

STRESSDETECTIE UIT SPRAAK

BE GROEP 1

Vic Degraeve, Sean Deloddere, Ernest Van Hoecke, Marie Vanzieleghem

Promotoren: Prof. dr. ir. Nilesh Madhu, Prof. dr. ir. Kris Demuynck, prof. dr. Marie-Anne Vanderhasselt, dr. Kristof Hoorelbeke

Begeleiders: Prof. dr. ir. Johan Bauwelinck, Prof. dr. ir. Dirk Stroobandt

Met dank aan: ir. Brecht Desplanques, ir. Bram Steenwinckel, dr. Olivier Janssens, prof. dr. ir. Sofie Van Hoecke

Project in het kader van het Vakoverschrijdend Projectvak in de Bachelors Computerwetenschappen en Elektrotechniek

Academiejaar: 2018 – 2019

Inhoudsopgave

1	Inleiding: Waarom stressdetectie uit spraak?	3
1.1	Stress	3
1.2	Stressdetectie uit spraak	3
1.3	Biometrische data	3
2	Technische uitvoering	4
2.1	Voorbereiding – Literatuurstudie	4
2.2	Datacollectie – Experimenten	4
2.2.1	Stressinductie en dekmantel	4
2.2.2	Proefpersonen ronselen	4
2.2.3	Verloop van het experiment	5
2.2.4	Ontwikkeling experiment-app	6
2.3	Dataverwerking	6
2.3.1	EDA	6
2.3.2	Finale dataset	7
2.4	Spraakverwerking	7
2.4.1	Stress in spraak	8
2.4.2	Mel-Frequency Cepstrum	8
2.5	Machinaal Leren	9
2.5.1	Parameters	9
2.5.2	Hyperparameters	9
2.5.3	Grid search	10
2.5.4	Overfitting	10
2.5.5	Cross-validation	10
2.5.6	Sensitiviteit en precisie	10
2.5.7	Classificatie met OpenSMILE-features	11
2.5.8	Classificatie met iVectors	15
2.5.9	Conclusie	20
2.5.10	Bemerkung bij cross-validation	20
2.6	Demo	20
3	Resultaten	21
3.1	Eerste opmerkingen: Bruikbaarheid van de data	21
3.2	Nauwkeurigheid van het model	21
3.2.1	Labels	21
3.2.2	Correlaties	22
3.3	Openstaande problemen en tekorten	23
3.4	Besluit	23
4	Praktisch	24
4.1	Planning en uitvoering	24
4.2	Communicatie, samenwerking en taakverdeling	24
A	Figuren	26
B	Protocol experiment	28
B.1	Cover story	28
B.2	Briefing	29
B.3	Randinformatie	29
B.4	Schuifregelaars – Visueel Analoge Schaal	29
B.5	Rust	29

B.6	Kalibratie – Fixed-form spraakdata	29
B.7	Montreal Imaging Stress Task (MIST) – Stressinductie	30
B.8	Feedback	30
B.9	Vragenlijsten	30
B.10	Debriefing	30
C	Android App	31
C.1	Opslag	31
C.2	Interface	31
D	Cross-validation	41
D.1	openSMILE	41
D.2	iVector	42
E	Informed consent	46
F	Invloed van de RRS en DASS schalen	47
F.1	Ruminative Response Scale (RRS)	47
F.2	Depression Anxiety Stress Scale (DASS)	47
G	Voorstel vakoverschrijdend project: Stemmingskwantificatie met NLP en wearable data	49
Lijst van Afkortingen		51
Bibliografie		53

1. Inleiding: Waarom stressdetectie uit spraak?

1.1 Stress

Stressgerelateerde klachten zijn een alsmaar toenemend fenomeen. Stress wordt door de WHO¹ zelfs *de* gezondheidsepidemie van de 21^{ste} eeuw genoemd [1]. Het aantal meldingen van overspannenheid en burn-out is dan ook de voorbije jaren flink gestegen [2]. Stress is noodzakelijk om te overleven, maar een te lange of te intense aanwezigheid kan leiden tot een pak problemen, waaronder bovengenoemde zaken, hoofdpijn, slaapproblemen [3], angststoornissen en depressie. Een efficiënte manier om stress te meten is dus aangewezen als eerste stap in zijn bestrijding. Bovendien is het wenselijk om er voor te zorgen dat het meten van deze stress zo weinig mogelijk moeite van de gebruiker vraagt. Stressdetectie uit spraak zou hiervoor een uitgelezen manier kunnen zijn.

1.2 Stressdetectie uit spraak

Stress wordt vandaag de dag al op verschillende manieren gemeten, voornamelijk aan de hand van vragenlijsten en biometrische data. Er zijn wel enkele nadelen verbonden aan deze methoden. Telkens lange vragenlijsten invullen is gewoonweg vervelend. Vragenlijsten zijn bovendien subjectief; personen kunnen bijvoorbeeld vragen verschillend interpreteren. Biometrische data – later, in sectie 1.3, meer over wat dit exact is – is dan weer objectief, maar er is vaak geavanceerde apparatuur nodig om de metingen uit te voeren. Mensen moeten langsgaan bij een expert, of vervelende wearables dragen. Stressdetectie uit spraak kent geen van deze nadelen. Een microfoon is tegenwoordig beschikbaar in iedere smartphone en af en toe iets inspreken is een pak vlotter en aangenamer dan herhaaldelijk lange vragenlijsten invullen.

Spraak is een enorm ingewikkeld proces, waarschijnlijk één van de lastigste taken die mensen uitvoeren zonder erbij stil te staan. Heel wat verschillende lichaamsdelen en het brein – later, in sectie 2.4, meer hierover – moeten samenwerken om te kunnen spreken. Hierdoor zijn er tijdens het spreken veel processen gaande waar de fysiologische reacties die stress teweeg brengt een invloed op kunnen hebben. Het is dus niet onwaarschijnlijk dat een accurate stressdetectie in spraak mogelijk is.

Een langetermijndoelstelling is om een app te maken die na bepaalde tijdsintervallen vraagt om iets in te spreken (of een persoon continu op te volgen en te herkennen wanneer er gesproken wordt), en aan de hand van die audio het stressniveau van de persoon te bepalen. Het zou dan mogelijk zijn om de evolutie van het stressniveau doorheen de tijd te bekijken. Een beeld krijgen van hoe stress verandert is uitermate interessant, zowel voor mensen met mentale gezondheidsproblemen, als bijvoorbeeld voor een bedrijfsleider die een idee wil krijgen van de stress van zijn of haar werknemers en dit wil vergelijken met hun productiviteit.

Vooraleer een dergelijke app ontwikkeld kan worden moet er natuurlijk eerst meer onderzoek gedaan worden naar de invloed van stress op spraak, en in welke mate deze gedetecteerd kan worden. Dit was de belangrijkste doelstelling van dit vakoverschrijdend project. Om dit te bereiken werd audio-, biometrische- enzelfrapportsdata, zowel voor als na een stressinductie, verzameld. De zelfrapportage werd gedaan aan de hand van VAS-schalen (Visueel Analoge Schalen, zie Appendix B.4).

1.3 Biometrische data

Na de literatuurstudie werd er besloten te focussen op twee soorten biometrische data: de geleidbaarheid van de huid (EDA) en het elektrocardiogram (ECG). Voor het verzamelen van deze data werd er gebruik gemaakt van twee wearables: de Imec ChillBand – die naast de geleidbaarheid van de huid ook de temperatuur meet en een accelerometer bezit – en een elektrocardiogramsensor die tevens door Imec voorzien werd. Deze ECG sensor klikt in een patch die op het borstbeen gekleefd wordt en bevat net zoals de polsbanden een accelerometer.

De ruwe, binaire data afkomstig van deze toestellen werd doorgestuurd naar een contactpersoon bij Imec, die ons dan de verwerkte features, uitgerekend op secondebasis, bezorgde. De software die hier verantwoordelijk voor is, is beschermd en kan niet door onszelf gebruikt worden. Deze features zijn bijvoorbeeld de *GSR_SCL* of de 'Galvanic Skin Response, Skin Conductance Level' en de *ECG_mean_heart_rate*, maar er zijn ook verschillende HRV (hartslagvariabiliteit) features.

¹Wereldgezondheidsorganisatie (Engels: World Health Organization)

2. Technische uitvoering

2.1 Voorbereiding – Literatuurstudie

Voordat de uitwerking van het project kon aanvatten, werd er nagekeken wat er in de literatuur al te vinden was in verband met stressdetectie uit spraak. Ons onderzoek bouwt deels verder op Imec's SWEET¹ onderzoek [4], waarbij de Imec ChillBands gebruikt werden voor het verzamelen van stressdata. Ook het idee van stressinductie (sectie 2.2.1) om aan stressdetectie te kunnen doen, kwam voort uit de literatuurstudie [5].

De literatuurstudie leverde verder ook op dat EDA een betrouwbare maat voor stressdetectie is [6] [7], en tevens dat de relatie tussen EDA en symptomen voor depressie reeds werd aangetoond [8]. Ook vonden we dat hartritmvariabiliteit (HRV) een mogelijke [9], hoewel omstreden, indicator voor stress kan zijn [10]. Daarnaast werd de combinatie van HRV, ECG en EDA met signalen als huidtemperatuur in het verleden al onderzocht via onder andere machinaal leren [11], [12]. In verband met detectie van emoties uit spraak werden tevens een aantal papers gevonden die met ons onderzoek in verband gebracht kunnen worden [13], [14].

Nog een interessant gegeven dat in de literatuur bestudeerd werd, is de asymmetrie in EDA [15]. Als aan beide polsen de EDA wordt opgemeten, zou dit asymmetrische signalen kunnen opleveren. Om verschillende redenen, waarvan ontbrekende EDA data voor één van beide polsbanden een belangrijke is, werd dit niet verder onderzocht. Ons onderzoek begon reeds een andere vorm aan te nemen en de extra polsband was handig om op terug te vallen in het geval van een slechte meting.

Verder werd er gezocht naar verschillende mogelijkheden die als manipulatiecontrole voor het experiment konden dienen [16]. Uiteindelijk kozen we voor de VAS, zie sectie 2.2.3. Ook wat betreft de stressinductieprocedure werden een aantal opties onderzocht en uiteindelijk werd voor de MIST [17] gekozen (zie sectie 2.2.1).

2.2 Datacollectie – Experimenten

2.2.1 Stressinductie en dekmantel

Om van de proefpersonen spraakfragmenten te verzamelen in zowel niet-gestresseerde als gestresseerde toestand, werd gebruik gemaakt van een stressinductiemethode. De stressinductieprocedure die geïmplementeerd werd is de korte versie van de 'Montreal Imaging Stress Task' (MIST, [17]). Deelnemers krijgen de taak om zo snel mogelijk rekensommen op te lossen, waarvan het antwoord altijd tussen nul en negen ligt. Ze kunnen niet rechtstreeks een antwoord aanduiden maar moeten met twee pijltjestoetsen een cijfer selecteren op een wiel, zie Appendix C.2 voor afbeeldingen. Tijdens deze sommen wordt bijgehouden hoe snel de deelnemer gemiddeld is, dit gebruiken we om de tijd per som tijdens de daadwerkelijke test te bepalen. Na ongeveer twee minuten eindigt de oefenronde en begint de echte test. Hier krijgt de deelnemer een timer te zien en een balk die aangeeft waar men scoort ten opzichte van de 'gemiddelde deelnemer'; een onhaalbare maatstaf. Op deze manier ontstaat een sociale stress, de deelnemer wil niet ondergemiddeld presteren en voor hen zit de afnemer van de test notities te maken. Samen met de hoge tijdsdruk en frustratie over de invoermethode induceert dit bij een meerderheid van de mensen stress.

Deelnemers mochten op voorhand niet weten dat we stress zouden induceren, want dit zou de stressinductie kunnen beïnvloeden. Om die reden werd een verhaal verzonnen dat als dekmantel tijdens het rekruteren van deelnemers verteld zou worden. Deelnemers kregen te horen dat het onderzoek in kader stond van snelle rekenvaardigheden die in verband gebracht zouden worden met biometrische signalen. De strak gedefinieerde leugen samen met de volledige procedure van het experiment is te vinden in Appendix B.

2.2.2 Proefpersonen ronselen

Een grote uitdaging was om genoeg deelnemers voor het experiment te vinden. Er werden twee Fnac-bonnen van 25 euro ter beschikking gesteld om mensen te lokken. Belangrijker was om duidelijk te vermelden dat er medische toepassingen konden zijn, wat mensen liever hoorden dan het feit dat ze moesten rekenen.

Er werd een reservatiesysteem opgezet via 'youcanbookme'; een systeem dat een handige synchronisatie met Google Calendar biedt. Twee mensen konden tegelijk het experiment komen afleggen, wat ook hielp met het overtuigen van potentiële proefpersonen. Deelnemers werden

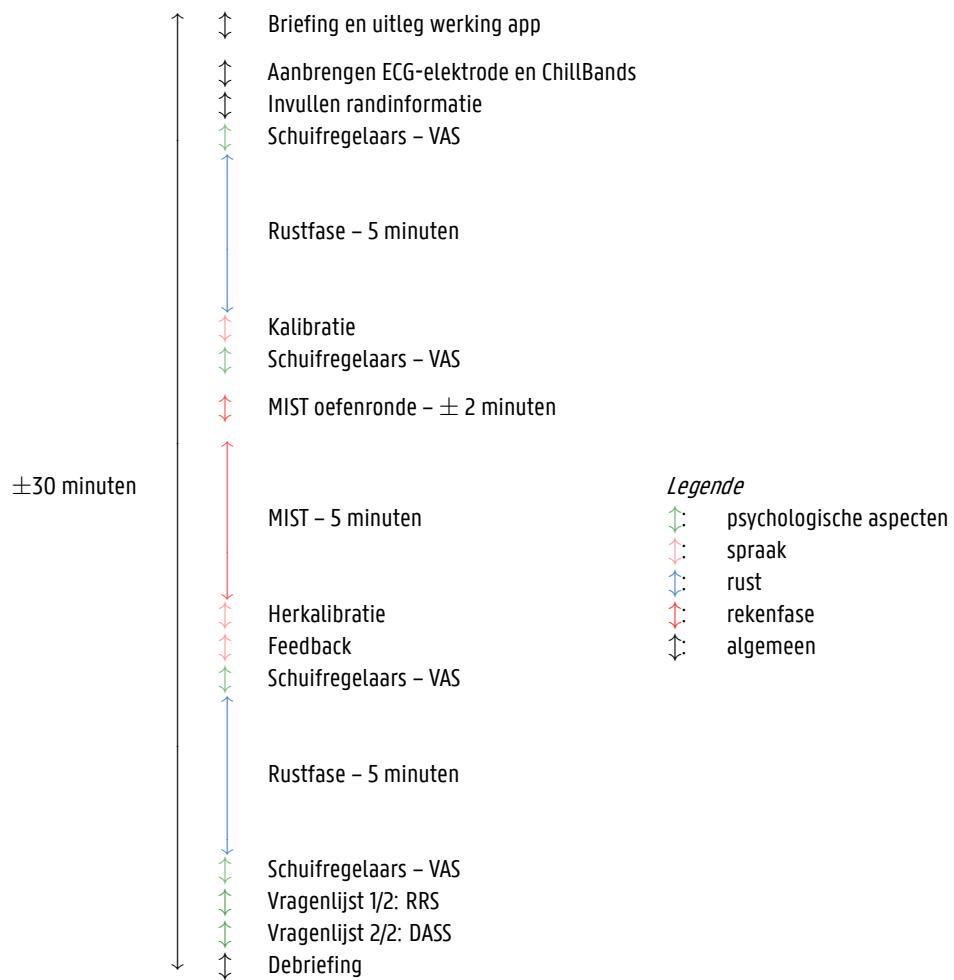
¹Stress in the Work Environment

gezocht op Twitter, Facebook, verschillende mailinglists van de universiteit en zelfs op straat. Iedereen die interesse toonde kreeg een link naar www.vopexperiment.be waar ze konden zien welke uren nog beschikbaar waren, en zich konden inschrijven. Het systeem stuurde automatisch herinneringsmails met links om een afspraak te verplaatsen of af te zeggen.

Slechts een handvol personen zijn niet komen opdagen. Uiteindelijk hebben 155 mensen deelgenomen, wat een mooi aantal is aangezien hun tijd niet vergoed werd.

2.2.3 Verloop van het experiment

Het experiment duurt ongeveer 30 minuten per persoon. Het volledige verloop van het experiment wordt weergegeven in Tabel 2.1 en wordt hieronder kort besproken. Om ervoor te zorgen dat de begeleiders op uniforme manier de experimenten afnamen, werd een gedetailleerd protocol opgesteld. Dit is te vinden in Appendix B.



Tabel 2.1: Verloop van het experiment: tijdlijn

Na binnenkomst kunnen de deelnemers tot rust komen, totdat een kamer vrij is. Vervolgens wordt hen, indien dit nog nodig is, het dekmantelverhaal verteld en wordt hen gevraagd of ze het sanitair willen gebruiken, waarna beide ChillBands en de ECG elektrodes worden aangebracht. De begeleider geeft vervolgens op de tablet een uniek id-nummer voor de deelnemer in, zodat anonimiteit gegarandeerd kan worden.

De deelnemer neemt plaats aan de tablet. Deze vult vervolgens randinformatie² in. Hierna wordt voor een eerste keer via schuifregelaars (VAS, Visueel analoge schaal) naar de gemoedstoestand gepeild. Vervolgens wordt gedurende vijf minuten, tijdens de rustfase, hun baseline gemeten. Dit is vooral belangrijk voor de hartslagvariabiliteit.

De test kan beginnen. Eerst wordt er gevraagd om een tekst voor te lezen. Er wordt verteld dat dit nodig is om het ASR (Automatic Speech Recognition) systeem te kalibreren. We vermelden dat we dit gebruiken om transcripties makkelijker te maken. Via dezelfde schuifregelaars als eerder wordt opnieuw aangeduid wat de huidige gemoedstoestand van de deelnemer is. De MIST wordt dan afgenoemd om stress te induceren, waarna we onmiddellijk weer dezelfde kalibratietekst laten voorlezen. Meteen hierna vragen we om feedback over de werking en

²leeftijd, geslacht, voorkeurshand, het al niet gebruiken van stimulerende middelen

interface van de app in te spreken. Dit laatste audiofragment wordt gebruikt als extra data om de algemeenheid van ons model te testen. Vervolgens kunnen de schuifregelaars opnieuw worden ingevuld, ook met zo weinig mogelijk pauze.

Na vijf minuten rust volgen de VAS nogmaals met tot slot de DASS (Depressie, Angst, Stress Schaal) en RRS (ruminative response scale). Dit zijn vragenlijsten om inzicht te krijgen in de mentale achtergrond van de deelnemer; de lijsten en hun invloed op ons onderzoek worden in detail besproken in Appendix F. Dit wordt achteraf gevraagd zodat het invullen van deze lijsten geen invloed zou hebben op de stressinductie. Ten slotte wordt men gedebriefed, waarbij verteld wordt dat we gelogen hebben en wat het echte doel van experiment is.

2.2.4 Ontwikkeling experiment-app

Om makkelijker data te verzamelen en bij te houden werd een Android app ontwikkeld in Java, in Android Studio. Deze app zorgt voor de structuur van het experiment, invoer en opslag van de VAS-schalen, timing van de rustperioden, opname van de spraakfragmenten en de afname van de MIST. Ook worden van zo goed als alle acties – verandering van activiteit, antwoord op een wiskunde vraag, fout of goed antwoord... – de precieze tijdstippen bijgehouden, om deze te kunnen correleren met de resultaten. Ook de twee psychologische vragenlijsten werden ingevuld en opgeslagen in de app.

De MIST werd door ons zelf geïmplementeerd. De gestelde rekenvragen dienen altijd een antwoord tussen 0 en 9 te hebben en worden telkens willekeurig genereerd. Het algoritme dat hiervoor werd geschreven voorziet parameters om de moeilijkheidsgraad aan te passen doorheen de test.

Alle informatie verzameld door de app werd geëxporteerd naar een JSON-bestand per deelnemer (zie Appendix C.1). De audiofragmenten werden opgeslagen in het verliesloze "WAV"-formaat met een samplingfrequentie van 48kHz. Deze instellingen werden bewust gekozen omdat het belangrijk is dat zo veel mogelijk informatie in de spraak behouden blijft. Een overzicht van de gebruikservaring van de app is te vinden in Appendix C.2.

