

Spraakherkenning

Bart Woelders
Klas 6W

Begeleider
Rob Hazelzet

Inhoudsopgave

Voorpagina	pagina 1
Inhoudsopgave	pagina 2
Inleiding	pagina 3
Hoofdstuk 1	pagina 5
- spraakherkenning	pagina 5
- Begrippen, Methoden en Theorieën	pagina 8
Hoofdstuk 2	pagina 14
- spraakherkenning met gebruik van computers	pagina 19
Hoofdstuk 3	pagina 21
- de invloeden van de variabelen	pagina 26
Bronnenlijst	pagina 27
Nawoord	pagina 28
Bijlage	pagina 29

Inleiding

Ik heb spraakherkenning tot onderwerp van mijn profielwerkstuk gekozen omdat het mij van de vele mogelijkheden het meest interesseerde. Van de door ons behandelde Natuurkunde spraken trillingen en golven mij überhaupt het meest aan. Dit was tevens een van de vele experimenten waarvoor begeleiding werd aangeboden door de Universiteit van Amsterdam. Behalve een duidelijke indicatie van wat dit profielwerkstuk inhield en geschikte informatie over het onderwerp was er ook de mogelijkheid om de experimenten op de Faculteit der Natuurwetenschappen, Wiskunde en Informatica uit te voeren.

Hier ben ik goed begeleid door de heer P.P.M. Molenaar, die zelf het profielwerkstuk heeft geschreven. Hij is gespecialiseerd in trillingen en geluidsgolven.

In het gebouw waar ik het experiment uitvoerde is de afdeling van de UvA gevestigd, die zich bezig houdt met het ontwikkelen van het programma Coach. De heer Molenaar is hier ook aan verbonden.

De metingen heb ik samen gedaan met twee scholieres uit Leiden. Hierdoor ontstond de mogelijkheid om behalve gegevens over de mannelijke stem ook gegevens over de vrouwelijke stem in te winnen, wat voor een uitgebreider onderzoek zorgt.

Dit zijn mijn hoofdvraag, deelvragen en hypothese:

Hoofdvraag

Hoe kunnen we op basis van de waarde van diverse variabelen, klanken herkennen en vastleggen?

Deelvragen

- 1 Wat is 'spraakherkenning, hoe kunnen wij dat toepassen en op welke natuurkundige begrippen en theorieën is spraakherkenning gebaseerd?
- 2 Hoe kunnen klanken herkend en vastgelegd worden, op basis van kenmerken die voor elk mens ongeveer identiek zijn?
- 3 Welke invloed hebben de verschillende variabelen, die van toepassing zijn op geluidstrillingen, op de herkenbaarheid van een uitgesproken klinker? Deze invloed wordt voor de diverse klinkers door middel van experiment gemeten.

Hypothese

De verschillen tussen de waarden van de diverse variabelen en de grootten van deze waarden, geeft het onderscheid tussen de diverse klanken aan. Door deze waarden te analyseren kan de desbetreffende klank worden herkend en vastgelegd.

De deelvragen worden uitgebreid behandeld in de volgende drie hoofdstukken. Elke deelvraag wordt in een afzonderlijk hoofdstuk behandeld. In hoofdstuk 3 wordt tevens de hypothese nagekeken: klopt deze of klopt deze niet?

In hoofdstuk een zal worden uitgelegd wat spraakherkenning is en hoe wij hier gebruik van kunnen maken in de maatschappij en in de industrie. Tevens zullen de natuurkundige begrippen, methoden en theorieën, en hun betrekking tot spraakherkenning, worden behandeld. Begrippen die aan bod komen zijn frequentie, amplitude en formanten. Bij spraakherkenning gebruikte methoden zijn Fourier analyse en lineaire predictie. Lineaire predictie is gebaseerd op de Fourier-wiskunde; ook deze komt kort aan bod.

In hoofdstuk twee zal worden behandeld hoe we Fourier analyse en Lineaire predictie kunnen gebruiken voor het herkennen van klanken. Er wordt uitgelegd hoe we deze methoden kunnen gebruiken om de Formanten te vinden en hoe we de gevonden Formanten kunnen toepassen om zo een letter (klinker) te vinden. Vervolgens wordt uitgelegt hoe een spraakherkenner woorden kan herkennen.

In het derde hoofdstuk bestuderen en vergelijken we de Formanten van de diverse klinkers. De verschillen en overeenkomsten tussen de Formanten van diverse klinkers tonen aan of we wel of niet een klinker kunnen onderscheiden van een andere klinker. Hierna wordt uiteengezet hoe we de juiste klinker kunnen vinden door gebruik te maken van alle eerder gebruikte technieken. Tot slot toetsen we de hypothese, en herschrijven we deze mogelijk.

Omdat volledige spraakanalyse erg ingewikkeld is, beperken we ons in dit onderzoek hoofdzakelijk tot het herkennen van een aantal verschillende klinkers.

Hoofdstuk 1

“Wat is ‘spraakherkenning’, hoe kunnen wij dat toepassen en op welke Natuurkundige begrippen en theorieën is spraakherkenning gebaseerd?”

Spraakherkenning is gebaseerd op een aantal Natuurkundige begrippen, methoden en theorieën die worden toegepast om spraak te herkennen. Later in dit hoofdstuk zullen alle toegepaste methoden en de belangrijkste begrippen uitgebreid besproken worden. Zo zal onder andere in worden gegaan op het begrip ‘Formanten’ en wordt er uitgelegt wat er precies met ‘Fourier Analyse’ en ‘Lineaire predictie’ wordt bedoeld. Allereerst is natuurlijk van belang om de volgende vraag te beantwoorden: “wát is spraakherkenning eigenlijk?” Ook zullen we kijken op wat voor manieren spraakherkenning wordt, of kan worden toegepast.

Spraakherkenning

‘Spraakherkenning, in de breedste zin van het woord, betekent ‘het herkennen en vastleggen van gesproken tekst (spraak)’. Dit klinkt behoorlijk nietszeggend, daarom zullen we dit verder verduidelijken: bij spraakherkenning is het de bedoeling dat de woorden (of klanken, letters, zinnen et cetera) worden herkent en vastgelegd. Dit doet men met gebruik van een computer. Deze analyseert spraak en bouwt de ontvangen geluidssignalen om in een andere vorm, bijvoorbeeld een ander signaal of een voor ons leesbare tekst.

Bij spraakherkenning wordt spraak dus allereerst herkend en vervolgens vastgelegd in een andere vorm zoals een elektrisch signaal of een tekst.

In onze huidige maatschappij hebben we al veel te maken met Spraakherkenning. Denk bijvoorbeeld aan telefonische informatielijnen: deze gebruiken spraakherkenning om te achterhalen wat de beller wil weten. De computer kan dan de gewenste informatie geven met behulp van digitale spraak, beter bekend als ‘spraaksynthese’.

Dit is bijvoorbeeld het geval bij informatielijnen die informatie geven over beurskoersen of over reistijden bij het openbaar vervoer. Een goed en eenieder bekend voorbeeld van spraaksynthese is de tijdmelding die men telefonisch kan bellen: de tijd wordt dan doorgegeven door een ‘computerstem’, oftewel digitale spraak: spraaksynthese!

In call centers wordt spraakherkenning vaak gebruikt om te achterhalen wat de beller wil weten, of wat zijn identiteit is: hij kan dan direct met de juiste persoon worden doorverbonden, wat een grotere efficiëntie tot gevolg heeft.

Ook in auto’s gebruikt men reeds spraakherkenning; er zijn bijvoorbeeld auto’s op de markt die werken met een navigatiesysteem met spraakherkenner. De bestuurder kan dan zeggen waar hij heen wil en het navigatiesysteem kan dan bijvoorbeeld met behulp van een scherm, maar natuurlijk nog beter met gebruik van spraaksynthese, aangeven welke kant de bestuurder op een gegeven moment op moet. En dan is er ook nog de mogelijkheid om 100% handsfree te bellen: spreek gewoon het telefoonnummer in!

Een andere mogelijkheid is het geven van commando's; bijvoorbeeld aan huishoudelijke apparaten of lampen. Een commando zou dan kunnen zijn: 'licht aan', of 'douche aan veertig graden'. Deze commando's worden dan door de spraakherkenner omgezet in signalen die worden vergeleken met eerder ingevoerde signalen. Komt een signaal overeen met een uit het geheugen van de spraakherkenner, dan zal de bij dit signaal passende opdracht worden uitgevoerd. Voorwaarde is wel dat in het betreffende apparaat een computer is ingebouwd of dat deze is aangesloten op een computer.

