, 2, 3]). Als deel van dit onderzoek wordt de Data Challenge georganiseerd.

- (aangedragen door het RIVM en de Arbo Unie)
- Het meten van kwaliteit van spraak en sociale interactie (aangedragen door SHARE)

Wij gingen aan de slag met het laatste onderwerp: het meten van kwaliteit van spraak en sociale interactie. De Survey of Health, Aging, and Retirement in Europe (SHARE) gebruikt op dit moment een vragenlijst om de kwaliteit van spraak te meten onder 50-plussers. Met interviews gehouden in 28 verschillende Europese landen en Israël is dit de grootste pan-Europese studie die het mogelijk maakt om microdata op het gebied van publieke gezondheid en sociaaleconomische levensomstandigheden longitudinaal en over landen heen te vergelijken. SHARE doet onderzoek naar vroege tekenen van cognitieve achteruitgang die bijvoorbeeld kunnen wijzen op voortekenen van dementie. Het doel van onze challenge was om een objectieve manier te vinden om spraakkwaliteit te schatten met behulp van een meting die geïntegreerd kan worden in het SHARE-interview. Als verdere eis werd gesteld dat de spraakkwaliteit-indicator onafhankelijk moet zijn van kennis, woordenschat en taalgebruik van de respondent.

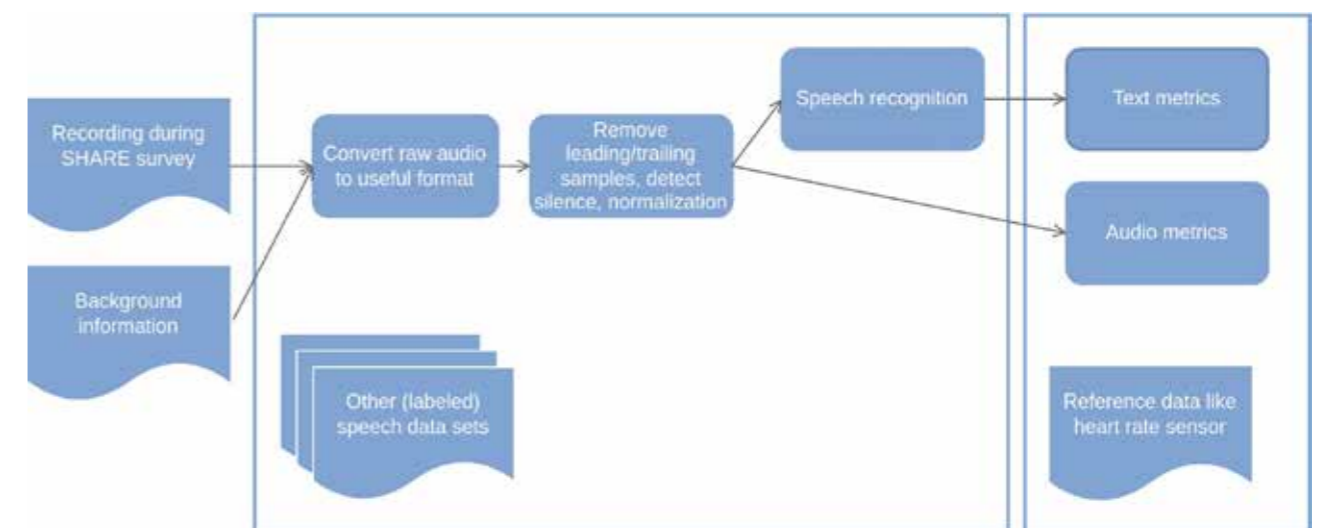
Challenges

In de challenge van 2021 werden de volgende vier onderwerpen aangedragen door een vijftal Nederlandse organisaties. Ze werden vervolgens verdeeld over de acht deelnemende groepen:

- Het meten van fysieke en mentale stress (aangedragen door het Ministerie van Defensie)
- Het meten van tijdsgebruik binnenshuis (aangedragen door het CBS)
- Het meten van blootstelling aan gevaarlijke stoffen

Opzet van de spraakkwaliteit-indicator

Het ontwikkelen van een spraakkwaliteit-indicator is een ingewikkelde taak. Gelukkig zijn er veel akoestische indicatoren beschikbaar om spraakkwaliteit te meten (zie bijvoorbeeld [4, 5]). Wij hebben een programma in de veelgebruikte computertaal Python geschreven dat geluid opneemt middels de microfoon van de laptop en vervolgens de snelheid van het spreken ('speech rate') en het



Figuur 1. Schema om de spraakkwaliteit-indicatoren te berekenen

gemiddelde aantal lettergrepen per seconde berekent. Hiervoor is een Speech-to-Text Classifier nodig. Vanwege de beperkte tijd hebben we de Google Cloud Speech API gebruikt, die in veel talen (waaronder Nederlands) beschikbaar is. De stappen om de indicatoren te berekenen zijn weergegeven in figuur 1. Naast deze twee indicatoren zouden zogenaamde *filled pauses* zoals ‘uhm’ en correcties tijdens het spreken kunnen duiden op cognitieve achteruitgang. We hebben een begin gemaakt aan het berekenen van de lengte en het aantal *filled pauses* door de geclassificeerde tekst te vergelijken met de geluidsgolven.

Naast het analyseren van de stem kan er ook worden gekeken naar andere lichamelijke factoren. Temperatuur, transpiratie en hartslag zijn factoren die kunnen veranderen wanneer iemand stress ervaart. Wanneer mensen moeite hebben met het lezen van een tekst of het beantwoorden van vragen, kan de lichaamstemperatuur omhooggaan, kunnen ze meer gaan transpireren of kan de hartslag omhoog gaan. Daarom gebruiken we sensor-gebaseerde metingen als aanvulling op de spraakwaliteit-indicatoren. We hebben daarom tijdens de challenge een handschoen samengesteld waarin zich sensoren bevinden die temperatuur, hartslag en galvanische huidreactie meten. De sensor die galvanische huidreactie meet detecteert transpiratie in de hand. Figuur 2 toont de handschoen die wij tijdens de challenge hebben gebruikt. De sensoren in de handschoen kunnen uitgelezen worden met een Arduino en de metingen kunnen bedraad



Figuur 2. Handschoen met sensoren voor het meten van temperatuur, hartslag en galvanische huidreactie

of wireless verstuurd worden naar een computer waar de data opgeslagen en geanalyseerd worden.