2.3 Dataverwerking

Na het verzamelen van alle biometrische data worden deze omgezet naar CSV-bestanden. Hierin zijn kolommen telkens een feature (bijvoorbeeld de mediaan van de hartslag) verwerkt per seconde. Een script snijdt dan de data bij zodat er per deelnemer een bestand voor de ECG data, de linker- en de rechterpolssband gegenereerd wordt. Zo wordt een dataset gebouwd die bestaat uit folders voor elk ID nummer, waarin zich al de verzamelde data van de persoon met dat ID bevindt.

Met alle data op de juiste plek kan een ander script nagaan wat er ontbreekt. Ditzelfde script bevat een heleboel functies om de data te onderzoeken. Zo kan ongeldige en ontbrekende data gemakkelijk gevonden worden en kan met een simpel commando een plot gegenereerd worden. In figuur A.2 worden enkele resulterende plots weergegeven.

De dataset wordt zowel manueel als automatisch gecontroleerd. Voor de automatische controle zijn een paar belangrijke heuristieken ontwikkeld. Eerst werden deze toegepast op de elektrodermische signalen.

2.3.1 EDA

Er wordt gekeken naar de totale energie van het signaal, als deze onder een bepaalde drempel ligt, betekent dit dat de meting hoogstwaarschijnlijk niet juist verlopen is. Deze heuristiek steunt op het feit dat de elektrodermische activiteit een simpele geleidbaarheidsmeting is. Wanneer de elektrodes slecht contact maakten met de pols zal de geleidbaarheid afnemen of zelfs naar nul gaan indien het contact volledig verloren is. Deze effecten laten de energie van het signaal sterk dalen.

Hierna worden de EDA signalen door een hoogdoorlaatfilter gestuurd. De breekfrequentie van deze filter is heuristisch bepaald door naar de energie van verschillende frequenties te kijken met een histogram en deze in verband te brengen met plots waarvan we de EDA wel en niet vertrouwen. Na het filteren wordt de energie van het signaal berekend en boven een bepaalde grens wordt besloten dat de data mogelijk ongeldig is. De hoge frequenties wijzen namelijk op ruis of discontinuïteiten.

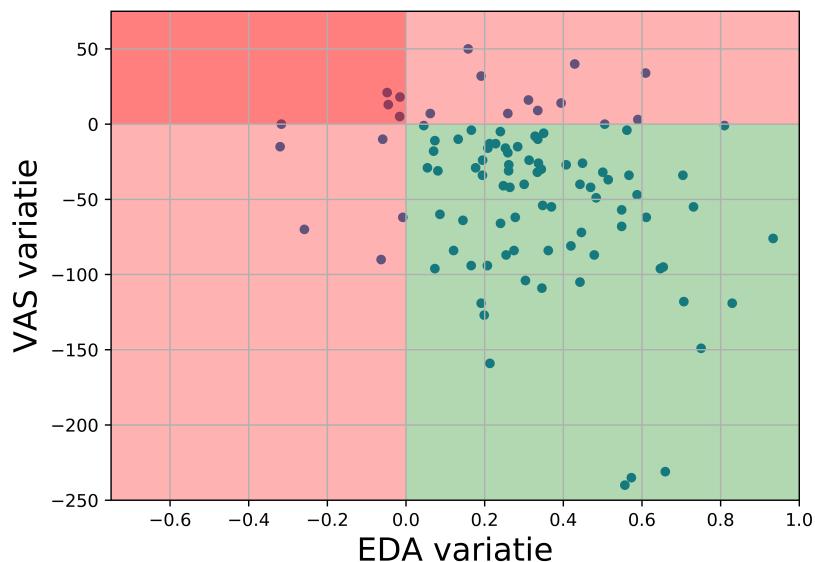
Ten laatste wordt er gekeken of er meerdere nullen aanwezig zijn in het signaal aangezien er in normale omstandigheden nooit een open kring zou mogen bestaan bij een EDA meting.

De functie `find_outliers_GSR` in `examine_population.py` implementeert deze heuristieken en laat weten welke deelnemers best nader bekijken worden om dan eventueel uit de dataset te verwijderen of te markeren als ongeldig.

2.3.2 Finale dataset

De ECG-data bleek al snel minder nuttig te zijn voor dit onderzoek, niet geholpen door het feit dat de gebruikte ECG patches niet bedoeld waren voor herhaald gebruik en bijgevolg ruis of foute metingen introduceerden. Er werd geen significante correlatie gevonden tussen de verschillende ECG features en de scores van de schuifregelaars. Met de EDA daarentegen is het duidelijk dat de geleidbaarheid van de huid toeneemt door de MIST, en dat de deelnemers zelf een meer negatieve gemoedstoestand aanduiden. Dit is te zien op figuur 2.1. De Pearson coëfficiënt hiervan is -0.31. Enkel de heuristisch geldig beschouwde data is afgebeeld op deze figuur.

De VAS score wordt – zoals verder toegelicht in Appendix B.4 – berekend door de positieve affect schuifregelaars bij elkaar op te tellen en de negatieve ervan af te trekken. De VAS variatie is dan deze score vlak na de MIST, min dezelfde score van vlak voor de MIST. Zo duidt een negatieve VAS-variatie op een daling in de gemoedstoestand terwijl een positieve variatie wijst op een stijging. De EDA-variatie werd bekomen door te kijken naar de EDA-waarde vlak na en vlak voor de MIST, en deze te normaliseren door te delen door de piekwaarde.



Figuur 2.1: Variatie van de EDA tegen de variatie van de gemoedstoestand tijdens de MIST

2.4 Spraakverwerking

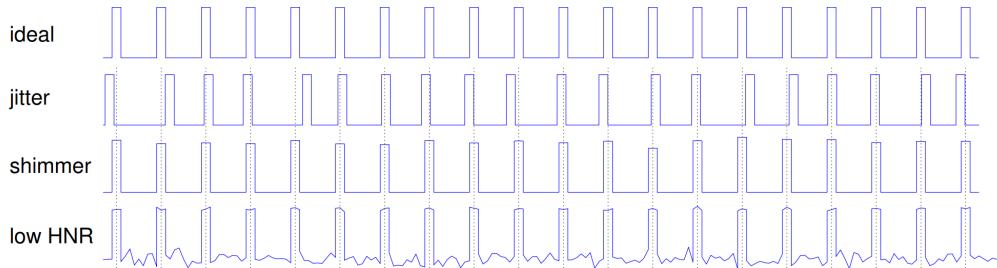
Het is nodig om te begrijpen hoe de mens spraak creëert en belangrijker nog, hoe deze efficiënt verwerkt kan worden. Aangezien 155 deelnemers hebben deelgenomen aan het experiment is de dataset niet heel erg groot en is een slimme verwerking noodzakelijk. Hiervoor worden in het volgende onderdeel verschillende technieken besproken, die al dan niet steunen op een aantal spraakverwerkingsconcepten die alsook toegelicht worden.

Er zijn veel kenmerken omvat in spraak omdat het tot stand komt door het samenwerken van meerdere lichaamsdelen en het brein. Zo zijn de lippen, tong, onderkaak, het gehemelte en het strotklepje voorbeelden van wat men articulatoren noemt omdat ze ofwel rechtstreeks of onrechtstreeks (bijvoorbeeld door de tong tegen het gehemelte te duwen) bestuurd kunnen worden. De stembanden behoren hier niet toe, ondanks het feit dat ze ook snelle bewegingen maken. De longen pompen lucht door het strottenhoofd en de luchtstroom doet de stembanden vibreren met een gegeven vibratiefrequentie. Deze frequentie kan veranderen door de lengte en spanning van de stembanden alsook het drukverschil aan te passen. Op die manier ontstaat de toonhoogte. De rest van de luchtpijp en mond gedraagt zich dan als filter en kan zo de andere eigenschappen van het geluid, voornamelijk de klankkleur, aanpassen.

Om effectief spraak te creëren kunnen de stembanden de luchtstroom snel aan- en afsluiten. Zo ontstaat een pulsentrein die door de filter vloeit, die als transferfunctie beschouwd kan worden, en bijgevolg gezien kan worden als een convolutie in het tijdsdomein en dus een vermenigvuldiging in het frequentiedomein. Aangezien de filter afstelbaar is kunnen zo een hele boel verschillende frequentiepatronen (klanken) gegenereerd worden. De lucht kan ook door een nauwe opening van de stembanden geduwd worden wat leidt tot turbulente stroom en dus een hoogfrequent, ruizig geluid. [18]

2.4.1 Stress in spraak

Er kan nog veel gezegd worden over spraak. Hierboven staat slechts een korte uitleg over wat het lichaam doet om te illustreren hoeveel factoren verantwoordelijk zijn voor de uiteindelijke klanken, woorden en zinnen. Spraak bevat dus vele kenmerken die in dit verslag features zullen genoemd worden, zoals in de literatuur over machinaal leren. Verbale informatie omvat de zinsstructuur, semantiek, context, fonetiek en prosodie van wat er gezegd wordt. Zeker via prosodie zal stress de stem beïnvloeden. De stem bevat ook veel niet-verbale informatie zoals de toonhoogte, het volume, tempo en modulatie die tevens sterk beïnvloed worden door de gemoedstoestand van de spreker. Op figuur 2.2 zijn voorbeelden te zien van afwijkingen op de ideale pulsentrein die een effect zullen hebben op de stem en een gevolg zijn van minder controle over de stembanden.



Figuur 2.2: Voorbeelden van niet-idealiteiten op de pulsentrein [18]

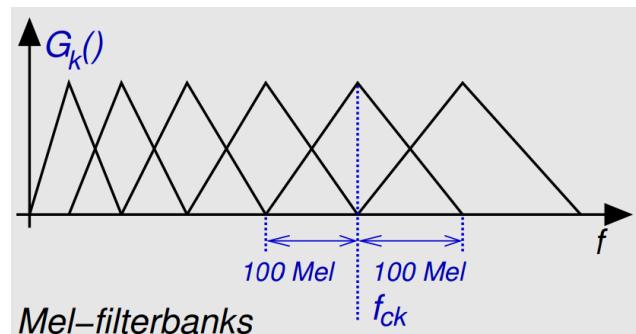
Doordat spraak deels periodiek opgebouwd is kan er veel informatie vergaard worden uit het frequentiedomein. In het tijdsdomein is enkel de amplitude en de aan- of afwezigheid van de pulsentrein makkelijk zichtbaar. Dit maakt een spectrogram een goede visualisatiekeuze, aangezien zo de aanwezige frequenties doorheen de tijd zichtbaar zijn.

2.4.2 Mel-Frequency Cepstrum

Het menselijk gehoor is afgestemd op de spraak. Dit zorgde voor de inspiratie om te kijken naar de limieten van het gehoor om spraakinformatie efficiënt voor te stellen. In het oor, meer bepaald in het slakkenhuis, exciteren verschillende golflengtes de membranen op een specifieke positie. Deze posities liggen meer op een logaritmische schaal. De Mel-frequentieschaal is een subjectieve benadering van de menselijke ervaring.

De MFCCs zijn de Mel-frequency cepstral coefficients. Samen vormen ze het Mel-frequency cepstrum (MFC). Het verschil tussen een cepstrum en het Mel-frequency cepstrum is dat de frequentie banden van het MFC even ver van elkaar liggen op de Mel-schaal, wat beter het menselijk gehoor benaderd dan de lineaire frequentiebanden van een normaal cepstrum. Deze voorstelling wordt vaak gebruikt om audio te comprimeren, maar de MFCCs zijn hier ook erg toepasselijk als features om mogelijk stress uit te herkennen. Dankzij de compactere voorstelling van informatie van deze methode hebben modellen minder vrije parameters en hebben ze tevens minder data nodig.

Deze coëfficiënten worden berekend door een Fourier transformatie uit te voeren en het vermogen van het spectrum af te beelden op de Mel-schaal via een driehoekige filterbank. Dan neemt men de logaritmes hiervan op elke Mel-frequentie en wordt een discrete cosinustransformatie uitgevoerd alsof de lijst van logaritmes een signaal is. De amplitudes van het resulterend spectrum zijn dan de MFCC's. De laatste transformatie wordt uitgevoerd om de individuele parameters te decorreleren. Zo is er weinig redundante informatie en kunnen ze gemakkelijk dienen als input voor verscheidene modellen. Deze methode wordt gebruikt om de features af te leiden maar hebben wij niet zelf geïmplementeerd (zie verder). Op figuur A.1 is te zien hoe Mel-cepstra opgebouwd zijn uit MFCCs en wat de invloed is van meer of minder coëfficiënten [18].



Figuur 2.3: Mel-filterbank [18]

De cepstrale coëfficiënten omsluiten het spectrale vermogen. De verandering van deze coëfficiënten in de tijd is belangrijk aangezien er anders

geen rekening gehouden wordt met de tijdsgebaseerde effecten. Shifted Delta Features helpen met taal- en sprekerherkenning [19], ze worden gecreëerd door afgeleides berekend op meerdere tijdstippen samen te nemen. Zo verstrekken ze veel tijdsafhankelijke informatie.

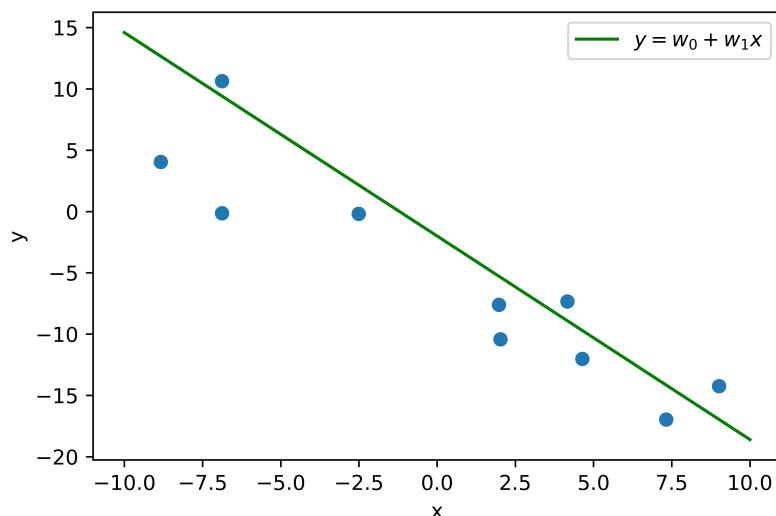
2.5 Machinaal Leren

Met de spraakfragmenten verzameld tijdens de experimenten kan nu een model getraind worden dat op basis van spraak kan voorspellen of een persoon al dan niet gestresseerd is. Over het algemeen is het fragment opgenomen vóór de stresserende wiskundetaak één van rustige spraak, en dat erna van gestresseerde spraak. Het herkennen van de categorie gestresseerd of rustig is een binair classificatieprobleem waarvoor data beschikbaar is, ideaal voor machinaal leren dus. Machinaal leren zoekt verbanden tussen deze spraakfragmenten en hun labels, en leert met deze verbanden ook nieuwe spraak te categoriseren.

Er zijn veel manieren om aan machinaal leren te doen, elk met hun eigen troeven en toepassingen. Voor deze toepassing werden twee verschillende aanpakken onderzocht.

2.5.1 Parameters

De courantste klasse schatters in machinaal leren zijn de parametrische modellen. Kort gezegd betekent dit dat deze methoden relevante informatie uit de invoerdata op een of andere manier encoderen naar een vast aantal parameters. Bij lineaire regressie tussen twee "features", de eenvoudigste vorm van machinaal leren, zijn dit simpelweg de coëfficiënten van een eerstegraadsfunctie. Het bepalen van de optimale parameters voor een gegeven dataset en labels wordt "trainen" genoemd.



Figuur 2.4: Simpele lineaire regressie. Hier zijn w_0 en w_1 zijn de interne parameters afgeleid uit de datapunten.

2.5.2 Hyperparameters

Andere parameters, hyperparameters, beïnvloeden dit trainingsproces en zijn resultaten, maar moeten dus op voorhand gekozen worden. Sommige hyperparameters kunnen meteen berecalculated en afgesteld worden aan de toepassing, maar vaak moet hiervoor ook een optimalisatie op basis van de dataset worden uitgevoerd. Dit proces heet "tunen". Het tunen van hyperparameters vereist een prestatiemaat om te optimaliseren. De makkelijkst interpreteerbare - maar zeker niet altijd de beste - maat voor klassificatie is de nauwkeurigheid of "accuracy". Dit is de verhouding tussen het aantal juist voorspelde klassen en de totale hoeveelheid voorspellingen. Omdat onze dataset even veel voorbeelden van gestresseerde en rustige spraak bevat, en (voorlopig) de kost van een vals positieve of vals negatieve voorspelling even groot is,³ werd er in alle optimalisaties die volgen gekozen voor de nauwkeurigheid als prestatiemaat.

³indien deze voorwaarden niet vervuld waren was de nauwkeurigheid een zeer slechte keuze geweest

2.5.3 Grid search

Het tunen van hyperparameters gebeurt doorgaans in een "grid search". Per hyperparameter dient er een reeks mogelijke waarden ingesteld te worden. Een grid search traint en evalueert een model voor alle combinaties van deze mogelijke waarden. De parameters die zorgen voor de beste prestatie, beschreven door de gekozen prestatia maat, zijn de optimale parameters.

Indien het model wordt geëvalueerd op dezelfde data waarmee het getraind is, zal deze prestatie echter zelden overeenkomen met het gedrag op nieuwe invoer. Het model heeft deze data al gezien en is erop afgesteld, dus zal waarschijnlijk een juiste voorspelling maken. Dit wordt opgelost door de dataset op te splitsen in een train- en validatieset. Het model wordt tijdens de gridsearch telkens getraind op de trainsubset, en geëvalueerd op de validatieset. Wanneer de optimale instellingen zijn gevonden kan het model uiteraard opnieuw getraind worden op de volledige dataset.

2.5.4 Overfitting

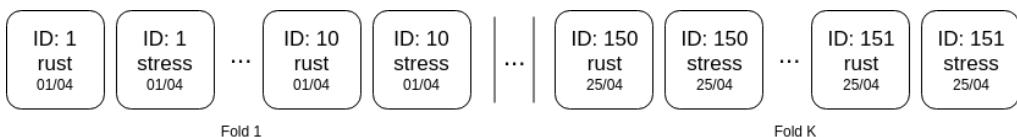
Een van de grootste uitdagingen in machinaal leren is het voorkomen van overfitting. Er wordt gezegd dat een model overfit op de traindata, als de prestatie op de validatieset niet overeenkomt met deze in de echte wereld. Een overfit model generaliseert niet naar ongeziene data. Wanneer de interne parameters van een model overfit zijn⁴, moet het geregulariseerd worden. Vaak kan dit door het zorgvuldig aanpassen van een regularisatiehyperparameter.

2.5.5 Cross-validation

Een model kan ook overfitten op de gekozen validatiedata. Dit gebeurt als een bepaalde configuratie hyperparameters toevallig goed werkt voor een gekozen validatieset die niet gelijk verdeeld is met de werkelijkheid, en dus niet goed generaliseert naar ongeziene data. Dit kan worden opgelost door elke configuratie in een grid search meerdere keren te evalueren met verschillende train- en validatiesets, en het gemiddelde te nemen van de resulterende prestaties als nieuwe prestatia maat: cross-validation. Bovendien moeten alle validatie- en trainsets gelijk verdeeld (i.i.d.⁵) zijn.

Wij gebruiken "KFold cross-validation" om hiervoor te zorgen. In KFold cross-validation wordt de gehele dataset opgedeeld in k even grote subsets of 'folds'. Deze subsets dienen gelijk verdeeld en onafhankelijk van elkaar te zijn. Een van deze subsets kan nu telkens gebruikt worden als validatieset, terwijl de unie van alle overige sets als traindata kan dienen. Op deze manier kan elke configuratie nu k keer geëvalueerd worden.

Voor onze toepassing is het belangrijk dat spraak van eenzelfde persoon ook in eenzelfde fold zit. Is dit niet het geval, dan zijn de folds niet langer onafhankelijk, en kan het zich voordoen dat een model dat bijvoorbeeld reeds getraind is op gestresseerde spraak van een persoon, tijdens cross-validation het rustige fragment van deze persoon moet evalueren. Deze datapunten in verschillende folds zijn niet onafhankelijk van elkaar. Het kan dan bijvoorbeeld zijn dat het model de neiging zal hebben te voorspellen dat dit een fragment van gestresseerde spraak is, omdat het de gestresseerde tegenhanger van deze persoon al kent, en de features van dit fragment mogelijks op die van het reeds gekende lijken. Het is dus nodig om het model te dwingen te letten op relevante verschillen, en hiervoor moeten geschikte folds gekozen worden. Eenzelfde redenering kan worden uitgezet voor het afdwingen dat data opgenomen op eenzelfde dag in eenzelfde fold zit (fragmenten opgenomen op dezelfde dag zijn waarschijnlijk niet volledig onafhankelijk van elkaar).