Momenteel wordt deze techniek nog bijna niet gebruikt. De benodigde technieken zijn er vrijwel niet en slechts een klein deel van de huishoudelijke apparaten wordt door computers bestuurd. Het is echter van belang dat deze technieken goed ontwikkelt worden: behalve dat hier een grote (algemene) markt voor is, zijn de toepassingen ook goed bruikbaar voor gehandicapten!

Ook computers kan je commando's geven, bijvoorbeeld om programma's te openen. Dit zal in de praktijk niet snel worden toegapst, omdat werken met een muis handiger en sneller is dan spraakherkenning. Dit is natuurlijk wel de ideale uitkomst voor mensen met een handicap. Een toepassing op de computer die wel van groot nut is, is het dicteren van teksten aan een computer. Met gebruik van een spraakherkenner kan spraak worden omgezet in een tekst op het beeldscherm. Dit heeft zo zijn voordelen, omdat de meeste mensen sneller praten dan typen. Ook zal het aantal RSI patiënten hiermee hevig afnemen; een toetsenbord wordt vrijwel overbodig. Nadelen zijn er echter ook: veertien keer backspace zeggen om een veertienletterig woord te wissen is niet echt handig. Andere opties zoals 'selecteren' en 'knippen' zijn waarschijnlijk ook makkelijker uit te voeren met gebruik van een toetsenbord.

Er zijn ook nog een aantal 'algemenere' nadelen.

Omdat elke stem anders is, moet een spraakherkenner getraind worden voordat je hem kan gebruiken. Dit betekent dat een toekomstig gebruiker allerlei woorden en cijfers moet uitspreken: de spraakherkenner kan dan wennen aan je stem.

Een ander probleem zijn de omgevingsgeluiden. Deze maken dat het door de spraakherkenner ontvangen geluidssignaal niet zuiver is. De computer moet verschil maken tussen deze geluiden en de gegeven commando's. Soms denkt de computer dat achtergrondgeluiden commando's van de gebruiker zijn. Dit ligt onder ander aan de verschillen in stemvolume en achtergrondgeluid en de microfoongevoeligheid. Bij sommige spraakherkenners wordt dit probleem opgelost door de spraakherkenner te ijken door het inspreken van cijfers. Deze problemen zullen in de toekomst zoveel mogelijk verholpen worden. Dit is mogelijk doordat er steeds betere apparatuur op de markt komt. Computers zijn veel sneller dan vroeger. Dit is nodig omdat spraakherkenners erg veel rekenkracht nodig hebben.

We hebben nu toepassing van spraakherkenning met gebruik van commando's toegelicht, evenals het omzetten van een geluidssignaal door een spraakherkenner: kijk bijvoorbeeld naar de computer die spraak in tekst omzet.

Er zijn echter nog andere toepassingen van spraakherkenning: verificatie en identificatie. Deze toepassingen worden gebruikt voor het herkennen van mensen.

Identificatie & Verificatie

Bij identificatie wordt gekeken wie de spreker is, deze wordt geïdentificeerd. Hierbij gebruikt men een methode die berust op vergelijking: het spraakfragment van de spreker wordt vergeleken met alle spraakfragmenten in de database. Hierbij wordt dus gekeken naar de verschillen tussen de stem van diverse personen. Identificatie is erg moeilijk: het spraakfragment moet met vele andere spraakfragmenten vergeleken worden. De technieken die hiervoor nodig zijn moeten nog worden ontwikkeld, of worden momenteel ontwikkeld. Dit kan men bijvoorbeeld gebruiken bij telefonisch bankieren. Zodra iemands stem overeenkomt met een stem uit het bestand, kan deze toe worden gelaten tot zijn spaarrekening of kan deze persoon worden doorverbonden met de juiste bankmedewerker, bijvoorbeeld zijn persoonlijke beleggings adviseur.

De tweede manier, verificatie, wordt nu al toegepast. Hierbij wordt er gekeken of iemand is wie hij zegt te zijn. Bij deze methode worden twee spraakfragmenten vergeleken. Van alle gebruikers van het systeem staat een spraakfragment opgeslagen in het bestand. Als iemand toegang wil krijgen tot het systeem moet deze persoon eerst zeggen wie hij is. Vervolgens moet hij precies hetzelfde spraakfragment uitspreken als het fragment dat in het bestand is opgeslagen. Deze twee fragmenten worden dan met elkaar vergeleken. Hieruit blijkt of de gebruikers is wie hij zegt te zijn: dit wordt geverifieerd. Het mag duidelijk zijn dat deze methode, verificatie, veel makkelijker is dan identificatie. Deze methode wordt dan ook al gebruikt. Het gaat hier wederom om de kenmerkende eigenschappen van iemands stem.

De computer snijdt het spraakfragment in stukken van 20 milliseconden, oftewel; elke seconde spraak wordt verdeeld in 50 delen. Van die stukken wordt een profiel gemaakt dat voor iedereen uniek is. Dit profiel worden bij verificatie vergeleken met het profiel van het spraakfragment van de spreker. Deze spraakherkenners zijn zo precies dat het bijna onmogelijk is om ze om de tuin te leiden.

Als toepassing moet je bijvoorbeeld kijken naar beveiliging van ruimtes en gebouwen: door middel van verificatie wordt gekeken of de persoon is wie hij zegt te zijn, vervolgens wordt gekeken of deze persoon toegang mag worden verschaft. Ook wordt de methode gebruikt door sommige bedrijven met een vast klantenbestand. Door middel van verificatie wordt gekeken of de beller een geregistreerde klant is. Na verificatie kan de beller zijn bestelling opgeven.

Het is belangrijk om op te merken dat er bij alle verschillende methoden en toepassingen die hierboven zijn genoemd nog een belangrijk onderscheid is te maken: het gebruik van verschillen in stem, of juist het gebruik van kenmerken die voor iedereen min of meer identiek zijn.

Bij identificatie en verificatie zal worden gekeken naar de verschillen in stem, bij commando's en gebruik op de tekstverwerker zal juist worden gekeken naar kenmerken die voor iedereen identiek zijn.

Begrippen, methoden en theorieën

In deze paragraaf zullen 'Frequentie', 'Geluidsintensiteit', 'Formanten', 'Fourier Analyse', 'Lineaire predictie' en de stelling van Fourier uitgebreid worden toegelicht.

Onder de frequentie van een (geluids)trilling verstaan we het aantal trillingen per seconden. De frequentie wordt uitgedrukt in hertz. (Hz)
Bij gesproken klanken ligt de frequentie tussen de 300 en de 4000 hertz. De oren van de mens zijn gevoelig voor trillingen met een frequentie die tussen de 20 en de 20.000 hertz ligt.

De trilling heeft ook een amplitude, in dit geval een geluidssterkte, die we hier verder de geluidsintensiteit zullen noemen. De verhouding van de intensiteit van de diverse frequenties tot elkaar is van groot belang: dit is de verhouding waarin de frequenties in een klank voorkomen.

Zo is in elke klank de geluidsintensiteit van een aantal frequenties het grootst. Deze frequenties komen 'harder' voor dan anderen en bepalen de klank dus in meerdere mate. De frequenties die het sterkst doorkomen noemen we Formanten. Deze Formanten karakteriseren een klank. De eerste drie Formanten, die dus het sterkst doorkomen, zijn voor ieder mens min of meer identiek. De overige Formanten zijn persoonlijk.

Dat sommige frequenties sterker doorkomen dan andere heeft te maken met de werking van onze keel en mond. De stembanden produceren grondtonen en diverse boventonen (het aantal boventonen bedraagt ongeveer 20). Deze worden versterkt door keel en mond: deze treden op als resonantiebuï. Hierbij moet vermeld worden dat elke boventoon steeds ongeveer 3 decibel zachter is.

De mond, tong en gehemelte werken echter ook als een filter; hierdoor komen sommige frequenties beter door dan andere, en hierdoor zijn wij ook in staat om vele diverse klanken te produceren.

Een klank bestaat dus uit ontzettend veel frequenties die allemaal in meer of mindere mate in een klank voorkomen. Willen we letters of woorden herkennen, dan moeten we de formanten van de letters zien te achterhalen: de eerste drie formanten bepalen immers wat voor een klank het is.

Methoden die wij hierbij gebruiken zijn Fourier Analyse en Lineaire predictie. Er zijn ook andere methoden, maar dit zijn de methoden die wij in Coach 5 konden gebruiken om een klank te analyseren. Voordat we verder ingaan op deze omzettingen, moet eerst de stelling van Fourier worden geponeerd en toegelicht.

Fourier Wiskunde

De stelling van Fourier luidt:

Elke periodieke trilling kan worden opgevat als een combinatie van een grondtoon en zijn boventonen.

