

“De trein naar Alkmáár en Àlkmaar”

Onderzoek naar de intonatie
van opsommingen in het
Nederlands

Naam	Sheean Spoel
Studentnummer	0556688
Begeleider	Paul Boersma
Module	Afstudeerscriptie Bachelor Taalwetenschap Universiteit van Amsterdam
Datum	13 augustus 2013

Samenvatting

Voor dit onderzoek zijn intonatiecontouren van Nederlandse opsommingen opgenomen en met behulp van een functionele data-analyse vergeleken. Bij deze analyse is gebruik gemaakt van een functionele principale-componentenanalyse om aan de contouren scores toe te wijzen. Deze scores zijn vervolgens met behulp van gangbare statistiek vergeleken. Om de invloed van sprekers en woordverschillen te minimaliseren zijn de intonatiecontouren omgezet naar semitonen en met behulp van de amplitude en normering gelijk getrokken. Er werden effect gevonden op de intonatie op basis van de positie van een woord in de opsomming, de vorm van het woord en de lengte van de opsomming. Er werd een duidelijk intonatieverschil gevonden tussen het eerste woord bij een opsomming van lengte 2 en die bij langere opsommingen. Het uiteindelijk gevonden model zou kunnen worden ingezet voor geautomatiseerde manipulatie en generatie van realistische intonatiecontouren in opsommingen.

Steekwoorden: intonatie, fonetiek, Nederlands, opsommingen, functionele data-analyse, downstep, FPCA

Abstract

In this study intonation contours of Dutch enumerations were recorded and compared using functional data analysis. In this analysis a functional principal component analysis was performed to assign scores to the different contours. These scores were then compared using conventional statistics. To minimize the influence of speakers and word differences the intonation contours were converted to semitones and aligned using the amplitude and a norming. Effects were found on the intonation by position of a word in an enumeration, the phonological properties of the word and the length of enumeration. A clear difference in intonation was found for the first word between enumerations of 2 and more words. The eventually found model could be used for automated manipulation and generation of realistic intonation contours of enumerations.

Keywords: intonation, phonetics, Dutch, enumerations, functional data analysis, downstep, FPCA

Inhoudsopgave

Inleiding	4
Theoretisch kader	4
Intonatiecontouren	4
Definities van opsommingen	6
Functionele data-analyse	7
Methode	12
Materialen	13
Respondenten	16
Procedure	16
Resultaten	18
Conclusies	23
Discussie en vervolgonderzoek	24
Praktische zaken	25
Toepassingen en verder onderzoek	26
Nawoord	26
Referenties	26
Appendix	28
Woordenbank	28
Draagzinnen	29
Begrippen	29
Uitgesloten woordrealisaties	30

Het opsommen van woorden komt in het dagelijks leven voor, van reclames waarin de ene na de andere superlatief genoemd wordt (“het beste, goedkoopste én eenvoudigste koffiezetapparaat”) tot het herhalen van het ingetoetste nummer in een telefooncentrale. Deze opsommingen komen vaak voor in een geautomatiseerde context zoals in die telefooncentrale, bij een stationsomroep of halteafroep in een tram. Bij veel van deze systemen zijn de benodigde woorden opgenomen in plaats van gegenereerd door middel van spraaksynthese. Door de juiste woorden achter elkaar te zetten kan vervolgens iedere gewenste opsomming worden gegenereerd.

Afhankelijk van de plek in een opsomming wordt een woord anders uitgesproken; het simpelweg achterelkaar plakken van woorden levert daarom weinig natuurlijke resultaten op. Bij het bouwen van een dusdanig systeem zijn er dan de volgende mogelijkheden: genoeg nemen met het onnatuurlijke resultaat, dezelfde woorden meerdere malen opnemen met verschillende uitspraken of opnames manipuleren zodat deze goed klinken. Door een accuraat model te hebben van de fonetische eigenschappen van een opsomming zou het mogelijk kunnen worden om met slechts één opname per woord te volstaan. Het woord zou dan op de gewenste manier kunnen worden gemanipuleerd zodat het op iedere mogelijke manier in een opsomming te gebruiken is. In dit onderzoek zal gekeken worden naar een model in termen van een functionele principale componentenanalyse van de intonatie in een opsomming; dit model is gebaseerd op intonatiecontouren van natuurlijke spraak en kan worden ingezet voor het manipuleren van bestaande woordopnames.

Theoretisch kader

Intonatiecontouren

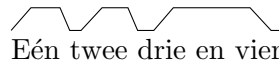
Volgens Caspers (2000) zijn er twee gangbare modellen van intonatie in het Nederlands: de Grammar of Dutch Intonation (GDI) van Van 't Hart & Collier (1975, geciteerd door Caspers) en het autosegmenteelmodel (Gussenhoven, 1991). Allebei zijn gekenmerkt door het opdelen van spraak in segmenten waarbij elk segment een bepaalde verandering (GDI) of hoogte (autosegmenteel) in intonatie weergeeft. Er wordt bij beide theorieën gebruik gemaakt van een vaste set van elementen die gecombineerd kunnen worden om intonatie te modelleren. GDI kent bijvoorbeeld vijf stijgende niveaus: 1-5 die gecombineerd kunnen worden met vijf dalende niveaus: A-E. Het autosegmenteelmodel (Goldsmith, 1976) werkt met meerdere niveau's: bijvoorbeeld fonemen, woorden en intonaties. Het aantal elementen in ieder niveau hoeft niet overeen te komen, het is bijvoorbeeld mogelijk dat een foneem of woord meerdere intonatie-elementen bevat. Een voorbeeld zoals door Goldsmith gegeven in het Sanskriet is:

$$\begin{array}{ccccccc}
 \text{b} & \text{a} & \text{l} & \text{i} & \text{a} & & \\
 & \swarrow & & \searrow & & & \\
 & \Sigma_1 & & \Sigma_{2,3} & & & \\
 & | & & \wedge & & & \\
 & \text{H} & & \text{L} & \text{H} & &
 \end{array}$$