Figuur 2.5: Illustratie van de gebruikte cross-validation-strategie

2.5.6 Sensitiviteit en precisie

Om de finale prestatie van statistische modellen beter te kunnen analyseren worden doorgaans de maten sensitiviteit en precisie aangewend.

De sensitiviteit wordt als volgt gedefinieerd:

$$\text{sensitiviteit} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.1)$$

Terwijl de precisie volgende betekenis heeft:

$$\text{precisie} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2.2)$$

⁴en zijn voorspellingen dus baseert op irrelevant variaties in de traindata die zich niet noodzakelijk voordoen in de echte wereld

⁵independent and identically distributed

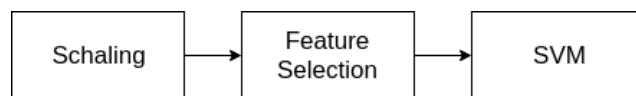
waarbij:

TP	Echtpositief
FP	Foutpositief
TN	Echtnegatief
FN	Foutnegatief

Een hoge sensitiviteit houdt dus in dat er slecht een kleine hoeveelheid voorspellingen valsnegatief zijn. De sensitiviteit is in feite een andere naam voor de waarschijnlijkheid dat positieve gedetecteerd worden. Een hoge precisie wijst op weinig valspositieven en is de kans dat een voorspellend positief in werkelijkheid ook een positief is.

2.5.7 Classificatie met OpenSMILE-features

De eerste en eenvoudigste aanpak die onderzocht werd maakt gebruik van een bestaande "feature extractor" genaamd openSMILE [20] die per spraakfragment een lijst interessante features (eigenschappen) samenstelt. Deze set van features wordt eerst herschaald, vervolgens met behulp van een analyse van hun covarianties uitgedund en tot slot wordt dit gevoed aan een Support Vector Machine.



Figuur 2.6: Overzicht van de openSMILE-architectuur

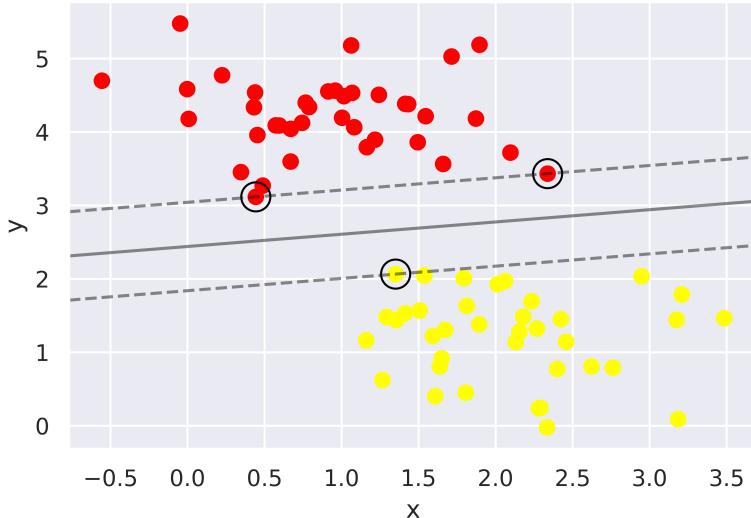
Features

OpenSMILE is een opensourcesoftware-project met als doel audiosegmenten op een zodanige manier te beschrijven dat traditionele classificatiemodellen hier makkelijk emotie in kunnen herkennen. Het openSMILE-project is gestart in 2008 en is nog steeds populair in het "Affective Computing"-onderzoeksgebied [20].

OpenSMILE biedt verschillende samenstellingen van features bedoeld voor verschillende doeleinden. Na de nodige experimentatie werd er gekozen voor de meegeleverde "emobase"-configuratie (Florian Eyben, 2009). Deze zet elk audiofragment om in een reeks van 988 features. Deze lijst omvat onder andere statistieken over de MFCC's, toonhoogtes en jitter of shimmer in de stembanden.

Support Vector Machines

Een Support Vector Machine (SVM) is een gesuperviseerd machinaal leren-model. De term gesuperviseerd slaat op het feit dat voor elk trainingspunt een label wordt meegegeven. Een SVM is een binair classificatiemodel en probeert twee labels lineair te scheiden met een hypervlak. Het optimalisatiecriterium zorgt voor een maximale afstand tussen dit hypervlak en de dichtse datapunten van elke klasse, of anders gezegd, een zo breed mogelijke marge. De datapunten op het hypervlak van elke klasse worden de support vectors genoemd. Deze support vectors karakteriseren het getrainde model volledig.



Figuur 2.7: Illustratie van het SVM optimalisatiecriterium met twee features, x en y . De volle lijn is het hypervlak en de stippellijn is de rand van de marge. De support vectors zijn omcirkeld.

Preprocessing

Gezien een SVM inherent steunt op de afstand van datapunten in de ruimte opgespannen door de features van de invoer, is het belangrijk dat één of een subset van de features deze maat niet domineert.

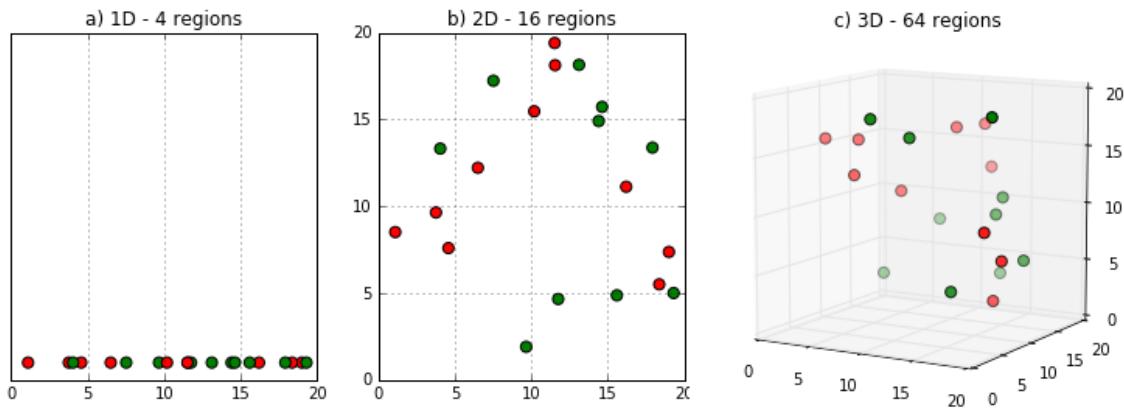
Elke feature moet dus geschaald worden zodanig dat deze allen binnen eenzelfde vooraf gekozen bereik vallen. Hiervoor wordt een aparte schaler "getraind" met de traindata. Voor elke dimensie van de invoer worden het gemiddelde μ en de standaardafwijking σ opgeslagen. Vooraleer deze aan het model door te geven wordt elke invoervector x nu eerst als volgt herschaald:

$$z = \frac{x - \mu}{\sigma} \quad (2.3)$$

Op deze manier zal elke feature een gemiddelde van 0 en variantie van 1 hebben, en dus meestal waarden aannemen in $[-1, 1]$.

Feature selection

De analyse van hoogdimensionale data brengt problemen met zich mee. Dit heeft in de literatuur de ietwat dramatische naam "Curse of Dimensionality" gekregen. Deze "vloek" beschrijft het feit dat wanneer de dimensionaleit van een ruimte stijgt, het volume ervan exponentieel toeneemt. In de context van machinaal leren houdt dit in dat de hoeveelheid data nodig om de gehele invoerruimte te classificeren, ook exponentieel stijgt. Met hoogdimensionale data en een te kleine dataset zal het overgrote deel van de ruimte leeg zijn, waardoor het label voor een groot aantal combinaties van input features waar het model geen voorbeelden van heeft gekregen, eigenlijk niet voorspeld kan worden.



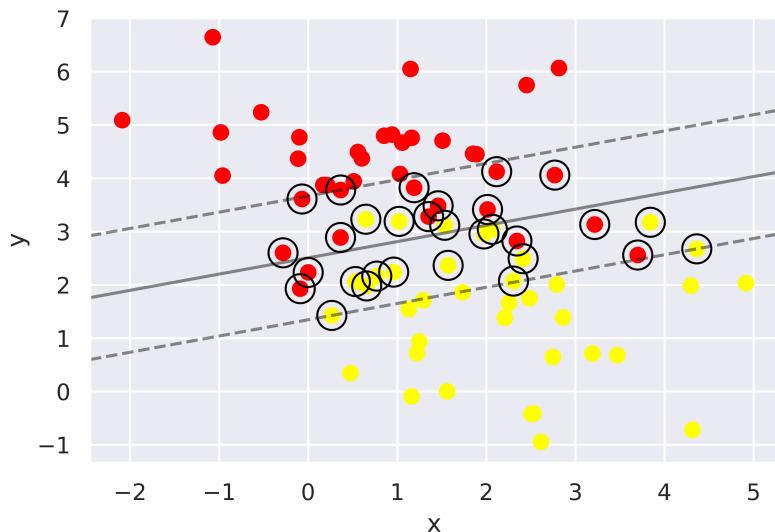
Figuur 2.8: Illustratie van the curse of dimensionality. Verdeelt men alle assen in de ruimte in 4 en telt men de resulterende regio's, dan stijgt dit exponentieel. Een steeds kleiner deel van de totale ruimte is ingenomen door datapunten [21].

Teneinde dit fenomeen tegen te gaan is het een goed idee de dimensionaliteit van de invoer zo laag mogelijk te houden. Hiervoor bestaan verscheidene "feature selection"-technieken. Voor dit project werd de "SelectKBest"-methode gehanteerd. SelectKBest rangschikt alle features volgens een gegeven prestatiemaat en selecteert enkel de K (dit is een tunable parameter) beste features.

Variantieanalyse of ANOVA (van het Engelse Analysis of variance) werd gekozen als prestatiemaat voor SelectKBest. Variantieanalyse vergelijkt de variantie binnen eenzelfde feature met de variantie van features tussen elkaar om te beslissen welke features het meest informatief zijn.

Regularisatie

Eerder werd vermeld dat een SVM de afstand tussen de dichtste datapunten van elke klasse en het hypervlak maximaliseert. Maar wat als de twee klassen niet linear scheidbaar zijn? In dit geval is er geen oplossing voor het minimalisatieprobleem. De optimalisatiecriteria zullen dus afgezwakt moeten worden. Het zal nu wel toegelaten zijn dat datapunten in de marge (tussen de support vectors) liggen: we spreken nu over een "soft margin SVM". De regularisatieparameter λ regelt hoe hard deze misclassificaties bestraft worden. Voor λ naderend naar oneindig zullen er geen samples in de marge worden toegelaten, wat overeenkomt met de eerder beschreven "hard margin SVM". Hoe dichter λ bij nul nadert, des te meer misclassificaties worden toegelaten.



Figuur 2.9: Regularisatie van een SVM laat toe dat er datapunten binnen die marge vallen, dit zijn dan ook support vectors die het model karakteriseren. Hier werd gekozen voor een λ van 0.1.

Het toelaten van foute voorspellingen regulariseert het model: er wordt minder naar de irrelevante details van de dataset gekeken en meer

naar de algemene trends in de data.

Kernel

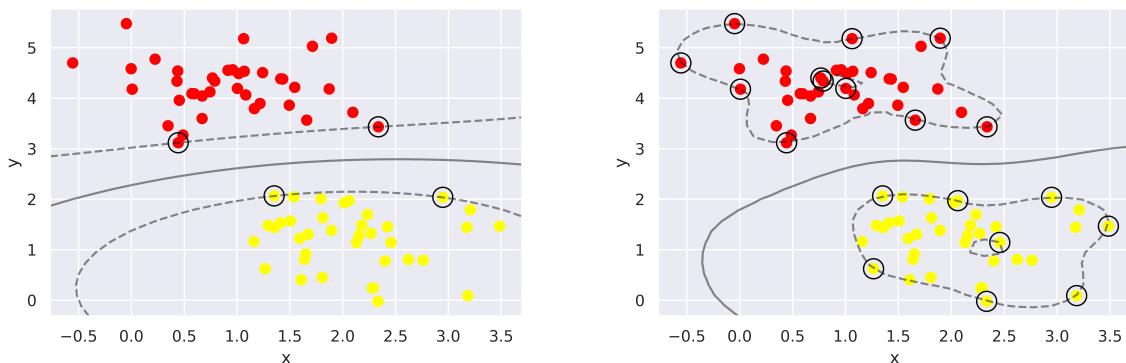
In veel gevallen zal de optimale scheiding tussen de klassen niet lineair zijn. Dit probleem zou omzeild kunnen worden door de invoerruimte eerst, via een niet-lineaire transformatie, om te vormen in een ruimte waar de klassen wel lineair van elkaar te onderscheiden zijn. Dit vereist mapping en opslag van elk datapunt naar een punt in deze nieuwe ruimte.

Gezien de werking van een SVM louter gebaseerd is op de afstandsmaat tussen verschillende monsters in de invoerruimte, volstaat het om een afstandsmaat in de nieuwe ruimte uit te drukken in functie van de oude features, en de Euclidische afstand te vervangen door deze "kernel"-functie, om niet-lineair gedrag te verkrijgen. Dit wordt in de literatuur "the kernel trick" genoemd.

Het is zelfs geen vereiste dat deze functie een echte afstandsmaat in een bestaande ruimte is; er bestaan zelfs kernels die een oneindig-dimensionale ruimte moeten voorstellen. De Radial Basis Function (RBF) kernel, de kernel gebruikt in dit project, is een voorbeeld hiervan:

$$K(x, x') = \exp\left(-\frac{\|x - x'\|}{2\sigma^2}\right) \quad (2.4)$$

Met behulp van de Taylor-expansie kan nagegaan worden dat deze functie een dotproduct in een oneindig-dimensionale ruimte voorstelt. Deze ruimte past zich zodanig aan dat hoofdassen lopen volgens de ruwe vorm van de data. De standaardafwijking σ beïnvloedt de mate waarin de nieuwe ruimte kan "plooien" naar complexere scheidingen. Een goede keuze van deze parameter is dus belangrijk om overfitting (de grens focust zich te veel op de gekozen trainingset) tegen te gaan.



Figuur 2.10: Beslissingsgrens van een SVM met de RBF kernel. Links is een σ van 10 gebruikt, rechts is de waarde hiervan 1.

Tuning

De opeenvolging van stappen uitgelijnd in de voorbije paragrafen werd geïmplementeerd in het Python-framework "scikit-learn" [22]. De besproken hyperparameters werden getuned in een 4-fold cross-validated grid search. Een bemerking hierbij is dat de parameter σ in scikit-learn voorgesteld wordt door zijn inverse; gamma. Parameter λ draagt de naam C in dit framework.

De cross-validation-curves in Appendix D.1 illustreren de invloed van de verschillende hyperparameters op de prestatie van het model.

C	gamma	K	kernel
10	0.001	110	rbf

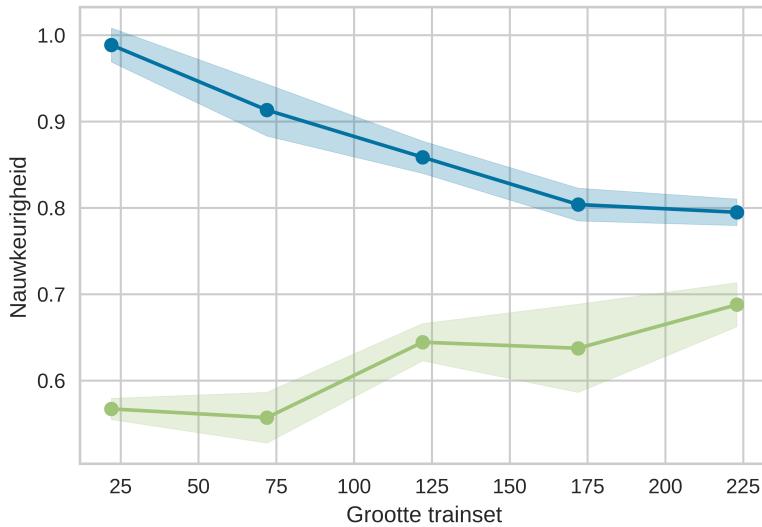
Tabel 2.2: Overzicht van de gevonden optimale hyperparameters

Bovenstaande hyperparameters werden bekomen door optimalisatie van de nauwkeurigheid. Dezezelfde waarden werden ook gevonden wanneer de precisie werd geoptimaliseerd.

Evaluatie

De finale op spraakdata getunedde en getrainde Support Vector Machine behaalt een cross-validation nauwkeurigheid van ongeveer 68%. Dit is een behoorlijk resultaat. Echter is deze aanpak sterk gelimiteerd door de features van openSMILE, die louter statistieken over een fragment

zijn. Bovendien is dit model niet heel robuust: het is zeer gevoelig aan de klanken die voorkomen in de spraak. De nauwkeurigheid op andere teksten dan de kalibratietekst ligt hoogstwaarschijnlijk veel lager.



Figuur 2.11: Leercurve van de geoptimaliseerde SVM

Uit een analyse van zijn leercurve, kan heel wat informatie over de SVM worden gedestilleerd. Zo valt af te lezen dat voor een kleinere trainset, de trainnauwkeurigheid extreem hoog is omdat het model met zo weinig data niet anders kan dan overfitten. Meer data zorgt ervoor dat de SVM "inziet" dat er overfitting aan de gang is: de nauwkeurigheid daalt tot een bepaalde vaste score wordt bereikt. Dit is de maximale onderscheidende kracht van het model in de gekozen configuratie.

Er valt ook af te lezen dat dit model beter zou presteren met meer data: de validatiennauwkeurigheid is nog niet afgevlakt. Meer data zou deze nauwkeurigheid dichter brengen naar die van de trainset, die wel al een stabiele waarde heeft bereikt.

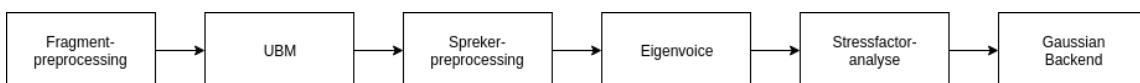
sensitiviteit	precisie	nauwkeurigheid
65.13%	70.27%	68.44%

Tabel 2.3: Overzicht van prestatie openSMILE

Bovenstaand overzicht licht de prestatie van het openSMILE + SVM-model toe. Een precisie van 70% wilt zeggen dat 7 van de 10 voorspellingen van stress ook effectief fragmenten waren opgenomen na de MIST. De sensitiviteit is de waarschijnlijkheid dat een gestresseerd spraakfragment ook gedetecteerd wordt, hier 65%. De nauwkeurigheid van het model, 68%, slaat op de fractie voorspellingen die overeenkomen met de meegegeven labels.

2.5.8 Classificatie met iVectors

Een tweede, robuustere, methode om stress en rust van elkaar te onderscheiden, is een slimme combinatie van ongesuperviseerd en gesuperviseerd machinaal leren. Ongesuperviseerd (Eng. unsupervised) leren vereist niet dat labels voor de datapunten meegegeven worden. Vaak gaat het dan om vormen van clusters van gelijkaardige data; een soort classificatie waarbij het model zelf beslist welke klassen er zijn. De iVector-methode [23] gebruikt de resultaten van een dergelijke clustering-operatie om variaties in uitspraak om te zetten in featurevectoren van een vaste lengte (dit zijn de iVectors).



Figuur 2.12: Overzicht van de iVector-architectuur

Features

Ook hier is er uiteraard een nood aan features afgeleid uit de spraakfragmenten. Per 10ms audio wordt een vector van de eerste n MFCC coëfficiënten (zie sectie 2.4) samengesteld. De hoeveelheid features per tijdseenheid n is een optimaliseerbare hyperparameter. Aan deze vector worden vervolgens Shifted Delta Coefficients (SDC) toegevoegd om informatie over veranderingen in de tijd mee te kunnen geven aan het model. Deze vorm van features zorgt voor een grote vrijheid in architectuur en meer mogelijkheden tot slimme optimalisaties dan bij de openSMILE-aanpak.

Preprocessing

De invoer voor het model is dus een matrix, met rijen van features per 10ms. Er zijn verschillende manieren om deze matrix te verwerken om telkens ietwat andere informatie te benadrukken.