Als we deze stelling omdraaien is hij goed toepasbaar op onze situatie:

Een trilling die een combinatie is van een grondtoon en zijn boventonen is een periodieke trilling.

Hieruit valt te concluderen dat de trilling bij een klank (een gesproken klank is samengesteld uit grondtoon en zijn boventonen) een periodieke trilling is. Oftewel; de trilling is samengesteld uit een aantal sinusoïden. Een trilling is dus ook samen te stellen uit een aantal sinusoïden. (sinussen en cosinussen)

De tak van de wiskunde die zich bezig houdt met het opbouwen van trillingen uit sinusoïden en alle gerelateerde onderwerpen noemen we de 'Fourier wiskunde'. Hierdoor komen we uit bij Fourier Analyse en Lineaire predictie.

Fourier Analyse

Als we in Coach Fourier Analyse uitvoeren wordt het U,t-diagram, omgezet in een I,freq-diagram. (U = spanning, I = geluidsintensiteit)

Bij Fourier Analyse wordt gepoogd uit een periodiek signaal de sinusoïden, (waaruit deze immers is opgebouwd) uit te rekenen (te benaderen) met een numerieke methode. Door gebruik te maken van deze sinusoïden kan worden berekend in welke mate een frequentie in de trilling voorkomt.

Nog even kort: Een periodieke trilling(klank) is samengesteld uit een grondtoon en zijn boventonen. Bij Fourier analyse wordt geprobeerd met een numerieke methode de sinusoïden waaruit deze periodieke trilling is opgebouwd te berekenen. Met gebruik van deze sinusoïden is te berekenen in welke mate de grondtoon en zijn boventonen (waaruit die trilling is opgebouwd) voorkomen. Er wordt dus berekend in welke mate de diverse frequenties voorkomen.

Het intensiteit,frequentie- diagram heeft voor de intensiteit een logistische schaalverdeling; geluidsintensiteit neemt immers logistisch toe:

$$L_p(\text{db}) = 10 * \log (I / I_0)$$

Hierin is L_p het geluidsniveau in decibel, I de geluidsintensiteit en I_0 de minimale geluidsintensiteit ($1,0 * 10^{-12}$) Wat wij (en Coach) dus de Intensiteit noemen (in dB) is eigenlijk het geluidsniveau. Intensiteit wordt namelijk gegeven in W/m^2

Vervolgens passen we Lineaire predictie toe: hierbij wordt achterhaald welke filters er zijn, en waar dus wel en waar niet wordt gefilterd.

Filter werking

Het filter, zoals eerder uitgelegd, dempt de ene frequentie sterker dan de ander. De ene boventoon wordt dus sterker gedempt dan de ander.

Als de sterkste geluidsintensiteit die het filter voor een bepaalde frequentie (bijvoorbeeld 1 kHz) doorlaat 30 dB is, en de geluidsintensiteit van de betreffende boventoon (dus de boventoon met een freq. van 1 kHz) is vóór het filteren 40 dB, dan zal deze boventoon worden gedempt. De geluidsintensiteit van de boventoon, zoals deze in de uitgesproken klank voorkomt is dan 30 dB.

Een andere boventoon, bijvoorbeeld met een frequentie van 2 kHz, heeft vóór het filteren een intensiteit van 20 dB. Als we stellen dat de sterkste geluidsintensiteit die het filter voor deze frequentie (2 kHz) doorlaat ook 30 is, dan wordt deze boventoon niet gedempt. De geluidsintensiteit van de boventoon vóór het filteren was immers maar 20 dB, en de boventoon wordt dus volledig 'doorgelaten'.

Lineaire predictie

Bij het uitvoeren van lineaire predictie stelt de computer een functie samen uit een aantal sinusoïden. Deze functie geeft de werking van het filter weer. Deze functie komt in een I, freq. diagram te staan: de waarde die de functie dan bij een bepaalde frequentie heeft, geeft de maximale intensiteit die het filter doorlaat weer. Bij het voorbeeld van bovenstaande tekst (boventoon met een frequentie van 1 kHz) zal de waarde van de functie dus 30 zijn. ($I = 30$ dB)

We laten de computer deze functie opstellen, omdat we aan de maxima van de functie kunnen zien welke frequenties het sterkst doorgelaten worden. Deze frequenties zijn de formanten van de klank

Als we het in onderstaande tekst over de functie hebben, dan gaat het altijd over de functie die de werking van het filter weergeeft. We gaan het proces van lineaire predictie nu stap voor stap doorlopen

1

Lineaire predictie is gebaseerd op het principe 'dat elk gegevenspunt van een trilling voorspeld kan worden uit een functie van alle vooraf berekende punten.'

We moeten dus een aantal punten van de functie berekenen. Vervolgens moeten we de functie opstellen die door deze punten gaat. Deze functie voorspeld ook alle andere gegevenspunten van de trilling.

Om de functie op te kunnen stellen moeten we dus allereerst een aantal punten van die functie berekenen.

2

Wat zijn hier punten van de functie? De functie moet de werking van het filter weergeven, oftewel de geluidssterkte die het filter voor elke frequentie doorlaat.

Een punt van de functie is dus een bepaalde frequentie (x-coördinaat) en de geluidssterkte die het filter voor die frequentie doorlaat. (y-coördinaat)
Door voor een aantal frequenties de geluidsintensiteit te berekenen komen we aan een aantal punten, waarna we de functie op kunnen stellen. (zie **1**)

(de x-coördinaat kan bijvoorbeeld 1 (kHz) zijn. Als we het voorbeeld hierboven aanhouden, dan is de bijbehorende y-coördinaat 30. Dit zijn de coördinaten van dit punt)

3

Hoe berekenen we voor een aantal frequenties hoe er daar gefilterd wordt? Dit doen we door de verhoudingen van de intensiteit van deze frequenties te vergelijken. Hierdoor kan berekend worden hoe er voor die frequenties gefilterd wordt. Zo komen we dus aan de geluidsterkte die het filter doorlaat, nodig voor de y-coördinaat van het punt.

Concluderend:

We hebben nu een aantal frequenties met bijbehorende waarde van de intensiteit die het filter doorlaat. Dit zijn een aantal punten van de functie.

We laten de computer nu de functie die door deze punten gaat berekenen. Dit is de functie die de werking van het filter weergeeft.

Deze functie is te tonen in het frequentiespectrogram. Hier vormt deze dan de omhullende curve. De reden dat deze curve omhullend is, is deze: frequenties waarvan de intensiteit kleiner is dan de intensiteit die het filter voor die frequentie doorlaat, vallen in het diagram 'weg' onder de curve. De frequenties waarvan de intensiteit groter is dan de intensiteit die het filter voor die frequentie doorlaat, worden gedempt en komen zo sterk door als dat het filter die frequenties maximaal door laat komen. Deze frequentie geven de maxima van de functie aan.

Samengevat:

met Fourier analyse worden de verschillende frequenties waaruit de klank is opgebouwd gescheiden. Dit gebeurt doordat men met een numerieke methode de sinusoiden berekent waaruit de trilling is opgebouwd. Aan de hand hiervan kan men ontdekken in welke mate een frequentie voorkomt.

Vervolgens wordt de werking van het filter door een functie weergegeven. Deze functie wordt samengesteld op basis van een aantal berekende punten en opgebouwd uit een aantal coëfficiënten: sinusoiden. Deze curve is de omhullende curve in het frequentiespectrogram. De maxima van deze curve geven aan welke frequenties de formanten zijn.

In het volgende hoofdstuk zal stap voor stap worden uitgelegd hoe wij bij de verwerking van onze metingen gebruik maakten van bovenstaande methoden. Ook zal de methode die men toepast in spraakherkenners uitvoerig worden uitgelegd.

Hoofdstuk 2

Hoe kunnen klanken herkend en vastgelegd worden, op basis van kenmerken die voor elk mens ongeveer identiek zijn?

In dit hoofdstuk zal allereerst worden getoond welke stappen wij hebben doorlopen voor het vinden van de Formanten van diverse klinkers; oftewel, hoe kunnen wij de kenmerken die voor elk mens min of meer identiek zijn vinden.

Vervolgens zal ook ontwikkelde spraakherkenning worden behandeld.

Voor het vinden van de eerste drie Formanten maakten we gebruik van een computer met daarop Coach 5 geïnstalleerd. Deze computer was verbonden met een Coachlab. Hierop was een geluidsensor aangesloten. In Coach 5 is gekozen voor de optie 'meten en analyseren'.

Telkens werd een klinker 'gezongen' dit zorgt voort een continu signaal.