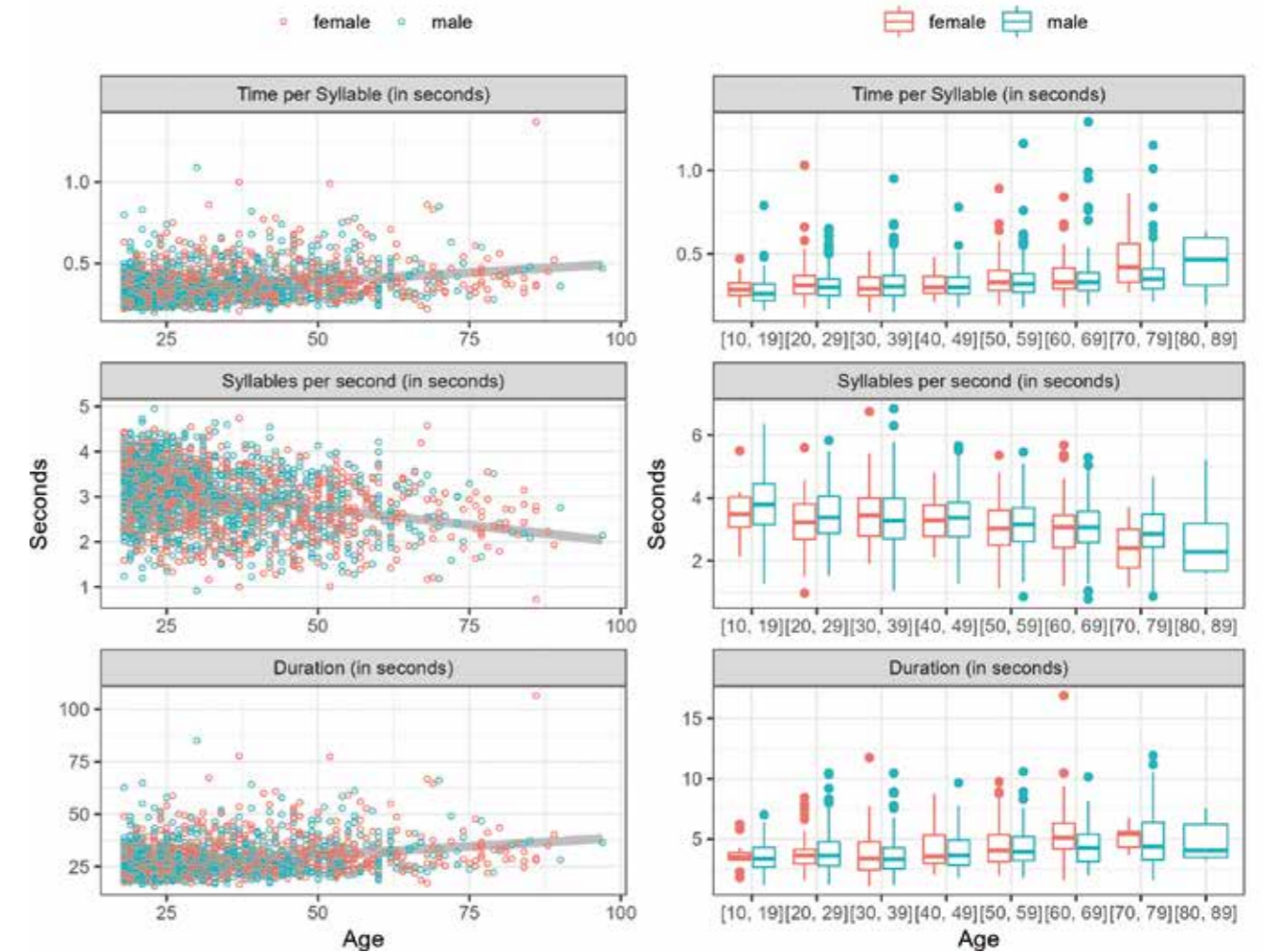
We stelden voor onze indicatoren in SHARE te integreren door deelnemers een korte passage te laten lezen en een kort verhaal te laten verzinnen (bijvoorbeeld door deelnemers te laten kiezen uit een reeks onderwerpen/afbeeldingen) zoals [6]. Dat laatste vereist het samenstellen en uitspreken van zinnen, en is daardoor cognitief gezien meer uitdagend. Op basis van de tekst en de opgenomen audio kunnen vervolgens de indicatoren berekend worden. Voor het SHARE interview begint, kan de desbetreffende persoon de handschoen aandoen, vervolgens kunnen er basiswaarden gemeten worden om later te vergelijken met de waarden die tijdens het interview gemeten zijn. Doordat de stem van de persoon en de sensoren allebei worden opgenomen kan de correlatie tussen de tekst, stem, temperatuur, hartslag en activiteit van de zweetklieren in de hand worden bepaald. De sensoren leveren op deze manier extra informatie over een eventuele cognitieve achteruitgang.