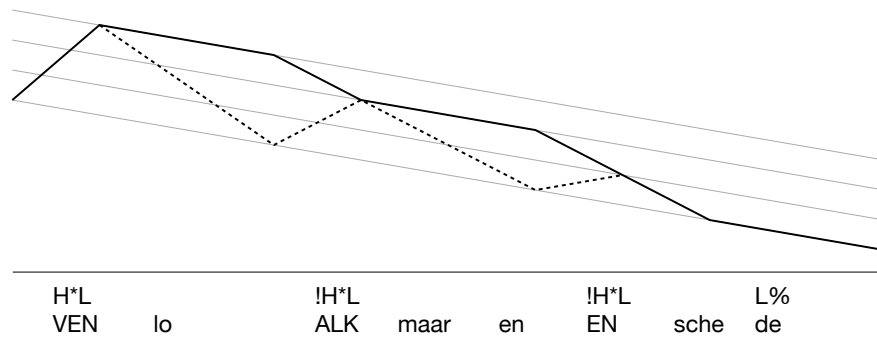
Hier zijn in de bovenste rij de fonemen van *bálà* te vinden, in de middelste rij de twee syllaben en in de onderste rij de twee hoge (H) tonen en één lage (L) toon. In het Nederlands is dit autosegmenteelmodel uiteindelijk verder ontwikkeld naar het Transcription of Dutch Intonation (ToDI) (Gussenhoven et al., 2005), in dit model wordt gebruik gemaakt van drie begintonen, drie eindtonen (waarvan er een het gebrek ervan is), vijf intonatieaccenten en twee modificaties (vertraagd of lager dan de vorige).

Keijsper (1984) heeft gekeken naar de betekenis van intonatiecontouren in het Nederlands, waaronder die bij de de intonatie van een opsomming. Zij heeft vastgesteld dat een dalende toon

een reeks sluit, terwijl bij een stijgende toon de reeks nog aan te vullen is. Hieruit kan worden verwacht dat bij een opsomming voor het eerste woord de toon een stijgende vorm heeft. Omdat de toon niet oneindig kan blijven stijgen wordt voor ieder nieuw woord weer vanaf beneden naar boven gestegen. Bij het laatste woord blijft de toon hoog en daalt deze vervolgens om de reeks af te sluiten. Dit wordt hieronder geïllustreerd:



Van Heuven (2004) heeft specifiek gekeken naar de planning van intonatie in opsommingen in het Nederlands, hier treedt downstep op (Gussenhoven et al., 2003). Een downstep is een modificatie in ToDI en wordt aangegeven met !H. Het staat voor een stijging die minder hoog is dan de direct voorafgaande hoge toon. Wanneer er meerdere downsteps achterelkaar worden gerealiseerd zullen deze tonen dus steeds lager eindigen. Uiteindelijk worden deze stijgingen en dalingen gerealiseerd als een dalend terraspatroon, waarbij er een dalende lijn is en er steeds in stapjes verder gedaald wordt. In figuur 1 is hier een voorbeeld van weergegeven.



Figuur 1. Een downstep in de intonatiecontour van een opsomming, met het onderliggende (gestreepte) contour en de oppervlaktecontour als vaste lijn.

Er werd door Van Heuven een effect gevonden van lengte van opsomming op de grootte van de eerste intonatedaling in de opsomming. Hiernaast werd gevonden dat de lengte van opsomming geen invloed had op de intonatiehoogte aan het begin en einde van de opsomming. Ook de grootte van de laatste intonatedaling bleef onveranderd bij een langere opsomming. Wat hieruit volgt is dat de daling bij ieder woord in de opsomming afhankelijk is van de opsommingslengte: hoe langer de opsomming hoe kleiner de stappen. Hieruit volgt dat de spreker planning moet uitvoeren om deze downstep goed te kunnen realiseren: het is nodig om te weten hoelang de opsomming is om hoog genoeg te beginnen en intonatiestappen te kunnen maken die zo klein of groot zijn als nodig. Overigens volgde uit een perceptieonderzoek in dit onderzoek van Van Heuven dat een luisteraar dit effect niet kan gebruiken voor het onderscheiden van afgekapte opsommingen: alleen het verschil tussen een opsomming van 2 woorden en langer kan worden gehoord.