De geluidsensor zet de verandering van de luchtdruk om in een wisselende spanning. De geluidsensor is zo ingesteld dat de spanning varieert tussen de min en de plus 0,5 volt. Dit is een analoge signaal. Dit analoge signaal wordt gesampled: omdat het niet mogelijk is de amplitude van het signaal constant te meten, zullen we deze amplitude om de zoveel tijd moeten meten. Wij hebben steeds gemeten gedurende een tijdsinterval van 500 milliseconden. De samplefrequentie bedroeg 10.000. Dit betekent dat de amplitude van het analoge signaal 10.000 keer per seconde werd gemeten.

Het is belangrijk dat de samplefrequentie minimaal twee keer zo groot is als de hoogste frequentie in het signaal, anders kunnen belangrijke eigenschappen van het signaal missen.

Vervolgens moet het signaal gedigitaliseerd worden, omdat computers enkel met digitale signalen kunnen werken.

Nu hebben we op het scherm een spanning-tijd diagram. Dit gaan we omzetten met Fourier Transformatie. Er wordt geprobeerd met een numerieke methode de sinusoiden waaruit de trilling is opgebouwd te berekenen. Hieruit kan dan de Intensiteit van elke frequentie worden berekend. Dit wordt weergegeven in een frequentiespectrogram.

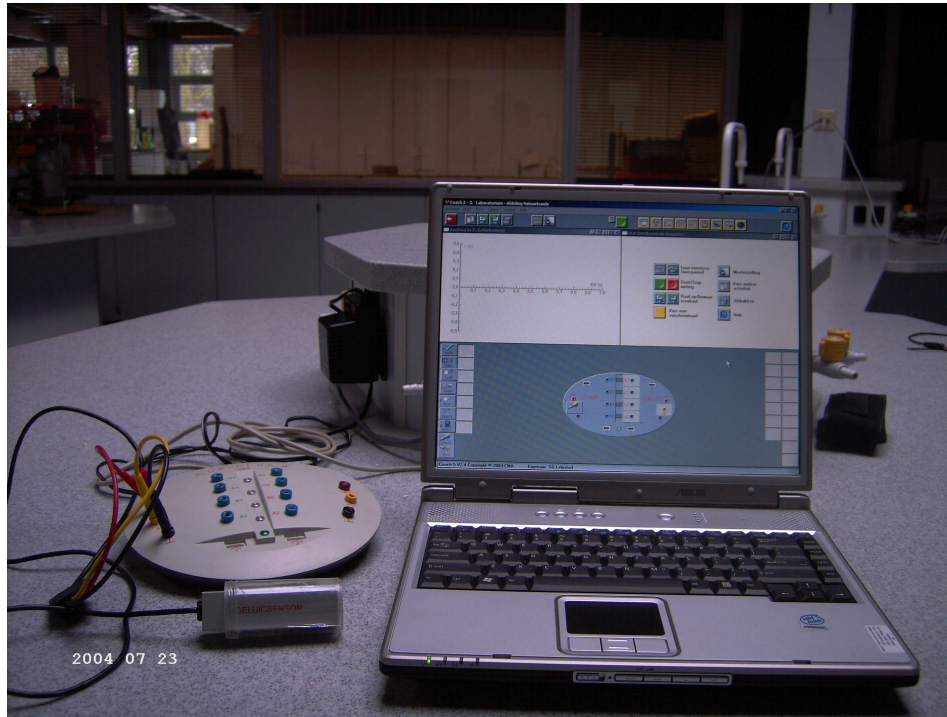
Vervolgens willen we de functie van de omhullende curve berekenen. Dit doen we met lineaire predictie. Deze functie wordt samengesteld op basis van een aantal vooraf gemeten punten en opgebouwd uit een aantal coëfficiënten: sinusoiden. De waarde van deze functie voor een bepaalde frequentie geeft aan of deze frequentie sterk, of minder sterk wordt doorgelaten.

Nu we deze omhullende curve in het frequentiespectrogram hebben staan kunnen we, na eventueel inzoomen, de maxima van de curve bepalen, met de functie 'lees uit'. De bijbehorende frequenties komen het sterkst door en zijn de Formanten van die klank. Bij de klinkers was het vooral belangrijk om de eerste drie Formanten te ontdekken: deze karakteriseren de klank, en zijn voor elk mens min of meer identiek.

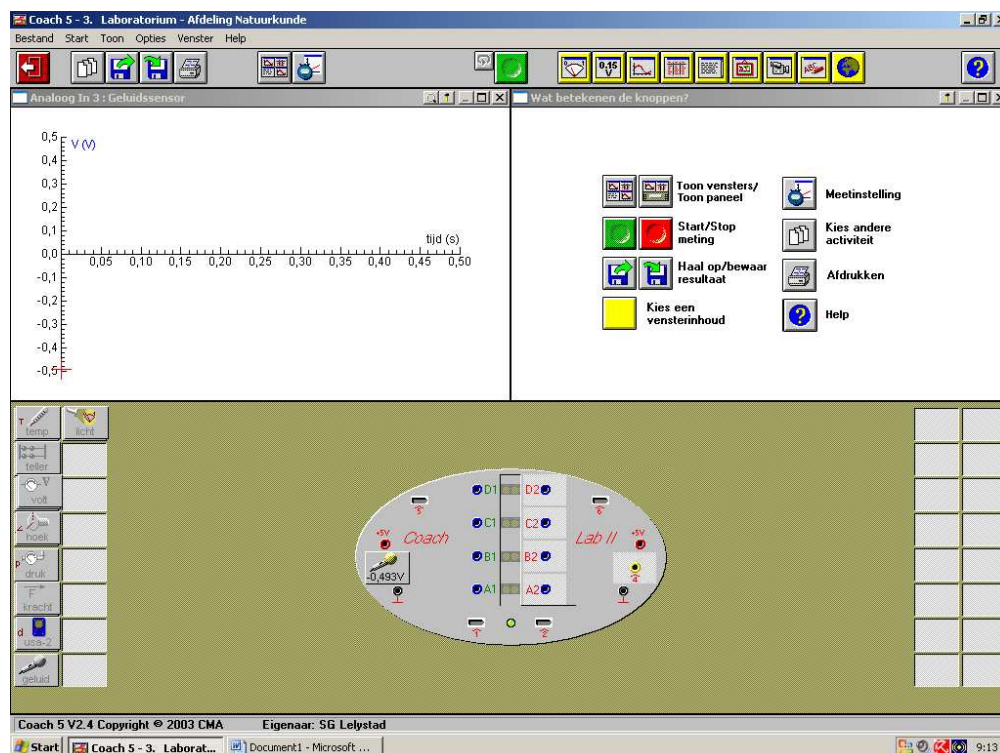
Op de volgende pagina zijn alle stappen visueel weergegeven.

De analyse van een willekeurige klinker.

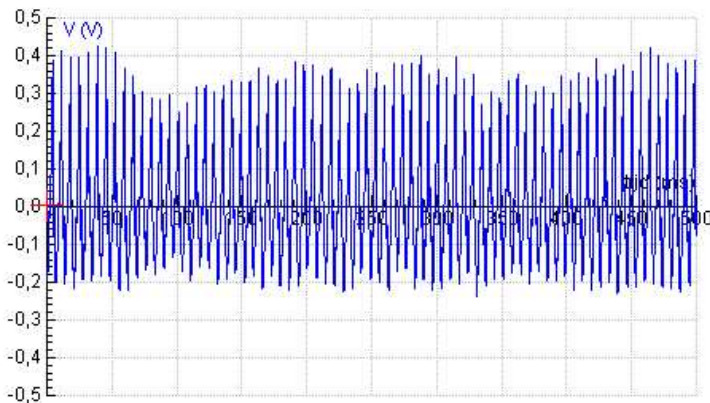
- 1 Meetopstelling
1: De computer met daarop geïnstalleerd Coach 5
2: Coachlab
3: geluidsensor