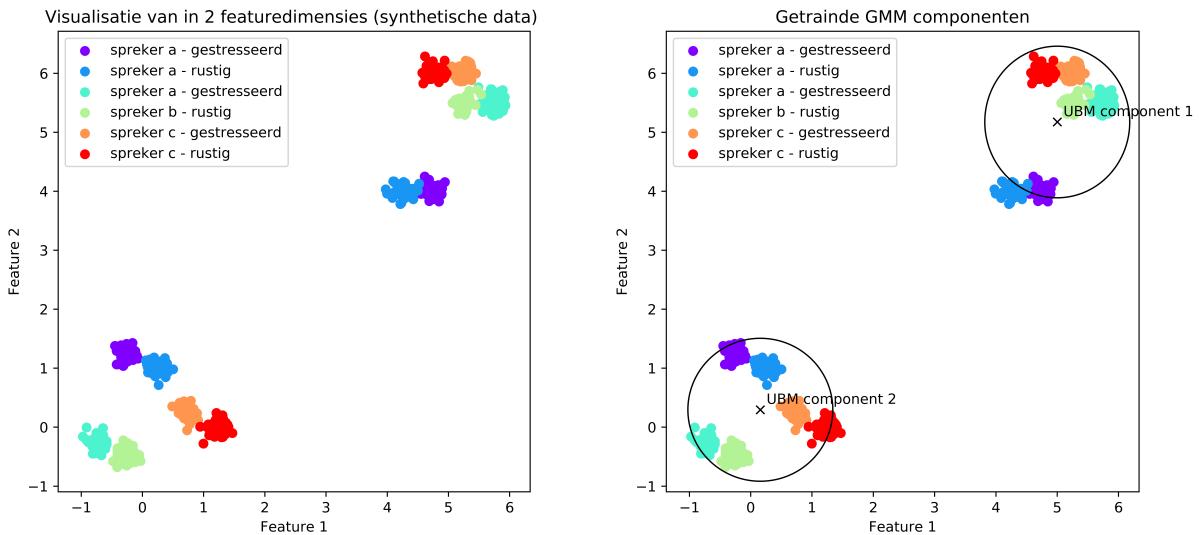
Met Cepstral Mean Subtraction (CMS) worden de rijen verminderd met hun gemiddelde over de tijd. Hierdoor worden spreker-specifieke verschillen in coëfficiënten gedeeltelijk verdoezeld: verschillende sprekers die hetzelfde zeggen zullen een minder verschillende invoermatrix verkrijgen met CMS-normalisatie.

Cepstral Mean and Variance Normalization (CMVN) deelt deze rijen vervolgens nog eens door hun standaardafwijking in de tijd. Op deze manier zullen de individuele coëfficiënten behalve hetzelfde gemiddelde (0), ook dezelfde standaardafwijking (1) hebben. Alle features zullen nu voor het grootste deel in het bereik $[-1, 1]$ vallen. Ook deze normalisatie verborgt spreker-specifieke variaties (gedeeltelijk). Een belangrijk verschil met de normalisatie bij de SVM-methode is dat hier genormaliseerd wordt binnen eenzelfde fragment en niet tussen verschillende datapunten.

Bij de iVector-methode kan er op verschillende momenten gekozen worden voor het al dan niet normaliseren van feature-matrices. Dit is telkens een te optimaliseren hyperparameter. Een eerste plaats waar dit kan is meteen na het extraheren van de matrix, dit heet dan "utterance normalisation".

Universal Background Model (UBM)

De eerste, en ongesuperviseerde, stap in deze methode is het trainen van een Gaussian Mixture Model (GMM) met de MFCC- en deltafeatures geëxtraheerd uit een database met gerelateerde spraak – lees: in dezelfde taal. Zo worden de verschillende fonemen die voorkomen in algemene spraak gegroepeerd. Het aantal groepen, of "mixtures", dat de GMM zal zoeken is een hyperparameter. Voor spraak zal de optimale waarde van deze parameter tussen de 32 en 256 liggen [24].



Figuur 2.13: Het Universal Background Model groept datapunten in componenten. Deze componenten stellen bepaalde klanken voor. Variabiliteit binnen deze klanken wordt gemodelleerd met factoranalyse. [25]

De locaties van deze mixtures in de ruimte opgespannen door de features worden samengevoegd in een vector en opgeslagen als m_{UBM} , de UBM means.

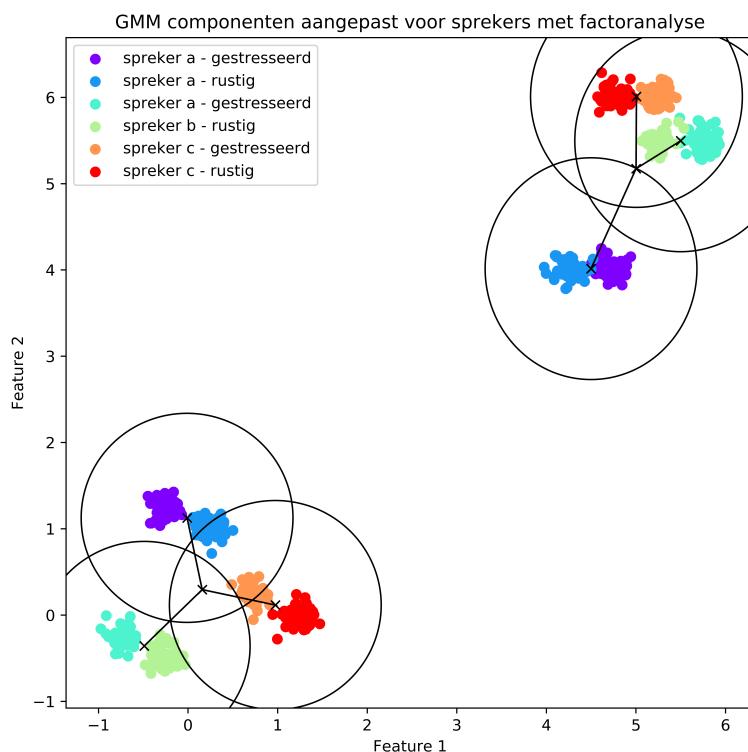
Eigenvoice

Een GMM met dezelfde parameters als het UBM wordt nu voor elke spreker afzonderlijk getraind. Er kan voor gekozen worden om de spraakfragmenten van de sprekers eerst samen te nemen en te normaliseren op één van de eerder besproken manieren. De componentgemiddelden worden op gelijkaardige wijze als het UBM bewaard als m_{spk} .

De verschuivingen van de gemodelleerde fonemen in de feature-ruimte worden nu gemodelleerd als een lineaire combinatie van "Eigenvoices" opgeliist in matrix U . De resulterende factoren y_{spk} zijn de coördinaten van de spreker in de Eigenvoice-ruimte en een volledige beschrijving van de spreker.

$$m_{\text{spk}} = m_{\text{UBM}} + U y_{\text{spk}} \quad (2.5)$$

De Eigenvoice-matrix U wordt ingevuld met behulp van de UBM-data, en daarbij toegevoegd, een identificatie van de spreker per fragment. Dit gebeurt onder andere met behulp van Principal Component Analysis (PCA), dat de richtingen met de grootste variatie in de dataset kan opsporen.



Figuur 2.14: Illustratie van de gevonden verschuivingen y_{spk}

De precieze werking van de algoritmes gebruikt om de Eigenvoice-matrix en sprekerfactoren af te leiden uit de data wordt beschreven in [25], en valt buiten het bestek van dit verslag. De rang r van matrix U , beïnvloedt op hoeveel manieren de uitspraak van verschillende fonemen kunnen veranderen bij verschillende sprekers, en is een instelbare hyperparameter.

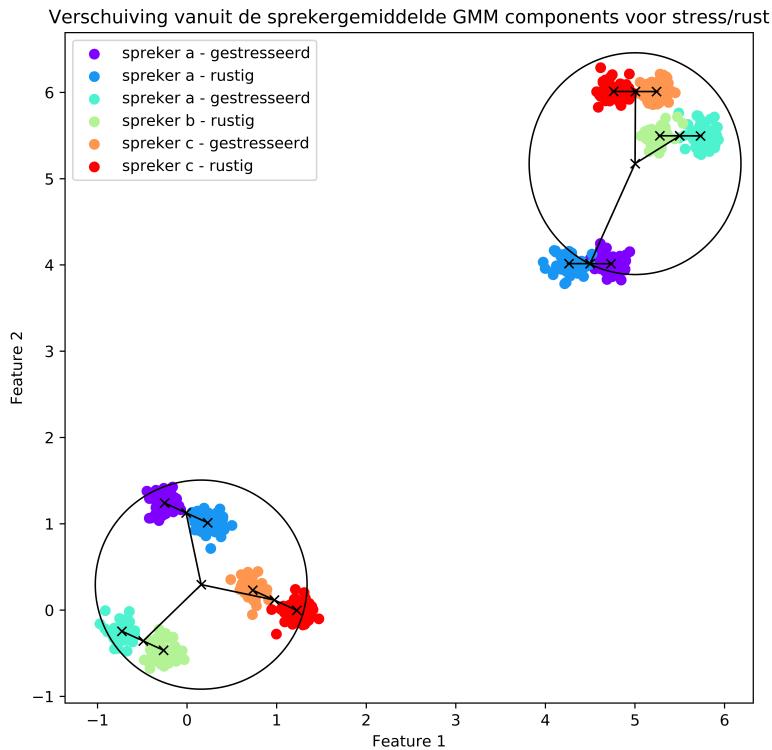
Stressfactoranalyse

Nu variabiliteit in uitspraak vanwege de spreker kan worden gecompenseerd door de sprekerfactoren, kan er - met deze dataset althans - vanuit gegaan worden dat de resterende variaties te maken zullen hebben met stress. Per fragment wordt een GMM getraind. De mixturegemiddelden m_{frag} kunnen als volgt beschreven worden:

$$m_{\text{frag}} = m_{\text{spk}} + V y_{\text{stress}} \quad (2.6)$$

De variabiliteitsmatrix V bevat opnieuw een aantal richtingen waarin de uitspraak kan variëren tussen rustige en gestresseerde fragmenten. V wordt ingevuld door voor elke klasse - in ons geval dus rustig en gestresseerd - vectoren in te vullen in de kolommen van de matrix. Deze

vectoren kunnen gevonden worden door voor de fragmenten uit elke klasse de gemiddelde verschuiving van de componentsgemiddelden m_{frag} tegenover deze waarbij gecompenseerd is voor de spreker m_{spk} te berekenen.



Figuur 2.15: Illustratie van de gevonden verschuivingen y_{stress}

Een bijkomende hyperparameter, de rang binnen de klasse (r_c), laat toe om op meerdere manieren te verschuiven naar de verschillende klassen. Intuïtief betekent dit dat het model toelaat dat mensen verschillend reageren op stress.

Gaussian Backend

De vectoren y_{stress} (iVectors) bevatten nu voor elk fragment de hoeveelheid verschuiving voor de richtingen gedefinieerd in V . Een voorspelling zou gemaakt kunnen worden door het maximum van deze verschuivingen te nemen en na te gaan bij welke klasse deze richting hoort.

Ons model opteert voor een iets geavanceerdere aanpak, zodat bij elke voorspelling ook een probabiliteit voor elke klasse ($p(c|X)$ met c de klasse en X het gegeven fragment) kan gegeven worden. Per klasse wordt een Gaussiaans model getraind op de geëxtraheerde y_{stress} -vectoren van elk fragment uit de trainset, dat de gekozen klasse onderscheid van de overige. Deze Gaussiaanse modellen geven op natuurlijke wijze de eerder beschreven probabiliteit als antwoord op de invoerdata. De iVectors worden dus geëvalueerd op elk model en de finale voorspelling is de klasse waarbij de waarschijnlijkheid het grootst is.

Tuning

De beschreven architectuur werd geïmplementeerd in Python met behulp van een framework dat ons vriendelijk verstrekkt is door ir. Brecht Desplanques. Alle hyperparameters werden getuned in een simpele gridsearch met 4-fold cross-validation.

Hoewel voor de UBM de enige vereiste is dat de spraak waarop deze getraind wordt gelijkaardig is aan deze van de effectieve traindata, en er dus grotere datasets beschikbaar zijn, werd ervoor gekozen om deze alsook en enkel op de verzamelde data te trainen. Een reden hiervoor is dat deze data met eenzelfde microfoon is opgenomen, en het op die manier zeker is dat er geen variatie in de data is afkomstig van het kanaal. Deze variatie kan later nog steeds in het model geïntegreerd worden, zodat er geëxperimenteerd kan worden met betere achtergrond-datasets.

Een belangrijke bemerking bij de cross-validation-strategie is dat per parameterconstellatie het UBM en het Eigenvoice-model telkens eerst op alle data werd getraind en enkel in de volgende stappen rekening werd gehouden met de train/validatie-splits. Dit is te verantwoorden door in te zien dat beide het UBM en het Eigenvoice-model geen gebruik maken van de labels over stress en louter dienen om de kenmerken van de spraak in kaart te brengen. Deze keuze werd gemaakt omdat het evalueren van de grid search anders dramatisch veel langer zou duren.

De cross-validation-curves in Appendix D.2 illustreren de invloed van de verschillende hyperparameters op de prestatie van het model

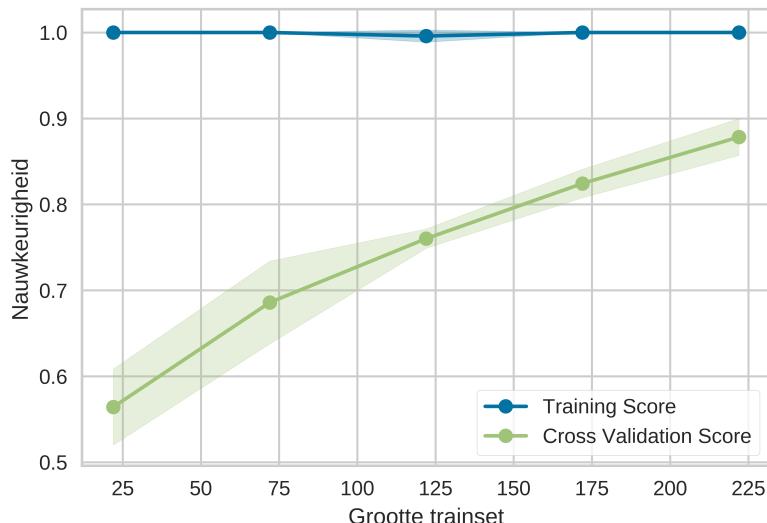
#mixtures	<i>n</i>	fragmentnormalisatie	sprekernormalisatie	r_c	<i>r</i>
128	25	CMVN	geen	3	80

Tabel 2.4: Overzicht van de gevonden optimale hyperparameters

Optimalisatie van deze parameters met de precisie als prestatia maat levert ook hier hetzelfde resultaat.

Evaluatie

Ook de leercurve van de iVector-aanpak bevestigt dat meer data geen overbodige luxe zou zijn. De vlakke trainnauwkeurigheid van bijna 100% maakt duidelijk dat het model meer dan krachtig genoeg is om te verschillen te modelleren, terwijl de nog steeds stijgende validatiescore erop wijst dat meer data waarschijnlijk een hogere score zou opbrengen. De kleine standaardafwijking van nauwkeurigheid tussen validatiesets bij hogere trainsetgroottes is een goede indicator dat er geen sprake is van overfitting.



Figuur 2.16: Leercurve van het geoptimaliseerde iVector-model

Het geoptimaliseerde iVector-model slaagt erin een cross-validation nauwkeurigheid van ongeveer 88% te behalen. Dit is beduidend hoger dan deze van openSMILE. Bovendien zou deze methode minder afhankelijk moeten zijn van de inhoud van het spraakfragment. Het iVector-model generaliseert dus beter naar natuurlijke spraak. Een overzicht van de resultaten is opgenoemd in onderstaande tabel:

sensitiviteit	precisie	nauwkeurigheid
85.81%	89.50%	87.84%

Tabel 2.5: Overzicht van prestatie iVector

De precisie van bijna 90% vertelt dat in 9 van de 10 gevallen, een voorspelling met stress daadwerkelijk overeenkomt met spraak na de MIST. Dit is ruwweg 20% beter dan de situatie bij de openSMILE-aanpak. De sensitiviteit wijst erop dat de iVector-aanpak met een probabiliteit van 86% een fragment ingesproken na een stresserende taak herkent. Ook deze prestatia maat en de nauwkeurigheid zijn 20% beter dan het openSMILE-model.

Een belangrijke bemerking hierbij is dat het trainen en evalueren van het iVector-model vele malen langer duurt dan het eenvoudigere SVM-model. Verdere mogelijke optimalisaties van deze laatste aanpak dienen dus niet zomaar verworpen te worden.

2.5.9 Conclusie

De resultaten van beide uitgeprobeerde modellen wijzen erop dat er wel degelijk detecteerbare verschillen zijn tussen de spraak voor en na een stresserende taak. Of deze informatie uit eerder welke spraak kan gehaald worden, en of de verschillen die de modellen merken volledig aan stress gerelateerd zijn, zijn onderwerpen voor verder onderzoek.

Bovendien is het duidelijk dat beide modellen beter zouden kunnen presteren met meer en beter gelabelde data. Met meer data zouden bovenbieden andere veelbelovende methoden zoals Deep Learning gebruikt kunnen worden.

2.5.10 Bemerkung bij cross-validation

De prestatie van de gebruikte modellen werd zoals eerder vermeld geëvalueerd met behulp van cross-validation. Bij machinaal leren is het aangewezen om naast cross-validation, een deeltje van de dataset apart te houden en te vergelijken met de cross-validation-score om na te gaan of er sprake is van overfitting op de train/validation splits. Wij hebben ervoor gekozen dit niet te doen omdat onze dataset niet heel groot is, en anders mogelijks kostbare data ongebruikt zou blijven. De standaardafwijking op de cross-validation-score is bovenbieden ook een maat voor de besproken overfitting.

2.6 Demo

De demo voor de posterbeurs is eigenlijk een vereenvoudigde versie van de app die gebruikt werd tijdens de experimenten. De demo-app vraagt eerst hetzelfde "Papa en Marloes"-tekstje voor te lezen, leidt de gebruiker vervolgens naar een verkorte versie van de MIST en vraagt opnieuw om hetzelfde tekstje in te spreken. Tenslotte worden deze spraakfragmenten doorgestuurd naar een Flask (Python) server die het op-alle-data-getrainde iVector-model gebruikt om te voorspellen of, en in welke mate, de gebruiker gestresseerd was voor en na de MIST. De server antwoordt met een JSON-array die deze scores bevat en de app geeft deze vervolgens weer op een resultatenscherm.

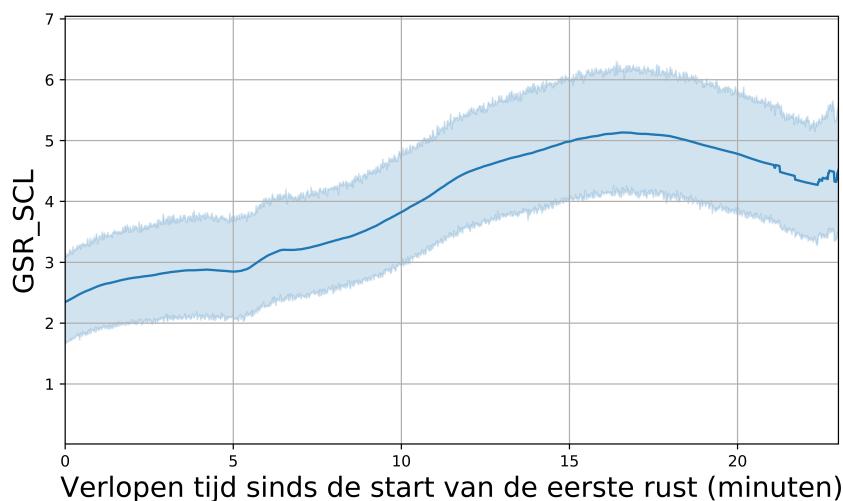
Deze client-server architectuur werd gekozen uit tijdsoverwegingen en om het veranderen van het gebruikte model makkelijker te maken. Bovendien is het niet triviaal deze modellen te porteren en uit te voeren in Java.

3. Resultaten

3.1 Eerste opmerkingen: Bruikbaarheid van de data

Na het verwerken van de data werd gekeken naar de plots en nagedacht over de beste aanpak voor het machinaal leren. Het werd al snel duidelijk dat de EDA data visueel het meest reageert op de stresstest. Hartslagdata daarentegen bevat bij velen erg veel ruis, vermoedelijk door het hergebruiken van de patches. Extra patches waren namelijk te duur en het was pas te laat duidelijk dat er niet genoeg ter beschikking waren. In de literatuur wordt HRV (hartslag variabiliteit) vaak gebruikt als biomarker voor stress [10]. Dit is echter geen eenduidig verband en in onze data was geen correlatie te vinden, vermoedelijk door de relatief korte testperiode.

Een illustratie hiervan voor een specifieke proefpersoon is te zien op Figuur A.2. Uit de ECG data (Figuur A.2a) zou enkel afgeleid kunnen worden dat de proefpersoon helemaal op het einde van het experiment, na de rust, gekalmeerd was. Deze vertelt ons heel weinig over het effect van de stressinductie. In de EDA-grafiek (Figuur A.2b) is voor deze persoon echter een duidelijke verhoging tijdens de stressinductie (rode arcering, de profronde van de MIST eindigt na de verticale stippellijn) merkbaar.