In het onderzoek van Van Heuven is bij het analyseren van de intonatiecontouren, zoals gebruikelijk, gebruik gemaakt van het analyseren van de intonatiehoogtes op gemarkeerde punten. De intonatiecontouren zijn vervolgens via eenvoudige lijnen door deze punten gerepresenteerd; een daadwerkelijk intonatiepatroon gedraagt zich uiteraard niet op een dusdanig strakke manier. Deze beschrijvingen zijn dan ook vooral gericht op het gebruik en perceptie van deze intonatiepatronen. Zo heeft bijvoorbeeld Caspers (2000) gekeken naar de *betekenis* van de verschillende

intonatiepatronen: bijvoorbeeld dat het patroon L*H H% (stijgende accentintonatie gevolgd door een hoge intonatie) betekent dat de spreker zich er niet op wil vastleggen dat deze waarde achtergrondinformatie is.

Desondanks is het autosegment/ToDI-model nauwkeurig genoeg om ingezet te kunnen worden voor het genereren van intonatiecontouren bij spraaksynthese (Van Leeuwen & Te Lindert, 1993; Terken, 1993). De strakke lijnen lijken geen gevolgen voor de perceptie te hebben (Mertens et al., 1997). Gubian (2011) stelt echter dat bij spraaksynthese deze ‘gestileerde’ aanpak een aantal problemen met zich brengt. Hij baseert zich hiervoor op eerder onderzoek (Dombrowski & Niebuhr, 2010; D’Imperio, 2001) waarbij subtiele effecten optreden die niet gevangen kunnen worden in een gesegmenteerd model. Naast deze beperking in het gesegmenteerde model wijst Gubian op mogelijke problemen bij het vaststellen van de relevante punten en niveaus omdat hiervoor het oordeel van opgeleide luisteraars nodig is. Hierdoor ontstaat het risico van bias en oordeelverschillen tussen de luisteraars. Om deze problemen te omzeilen gebruikten Gubian (2011) en Aston et al. (2010) voor het analyseren en modelleren van intonatiecontouren de statistische methodes uit de functionele data-analyse (Ramsay et al., 2009). Hierdoor kunnen de curves zelf direct worden gemanipuleerd en statistisch worden vergeleken op een manier die minder afhankelijk is van het vooraf opgestelde model en oordeel van de onderzoeker zelf. Later zal de werking van deze analysemethode worden toegelicht.

Gubian et al. (2011) hebben deze methode toegepast voor het analyseren en manipuleren van intonatie in het Napolitaans Italiaans. In deze taal is de vraag-stellingonderscheid met alleen een intonatieverschuiving te realiseren. Zo wordt bij de zin ‘Milena lo vuole amaro(?)’ (Milena drinkt het [koffie] zwart) het verschil tussen de vraag ‘Drinkt Milena het zwart?’ en de stelling ‘Milena drinkt het zwart’ gerealiseerd door een piek in de intonatie aan het begin van de vraag, bij een stelling is ook een piek maar die is 0,1 seconde later. Door zinnen zoals deze op te nemen met dezelfde structuur, maar een andere focus konden Gubian et al. met behulp van de functionele principale-componentenanalyse (FPCA) de verschillen classificeren en modelleren en inzetten om bestaande opnames om te zetten van vraag naar stelling en vice versa.

Aston et al. (2010) hebben FPCA en een mixed model gebruikt om de intonatie van de toontaal Qiang op woordniveau te modelleren. De FPCA werd gebruikt om aan de verschillende intonatiecontouren scores toe te wijzen, deze scores konden vervolgens statistisch vergeleken worden in het mixed model. Er is op die manier bekeken welke factoren de intonatiecontour van een klinker beïnvloeden. Zo heeft in deze toontaal de uitgesproken klinker invloed op de intonatiecontour, overeenkomend met de theorieën over deze taal. Hiernaast werden subtielere effecten van geslacht, spreker en omringende klinkers ontdekt.

In dit onderzoek zal op een vergelijkbare manier als bij Gubian et al. (2011) en Aston et al. (2010) worden gekeken naar de intonatie van een opsomming. In plaats van het analyseren van zinnen of klinkers zal ik mij echter richten op de intonatie van woorden in Nederlandse opsommingen. Hierbij zullen gehele intonatiecontouren van woordrealisaties vergeleken worden in plaats van geselecteerde punten zoals bij Van Heuven (2004).

Definities van opsommingen

Om deze opsommingen te kunnen analyseren zal er allereerst worden begonnen met het omschrijven van wat een opsomming precies inhoudt. Als uitgangspunt kan gekeken worden naar de essentie van een opsomming op woordniveau:

op·som·men *somde* -, *h-gesomd* op de rij af noemen (Van Dale, 2013)

Een opsomming is op te vatten als een set van woorden met een bepaalde volgorde, hiervoor zal notatie 1 worden gehanteerd:

$$S(n) := x_1, x_2, \dots, x_n : x_i \in \Omega \quad (1)$$

Hierbij is $S(n)$ een opsomming van lengte n woorden, waarbij ieder woord x_i in de opsomming uit de verzameling van alle woorden Ω wordt gehaald. Op deze manier kan nu over de positie en lengte van een opsomming worden gesproken. Het gebruik hiervan zal het positiekenmerk worden genoemd (zie woordenbank in appendix) en kan worden gevat in notatie 2.