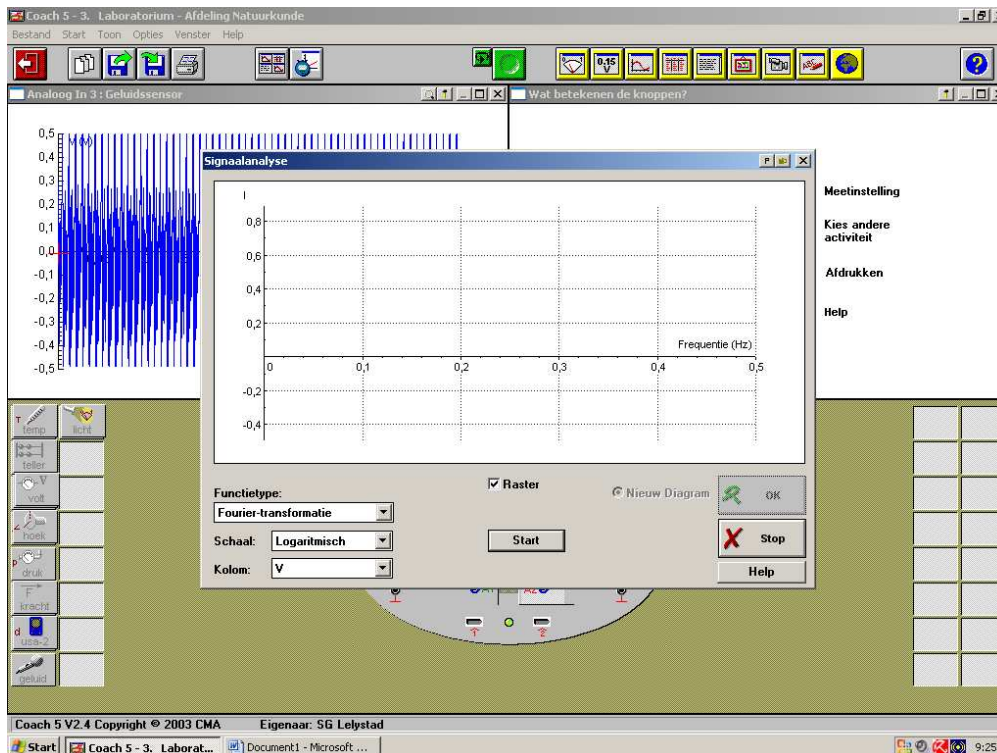
- 2 Het beginscherm van Coach. Gekozen is voor 'meten en analyseren'. Links boven staat het lege U,t -diagram, onder staat het Coach Lab II weergegeven, met daarop aangesloten een geluidsensor



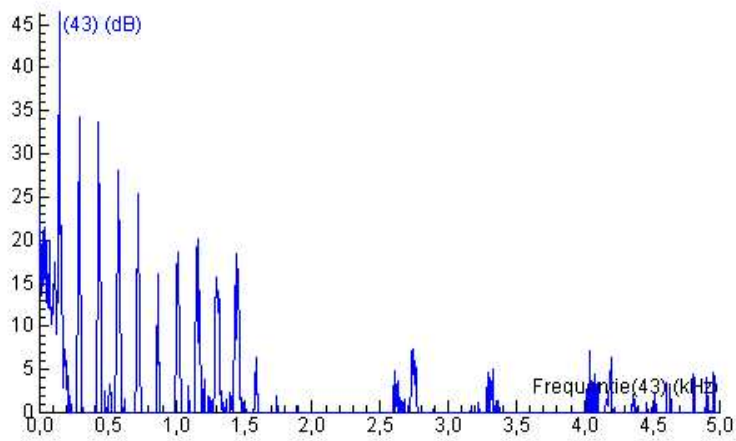
- 3 Het spanning-tijd diagram. Het interval waarin is gemeten bedraagt 500 ms. De spanning varieert tussen de plus en de min 0,5 volt.



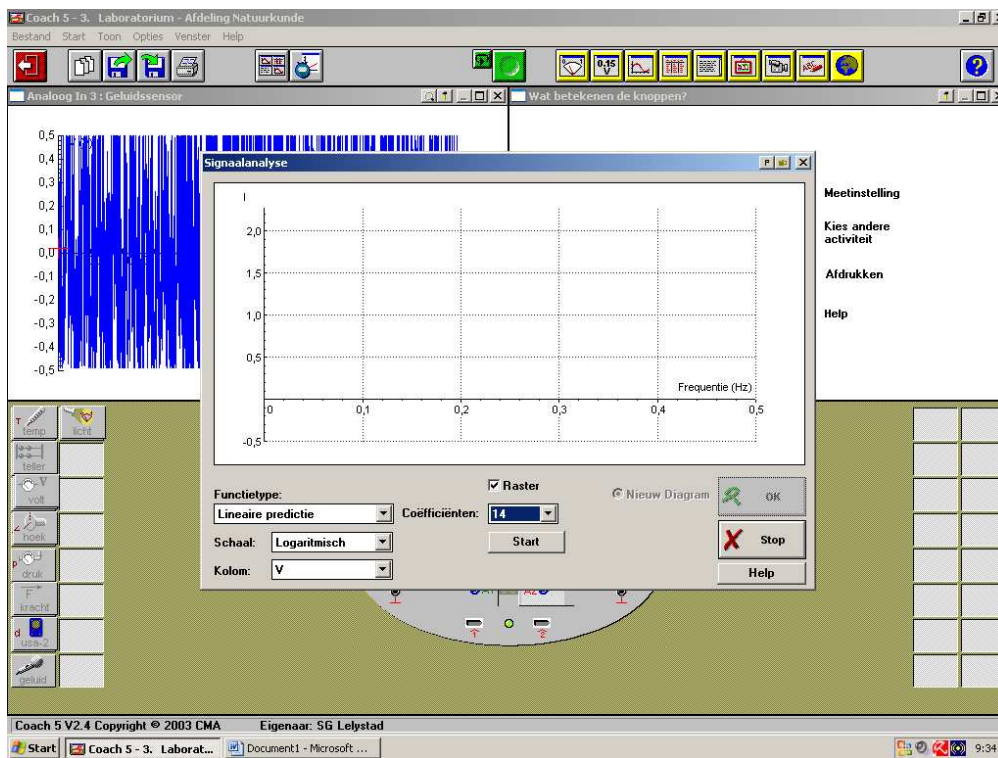
- 4 Het scherm dat verschijnt als je Fourier Analyse wil uitvoeren.(na invoer) Hier kom je door op het diagram te klikken en te kiezen voor 'analyseren, geluidsanalyse'



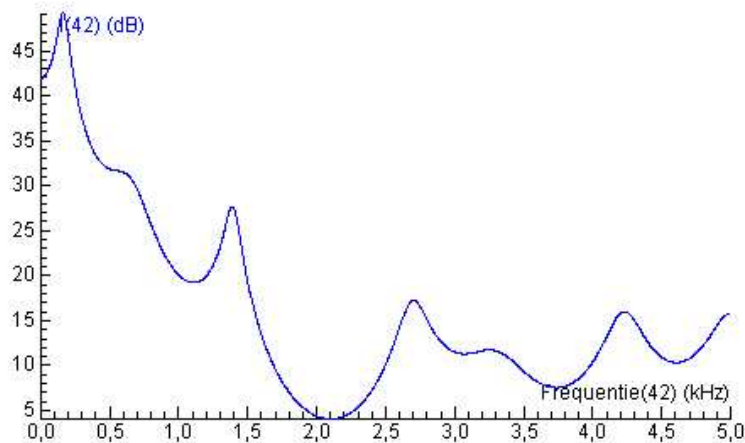
5 Het frequentiespectrum.



6 Het scherm dat verschijnt als je Lineaire predictie wil uitvoeren.(na invoer)



7 De omhullende curve, die ontstaat na het uitvoeren van lineaire predictie



Door vervolgens voor 'lees uit' te kiezen, kan je makkelijk de coördinaten aflezen, oftewel de frequentie en de intensiteit. Door dit te doen voor de maxima, kan je de frequentie en intensiteit van de formanten vinden.

We hebben nu dus de eerste drie formanten gevonden. De gevonden waarden kunnen gebruikt worden als vergelijkingsmateriaal. Een onbekende klank, in dit geval een klinker, kan na analyse worden vergeleken met met de klinkers die wel bekend zijn. De formanten van de onbekende klinker worden vergeleken met de Formanten van de bekende klinkers. Als deze Formanten overeen komen, kan worden bepaald welke klinker het is.

Het mag duidelijk zijn dat dit in de praktijk niet zo gaat. De manier waarop computers spraak herkennen wordt in de volgende paragraaf besproken. In Hoofdstuk 3 zullen we verder op de experimenteel verkregen meetgegevens ingaan.

Spraakherkenning met gebruik van Computers

Bij spraakherkenning wordt wederom spraak in een wisselende spanning omgezet met behulp van een microfoon of geluidsensoren. Dit analoge signaal wordt gesampled. Daarbij moet aan eerder genoemde voorwaarden worden voldaan. De verkregen analoge waarden moeten worden omgezet in digitale waarden, in verband met het gebruik van de computer.

Dit signaal gaat eerst door een filter, waardoor erg lage en erg hoge frequenties extra energie krijgen; dit gebeurt in het menselijk oor namelijk ook.

Vervolgens wordt er Fourier Transformatie uitgevoerd. Hiervoor wordt om de 10 milliseconden een stuk van 25 ms van het signaal geanalyseerd door een aantal filters. Dit houdt in dat de stukken elkaar overlappen, wat voor een grotere zekerheid zorgt.