Figuur 3.1: Gemiddelde EDA variatie doorheen het experiment

Figuur 2.1 toonde reeds aan dat de EDA en VAS schalen reageren op het experiment. Figuur 3.1 toont hoe alle geldige EDA data gemiddeld reageert doorheen het experiment met een betrouwbaarheidsinterval. Aangezien niet iedereen het experiment even snel doorloopt zijn geen regio's aangeduid. De rustperiodes duren exact vijf minuten, maar deelnemers kunnen getreuzeld hebben bij het invullen van de randinformatie. Toch is er een duidelijke stijging in EDA na vijf minuten, terwijl het daarvoor iets wat afvlakte. De stressstaak duurt ongeveer zeven minuten en er is een duidelijke stijging met daaropvolgende daling te zien. Dit geeft vertrouwen in het gebruik van EDA om stress te meten. Op het einde van deze figuur zijn er lichte artefacten door de verschillende lengtes van experimenten. Correlaties met verschillende ECG-features waren niet zomaar te vinden.

3.2 Nauwkeurigheid van het model

3.2.1 Labels

Stress definitief bepalen aan de hand van biometrische signalen of zelfrapportering is niet eenvoudig. Het is onmogelijk om een exacte stress-score toe te kennen aan elk spraakfragment, en daar schuilt een grote moeilijkheid van dit project. De modellen zijn getraind met de aannname dat elke deelnemer bij de eerste spraakopname niet gestresseerd is en bij de tweede spraakopname wel, en dit altijd in dezelfde mate. Natuurlijk is dit niet het geval. Niet iedereen begint met dezelfde hoeveelheid stress of reageert op de stressinductie. Deze labelingsstrategie

gaf echter verbazend goede resultaten tijdens de machinaal leren-fase en is zeker het eenvoudigst om mee te werken. Aangezien perfecte labels voor de data niet bestaan, genereren we labels met een heuristiek en labelen we ook manueel. De modellen worden dan gebruikt om zelf stress scores aan de dataset te geven en zo de labels te voorspellen¹. Op die manier kan een indicatie van de nauwkeurigheid bekomen worden. Het resultaat hiervan is te zien in tabel 3.1. In deze tabel staat N voor nauwkeurigheid, P voor precisie en S voor sensitiviteit.

	Gegenereerd						Manueel					
	Voor			Na			Voor			Na		
	N	P	S	N	P	S	N	P	S	N	P	S
iVector	62.9%	31.3%	13.5%	62.9%	81.0%	71.6%	82.1%	15.0%	37.5%	70.7%	92.2%	74.1%
OpenSMILE	61.2%	40.0%	43.2%	63.8%	83.5%	69.5%	60.2%	2.3%	12.5%	63.4%	90.4%	67.0%

Tabel 3.1: Nauwkeurigheid, precisie en sensitiviteit van de voorspelde labels

De heuristiek om labels te genereren bekijkt eerst of de EDA-data geldig is. Zo ja, dan wordt bekeken of er een significante stijging was in EDA tijdens de eerste rust, wat zou kunnen betekenen dat de persoon gestresseerd begint aan de MIST. Als tijdens de MIST, de EDA stijgt én de zelf ingevulde gemoedstoestand daalt, dan wijst dit op een gestresseerde toestand na de MIST. Dit steunt op figuur 2.1. De manuele labeling kijkt ook voornamelijk naar de EDA data maar is natuurlijk subjectiever.

Het is duidelijk dat de precisie en sensitiviteit steeds een pak beter zijn na de MIST dan ervoor. Dit zou een gevolg kunnen zijn van het feit dat de modellen getraind zijn op basis van spraak voor of na de MIST in plaats van met echte stresslabels. De nauwkeurigheid varieert minder sterk voor en na de MIST wat dit verder bevestigt. Het is echter belangrijk om te onthouden dat zowel de heuristisch gegenereerde labels als de manueel gemaakte labels niet noodzakelijk erg accuraat zijn.

De hoogste nauwkeurigheid, precisie en sensitiviteit vindt plaats bij het vergelijken van de iVector voorspelde labels met de manueel bepaalde labels. Er is hier ook een duidelijk verschil tussen de iVector- en OpenSMILE-methode. Aangezien de iVector-techniek reeds betere resultaten leek te leveren geeft dit desondanks de subjectieve aard van de manuele labels toch vertrouwen in het iVector-model en de manuele labeling. Gezien de erg lage precisie voor de MIST doet dit vermoeden dat deze niet accuraat gelabeled is, of dat het model beter is in voor- en na onderscheiden om een verschillende reden dan stress.

Het OpenSMILE-model voorspelt dat 85.2% van de deelnemers niet gestresseerd waren tijdens het feedback gedeelte van de app, het iVector-model voorspelt dat dit 82.4% was. Dit is mogelijk aangezien we merkten dat veel deelnemers kalmeerden door vrij te mogen praten. De nauwkeurigheid van de voorspellingen van beide modellen vergeleken met beide labeling methodes is steeds onder de 35% met een precisie lager dan 25%. Dit kan wijzen op slechte feedbacklabels en een lagere robuustheid van de modellen bij vrije spraakinformatie.

Dit alles toont het belang van betere labeling in de toekomst, zeker omdat de manuele labels meer overeenkomen met de voorspellingen van de modellen. Betere heuristieken of algoritmes moeten ontwikkeld worden om de subjectieve aard van het labelen te verwijderen. Het is ook niet zeker of iedereen reageert op een detecteerbare manier.

Indien er voor elk spraakfragment een accurate stress-score zou bestaan, zou dit in plaats van de harde grens tussen helemaal niet en volledig gestresseerd aan het model meegegeven kunnen worden. Op die manier zou het model ook kunnen leren de mate waarin iemand gestresseerd is beter te voorspellen.

3.2.2 Correlaties

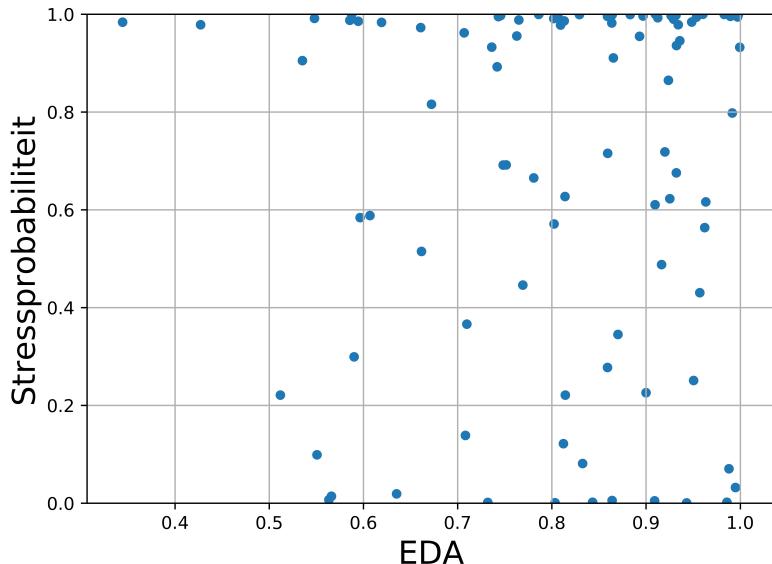
Verder is het ook erg interessant om te kijken of de 'stress-scores' afkomstig van de modellen een correlatie vertonen met een biometrische maatstaf of de zelf gerapporteerde gemoedstoestand. Dit bleek niet zo vanzelfsprekend te zijn. Er is grondig gezocht naar correlaties tussen verschillende data en de probabiliteiten geleverd door de modellen. Vooral de EDA- en VAS-scores leken veelbelovend. Om een verband tussen de geleidbaarheid van de huid en de probabiliteiten te zoeken is er gekeken naar de variatie van het signaal tijdens de spraakopnames en tijdens de rust, genormaliseerd en niet genormaliseerd, maar dit leverde niet veel op. Ook het gemiddelde EDA-niveau vergeleken met de piek bracht geen verbetering, en een combinatie van dit genormaliseerd gemiddelde met de variatie hielp ook niet.

De VAS-scores stootten op gelijkaardige problemen. Na geen correlatie waar te nemen met de gemoedstoestand en de 'stress-score' werd gekeken naar de variatie van de VAS-scores en de variatie van de stress-scores. Ook dit bracht niet veel aan het licht. Het is geen fantastisch nieuws dat correlaties lastig op te sporen waren, maar dat betekent niet dat er geen bestaan. De modellen zijn nog niet perfect, net zoals de zelfrapportering van de stress. Wel verwachten we meer correlatie met de geleidbaarheid van de huid, het ontbreken hiervan kan wijzen op een slechte verwerking van de EDA of onnauwkeurigheid van de modellen.

Er zijn echter veel factoren die deze correlatie niet eenduidig maken.

Ter illustratie van de moeilijkheden toont figuur 3.2 de gemiddelde geleidbaarheid van de huid tijdens de tweede spraakopname, gedeeld door de piekwaarde, op de horizontale as en de stressprobabiliteit na de MIST van het iVector model op de verticale as. Deze figuur toont wel dat de stressprobabiliteit hoog is na de MIST, en dat de EDA dicht bij de piek ligt, maar er is geen duidelijk verband.

¹dit gebeurt met KFold cross-validate; de labels van elke testset worden voorspeld met een model getraind op de overige data, totdat elk fragment een label heeft



Figuur 3.2: Gemiddelde EDA t.o.v. de piek tegen stressprobabiliteit, na de MIST

Let wel dat de probabiliteit die de modellen geven niet noodzakelijk overeenkomt met de hoeveelheid stress, en dat het gebrek aan correlatie dus niet geheel onverwacht is. Om een betekenis als deze af te dwingen, moeten er zoals eerder vermeld nauwkeurige stress-scores per fragment beschikbaar zijn, en moeten de modellen getraind worden voor regressie in plaats van classificatie. Dit laatste zou in verder onderzoek al uitgetest kunnen worden op genormaliseerde verschillen in EDA.

3.3 Openstaande problemen en tekorten

Een opmerking die we kunnen maken bij de volgorde van het experiment, en het opnemen van (niet-)gestresseerde spraak is dat, omdat proefpersonen zich begeven in een voor hen onbekende omgeving, voor sommige proefpersonen bij het binnengaan (en soms na de eerste rust nog) al wat stress geïnduceerd wordt. Voor toekomstige experimenten kan het daarom beter zijn om eerst, na de MIST, gestresseerde data op te nemen, en dan pas, na een rustfase, de niet-gestresseerde spraak. Ook lijkt het een goed idee om de proefpersonen de tekst voluit te laten voorlezen nog voor ze beginnen aan de eerste opname. Nu is er een risico dat de proefpersonen na de stresstest de tekst anders voorlezen omdat ze er beter vertrouwd mee zijn.

Zoals reeds meermaals aangehaald is betere labeling ook iets waar verder onderzoek naar kan kijken. Het kwantificeren van stress is niet simpel en elke verbetering hierin zal de toepasbaarheid van de modellen op vrije spraak sterk verbeteren.

De ECG patches zijn duur. Deze zijn dan ook gemaakt voor langdurig gebruik door één persoon en niet voor een kortstondig, stationair experiment met vele deelnemers. Voor de SWEET study [4] was dit een ideale keuze, maar voor dit onderzoek was een stationaire ECG oplossing een passendere keuze geweest. Dan hoeven slechts de pleisters van de aparte elektrodes vervangen te worden in plaats van een unieke patch, en zijn er geen zorgen over batterijen of het hergebruiken van patches.

Ook voor de geleidbaarheid van de huid waren er mogelijks betere keuzes. De bands van Imec zijn handig maar nog een prototype. Ze waren duidelijk niet ontworpen om herhaaldelijk aan en uit te doen, wat ze beschadigde. Aangezien voornamelijk de geleidbaarheid nuttig leek voor ons, zouden twee simpele elektrodes waarschijnlijk volstaan, en is geen wearable nodig.

3.4 Besluit

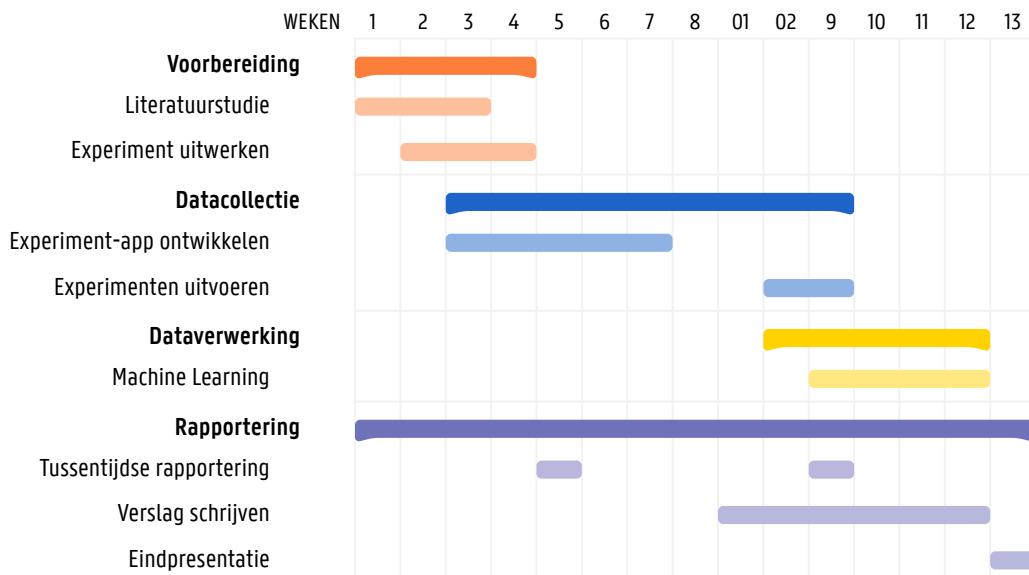
Stressdetectie uit spraak lijkt zeker mogelijk. De getrainde modellen slagen er effectief in verschillen te vinden, en met de iVector-techniek worden stress-scores bekomen die indicatief lijken bij zelf-ingesproken tekst. Er is echter zeker nog werk te doen: betere labels zijn belangrijk zodat de toepasbaarheid van de modellen beter getest kan worden en zodat ze op meer relevante data getraind kunnen worden.

4. Praktisch

4.1 Planning en uitvoering

Het doel van het project is sinds het indienen van het projectvoorstel in januari (zie Appendix G) behoorlijk gewijzigd. Oorspronkelijk zou er gebruik gemaakt worden van NLP¹ voor het verwerken van de spraakdata en zouden we zelf wearables ontwikkelen. Na een aantal vergader-sessies met onze promotoren in de eerste weken van het semester, nam het project zijn huidige vorm aan.

De planning die voor de eerste presentaties werd opgesteld wordt weergegeven in onderstaande Gantt-grafiek. De weken zijn genummerd zoals in de academische kalender van de UGent, waarbij de weken met labels '01' en '02' respectievelijk de eerste en tweede week van het paasreces voorstellen.



De experimenten zijn van start kunnen gaan van zodra we al het nodige materiaal (ChillBands, ECG-patches, tablets en koptelefoons en uiteraard de experiment-app) ter beschikking hadden. De eigenlijke machine learning is ongeveer een week later gestart dan oorspronkelijk het plan was, omdat we nog niet voldoende verwerkte data ter beschikking hadden in week 9, wanneer we nog volop experimenten aan het afnemen waren. De andere onderdelen van het project verliepen ongeveer zoals gepland.

4.2 Communicatie, samenwerking en taakverdeling

De promotoren en begeleiders werden via mail gecontacteerd. Bijna elke week, behalve in de weken waarin de experimenten werden uitgevoerd, vond een meeting met onze promotoren plaats. De communicatie binnen de groep verliep voornamelijk via Slack. De samenwerking verliep vlot en er waren geen onoverkomelijke discussiepunten tussen groepsleden.

De taakverdeling binnen de groep is te zien op tabel 4.1. De ontwikkeling van de experiment-app werd door Vic en Sean gedaan. Het machinaal leren werd door Vic gedaan, aangezien hij daar de meeste ervaring en achtergrondkennis over heeft en er een sterke tijdsdruk was na de experimenten. Dankzij de ervaring met (en voorbeeldcode voor) de iVector-methode die de promotoren met ons deelden, is dit toch op tijd gelukt.

Ondertussen focuste de rest zich op het verslag en schreef Ernest verschillende heuristieken en plotfuncties om data te onderzoeken en vergelijken met de output van de modellen. Sean en Marie bekeken ook veel data en schreven een aantal functies om verbanden te zoeken. Het

¹Natural Language Processing

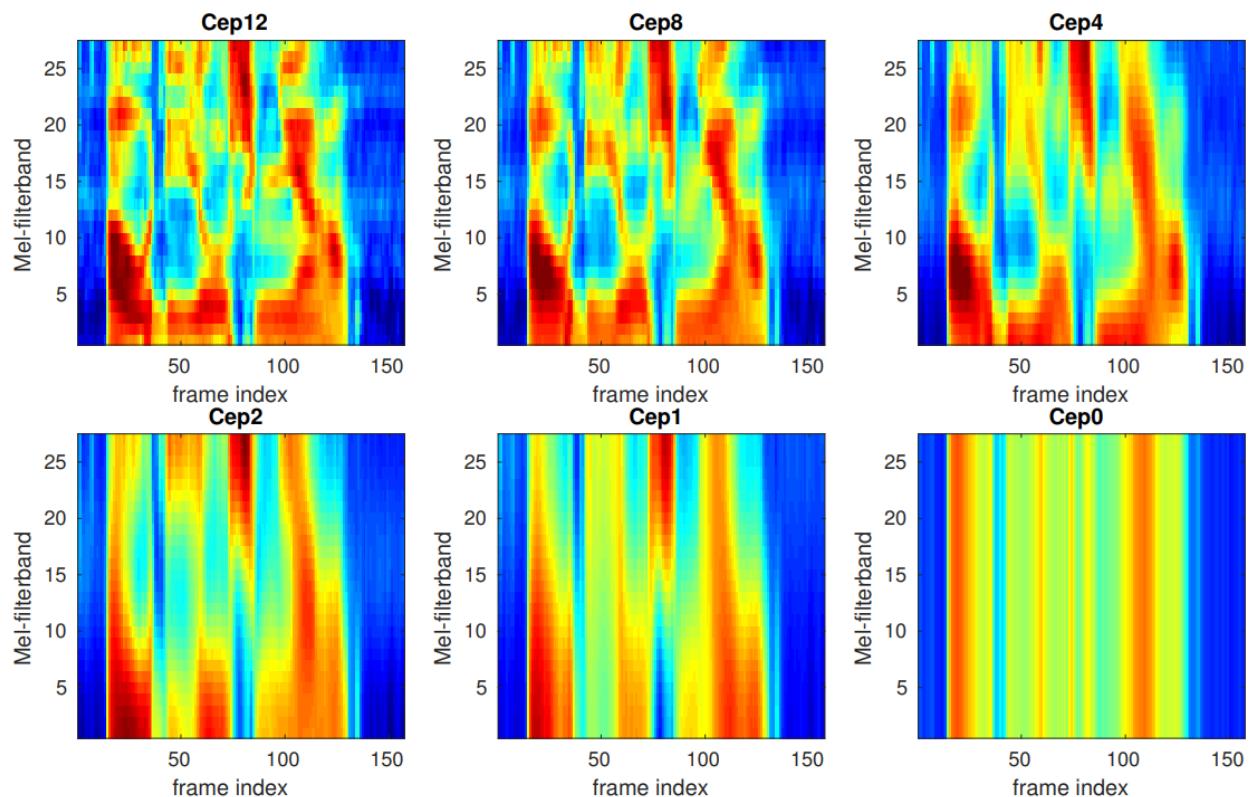
oorspronkelijk plan was om zelf wearables te gaan ontwikkelen, wat Ernest (de elektrotechnieker van de groep) voor zijn rekening ging nemen. Toen dit deel van het project wegviel, zette hij zich extra in voor praktische zaken.

Het afnemen van de experimenten tijdens de twee weken die daarvoor voorzien waren, is iets waar alle groepsleden zich voor ingezet hebben, evenals algemene zaken zoals de literatuurstudie en de rapportering (verslag en presentaties). De experimenten afnemen en proefpersonen zoeken was meer dan twee weken voltijds werk waar iedereen even veel aan heeft meegewerkt.