$$x_i \in S(n) \quad (2)$$

Hierin bepaalt i in x_i de positie in een opsomming van lengte n . Het één na laatste woord in een opsomming kan dan worden aangegeven met: $x_{n-1} \in S(n)$. Zo kan bijvoorbeeld de opsomming $S(3)$ ‘appels, peren en ijsjes’ zijn waarbij $x_2 \in S(3)$ *peren* is. Een ander aspect waarnaar gekeken wordt zijn de eigenschappen van het woord x_i zelf. De eigenschappen die worden meegenomen zijn het aantal syllaben van het woord c.q. de lengte L en de positie van de klemtoon k . Tezamen worden zij voor het gemak het woordkenmerk genoemd, gevat in notatie 3.

$$(L, k) \quad (3)$$

Ieder woord heeft nu een bepaald woordkenmerk, zo hebben bijvoorbeeld *Venlo* en *Haarlem* allebei het woordkenmerk $(2, 1)$. Deze notatie zal ook worden gebruikt bij het toelichten van het aanmaken van stimuli in *Materialen*.

Functionele data-analyse

De functionele data-analyse (Ramsay et al., 2009) omvat een set aan statistische functies die vanuit R (R Core Team, 2013) en MATLAB kunnen worden aangeroepen. Op deze manier kan geavanceerde statistiek worden aangeroepen zonder een diegravende kennis van de precieze wiskundige en computationele aspecten te vereisen. Dit is vergelijkbaar met hoe doorgaans bijvoorbeeld een t-toets of ANOVA-toets zou worden gebruikt. Omdat de gebruikte functies minder gangbaar zijn dan bijvoorbeeld de hiervoor genoemde t-toets en ANOVA, zal er kort worden stilgestaan bij de werking en eigenschappen van deze functies en functionele data-analyse.

Bij gangbare statistiek worden datapunten met elkaar vergeleken of beschreven, bijvoorbeeld de leeftijd van de proefpersonen. De data is in dit voorbeeld per respondent één punt: het is één waarde. Interessanter wordt het wanneer er meerdere aspecten worden bekeken van dezelfde proefpersoon. Dit zou bijvoorbeeld de lengte en het geslacht van de proefpersoon kunnen zijn. Nu hebben we twee ratiovariabelen met de lengte en leeftijd per respondent en een nominale variabele met het geslacht (man of vrouw). Er zou nu kunnen worden gekeken naar correlaties of verschillen tussen groepen etc. Stel we zouden geïnteresseerd zijn in de lengtegroei van kinderen en hebben een longitudinaal onderzoek uitgevoerd waarbij op verschillende punten in tijd de lengte is gemeten van een respondent. In dit voorbeeld van Ramsay et al. wordt voor deze situatie een Functionele principale-componentenanalyse (FPCA) ingezet.

Deze techniek is een uitbreiding van de principale-componentenanalyse (PCA). Bij een PCA wordt een matrix van meetgegevens omgezet naar componenten. De meetgegevens zijn samengevoegd tot één matrix x . Deze matrix x bestaat uit N rijen voor iedere respondent en P kolommen voor ieder meetpunt. Laten we beginnen met het eerste component: dit is de componentscore s_1 welke bepaald wordt met behulp van de meting x en de gewichtsvector w_1 . Voor de gewichtsvector wordt definitie 4 gehanteerd (Fodor, 2002):

$$w_1 = \arg \max_{|w|=1} \text{Var}(x^T w) \quad (4)$$

Hierbij wordt gezocht naar de gewichtsvector w_1 van lengte 1 die, vermenigvuldigd met de getransponeerde (gekantelde) meting x^T een zo hoog mogelijke variantie heeft tussen de datapunten. Op eenzelfde manier wordt naar de volgende componenten gezocht s_2, \dots, s_k , die de hoogste variantie verklaren van wat er nog niet door de voorgaande componenten is verklaard. De componentenscore s_1 wordt bepaald door definitie 5.

$$s_1 = x^T w_1 \quad (5)$$

Door deze componentenscores te vermenigvuldigen met de gewichtsvectoren kan een benadering van de oorspronkelijke meting worden verkregen. Omdat de meeste variantie in het eerste component zit kan het aantal dimensies worden verminderd (c.q. componenten weglaten) met zo min mogelijk effect op de data (Fodor, 2002). De overgebleven scores kunnen vervolgens gebruikt worden voor eenvoudigere analyse, ook bij taalwetenschappelijk onderzoek wordt dit principe toegepast. Zo hebben Jacobi et al. (2005) een PCA gebruikt voor onderzoek naar de aspecten bij klankuitspraak in het ‘Poldernederlands’, deze componentenscores werden gebruikt om een eenvoudiger beeld te krijgen van de beweging van klinkers door de tijd. Boersma & Kovacic (2006) hebben met een PCA op een formantanalyse verschillende Kroatische muziekstijlen van elkaar kunnen onderscheiden.