Indicatoren

In deze paragraaf illustreren we onze indicatoren met behulp van twee openbaar beschikbare datasets. De eerste dataset bestaat uit opnames van personen die allen dezelfde korte passage lezen, de tweede dataset bestaat uit opnames van personen die verschillende zinnen met een verschillend aantal woorden en verschillende inhoud lezen (de dataset zijn beschikbaar via [7, 8]). De resultaten van de indicatoren zijn weergegeven in figuur 3. De drie spraakwaliteit-indicatoren zijn afgezet tegen de leeftijd van respondenten, uitgesplitst naar geslacht. In beide datasets is te zien dat het aantal lettergrepen per seconde afneemt met de leeftijd, terwijl de tijd per lettergreep en de lengte van de spreektijd een lichte toename vertonen.

We concluderen dat onze methodologie in staat is onderscheid te maken tussen spraakwaliteit van verschillende leeftijdsgroepen in de dataset. Hoewel het nog verder ontwikkeld en gevalideerd dient te worden, concluderen we dat onze indicatoren potentie hebben om te gebruiken voor het meten van in ieder geval dit aspect van kwaliteit van spraak. De betrokkenen partijen in SHARE bleken geïnteresseerd te zijn in onze oplossing en vroegen om verdere uitwerking.

De eerste plek van de challenge was behaald door een team van studenten dat voorstelde NFC-chips te gebruiken om tijdsgebruik binnenshuis te meten. Het elegante



Figuur 3. De drie spraakwaliteit-indicatoren naar leeftijd en geslacht op basis van de eerste dataset (links) en de tweede dataset (rechts)

idee, samen met het werkende prototype en de presentatie maakte dit team tot winnaar. Het CBS zal dit idee verder oppakken en de mogelijkheden onderzoeken om het ook in de praktijk in te zetten.

Onze dank gaat uit naar Gerard Draadler (Student Bedrijfskunde, Haagse Hogeschool), die deel uitmaakte van het team tijdens de Sensor Data Challenge.

REFERENTIES

- [1] Klingwort, J., Buelens, B., & Schnell, R. (2019). Capture-recapture techniques for transport survey estimate adjustment using permanently installed highway sensors. *Social Science Computer Review*, 39(4), 527–542. doi: 10.1177/0894439319874684.
- [2] Puts, M. J. H., Daas, P. J. H., Tennekes, M., & de Blois, C. (2019). Using huge amounts of road sensor data for official statistics. *IMS Mathematics*, 4(1), 12–25. doi: 10.3934/Math.2019.1.12.
- [3] Consten, A., Puts, M., de Witt, T., Bisoti, E., Papandreou, C. P. K., Bis, M., Bliska, A., & Langsrud, Ø. (2017). *ESSnet Big Data, Work Package 4 AIS Data, Milestone 4.10 Progress and technical report of 1st internal WP-Meeting*. https://ec.europa.eu/eurostat/cros/sites/default/files/WP4_Milestone_4.10_2017_11_13.pdf.
- [4] Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145–164.
- [5] Weiner, J., & Schultz, T. (2018). Selecting features for automatic screening for dementia based on speech. *International Conference on Speech and Computer*, 747–756. Springer, Cham.
- [6] Rodgers, J. D., Tjaden, K., Feenaughty, L., Weinstock-Guttman, B., & Benedict, R. H. (2013). Influence of cognitive function on speech and articulation rate in multiple sclerosis. *Journal of the International Neuropsychological Society*, 19(2), 173.
- [7] Tatman, R. (2017). *Speech Accent Archive: Parallel English speech samples from 177 countries*. <https://www.kaggle.com/ratman/speech-accent-archive>.
- [8] Mozilla. (2017). *Common Voice*. <https://www.kaggle.com/mozillaorg/common-voice>

RIANNE DRIJVER (riannedrijver@gmail.com) studeert Sensor Technology – major elektrotechniek en is honours student bij de Hanzehogeschool. SIGRID VAN HOEK (st.vanhoek@cbs.nl), JONAS KLINGWORT (j.klingwort@cbs.nl) en ROB WILLEMS (rma.willems@cbs.nl) methodologen bij het Centraal Bureau voor de Statistiek.