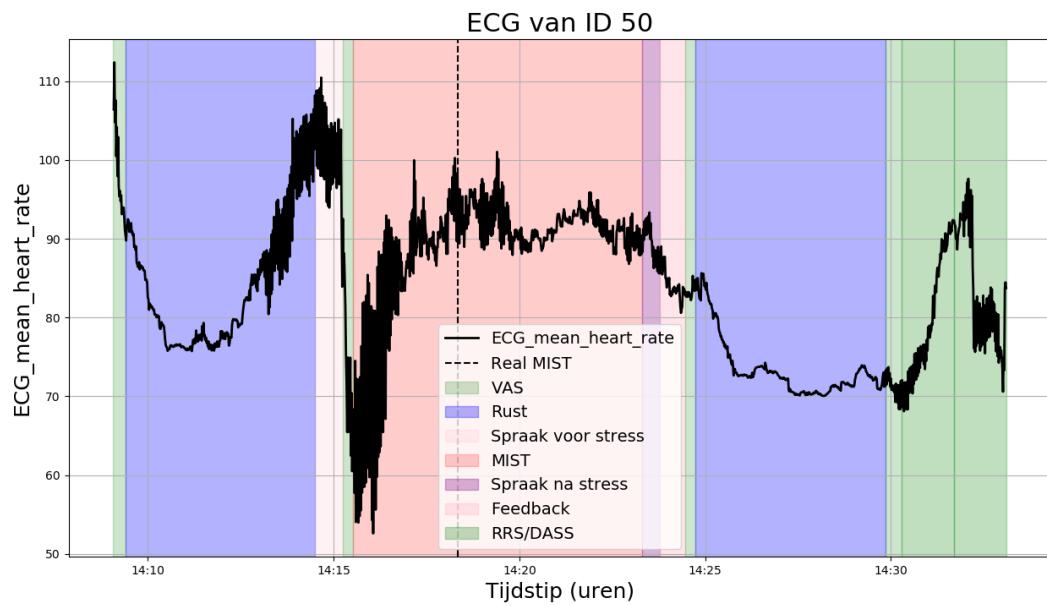
Experiment afname	Iedereen
Literatuurstudie	Iedereen
Proefpersonen ronselen	Iedereen
Machine learning	Vic
App: MIST, spraakopname	Vic
App: schuifregelaars, vragenlijsten	Sean
Experiment protocol	Iedereen, Ernest
Inrichting en citaties verslag	Marie
Verslaggeving	Iedereen
Dataverwerking	Ernest
Data labeling	Sean
Invloed RRS en DASS	Marie, Sean

Tabel 4.1: Tabel van de uiteindelijke taakverdeling

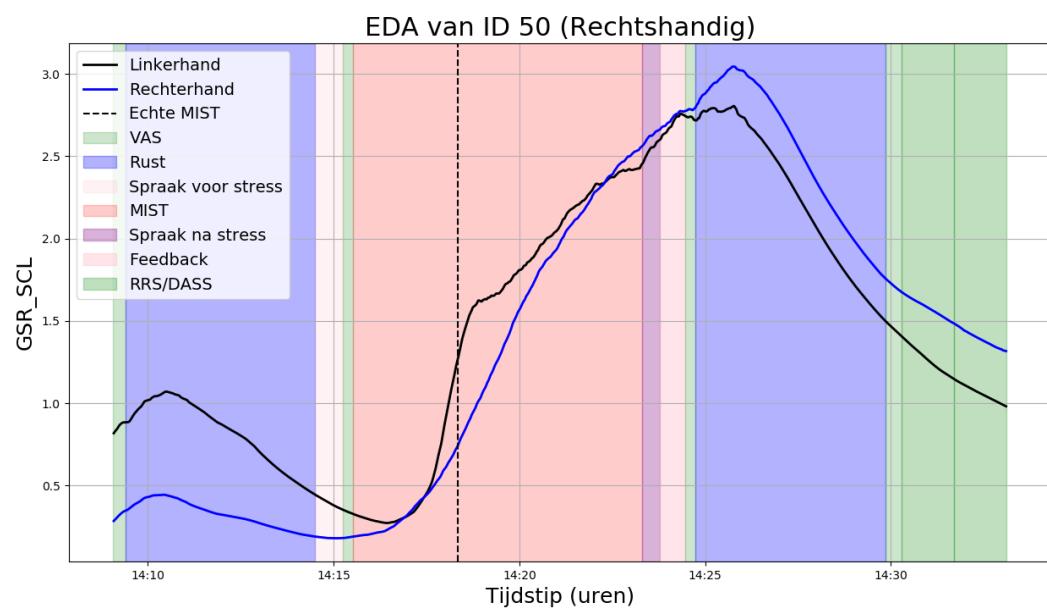
A. Figuren



Figuur A.1: Mel-cepstra



(a) Ruizige ECG van ID 50

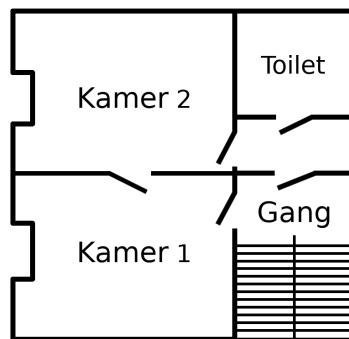


(b) EDA van ID 50, plot van Skin Conductance Level (SCL)

Figuur A.2: Biometrische signalen van proefpersoon met ID 50.

B. Protocol experiment

Om ervoor te zorgen dat we aan al onze deelnemers dezelfde valse info gaven, zodat zo weinig mogelijk vooringenomenheid geïntroduceerd werd, en ook zodat het experiment elke keer op dezelfde manier verliep, stelden we een protocol op dat vooraf en tijdens het afnemen van de experimenten door elk van de begeleiders gevuld diende te worden. Het verloop van het experiment, zoals in Tabel 2.1 wordt weergegeven, wordt in deze appendix in detail uiteengezet. Appendix C.2 geeft een weergave van dit verloop, zoals deelnemers het op de tablet te zien kregen.



Figuur B.1: Indeling van de experimentruimte.

De experimenten werden uitgevoerd in twee relatief uniforme ruimtes, die zich bevinden op de tweede verdieping van een rijhuis dichtbij het station van Gent Sint-Pieters. Figuur B.1 geeft de indeling van deze verdieping weer. We konden in beide kamers tegelijkertijd een deelnemer de test laten ondergaan.

B.1 Cover story

Tijdens het rekruteren van deelnemers werd een leugen verteld, omdat het belangrijk was dat ze niet op voorhand wisten dat we gebruik maakten van stressinductie. Deze luidt als volgt:

"We willen de snelle rekenvaardigheden van studenten en afgestudeerden testen. Op basis hiervan willen we bestuderen wat de invloed van een heleboel biometrische signalen is op de behaalde prestaties, maar ook hoe verschillende richtingen presteren en hoe we scoren in vergelijking met andere universiteiten en populaties."

Ook mogen de deelnemers op voorhand niet weten waar de opgenomen spraakfragmenten precies voor gebruikt zullen worden. Het opnemen van de kalibratietekst verklaren we weg als volgt:

"Onze supervisoren willen dat deze experimenten een transcript hebben, wat toelaat te controleren of we de verschillende procedures getrouw doorliepen en jullie vragen uniform beantwoordden. Dit is belangrijk aangezien de experiment-afnames geleid worden door verschillende studenten. Hiervoor gebruiken we ASR (Automatic Speech Recognition), wat zeer goed is in het automatisch neerschrijven van al wat gezegd wordt. Om hiervan gebruik te kunnen maken, dienen we echter alle participerende stemmen te kalibreren. Mijn stem werd reeds gekalibreerd in de app bij het opzetten van dit experiment. Jouw stem dienen we zo meteen te kalibreren. Hiertoe zullen we je een korte tekst laten voorlezen als kalibratie."

Bij de tweede 'kalibratie' zeggen we simpelweg dat we het systeem willen controleren en wordt geen verdere info gegeven. Als deelnemers verdere vragen stellen over deze coverstory, krijgen ze te horen dat ze geen extra informatie krijgen, maar dat hen achteraf, na het experiment, alle info gegeven zal worden.

B.2 Briefing

Waarschijnlijk heeft de deelnemer op dit punt de cover story al gehoord. Zo niet, wordt die nu gegeven. Daarnaast wordt tijdens de briefing uitgelegd hoe de MIST zal verlopen, dat men vooraf en achteraf een aantal keer schuifregelaars moet gebruiken om de gemoedstoestand aan te duiden en dat men dan twee vragenlijsten moet invullen. Na deze uitleg wordt gevraagd om de informed consent (zie Appendix E) te ondertekenen. De gegeven uitleg verloopt ongeveer als volgt:

"Bedankt voor jouw interesse in ons onderzoek. Ik ga je een korte uitleg geven over wat je te wachten staat. Voor we van start gaan: als je naar de wc moet gaan, is nu een goed moment. Het is belangrijk dat je rustig kan beginnen aan de test, na mijn uitleg krijg je nog een paar minuten rust."

"Doorheen dit experiment verzamelen we biometrische signalen. Dit doen we om inzicht te verwerven in hoe men omgaat met de rekentaak. Daarvoor moeten deze wearables strak rond elke pols, en moeten we de ECG elektrodes bevestigen. Alle data wordt gekoppeld aan een ID nummer, jouw naam wordt enkel bijgehouden tot het einde van het experiment waarna hij onmogelijk te koppelen is met jouw ID nummer."

"Tijdens het experiment maken we gebruik van automatic speech recognition, hiervoor ga je gevraagd worden een tekstje voor te lezen ter kalibratie. Daarna wordt gepeild naar hoe je je voelt via een aantal schuifregelaars. Dan begint de daadwerkelijke test. Hier wordt gepeild naar je rekenvaardigheid. Eerst krijg je een paar rekenvraagjes om je vertrouwd te maken met het systeem. Doe deze zo snel en accuraat mogelijk zodat je goed geoefend bent. Na de volledige test zal je opnieuw dezelfde kalibratie en schuifregelaars tegenkomen, waarna je weer tot rust kan komen en na een korte vragenlijst verlost wordt van de apparatuur."

"Veel succes en bedankt voor je deelname."

Na de briefing neemt de deelnemer plaats aan de tablet. Vervolgens geeft de begeleider een uniek id-nummer voor de deelnemer in, zodat anonimiteit gegarandeerd kan worden.

B.3 Randinformatie

Voor de test begint wordt wat randinformatie gevraagd. Meer specifiek wordt er gevraagd naar leeftijd, geslacht, voorkeurshand en het al dan niet ingenomen hebben van stimulerende middelen (nicotine/cafeïne) in de afgelopen twee uur. Deze info zal gebruikt worden om steekproefstatistieken op te stellen.

We zijn specifiek geïnteresseerd in de voorkeurshand (i.e. het linkshandig of rechtshandig zijn) van de deelnemers, omdat dit invloed kan hebben op de accelerometerdata van de twee Chill Bands die ze gedurende het experiment dragen. Het ingenomen hebben van stimulerende middelen kan op zijn beurt invloed hebben op het stressniveau van de deelnemers.

B.4 Schuifregelaars – Visueel Analoge Schaal

De schuifregelaars – Visueel analoge schaal (VAS) – peilen op vier momenten tijdens het experiment naar de gemoedstoestand van de deelnemer. Er wordt gevraagd om op een schaal van 0 ('Helemaal niet') tot 100 ('Heel erg') aan te geven hoe 'moe', 'krachtig', 'boos', 'tevreden', 'gespannen', 'neerslachtig' en 'prettig' ze zich voelen. De antwoorden op deze VAS worden gebruikt als manipulatiecontrole voor de stressinductie, gebruik makend van positief ('krachtig', 'tevreden', 'prettig') en negatief affect ('boos', 'gespannen', 'neerslachtig') en arousal ('moe'). Als de deelnemer lagere scores aanduidt op de positieve emoties en hoger op de negatieve na de stressinductie, dan is dit een goede indicatie dat de stressinductie gewerkt heeft.

B.5 Rust

De gecontroleerde rustfase bestaat uit vijf minuten rustig zitten. We vragen de deelnemers niet te praten. Op het scherm van de tablet is niets te zien. De tablet signaleert wanneer de tijd om is.

B.6 Kalibratie – Fixed-form spraakdata

Zoals hierboven vermeld, wordt aan de deelnemers verteld dat ze een kalibratietekst dienen voor te lezen om het ASR (Automatic Speech Recognition) systeem te kalibreren. We vermelden dat we dit gebruiken om transcripties makkelijker te maken. In werkelijkheid gebruiken we deze 'kalibratietekst' om fixed-form spraakdata te verzamelen voor en na de stressinductie.

Als tekst werd 'Papa en Marloes' [26] gekozen. Dit is een tekst die gebruikt wordt om de stemkwaliteit te testen omdat deze fonetisch gebalanceerd is – de fonemen die voorkomen zijn in verhouding met hoe ze in spontane spraak voorkomen – en omdat de tekst relatief eenvoudig is om voor te lezen.

B.7 Montreal Imaging Stress Task (MIST) – Stressinductie

Voor de stressinductie werd gebruik gemaakt van de Montreal Imaging Stress Task (MIST) [17]. Dit is een reeks rekenvragen die ontworpen is als stressinductietak. Na de oefenronde van de MIST moet benadrukt worden dat de deelnemer goed moeten scoren.

"Nu begint de daadwerkelijke test, houd er rekening mee dat je goed genoeg moet scoren opdat onze resultaten bruikbaar zijn. Het gemiddelde van de andere deelnemers ligt tussen 70 en 85 procent. De data mag niet teveel afwijken voor ons onderzoek."

Na de MIST vindt een 'herkalibratie plaats, waarbij nog eens dezelfde tekst dient voorgelezen te worden als in §B.6.

B.8 Feedback

Na de tweede 'kalibratie' wordt naar de mening van de deelnemer gevraagd, zodat men het niet vreemd vindt deze luidop te delen. Doe dit voordat men de schuifregelaars invult, een scherm op de app zal hier ook aan herinneren.

"Het is de bedoeling dit onderzoek later op veel grotere schaal te voeren. Om het experiment zo aangenaam mogelijk te maken voor een grote groep testpersonen, willen we verzekeren dat onze onderzoeksapp de best mogelijke gebruikservaring biedt. Hiertoe vragen we jouw mening over onder andere:

- De algemene look-and-feel van de app
- Jouw interactie met de app
- De werking van het wiskunde-experiment

Alle feedback is welkom."

B.9 Vragenlijsten

Op het einde van het experiment wordt de proefpersonen gevraagd om twee vragenlijsten in te vullen. Dit zijn de Ruminative Response Scale (RRS) en de Depressie Angst Stress Schaal (DASS). Voor ons experiment worden deze schalen gebruikt als controle voor individuele verschillen tussen testpersonen wat betreft de mate van repetitief negatief denken en depressie-, angst-, en stressklachten, aangezien deze zaken invloed kunnen hebben op hoe de participanten reageren op de stressinductie. Ook is het goed hier zicht op te hebben, omdat voor de langetermijndoelen van het project interessant kan zijn na te gaan in welke mate spraakdata en biometrische parameters indicatief zijn voor hoge stressniveaus in de context van depressie. In Appendix F.2 worden deze schalen en de invloed van de antwoorden erop op ons experiment verder toegelicht.

B.10 Debriefing

"Je beëindigde zojuist ons volledige experiment. Hierin zat een taak waarop je te zien kreeg dat je ondergemiddeld presteerde. Dit was echter valse feedback, die louter diende om stress te induceren. De taak is zo ontworpen dat je slecht scoort, je prestatie werd gemeten tijdens de oefenronde en maakte zichzelf moeilijker wanneer nodig. De gemiddelde score van de andere deelnemers was een leugen. We zijn namelijk geïnteresseerd in stress en de invloed ervan op de data die we verzamelden, voornamelijk de spraakfragmenten die je moet inspreken. Het spreekt voor zich dat we niet willen dat andere deelnemers dit te weten komen, dus gelieve niets uit deze debriefing door te vertellen."

"Op deze manier verzamelen we data, gekoppeld aan stresserende momenten, die we zullen gebruiken in machine learning modellen om te onderzoeken hoe stress gemeten en eventueel voorspeld kan worden. Bedankt voor de deelname, en stel gerust eventuele vragen."

C. Android App

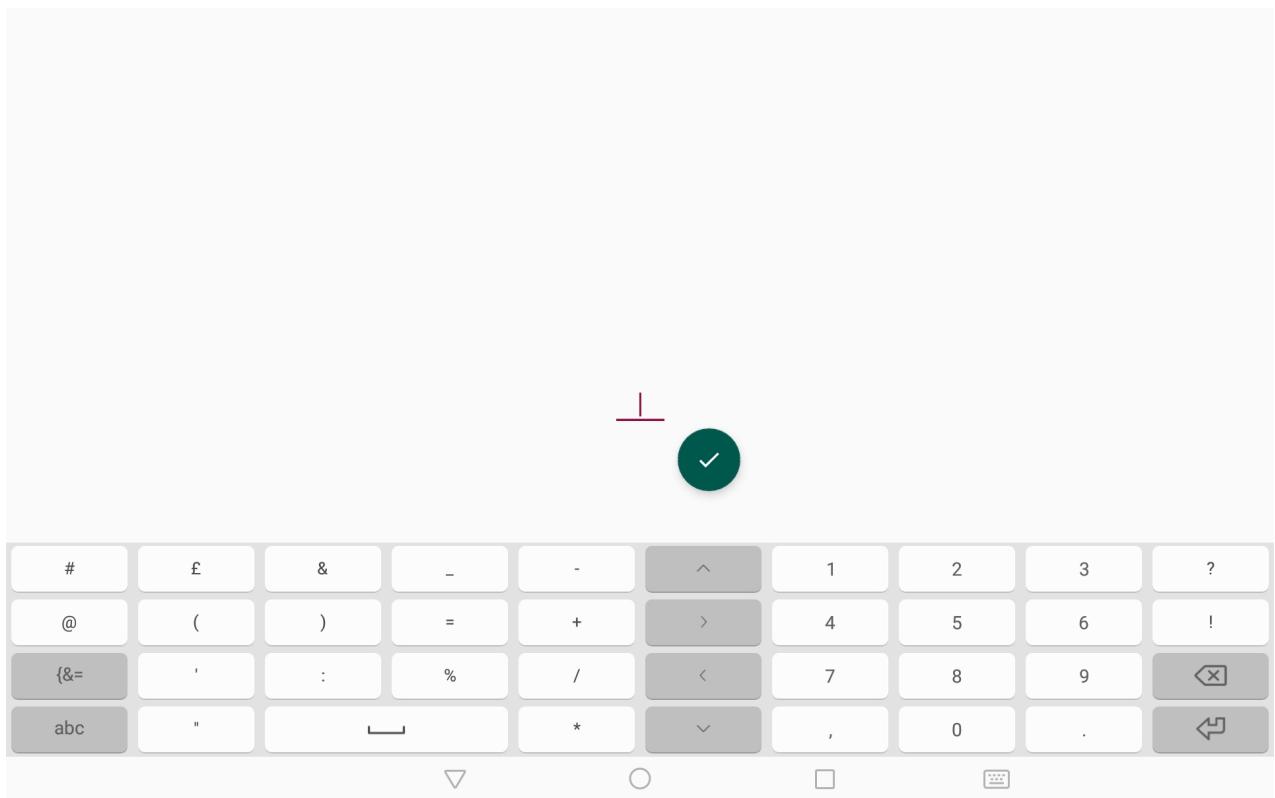
C.1 Opslag

```
1  {
2      "activityTimestamps": [
3          {
4              "enter": 1555492747415, "leave": 1555493750420, "name": "ID"
5          }, ...
6      ],
7      "actualQuestions": [
8          {
9              "answer": 7, "correct": true,
10             "difficulty": "EASY", "equation": "-22 + 9 + 20",
11             "finish": 1555494380503, "start": 1555494374138,
12             "submitted": 7, "timeout": false
13         }, ...
14     ],
15     "age": 18,
16     "controlQuestions": [
17         {
18             ...
19         }, ...
20     ],
21     "dass": [
22         0, ...
23     ],
24     "gender": "Man",
25     "id": 43,
26     "knows": true,
27     "panas": [
28         0, ...
29     ],
30     "rightHanded": false,
31     "sliders": [
32         [
33             16, ...
34         ], ...
35     ],
36     "stimulants": true
37 }
```

Listing 1: Voorbeeld van (verkorte en bewerkte) JSON-file voor een proefpersoon

C.2 Interface

Onderstaande screenshots illustreren de opeenvolging van schermen en functionaliteiten die de app bood tijdens de experimenten.