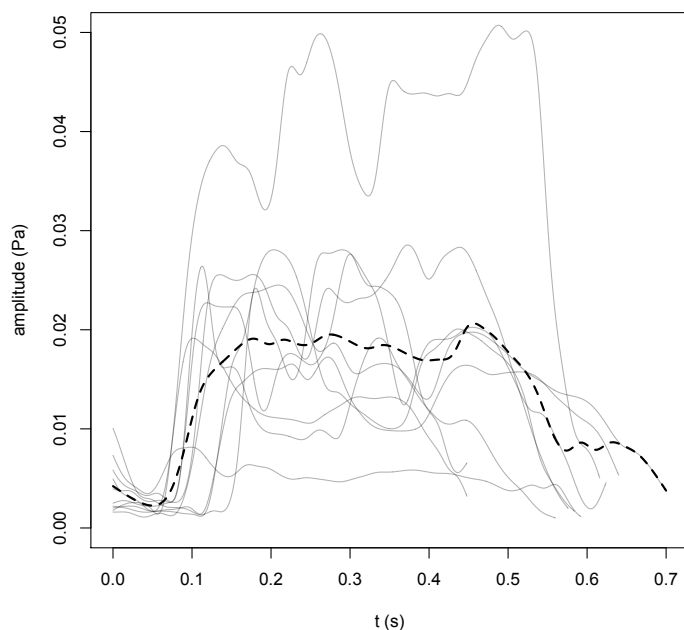
Bij een FPCA (Ramsay et al., 2009) wordt het principe uitgebreid van vectoren naar functies; er worden nu niet gewichtsvectoren verkregen maar functionele componenten (zogeheten harmonics) die ook zoveel mogelijk variatie verklaren en net zoals de componenten in een PCA vermenigvuldigd kunnen worden met componentenscores. De componenten van de functie f_i worden in stelling 6 (Ramsay et al., 2009, blz 102) als volgt gegeven:

$$s_{ij} = \rho_{\zeta_j}(f_i - \bar{f}) \equiv \int \zeta_j(t)[f_i(t) - \bar{f}(t)]dt \quad (6)$$

De meting bestaat nu niet uit een matrix $N \times P$ maar uit een verzameling van N functies over hetzelfde domein, bijvoorbeeld t van 0 tot 1. De functies f_1, \dots, f_N zijn hierbij genormeerd ten opzichte van het gemiddelde \bar{f} . Hierbij zijn $s_{i1}, s_{i2}, \dots, s_{ik}$ de k -componentenscores die worden teruggegeven door de zoekfunctie ρ_{ζ_j} voor de gemeten functies $f_{i \dots n}$. $\zeta_1(t), \dots, \zeta_n(t)$ zijn de N componentfuncties (harmonics), waarbij $\zeta_1(t)$ de meeste variantie verklaart. Een benadering van de oorspronkelijke functie f_i kan worden verkregen met behulp van de componentenscores en componentfuncties met stelling 7 (Gubian, 2011). Het fundamentele verschil hierbij met een niet functionele PCA is dat er wordt gewerkt met een integraal in plaats van een somming van de kolommen van de meting.

$$\hat{f}_i(t) = \bar{f}(t) + \sum_{j=1}^N s_{ij} \zeta_j(t) \quad (7)$$

Hier is te zien dat de informatie van een specifiek intonatiepatroon f_i te herleiden is tot een kleine set van numerieke componentenscores s_{ij} . Hiernaast is eenmalig de functie $\bar{f}(t)$ nodig die een gemiddelde representeert van de curves en een functie $\zeta_j(t)$ voor ieder component j . De reconstructie van het patroon gebeurt dan vervolgens door deze computationeel triviale stappen in 7: optellen en vermenigvuldigen. De uitkomsten van deze FPCA kunnen niet alleen numeriek maar ook kwalitatief worden geïnterpreteerd; de vormen van de curves kunnen bijvoorbeeld bekeken worden bij verschillende subselecties van f_i . Hiernaast is het mogelijk om de componentenscores te gebruiken om de *verschillen* tussen de functies onderling te kwantificeren. De



Figuur 2. Een voorbeeld van 10 ongenormeerde amplitudecontouren van realisaties van het woord ‘Venlo’. De grijze lijnen geven de individuele realisaties weer, de donkere stippellijnen een gemiddelde hiervan.

verschillende scores kunnen ook gebruikt worden om te ontdekken welke groepen er zijn, zonder vooraf een splitsing te hoeven te maken in de data.

Omdat de functies f_i bij FPCA over hetzelfde tijdsdomein worden vergeleken (Ramsay et al., 2009) is het nodig dat de verschillende contouren uitgelijnd moeten worden. Dit is duidelijk te illustreren met behulp van een aantal van de amplitudecontouren, in figuur 2 zijn 10 verschillende realisaties van het woord *Venlo* te zien. De sprekers verschillen onderling in luidheid en intonatie: mannen spreken bijvoorbeeld doorgaans lager dan vrouwen en niet iedereen spreekt precies even snel en luid. Dit kan gedeeltelijk opgevangen worden met een normering zoals in figuur 3a, hier wordt dan op globaal niveau overeenstemming bereikt: iedereen praat nu even snel, even luid en ook met dezelfde toonhoogte. Desondanks blijven er verschillen bestaan die te maken hebben met de verschillen in de specifieke realisatie van de woorden, het gaat hier om het waarin specifieke klanken in *Venlo* worden gerealiseerd. Zo wordt bijvoorbeeld de $/\varepsilon/$ in *Venlo* nooit precies even lang uitgesproken.