Met een Fourier Transformatie wordt het spraaksignaal omgezet in verschillende sinusoiden. Hierdoor kan het signaal gescheiden worden in de verschillende frequenties waaruit het is opgebouwd.

Hier kan vervolgens een frequentiespectrogram van worden gemaakt.

Met behulp van deze frequentiespectrogrammen worden klanken herkend.

De Nederlandse taal heeft 37 klanken. Voor elk van deze klanken wordt de kans berekend dat ze overeenkomen met een deel van het spraaksignaal. De klank die het meest overeenkomt met het betreffende deel van het spraaksignaal zal dan worden gekozen om dat deel van het signaal te representeren.

De formule waarmee deze kans wordt berekend luidt:

$$P(S=K) = \frac{P(K=S) \times P(K)}{P(S)}$$

$P(S=K)$ is de kans dat het betreffende spraaksignaal die klank representeert.

$P(K=S)$ is de kans dat die klank wordt gerepresenteerd door dat spraaksignaal.

$P(K)$ is de kans op die klank.

$P(S)$ is de kans op dat spraaksignaal.

Oftewel de kans dat het spraaksignaal die klank (bijvoorbeeld een 'oo') representeert is gelijk aan de kans dat die klank wordt gerepresenteerd door dat spraaksignaal, maal de kans op die klank, gedeeld door de kans op dat spraaksignaal.

Hierin is de kans op de betreffende klank bekend, evenals de kans op het betreffende spraaksignaal. De kans op een 'e' is bijvoorbeeld groter dan de kans op een 'oe' de grootte van deze kans is al vastgelegd. Hetzelfde geldt dus voor spraaksignalen. Ook daarvan weten we hoe vaak ze gemiddeld voorkomen.

De kans dat een klank wordt gerepresenteerd door het betreffende spraaksignaal wordt berekend door vergelijking van de frequentiespectrogrammen van deze klank en het spraaksignaal.

Hieruit volgt dat hoe meer de frequentiespectrogrammen overeenkomen, des te groter is de kans dat het spraaksignaal die klank representeert. Ook als de kans op die klank groter is, neemt de kans dat het spraaksignaal die klank representeert toe. Als de kans op het spraaksignaal echter toeneemt, dan neemt de kans dat dit spraaksignaal de klank representeert af.

Dit is makkelijk uit te leggen: als de kans op zowel die klank ongeveer evengroot is als de kans op dat spraaksignaal, dan zal $P(K=S)$ bijna geheel uitmaken hoe groot de kans is dat dat spraaksignaal die klank representeert. Als de kans op een klank toeneemt, terwijl de kans op dat spraaksignaal gelijk blijft, of afneemt, dan neemt $P(S=K)$ ook toe. Het spraaksignaal zal dan immers eerder die klank representeren dan een andere klank.

Als de kans op dat spraaksignaal toeneemt, terwijl de kans op die klank gelijkblijft of zelfs afneemt, zal $P(S=K)$ afnemen; het spraaksignaal zal namelijk minder snel die klank vertegenwoordigen.

De woorden die de spraakherkenner kan herkennen zijn opgeslagen als een rij klanken. De rij herkende klanken wordt vergeleken met de woorden in de 'vocabulaire', zoals dat heet, en zo worden de woorden herkend. Aangezien we de woordgrenzen niet weten, zijn er verschillende woorden mogelijk. Daarom maakt de spraakherkenner altijd een aantal mogelijke zinnen. Om de juiste zin te selecteren, wordt gebruik gemaakt van een taalmodel. Hierin staat welke woorden waarschijnlijk op elkaar zullen volgen. Dit maakt de spraakherkenner erg betrouwbaar.

Hoofdstuk 3

Welke invloed hebben de verschillende variabelen, die van toepassing zijn op geluidstrillingen, op de herkenbaarheid van een uitgesproken klinker? Deze invloed wordt voor de diverse klinkers door middel van experiment gemeten.

Hierbij kijken we vooral naar de Formanten, deze zijn immers kenmerkend voor een klinker.

We hebben de formanten bepaald op de manier die beschreven staat in Hoofdstuk 2
We hebben de formanten bepaald voor de volgende klinkers:

‘e’, ‘ee’, ‘a’, ‘aa’, ‘o’, ‘oo’, ‘u’, ‘uu’, ‘i’, ‘ie’, ‘ui’, ‘oe’, ‘au’ en ‘ei’.

Dit hebben we gedaan voor zowel de mannelijke als de vrouwelijke stem.

Beide proefpersonen waren 17 jaar.

Elke meting is dubbel uitgevoerd.

Telkens is er gemeten gedurende een tijdsinterval van 500 milliseconden. De samplefrequentie bedroeg 10.000. Elke klinker werd ‘gezongen’, wat ervoor zorgt dat het opgenomen signaal continu is.

Hieronder staan de meetgegevens.

Bij elke klinker staat M1, M2, V1 en V2.

M1 staat voor de eerste meting aan de mannelijke stem, M2 staat voor de tweede meting. V1 staat voor de eerste meting aan de vrouwelijke stem, V2 staat voor de tweede meting.

Onder elke formant staat de frequentie en de geluidsintensiteit voor die frequentie. De frequentie is uitgedrukt in kiloHertz (kHz). De geluidsintensiteit is uitgedrukt in decibel (dB) .

E

Freq(kHz) – I(dB)	Formant 1	Formant 2	Formant 3
M1	0,63 – 47,76	1,72 – 40,71	2,39 – 28,68
M2	0,62 – 48,51	1,71 – 40,74	2,39 – 32,58
V1	0,81 – 53,79	1,80 – 43,52	3,09 – 38,08
V2	0,82 – 51,29	1,90 – 39,63	3,05 – 33,96

EE

Freq (kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,46 – 50,36	1,94 – 37,36	3,68 – 37,11
M2	0,44 – 47,70	1,88 – 39,85	3,59 – 36,43
V1	0,43 – 47,17	2,59 – 34,35	3,24 – 31,36
V2	0,46 – 49,07	2,48 – 40,05	3,14 – 37,33

A

Freq(kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,67 – 48,06	1,04 – 45,49	3,53 – 27,68
M2	0,65 – 46,65	1,08 – 44,50	3,45 – 26,26
V1	0,27- 43,00	1,06 – 42,29	3,74 – 11,32
V2	0,30 – 39,17	1,07 – 41,79	3,80 – 19,41

AA

Freq(kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,76 – 48,94	1,29 – 42,48	2,60 – 36,90
M2	0,70 – 47,87	1,27 – 46,31	2,53 – 36,55
V1	0,81 – 47,97	1,33 – 42,38	3,80 – 32,64
V2	0,78 – 47,57	1,28 – 41,78	3,80 – 35,52

O

Freq(kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,41 – 55,61	3,13 – 27,87	2,34 – 25,40
M2	0,41 – 54,00	3,23 – 23,83	2,33 – 21,18
V1	0,29 – 42,70	3,77 – 21,30	3,07 – 11,24
V2	0,25 – 45,11	3,75 – 26,66	3,14 – 18,89

OO

Freq(kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,49 – 50,61	0,98 – 44,01	2,28 – 28,77
M2	0,49 – 50,41	1,00 – 43,64	2,21 – 28,30
V1	0,52 – 51,23	1,04 – 39,63	3,00 – 4,81
V2	0,51 – 56,82	1,03 – 47,07	3,03 – 16,65

U

Freq(kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,40 – 50,28	1,41 – 27,60	3,54 – 37,31
M2	0,40 – 48,81	1,40 – 33,10	3,45 – 31,74
V1	0,46 – 40,38	1,77 – 28,90	2,83 – 25,21
V2	0,46 – 43,08	1,62 – 26,31	2,91 – 21,23

UU

Freq(kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,29 – 45,96	1,90 – 34,54	3,27 – 33,48
M2	0,30 – 50,60	1,98 – 25,16	3,36 – 24,70
V1	0,28 – 25,61	1,91 – 33,96	3,85 – 32,70
V2	0,25 – 43,44	1,90 – 29,48	3,87 – 29,30

I

Freq(kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,38 - 55,63	1,91 – 39,10	2,53 – 30,34
M2	0,38 – 59,07	1,93 – 40,19	2,53 – 32,59
V1	0,32 – 40,57	1,76 – 22,71	2,72 – 17,63
V2	0,33 – 56,72	1,88 – 31,28	2,59 – 28,99

IE

Freq(kHz) - I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,24 – 46,36	3,60 – 33,92	2,10 – 33,43
M2	0,29 – 43,93	3,63 – 31,06	2,11 – 30,24
V1	0,30 – 69,45	2,40 – 35,09	3,73 – 23,60
V2	0,29 – 67,17	2,53 – 32,01	3,73 – 26,50