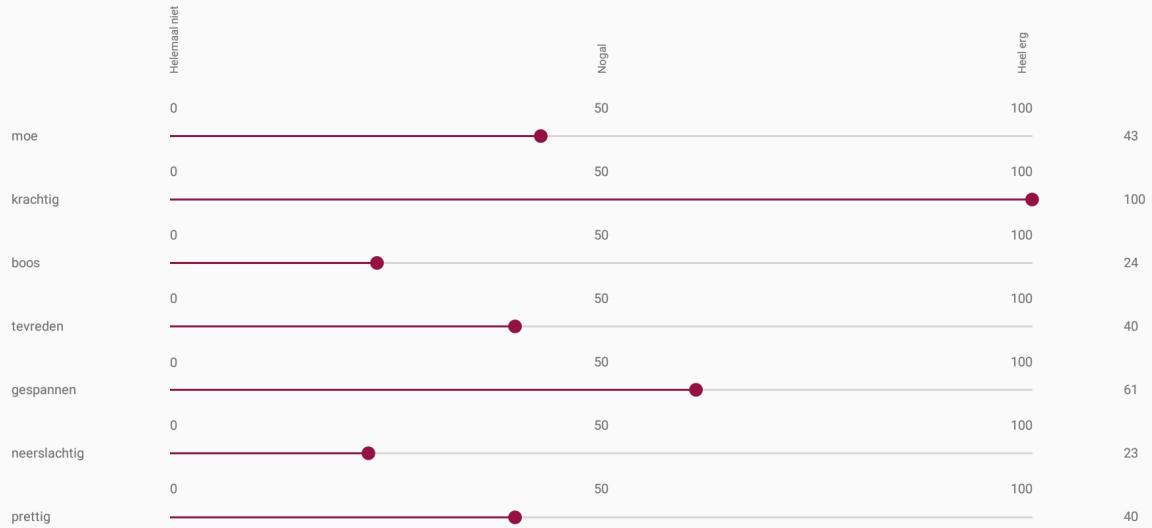


Voor we beginnen

Leeftijd	16	▼
Geslacht	Man	▼
Voorkeurshand	Links	▼
Ik heb binnen de twee uur voorafgaand aan deze test stimulerende middelen zoals caffeine of nicotine gebruikt	Ja	▼



Ik voel me



Klaar

Tik om verder te gaan

Kalibratie

De universiteit wil dat deze experimenten een transcript hebben. Hiervoor gebruiken we ASR (Automatic Speech Recognition), wat zeer goed is in het automatisch neerschrijven van al dat gezegd wordt. We vragen enkel om onderstaande tekst voor te lezen als kalibratie

Papa en Marloes staan op het station.
Ze wachten op de trein.
Eerst hebben ze een kaartje gekocht.
Er stond een hele lange rij, dus dat duurde wel even.
Nu wachten ze tot de trein eraan komt.
Het is al vijf over drie, dus het duurt nog vier minuten.
Er staan nog veel meer mensen te wachten.
Marloes kijkt naar links, in de verte ziet ze de trein al aankomen.



Oefenronde

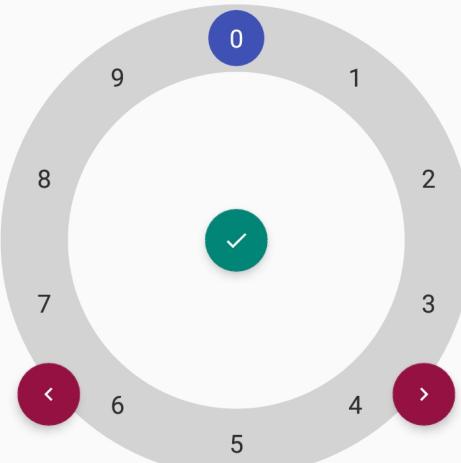
Bestuur het wiel met de pijltjes om de rekenvragen te beantwoorden

Bevestig met de groene knop in het midden

Tik om te beginnen



$4 + 2 + 1 = ?$



Nu voor echt

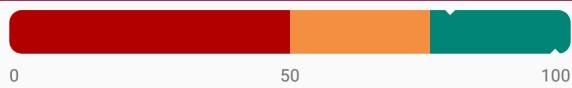
Vanaf nu zal er een tijdslimiet op de vragen staan.

Ook zal je nu zien hoe je scoort tegenover het gemiddelde.

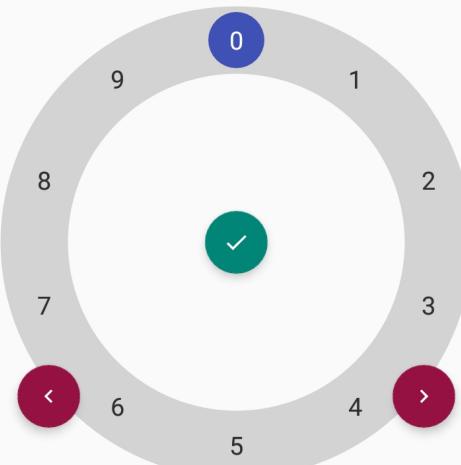
Bovenste pijltje is de score van de gemiddelde deelnemer, onderste die van jou.

Tik om verder te gaan



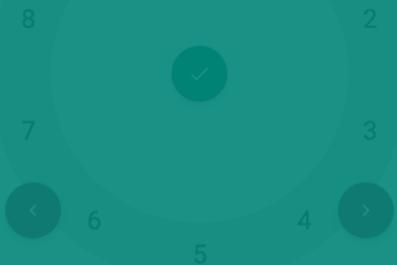


$$1 + 12 - 11 = ?$$



$$-12 + 4 + 8 = ?$$

CORRECT



$$-9 + 3 + 9 = ?$$

INCORRECT



$$-22 + 10 + 17 = ?$$

TIMEOUT



Herkalibratie

De universiteit wil dat deze experimenten een transcript hebben. Hiervoor gebruiken we ASR (Automatic Speech Recognition), wat zeer goed is in het automatisch neerschrijven van al dat gezegd wordt. We vragen enkel om onderstaande tekst voor te lezen als kalibratie

Papa en Marloes staan op het station.
Ze wachten op de trein.
Eerst hebben ze een kaartje gekocht.
Er stond een hele lange rij, dus dat duurde wel even.
Nu wachten ze tot de trein eraan komt.
Het is al vijf over drie, dus het duurt nog vier minuten.
Er staan nog veel meer mensen te wachten.
Marloes kijkt naar links, in de verte ziet ze de trein al aankomen.



Feedback

Het is de bedoeling dit onderzoek later op veel grotere schaal te voeren. Om het experiment zo aangenaam mogelijk te maken voor een grote groep testpersonen, willen we verzekeren dat onze onderzoeksapp de best mogelijke gebruikservaring biedt. Hiertoe vragen we jouw mening over onder andere:

- De algemene look-and-feel van de app
- Jouw interactie met de app
- De werking van het wiskunde-experiment

Alle feedback is welkom.



Vragenlijst 1/2

Gelieve elk van de onderstaande uitspraken te lezen en aan te geven of je bijna nooit, soms, vaak, of bijna altijd datgene denkt of doet wat in elke uitspraak staat beschreven, wanneer je droevig bent, je neerslachtig of depressief voelt. Gelieve aan te geven wat je dan doorgaans doet, niet wat je denkt dat je zou moeten doen.

- 0 = bijna nooit
- 1 = soms
- 2 = vaak
- 3 = bijna altijd

Ik denk na over hoe alleen ik me voel.	<input type="radio"/> 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3
Ik denk na over de vermoeidheid en de pijn die ik voel.	<input type="radio"/> 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
Ik denk na over hoe moeilijk het is me te concentreren.	<input type="radio"/> 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3
Ik denk na over hoe passief en ongemotiveerd ik me voel.	<input type="radio"/> 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3
Ik analyseer recente gebeurtenissen om te proberen te begrijpen waarom ik neerslachtig/depressief ben.	<input type="radio"/> 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3
Ik denk na over hoe ik niets meer lijkt te voelen.	<input type="radio"/> 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3
Ik denk "Waarom kom ik maar niet op gang?"	<input type="radio"/> 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3
Ik denk "Waarom reageer ik altijd op deze manier?"	<input type="radio"/> 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3



Vragenlijst 2/2

Geeft voor ieder van de onderstaande uitspraken aan in hoeverre de uitspraak de afgelopen week voor u van toepassing was door een nummer aan te duiden. Er zijn geen goede of foute antwoorden. Besteed niet te veel tijd aan iedere uitspraak, het gaat om uw eerste indruk.

- 0 = Helemaal niet of nooit van toepassing
- 1 = Een beetje of soms van toepassing
- 2 = Behoorlijk of vaak van toepassing
- 3 = Zeer zeker of meestal van toepassing

Ik vond het moeilijk mezelf te kalmeren	<input type="radio"/> 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
Ik merkte dat mijn mond droog aanvoelde	<input type="radio"/> 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
Ik was niet in staat om ook maar enige postief gevoel te ervaren	<input type="radio"/> 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3
Ik had de neiging om overdreven te reageren op situaties	<input checked="" type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
Ik vond het moeilijk om initiatief te nemen om iets te gaan doen	<input type="radio"/> 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3
Ik had de neiging om overdreven te reageren op situaties	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input checked="" type="radio"/> 3
Ik merkte dat ik beefde (bijv. met de handen)	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input checked="" type="radio"/> 3



Bedankt voor je deelname

Ik was op voorhand al op de hoogte van het echte doel van de studie (stressinductie en niet inspanningen/wiskunde)

AFRONDEN

D. Cross-validation

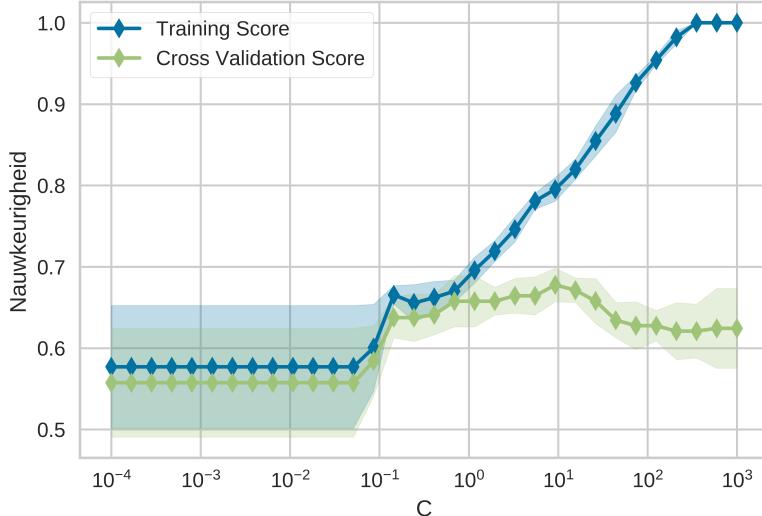
D.1 openSMILE

Onderstaande figuren geven de train- en validatiescores weer in functie van een aantal geoptimaliseerde hyperparameters. Deze curves werden gegenereerd door alle hyperparameters, behalve deze die gevarieerd werd, op hun optimale waarde in te stellen.



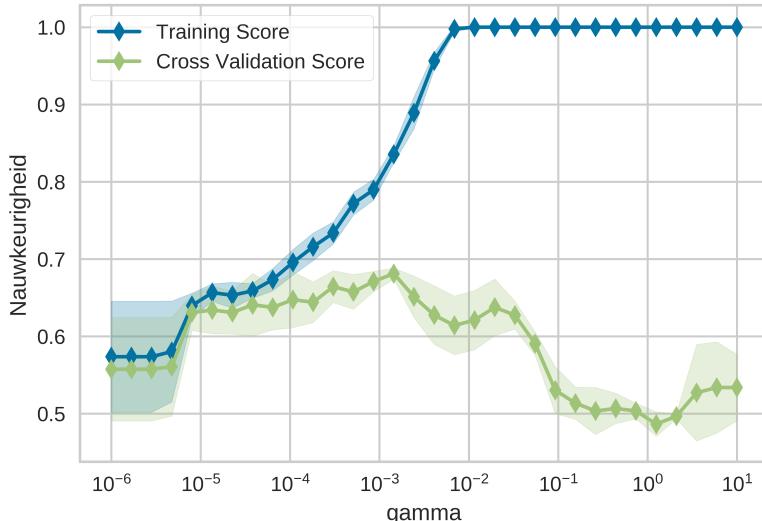
Figuur D.1: Cross-validation voor parameter k

Het aantal geselecteerde features, k , toont een duidelijk maximum rond 110. Hogere waarden zorgen voor een stijging van de trainnauwkeurigheid en een daling van de validatienauwkeurigheid: het model heeft niet voldoende data voor de hoeveelheid features en begint te overfitten. Dit is een duidelijke illustratie van "the curse of dimensionality".



Figuur D.2: Cross-validation voor parameter C

Het is duidelijk dat de SVM te veel misclassificaties toelaat voor hele lage waarden van de regularisatieparameter C en langzaamaan begint te overfitten vanaf een waarde van ongeveer 10. Er wordt een optimum bereikt voor een C van 0.1.

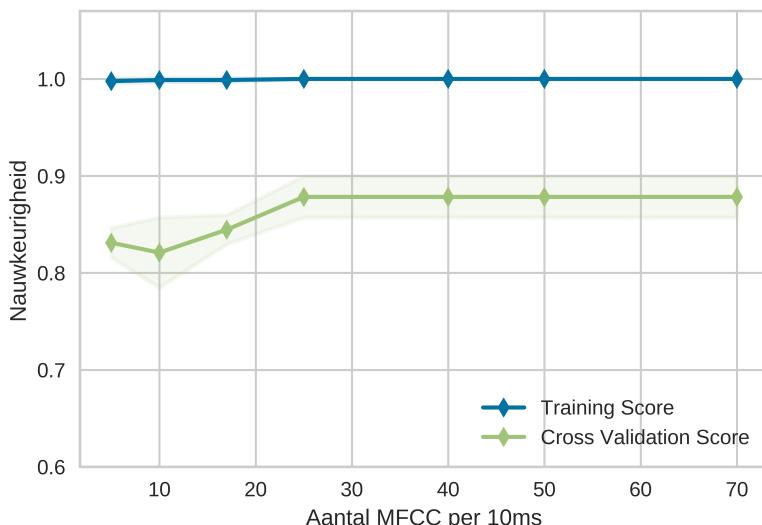


Figuur D.3: Cross-validation voor parameter gamma

Een te lage gamma (of te hoge σ) zorgt voor een te simpele grens en benadert de lineaire kernel. Het is duidelijk dat het model in dat geval underfit gezien de train- en validatiescores ongeveer gelijk – en laag – zijn. Een gamma van 0.001 leidt tot de grootste validatienuwkeurigheid. Grottere waarden zorgen alweer voor overfitting.

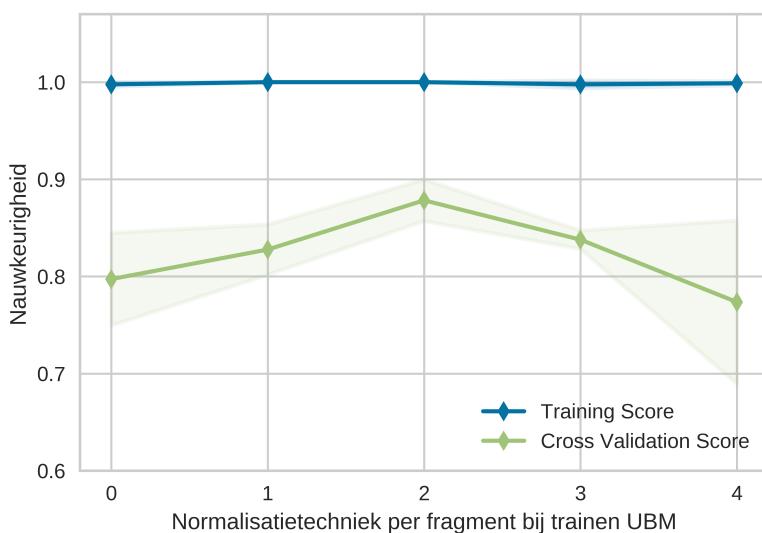
D.2 iVector

Onderstaande figuren illustreren de invloed van de eerder besproken hyperparameters op de prestatie van het model. Deze grafieken werden net zoals bij de SVM-aanpak gegenereerd door alle parameters behalve de onderzochte constant te houden op het gevonden optimum.



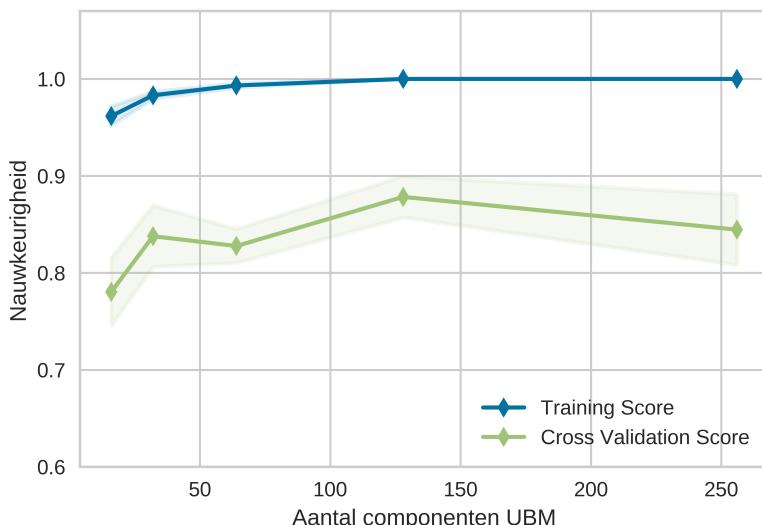
Figuur D.4: Cross-validation voor het aantal MFCC-features dat gebruikt wordt per 10ms

Zoals te zien op figuur D.4, lijkt de precieze hoeveelheid MFCC-features geen invloed te hebben op het model, zolang het er meer dan 25 zijn.



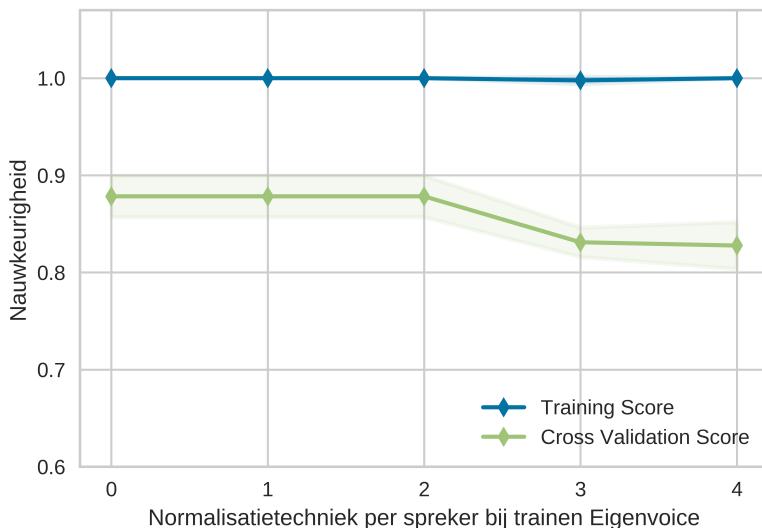
Figuur D.5: Cross-validation voor variërende fragmentnormalisatie-instellingen.

Variatie van de normalisatietechniek gebruikt vooraleer het UBM getraind wordt levert een duidelijk maximum. In figuur D.5 verwijst 0 naar geen normalisatie, 1 naar CMS, 2 naar CMVN, 3 naar “feature warping” [27] en 4 naar de methode `Normalize()` uit `sklearn` [28].



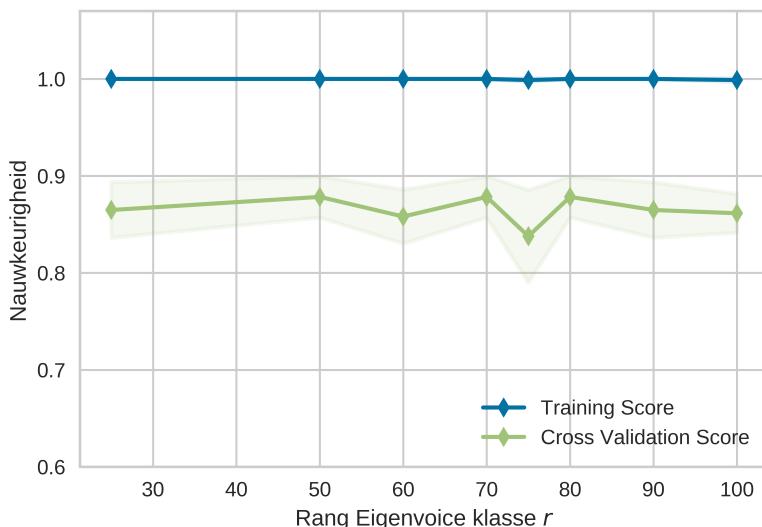
Figuur D.6: Cross-validation voor het aantal componenten of mixtures van het UBM

Het model scoort, in deze configuratie, duidelijk het best met 128 componenten. Minder mixtures zullen verschillende fonemen groeperen als een, meer zullen groepen zoeken binnen fonemen. Beide situaties hinderen de goede werking van de iVector-principes.