Het uitlijnen van contouren is mogelijk met feature alignment of met dataregistratie (Ramsay et al., 2009). In alle gevallen wordt een time-warpfunctie $h_i(t)$ op iedere functie f_i toegepast waardoor de functie f_i sneller of juist langzamer loopt op ieder moment in het tijdsinterval. Zo zou bijvoorbeeld de oorspronkelijk functie f_i kunnen worden samengedrukt rond het begin en het eind van het tijdsinterval en dus juist worden uitgerekt in het midden van het tijdsinterval. Het is gangbaar dat het tijdsinterval $[0, 1]$ wordt gebruikt en het is dan ook gangbaar dat $h_i(0) = 0$ en $h_i(1) = 1$.

Wanneer er *feature alignment* wordt gebruikt, worden de functies uitgelijnd op bepaalde vastgestelde kenmerken in de tijd; bijvoorbeeld wanneer een bepaalde klank wordt uitgesproken. Er wordt dan voor iedere functie f_i een transformatie $h_i(t)$ bepaald die ervoor zorgt dat die

kenmerken voor alle functies allemaal op dezelfde manier getimed zijn. Gubian et al. (2011) gebruiken hiervoor bijvoorbeeld een op hidden-Markovmodellen gebaseerde spraakherkenner om de tijdsgrenzen van de klanken te verkrijgen. Bij eerder onderzoek van Gubian et al. (2010) werd gebruik gemaakt van handmatig geannoteerde intervals van bepaalde klinkers met focus. Bij *dataregistratie* wordt het gemiddelde $\bar{f}(t)$ tegen een getime-warpte $f_i(t)$ geplot, uiteraard is $h_i(t)$ op het begin nog onbekend. Het doel is om een $h_i(t)$ te vinden waarbij deze plot zoveel mogelijk een rechte lijn benadert. De twee functies zijn dan proportioneel ten opzichte van elkaar, dit is als volgt geformuleerd (Ramsay & Silverman, 2005):

$$C(h_i) = \left[\begin{array}{cc} \int \{\bar{f}(t)\}^2 dt & \int \bar{f}(t) f_i[h_i(t)] dt \\ \int \bar{f}(t) f_i[h_i(t)] dt & \int \{f_i[h_i(t)]\}^2 dt \end{array} \right] \quad (8)$$

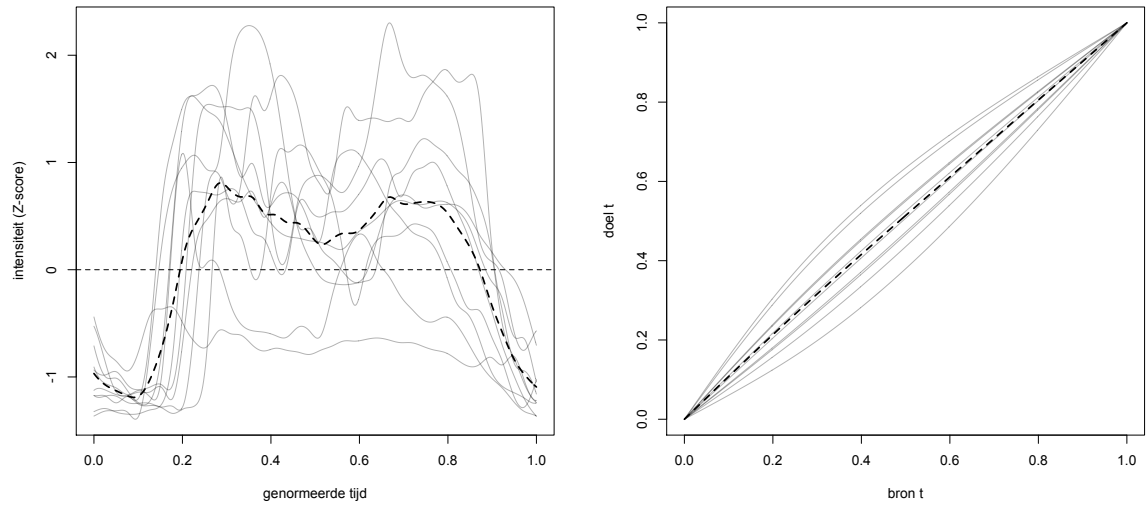
De functie $h_i(t)$ wordt dan zo geschat dat de eigenwaarde van $C(h_i)$ zo klein mogelijk is. Dit is een computationeel intensieve procedure, die zelfs op een snelle computer een paar uur rekentijd nodig kan hebben. Overigens hoeft het geen keuze te zijn tussen feature alignment en dataregistratie, de dataregistratie kan worden toegepast op de resultaten van de feature alignment om deze verder te verfijnen. In figuur 3b zijn de time-warpfuncties te zien voor de verschillende amplitudecontouren, in figuur 3c de contouren waarbij $h_i(t_p) = t_i$ en t_p de gemiddelde tijd is waar de eerste grote piek in het amplitudecontour optreedt, vervolgens is dit verder verfijnd met dataregistratie in figuur 3d.