UI

Freq(kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,27 – 52,20	2,12 – 31,43	1,69 – 23,82
M2	0,30 – 54,76	2,16 – 35,72	1,62 – 33,78
V1	0,29 – 50,00	1,83 – 35,67	2,70 – 25,65
V2	0,28 – 49,68	1,82 – 36,66	2,68 – 22,24

OE

Freq(kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,30 – 58,71	0,76 – 40,88	3,60 – 22,36
M2	0,32 – 56,34	0,78 – 31,19	3,68 – 28,52
V1	0,31 – 52,53	0,73 – 57,58	2,72 – 12,24
V2	0,29 – 48,18	0,75 – 51,00	2,76 – 10,66

AU

Freq(kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,75 – 47,79	1,23 – 40,09	3,56 – 26,33
M2	0,76 – 44,38	1,20 – 43,00	3,40 – 25,90
V1	0,70 – 49,50	1,36 – 43,77	2,85 – 20,75
V2	0,70 – 52,09	1,39 – 43,55	2,81 – 27,64

EI

Freq(kHz) – I(dB)	Formant 1(Hz)	Formant 2(Hz)	Formant 3(Hz)
M1	0,73 – 46,70	1,53 – 36,95	3,72 – 26,26
M2	0,70 – 48,11	1,52 – 36,62	3,51 – 30,12
V1	0,71 – 51,82	1,80 – 37,74	2,96 – 22,56
V2	0,75 – 50,87	1,79 – 43,58	2,91 – 40,76

Om verder handiger te kunnen werken, nemen we voortaan het gemiddelde van de eerste en de tweede meting. De letter wordt gevolgd door het geslacht van de spreker. Daarachter staat de frequentie (kHz) en de geluidsintensiteit (dB) per formant.

Klinker - Sexe	Formant 1	Formant 2	Formant 3
E – M	0,63 – 48,14	1,72 – 40,73	2,39 – 30,63
E – V	0,82 – 52,54	1,85 – 41,58	3,07 – 36,02
EE – M	0,45 – 49,03	1,91 – 38,61	3,64 – 36,77
EE – V	0,45 – 48,12	2,54 – 37,20	3,19 – 34,35
A – M	0,66 – 47,36	1,06 – 45,00	3,49 – 26,97
A – V	0,29 – 41,09	1,07 – 42,04	3,77 – 15,37
AA – M	0,73 – 48,41	1,28 – 44,40	2,57 – 36,73
AA – V	0,80 – 47,77	1,31 – 42,08	3,80 – 34,08
O – M	0,41 – 54,81	3,18 – 25,85	2,34 – 23,29
O – V	0,27 – 43,91	3,76 – 23,98	3,11 – 15,07
OO – M	0,49 – 50,51	0,99 – 43,83	2,25 – 28,54
OO – V	0,52 – 54,03	1,04 – 43,35	3,02 – 10,73
U – M	0,40 – 49,55	1,41 – 35,35	3,50 – 34,53
U – V	0,46 – 41,73	1,70 – 27,61	2,87 – 23,22
UU – M	0,30 – 48,28	1,94 – 29,85	3,32 – 29,09
UU – V	0,27 – 34,53	1,91 – 31,72	3,86 – 31,00
I – M	0,38 – 57,35	1,92 – 39,65	2,53 – 31,47
I – V	0,33 – 48,65	1,82 – 27,00	2,66 – 23,31
IE – M	0,27 – 45,15	3,62 – 32,49	2,11 – 31,84
IE – V	0,30 – 68,31	2,47 – 33,55	3,73 – 25,05
UI – M	0,29 – 53,48	2,14 – 33,58	1,66 – 28,80
UI – V	0,29 – 49,84	1,83 – 26,17	2,69 – 23,95
OE – M	0,31 – 57,53	0,77 – 36,04	3,64 – 25,44
OE – V	0,30 – 50,36	0,74 – 54,29	2,74 – 11,45
AU – M	0,76 – 46,09	1,22 – 41,55	3,48 – 26,12
AU – V	0,70 – 50,80	1,38 – 43,66	2,83 – 24,20
EI – M	0,72 – 47,41	1,53 – 36,79	3,62 – 28,19
EI – V	0,73 – 51,35	1,80 – 40,66	2,98 – 31,66

Na uitvoerige vergelijking kunnen we vaststellen in hoeverre de klinkers overeenkomen. Hierbij moet worden gelet op de volgende aanname: Twee formanten zijn gelijk, als het frequentieverschil niet groter dan 100 hertz is. Als we naar de tabel waarin de dubbele metingen staan (M1, M2 en V1, V2) kijken, kunnen we zien dat het verschil in frequentie van de formanten niet groter is dan 100 Hz. Meestal ligt dit verschil rond de 30 Hz. De kans dat twee Formanten van dezelfde orde (1^e, 2^e, 3^e.... formant) van dezelfde klinker zijn als het verschil in frequentie van deze twee formanten 100 Hz of groter is, is erg klein. Daarom is gesteld dat twee formanten ongelijk zijn aan elkaar als het verschil in frequentie tussen twee formanten groter is dan 100 Hz, oftewel 0,1 kHz.

Na vergelijking van de 1^e Formanten bleek dat de eerste Formant van elke klinker overeenkwam met de eerste Formant van een aantal andere klinkers. Dit aantal andere klinkers verschilde van vier tot veertien.

Dit betekent dus concreet dat we een gegeven klinker niet kunnen herkennen op basis van de waarde van de eerste Formant, aangezien deze waarde gelijk is aan 4 tot 14 waarden van de eerste Formant van andere klinkers.

Vervolgens gingen we voor elke klinker de tweede Formant vergelijken met de tweede Formant van die klinkers, waarvan de eerste Formant overeenkwam met de eerste Formant van de gekozen klinker. Oftewel; is er geen onderscheid te maken op basis van de eerste Formant, probeer dan onderscheid te maken op basis van de tweede Formant.

Als we alle klinkers in bovenstaande tabel van boven naar beneden nummeren van 1 tot 28, dan waren ook de tweede Formanten van klinkers 1, 2, 3, 7, 8, 11, 12, 15, 16, 17, 18, 22, 23, 24, 25, 26 en 28 overeenkomstig met de tweede Formant van een aantal andere klinkers. Voor deze klinkers geldt dus, dat zowel de eerste als de tweede Formant overeenkomen met de eerste en tweede Formanten van een aantal andere klinkers. Dit aantal verschilde van 1 tot 5 andere klinkers.

Voor de klinkers, waarvan zowel de eerste als de tweede Formant overeenkomen, kijken we naar de derde Formant. De derde Formant van een bepaalde klinker vergelijken we met de derde Formant van die klinkers, waarvan zowel de eerste als de tweede Formant gelijk is aan de eerste respectievelijk de tweede Formant van de klinker die we onderzoeken.

Uiteindelijk blijven er nog een paar klinkers over waarvan alle Formanten gelijk worden gerekend aan de Formanten van een aantal andere klinkers.

$E - V$ komt overeen met $E_i - V$

$I - V$ komt overeen met $U_i - V$

Dit is wat we te weten komen als we enkel naar de waarde van de Frequentie bij de Formanten kijken. Als we naar de verhoudingen van de Intensiteit van deze Frequenties kijken kunnen we helaas over $E - V$ en $E_i - V$ niet veel te weten komen: De intensiteit van de frequenties bij beide klinkers is steeds ongeveer even sterk. Ook bij $I - V$ en $U_i - V$ kunnen we geen onderscheid maken door te letten op het verschil in intensiteit van de frequenties: ook hier is de intensiteit van de frequenties bij beide klinkers steeds ongeveer even sterk.

In het echt kan dit opgelost worden door het gebruik van precieze apparatuur en het zorgvuldig bepalen van de frequenties en intensiteit van de Formanten. Nu is deze bepaald door het gemiddelde van twee metingen te nemen. Dit kan preciezer worden berekend door het gemiddelde van een groot aantal proefpersonen te nemen. De kans dat frequenties afwijken van deze gemiddelde frequenties is dan erg klein. Ook is het zo dat twee personen (slechts een voor elke sexe) niet erg representatief zijn voor de bevolking.

Een meting met bij een grote groep met verschillende leeftijd zal dus een beter en preciezer resultaat leveren, waarbij het onderscheid tussen alle klinkers wél duidelijk zichtbaar is.

In bijlage (1) staan de kolommen die werden gebruikt bij het vergelijken van de Formanten van de klinkers.