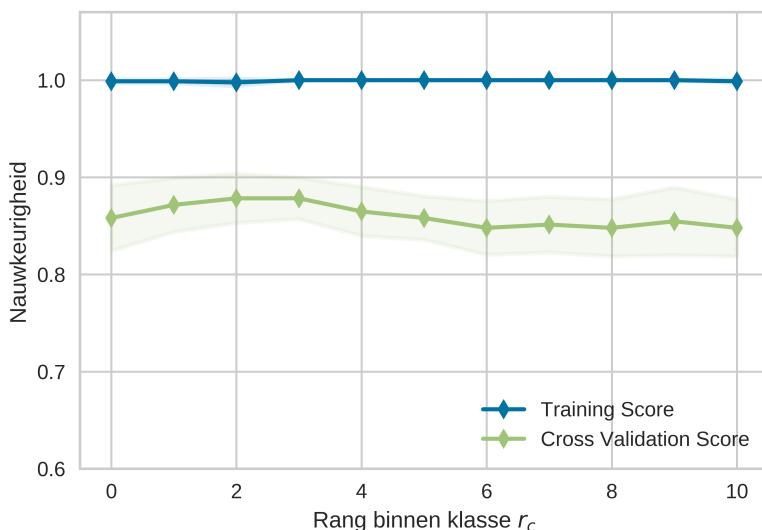
Figuur D.7: Cross-validation voor variërende sprekernormalisatie-instellingen

De normalisatietechniek gebruikt bij het trainen van het Eigenvoice-model voor sprekervariatie lijkt geen verschil te maken, zolang er geen gebruik gemaakt wordt van “feature warping” of de sklearn-methode. Om deze reden werd ervoor gekozen om op dit punt geen normalisatiestap in te voeren. In figuur D.7 verwijst 0 naar geen normalisatie, 1 naar CMS, 2 naar CMVN, 3 naar “feature warping” en 4 naar de methode Normalize() uit sklearn.



Figuur D.8: Cross-validation voor parameter r dat de gewenste rang van de Eigenvoicematrix bepaalt

Een Eigenvoicerank r van 80 leidt tot de beste prestatie van het model. Het verschil de prestatie voor andere waarden is echter niet groot. Merkwaardig is dat er voor een r van 75 een dip in nauwkeurigheid is.



Figuur D.9: Cross-validation voor parameter r_c dat het aantal manieren waarop iemand gestresseerd kan zijn beïnvloedt

Er wordt een optimum bereikt voor een r_c van 3. Er zullen dus 3 richtingen gezocht worden waarin stress in de stem kan verschillen met gemiddelde uitspraak.

E. Informed consent



INFORMED CONSENT

Ik, ondergetekende, verklaar hierbij dat ik, als proefpersoon bij een experiment aan de Vakgroep Ingenieurswetenschappen en architectuur van de Universiteit Gent,

1. de uitleg over de aard van de vragen, taken, opdrachten en stimuli die tijdens dit onderzoek zullen worden aangeboden, heb gekregen en dat mij de mogelijkheid werd geboden om bijkomende informatie te verkrijgen
2. totaal uit vrije wil deelneem aan het wetenschappelijk onderzoek
3. de toestemming geef aan de proefleider om mijn resultaten op vertrouwelijke wijze te bewaren, te verwerken en anoniem te rapporteren
4. de toestemming geef aan de proefleider om ganonimiseerde data online beschikbaar te stellen voor wetenschappelijke doeleinden
5. op de hoogte ben van de mogelijkheid om mijn deelname aan het onderzoek op ieder moment stop te zetten
6. indien ik deelneem in het raam van mijn opleiding: weet dat niet deelnemen of mijn deelname aan het onderzoek stopzetten op geen enkele manier invloed heeft op eventuele evaluatie en/of studiebegeleiding
7. ervan op de hoogte ben dat ik op aanvraag een samenvatting van de onderzoeksbevindingen kan krijgen

Gelezen en goedgekeurd op,

Handtekening

De proefpersoon



F. Invloed van de RRS en DASS schalen

Zoals vermeld in sectie 2.2.3 en in het protocol in Appendix B werd de deelnemers aan het experiment gevraagd om twee vragenlijsten in te vullen die peilen naar hun psychologische toestand. In deze appendix worden zowel de lijsten zelf als de invloed van de antwoorden op deze lijsten besproken.

F.1 Ruminative Response Scale (RRS)

De Ruminative Response Scale (RRS, door [29]) is een schaal die de neiging van de proefpersonen om te rumineren inschat. Er zijn 22 vragen (bv. "Ik denk na over hoe alleen ik me voel."), die beantwoord worden met een cijfer tussen 0 ("bijna nooit") en 3 ("bijna altijd").

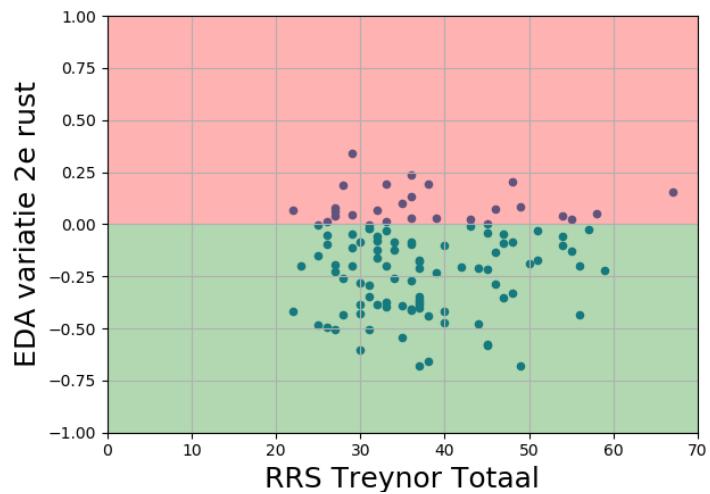
Rumineren is het herhaaldelijk langdurig reflecteren en piekeren over je gevoelens en problemen. Een gevolg hiervan zou kunnen zijn dat mensen die dit doen een verhoogde stress vertonen tijdens de rustfases. Daarom werd er ook onderzocht of er een verband is tussen de verandering in EDA tijdens een rustperiode en de ruminatie score van een proefpersoon. Ook het verband tussen het verschil in VAS scores voor en na een rustperiode en ruminatie werd bekeken. Er werd gekozen voor de tweede rustfase, aangezien er verwacht wordt dat een persoon met aanleiding tot ruminatie minder snel herstelt van een negatieve ervaring, in dit geval de stressinductie, en dit mogelijk merkbaar zou zijn in de EDA verandering en VAS scores.

Op figuur F.1 is het verband tussen het verschil in EDA bij het begin en einde van de tweede rust, en de ruminatie score zichtbaar. Over het algemeen dalen de EDA waarden, wat in lijn ligt met wat we verwachten te zien tijdens een rustperiode. Echter ondanks onze verwachtingen dat proefpersonen met een hogere ruminatie score minder zouden dalen in EDA of mogelijk zelfs stijgen, is er geen duidelijk verband te merken. Bij het vergelijken van de VAS scores met de ruminatie score werd een gelijkaardige plot bekomen.

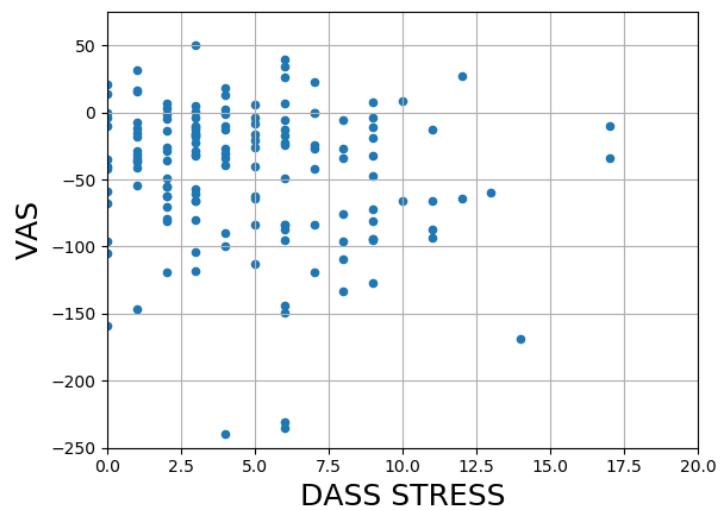
F.2 Depression Anxiety Stress Scale (DASS)

De Depressie, Angst en Stress Schaal (Eng: Depression, Anxiety and Stress Scales (DASS-21), door [30]) wordt gebruikt om te meten in welke mate de proefpersonen depressieve (bv. "Ik had het gevoel dat mijn leven geen zin had"), angstige (bv. "Ik had het gevoel dat ik bijna in paniek raakte") of stressgerelateerde (bv. "Ik merkte dat ik nogal licht geraakt was") symptomen ervaren. De drie subschalen (Depressie, Angst, Stress) hebben elk 7 vragen, met antwoordopties tussen 0 ("Helemaal niet of nooit van toepassing") en 3 ("Zeer zeker of meestal van toepassing").

Figuur F.2 geeft het verband tussen de DASS stress-schaal en het verschil in VAS scores vlak voor en na de stressinductie. We verwachten hier een verband in te zien, aangezien de DASS stress-schaal de stressgevoeligheid van de deelnemer opmeet. Er is echter geen duidelijk verband merkbaar. De VAS scores nemen over het algemeen meer af na de stressinductie (groter negatief effect), wat te verwachten is, maar het is niet zo dat voor mensen die hoger scoren op de DASS stress-schaal ook een groter negatief effect bij de VAS zichtbaar is. Aangezien er wel een linear verband is tussen de EDA en de VAS (zie Figuur 2.1), maar geen verband tussen de VAS en DASS stress-score, kunnen we verwachten dat er tussen de EDA en de DASS stress-score geen verband te vinden zal zijn. Voor de andere schalen van de DASS werden gelijkaardige resultaten gevonden.



Figuur F.1: Verband tussen het verschil in EDA bij begin en einde tweede rust, en de ruminatie score op de RRS



Figuur F.2: Verband tussen DASS stress-score en het verschil in VAS scores vlak voor en na de stressinductie

G. Voorstel vakoverschrijdend project: Stemmingkskantificatie met NLP en wearable data

Groepsleden

- Cedric Anné (computerwetenschappen)
- Ernest Van Hoecke (elektrotechniek)
- Vic Degraeve (computerwetenschappen)

Probleemstelling

Eenduidig gekwantificeerde data van de psychologische toestand van een persoon zou van onmeetbare waarde zijn voor bijvoorbeeld patiënten die lijden aan een bipolaire stoornis. Op die manier zouden therapeuten en de patiënt zelf gewaarschuwd kunnen worden dat het bijvoorbeeld even net iets te goed gaat en het aangewezen is het even kalmer aan te doen. Ook kan de invloed van een bepaalde therapie zo gemakkelijk gemeten worden. Er bestaan reeds applicaties met als doel je eigen gemoedstoestand bij te monitoren. Echter zijn de metrieken die hierbij gebruikt worden te weinig granulair (bijvoorbeeld een lachend of wenend gezichtje) of niet heel informatief.

Doelstelling

Gebruikmakend van het vakgebied "natural language processing" - waar in de afgelopen jaren enorme vooruitgangen in zijn gemaakt - zou uit een dagelijks dagboekfragment (of een analyse van tweets) al heel wat informatie gedestilleerd kunnen worden. Met data zoals de hartslag en geleidbaarheid van de huid - dit laatste blijkt veel informatiever te zijn dan men zou denken - zou bovendien een cijfer op de hoeveelheid stress kunnen worden geplakt. Dit is meteen ook een mogelijke overlap met de richting elektrotechniek. Het is de bedoeling om dit "stemmingsregressie" algoritme vervolgens te implementeren in een gebruiksvriendelijke app die mogelijks zelf ook enkele eenvoudige analyses doet op de data, en deze mooi uitzet in een grafiek. Hieruit zouden al veel inzichten verworven kunnen worden (vergelijk met grafieken lezen van aandelen, deze "crashen" ook als het even te goed gaat.)

Uitbreidingen

Door de app uit te breiden met een sociaal netwerk van patiënten en therapeuten zou een therapeut real-time feedback kunnen krijgen over de werking van zijn behandeling. Op basis van het profiel en de aandoening van de patient zouden gelijkaardige gebruikers gekoppeld kunnen worden, en zouden bepaalde behandelingen aangeraden kunnen worden aan bepaalde groepen mensen (zoals een gewoon recommender system) op basis van wat eerder werkte op gelijkaardige patienten. Bovendien zouden therapeuten uit deze data ook meteen een beter beeld krijgen over de toepasbaarheid van verschillende therapieën.

Vakoverschrijdend karakter

- Statistiek (machine learning)
- Programmeren (implementatie)
- Softwareontwikkeling (implementatie interface)
- Multimediatechnieken (eventuele websites om data te verzamelen, er zijn echter al gerelateerde datasets te vinden)
- Signaalverwerking (wearables voor stress data)
- Elektrische netwerken (wearables)

Opmerkingen in verband met haalbaarheid

Ikzelf (Vic Degraeve) heb reeds het mastervak "Machinaal Leren" (E061330) afgerond. Ernest Van Hoecke is al geslaagd voor het vak signaalverwerking. Bovendien hebben we dit idee voorgelegd aan enkele vrienden uit de master psychologie, die graag hulp zouden bieden voor de menswetenschappelijke kant van het project.

Lijst van Afkortingen

ANOVA	Variantieanalyse (Analysis of variance)
CMS	Cepstral Mean Subtraction
CMVN	Cepstral Mean and Variance Normalization
DASS	Depressie Angst Stress Schaal (vertaling van het Engelse 'Depression Anxiety Stress Scale')
ECG	Electrocardiogram
EDA	Electrodermal Activity, zie ook GSR
FN	Foutnegatief (False Negative)
FP	Foutpositief (False Positive)
GMM	Gaussian Mixture Model
GSR	Galvanic Skin Response, dit is een andere benaming voor EDA
HRV	Hartritmevariabiliteit
i.i.d	Onafhankelijk en identiek verdeeld (independent and identically distributed)
MFC	Mel-frequency cepstrum
MFCC	Mel-frequency cepstral coefficient
MIST	Montreal Imaging Stress Task
NLP	Natural Language Processing
PCA	Principal Component Analysis
RBF	Radial Basis Function, dit is een SVM kerneltype
RRS	Ruminative Response Scale
SCL	Huidgeleidbaarheidsniveau (van het Engelse 'Skin Conductance Level')
SDC	Shifted Delta Coefficients
SVM	Support Vector Machine
TN	Echtnegatief (True Negative)
TP	Echtpositief (True Positive)
UBM	Universal Background Model
VAS	Visueel Analoge Schaal
WHO	World Health Organization

Bibliografie

- [1] Jenette L Smith and Miguel A Perez. The Importance of Stress Management in Today's Society. Technical Report 1, Rev. Cient. Sena Aires., 2018.
- [2] Volksgezondheidenzorg.info. Overspannenheid en burn-out. <https://www.volksgezondheidenzorg.info/onderwerp/overspannenheid-en-burn-out/cijfers-context/trends>.
- [3] Ana Paula Amaral, Maria João Soares, Ana Margarida Pinto, Ana Telma Pereira, Nuno Madeira, Sandra Carvalho Bos, Mariana Marques, Carolina Roque, and António Macedo. Sleep difficulties in college students: The role of stress, affect and cognitive processes. *Psychiatry Research*, 260:331–337, 2 2018.
- [4] Elena Smets, Emmanuel Rios Velazquez, Giuseppina Schiavone, Imen Chakroun, Ellie D'Hondt, Walter De Raedt, Jan Cornelis, Olivier Janssens, Sofie Van Hoecke, Stephan Claes, Ilse Van Diest, and Chris Van Hoof. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. *npj Digital Medicine*, 1(1):67, 12 2018.
- [5] Elena Smets. *Towards large-scale physiological stress detection in an ambulant environment*. PhD thesis, Ghent University, 2018.
- [6] Javier Hernandez. *Towards Wearable Stress Measurement*. PhD thesis, Massachusetts Institute of Technology, 2015.
- [7] Lisetti C. and Nasoz F. Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals. *EURASIP Journal on Applied Signal Processing*, 11:1672–1687, 2004.
- [8] Asma Ghandeharioun, Szymon Fedor, Lisa Sangermano, Dawn Ionescu, Jonathan Alpert, Chelsea Dale, David Sontag, and Rosalind Picard. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In *2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017*, pages 325–332, 2017.
- [9] Foteini Agrafioti, Dimitris Hatzinakos, and Adam K. Anderson. ECG pattern analysis for emotion detection. *IEEE Transactions on Affective Computing*, 3(1):102–115, 1 2012.
- [10] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investigation*, 15(3):235–245, 3 2018.
- [11] Johannes Wagner, Johannes Wagner, Jonghwa Kim, and Elisabeth Andre. From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification. In *IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA & EXPO (ICME 2005)*, pages 940–943, 2005.
- [12] Cong Zong and Mohamed Chetouani. Hilbert-Huang transform based physiological signals analysis for emotion recognition. *IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2009*, pages 334–339, 2009.
- [13] OW Kwon, K Chan, J Hao, and TW Lee. Emotion Recognition by Speech Signals. *Conference: 8th European Conference on Speech Communication and Technology, EUROSPEECH 2003 - INTERSPEECH 2003, Geneva, Switzerland*, pages 125–128, 2003.
- [14] Jun Deng, Zixing Zhang, Florian Eyben, and Bjorn Schuller. Autoencoder-based unsupervised domain adaptation for speech emotion recognition. *IEEE Signal Processing Letters*, 21(9):1068–1072, 2014.
- [15] Rosalind W. Picard, Szymon Fedor, and Yadid Ayzenberg. Multiple Arousal Theory and Daily-Life Electrodermal Activity Asymmetry. *Emotion Review*, 8(1):62–75, 2016.
- [16] Leah-Marie Byrion and Matthew T. Feldner. Self-Assessment Manikin. In *Encyclopedia of Personality and Individual Differences*, pages 1–3. Springer International Publishing, Cham, 2017.
- [17] K Dedovic, R Renwick, N K Mahani, V Engert, S J Lupien, and J C Pruessner. The Montreal Imaging Stress Task. *Journal of Psychiatry and Neuroscience*, 30(5):319, 2005.
- [18] Kris Demuynck. Voice Analytics: Speech processing crash course, 2019.
- [19] Jose R Calvo, Rafael Fernández, and Gabriel Hernández. Application of Shifted Delta Cepstral Features in Speaker Verification. Technical report, Advanced Technologies Application Center, CENATAV, 2007.

- [20] Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller. Recent developments in openSMILE, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM international conference on Multimedia - MM '13*, pages 835–838, New York, New York, USA, 2013. ACM Press.
- [21] Clever Owl. Curse of Dimensionality Explained. <http://cleverowl.uk/2016/02/06/curse-of-dimensionality-explained/>, 2016.
- [22] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, David Cournapeau, Alexandre Passos, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. In *Journal of Machine Learning Research*, volume 12, pages 2825–2830, 2011.
- [23] Najim Dehak, Patrick J Kenny, Réda Dehak, Pierre Dumouchel, and Pierre Ouellet. Front-End Factor Analysis for Speaker Verification. *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, 19(4), 2011.
- [24] Brecht Desplanques and Kris Demuynck. Cross-lingual Speech Emotion Recognition through Factor Analysis. Technical report, Ghent University - imec, IDLab, 2018.
- [25] Ondrej Glembek, Lukas Burget, Pavel Matejka, Martin Karafiat, and Patrick Kenny. Simplification and optimization of i-vector extraction. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4516–4519. IEEE, 5 2011.
- [26] J. van de Weijer and I. Slis. Nasaliteitsmeting met de nasometer. *Logop Foniatr.*, 63:97–101, 1991.
- [27] Jason Pelecanos and Sridha Sridharan. Feature Warping for Robust Speaker Verification. Technical report, ISCA Archive, 2001.
- [28] Documentation Scikit-learn 0.21.1. normalize. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.normalize.html>.
- [29] Wendy Treynor, Richard Gonzalez, and Susan Nolen-Hoeksema. Rumination Reconsidered: A Psychometric Analysis. *Cognitive Therapy and Research*, 27(3):247–259, 2003.
- [30] P.F. Lovibond and S.H. Lovibond. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33(3):335–343, 3 1995.