Het uitlijnen is niet altijd noodzakelijk om de contouren uit te lijnen over tijd bij het gebruik van FPCA, zo hebben Aston et al. (2010) dit probleem omzeild door alleen de intonatie van de klinkers te meten en die te normeren. Het is hiervoor wel nodig deze klinkers te herkennen of handmatig te markeren. Het onderzoek van Aston et al. was gericht op het modelleren van de intonatiecontouren van de klinkers in een toontaal terwijl ik bij dit onderzoek vooral geïnteresseerd ben in de intonatiecontouren van gehele woorden. Hierom zullen de intonatiecontouren uiteindelijk worden uitgelijnd met behulp van hun samenhangende amplitudecontour. De hypothese hierbij is dat bij het ‘verplaatsen’ van woorden x_i binnen een opsomming $S(n)$ de vorm van de woordamplitudes minder veranderen dan de intonatiecontouren.

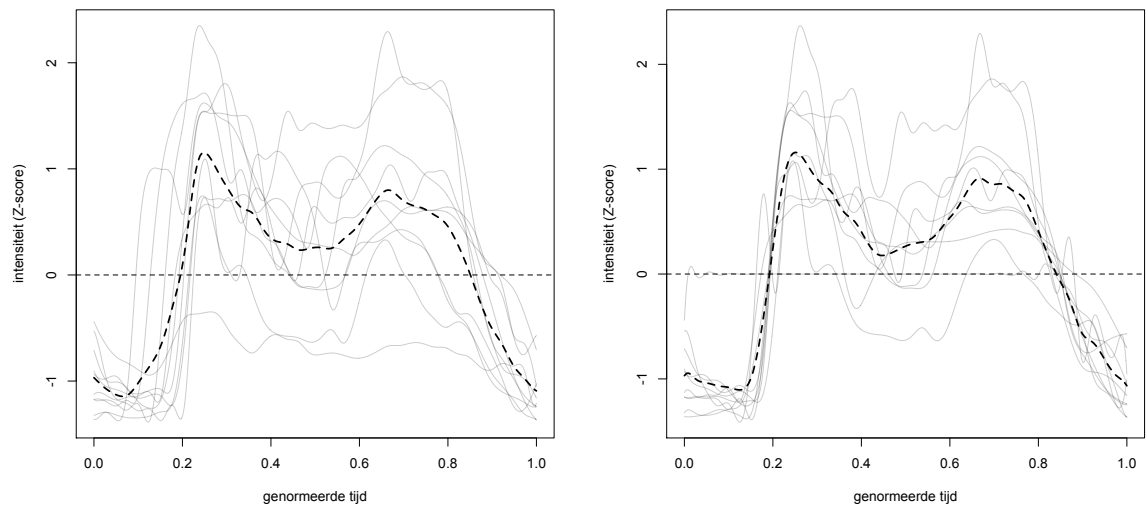
Hierboven is als vanzelfsprekend gesproken over het benaderen van amplitude- en intonatiecontouren als zijnde functies. Functies kunnen echter zelden tot nooit direct gemeten worden maar worden afgeleid uit losse datapunten gemeten in de tijd. Bij het meten van fenomen komen zaken als ruis en onzekerheid naar voren die op een goede manier moeten passen in het model. Hiernaast is er het vraagstuk wat er gedaan moet worden met de ruimte tussen de gemeten datapunten: moeten daar rechte lijnen tussen worden getrokken of is er misschien een ‘mooiere’ functie die beter recht doet aan de gemeten data?

De gangbare oplossing (Ramsay & Silverman, 2005) bij een functionele data-analyse is om een bepaalde basisfunctie $\hat{f}_i(t)$ te kiezen die aan de gemeten data $y_i(t) : t \in (0, \dots, T)$ wordt gefit. Een eenvoudige fit zou bijvoorbeeld kunnen zijn om $\hat{f}_i(t)$ het gemeten datapunt te laten teruggeven die het dichtst bij t zit, dit zou resulteren in een getrapte functie waarbij er bij iedere gemeten $t \in (0, \dots, T)$ een stap omhoog (of omlaag) wordt gedaan afhankelijk van de gemeten uitkomst. Een andere eenvoudige oplossing zou een lineaire interpolatie te doen, dat wil zeggen dat de twee datapunten die het dichtst bij t liggen worden gemiddeld.

Bij complexere functies kan het voorkomen dat de uiteindelijke benadering $\hat{f}_i(t)$ niet precies door alle gemeten datapunten lopen. Hiervoor kan worden gekozen als de formule al vaststaat (bijvoorbeeld bij een fysisch proces) en dat het vooral interessant is om de gehele basisfunctie de parameters te geven waarbij deze zo dicht als mogelijk bij alle datapunten loopt. Dit kan worden geformuleerd in een veralgemeniseerde formule 9 van Eilers & Marx (1996) worden gevat.



(a) Amplitudes genormeerd naar $t = [0, 1]$ met een Z-normering voor de $f_i(t)$ -waarde. (b) De time-warpfuncties $h_i(t)$ bepaald a.h.v. de eerste grootste piek in iedere realisatie in 3a.