De invloeden van de variabelen

We kunnen een klinker herkennen op basis van de frequentiewaarden en de Intensiteit van de Formanten. De eerste drie Formanten karakteriseren de klinker, de overige Formanten zijn persoonlijk. Het gaat ons dus om de frequentiewaarden en de Intensiteit van de eerste 3 Formanten. Deze zijn afhankelijk van een aantal factoren.

De stembanden produceren een grondtoon en zijn boventonen bij het vormen van een klank. Dit kan al verschil aanbrengen in de frequenties bij mannen en bij vrouwen. Vervolgens worden deze trillingen versterkt: keel en mond treden als resonantiebuï op. De mond, het gehemelte en de tong werken echter ook als een filter. Deze dempen sommige frequenties meer dan andere. Welke frequenties gedempt worden is afhankelijk van de positie van de tong en de vorm van mond en gehemelte.

Uiteindelijk zijn er een aantal frequenties die het sterkst doorkomen, dit zijn de genoemde Formanten.

De opgestelde hypothese luidde:

De verschillen tussen de waarden van de diverse variabelen en de grootten van deze waarden, geeft het onderscheid tussen de diverse klanken aan. Door deze waarden te analyseren kan de desbetreffende klank worden herkend en vastgelegd.

Zoals hierboven uitgelegt, kan een klank herkend worden op basis van de Formanten. De grootten van, en de verschillen tussen de frequentiewaarden en de intensiteit van de waarden geeft het onderscheid tussen de diverse klanken aan. Deze frequentiewaarden en de intensiteit van deze Formanten is afhankelijk van een aantal variabelen en een aantal factoren. Hieronder verstaan we onder andere de intensiteit van grondtoon en boventonen, en de werking van het filter, oftewel de positie van de tong in de mond, en de vorm van mond en gehemelte.

Met gebruik van Fourier Analyse en Lineaire predictie kunnen we tonen analyseren en de verschillende waarden die van belang zijn (frequentie, Intensiteit) ontdekken en deze gebruiken om de desbetreffende klank vast te leggen. De bovenstaande hypothese zit dus bijna goed, maar mist de juiste formulatie.

We kunnen concluderen dat:

De verschillen tussen de waarden van de diverse variabelen en de grootten van deze waarden, geven het onderscheid tussen de diverse klanken aan. Door klanken te analyseren kunnen belangrijke gegevens worden ontdekt. Deze gegevens kunnen worden gebruikt om een klank te herkennen en vast te leggen.

Bijlage (1)

Hieronder staan de kolommen die ik gebruikte om de verschillen tussen de Formanten te onderzoeken. In deze kolommen staat elk cijfer voor een klinker in de tabel, zoals staat aangegeven in Hoofdstuk 3. De letters achter elk cijfer staan voor de klinker waarvan de Formant overeenkomt met de Formant van de klinker waar het cijfer voor staat.

Bijvoorbeeld:

1 Aav

staat voor: de 'Nde' formant van de vrouwelijke 'Aa' komt overeen met de 'Nde' formant van de mannelijke 'e'. (E – M)

Dit is de lijst

Eerste formant

1	Am	Aam	Auv	Eim	Eiv							
2	Aam	Aav	Aum	Eim	Eiv							
3	Eev	Om	Oom	Oov	Um	Uv	Im					
4	Eem	Om	Oom	Oov	Um	Uv	Im					
5	Em	Aam	Aum	Auv	Eim	Eiv						
6	Ov	Uum	Uuv	Im	Iv	lem	lev	Uim	Uiv	Oem	Oev	
7	Em	Ev	Am	Aav	Aum	Auv	Eiv					
8	Ev	Aam	Aum	Auv	Eim	Eiv						
9	Eem	Eev	Om	Oom	Um	Uv	Im	Iv	Oem			
10	Av	Uum	Uuv	Iv	lem	lev	Uim	Uiv	Oem	Oev		
11	Eem	Eev	Om	Oov	Um	Uv						
12	Eem	Eev	Oom	Uv								
13	Eem	Eev	Om	Oom	Uv	Uum	Im	Iv	lev	Oem	Oev	
14	Eem	Eev	Om	Oom	Oov	Um	Im					
15	Av	Ov	Um	Uuv	Im	Iv	lem	lev	Uim	Uiv	Oem	Oev
16	Av	Ov	Uum	Iv	lem	lev	Uim	Uiv	Oem	Oev		
17	Eem	Eev	Av	Om	Um	Uv	Uum	Im, Iv, lev, Uim, Uiv, Oem, Oev				
18	Av	Om	Ov	Um	Uum	Uuv	Im, Iv, lem, lev, Uim, Uiv, Oem, Oev					
19	Av	Ov	Uum	Uuv	Iv	lem	lev	Uim	Uiv	Oem	Oev	
20	Av	Ov	Um	Uum	Uuv	Im	Iv, lem, Uim, Uiv, Pem, Oev					
21	Av	Ov	Uum	Uuv	Im	Iv	lem, lev, Uiv, Oem, Oev					
22	Av	Ov	Uum	Uuv	Im	Iv	lem, lev, Uim, Uiv, Oem, Oev					
23	Av	Om	Ov	Um	Uum	Uuv	Im, Iv, lem, lev, Uim, Oev					
24	Av	Ov	Um	Uum	Uuv	Im	Iv, lem, lev, Uim, Uiv, Oem					
25	Ev	Am	Aam	Aav	Auv	Eim	Eiv					
26	Em	Am	Aam	Aav	Aum	Eim	Eiv					
27	Em	Ev	Am	Aam	Aav	Aum	Auv	Eiv				
28	Em	Ev	Am	Aam	Aav	Aum	Auv	Eim				

Eerste en Tweede Formant

1	Eiv					
2	Eiv					
3	Im					
4	-					
5	-					
6	-					
7	Aav	Auv				
8	Aam	Aum	Auv			
9	-					
10	-					
11	Oov					
12	Oom					
13	-					
14	-					
15	Uuv	Im				
16	Uum	Iev	Uiv			
17	Eev	Eem	Uum	Iv	Uiv	
18	Uuv, Im, Uiv					
19	-					
20	-					
21	-					
22	Uuv	Im	Iv			
23	Oev					
24	Oem					
25	Aam	Aav				
26	Aam	Aav				
27	-					
28	Em	Ev				

Eerste, Tweede en Derde Formant

2	Eiv
18	Uiv
22	Iv
28	Ev

Waaruit volgt dat alleen van klinkers '2, 18, 22, 28' alle drie de Formanten met alle drie de Formanten van een andere klinker overeenkomen.

Bronnenlijst

- 1 www.studeren.uva.nl/informatiekunde/object.cfm/objectID=E10CA58E-8298-4834-850B687480B3F24F

Hier is de informatie over het profielwerkstuk gegeven. Er staat een link op die ik heb gebruikt. De link luidt "aanvullende informatie" en opent een wordbestand.

- 2 <http://mediatheek.thinkquest.nl/~llb356/pagina.php?taal=nl&ow=3&pag=5>

Hier is erg bondig alle informatie over spraakherkenning te vinden.

- 3 www.cma.science.uva.nl/support/webhelp/c5/c5hlp/diagram_signal_analysis.htm

De tekst op deze site is geschreven door de heer Molenaar, tevens schrijver van het profielwerkstuk en mijn begeleider van de UvA

Behalve informatie van deze 3 internet bronnen, heb ik ook een groot aantal vragen gesteld aan de heer Molenaar met gebruik van e-mail.

Verder is ook Lesboek deel 1B van Natuurkunde 1 gebruikt voor het opzoeken van enige begrippen en Formules.

Tot slot moet ook Binas worden vermeld. Tabel 98 geeft informatie over Formanten. Deze informatie is echter niet gebruikt bij het schrijven van dit profielwerkstuk en heeft enkel ter ondersteuning gediend.

Nawoord

Ik heb met veel plezier aan dit profielwerkstuk gewerkt. Toen ik een keuze voor een onderwerp moest maken sprak Spraakherkenning me al erg aan, en dat is er niet minder op geworden. Dat ik het onderwerp interessant vond, heeft ervoor gezorgd dat ik een sterke drang had om alles te begrijpen.

Ik wil allereerst de heer Molenaar bedanken voor zijn hulp; zowel door de assistentie bij de uitvoering van het experiment bij de Faculteit der Natuurwetenschappen, Wiskunde en Informatica, als door het beantwoorden van mijn vragen heeft hij mij zeer geholpen.

Verder wil ik Sanne en Ellen uit Leiden bedanken voor de samenwerking bij het experiment. Door deze samenwerking is de hoeveelheid meetgegevens erg uitgebreid geworden.

Al met al ben ik er erg tevreden over en ben ik blij dat ik over dit onderwerp meer heb geleerd door het maken van dit Profielwerkstuk.