(c) Amplitude na het toepassen van de time-warpfuncties $h(t)$ van 3b. (d) Geregisteerde amplitudes na het uitvoeren van dataregistratie op de amplitudes van 3c

Figuur 3. Normering en dataregistratie toegepast op 10 amplitudecontouren van realisaties van het woord ‘Venlo’. De grijze lijnen geven individuele realisaties weer, de donkere stippellijnen de gemiddelden hiervan.

$$\hat{f}_i(t) = \arg \min_p \sum_{t=0}^T (y_i(t) - \hat{f}_p(t))^2 \quad (9)$$

Hierbij is p een instelparameter voor een basisfunctie $\hat{f}_p(t)$. Bij functionele data-analyse is het gangbaar (Ramsay & Silverman, 2005) om voor deze basisfunctie een Fourier-functie te kiezen bij periodieke data en B-splines voor niet-cyclische data. De Fourier-functies zullen verder buiten beschouwing blijven, gezien ik bij dit intonatie-onderzoek vooral in de niet-cyclische aspecten geïnteresseerd ben.

Bij B-splines (Eilers & Marx, 1996; Foley et al., 1996) wordt een functie of dataset $f_i(t)$ benaderd of gemodelleerd door een aantal B-splinesfuncties in $\hat{f}_i(t)$ (formule 10).

$$\hat{f}_i(t) = \sum_{j=1}^n a_j B_j(t, q) \quad (10)$$

Elk van deze B-splinefuncties $B_j(t, q)$ bestaat uit $q + 1$ segmenten die worden gedefinieerd door een graad q polynomiale functie, waarbij q de ‘graad’ weergeeft van de B-spline-functie. De segmenten staan tussen $q+2$ knopen op het interval $[-\frac{q}{2}, \frac{q}{2}]$. Alleen in de $q+2$ knopen daartussen is de functie B_j positief en alleen door die punten wordt de vorm van deze segmentfunctie bepaald. Hierdoor kunnen de B-splines alleen op dat ene punt worden aangepast, zodat deze overeenkomt met de gemeten data. Dit wordt gedaan door de waarde a_j aan te passen. Door alle waarden a_j te manipuleren kan de uiteindelijke B-splinesfunctie $\hat{f}_i(t)$ worden verkregen die de data $f_i(t)$ benaderdt. Omdat het aanpassen van een knoop a_j slechts een lokaal effect heeft is deze benadering door een computer snel uit te voeren.

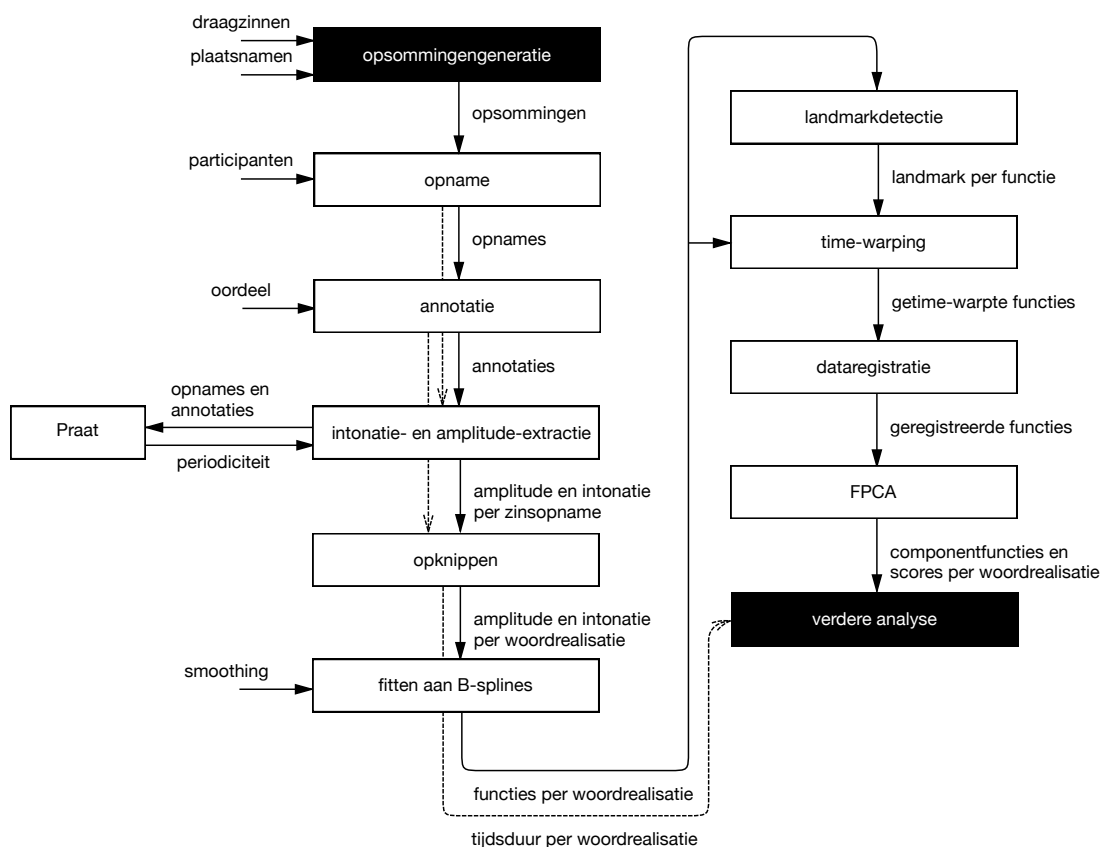
Naast het zo nauw mogelijk fitten van deze functie op de datapunten komt er doorgaans een ander punt bij kijken: de ruwheid van de uiteindelijke functie. Een ruwe functie springt net als de ruis op en neer, terwijl een gladde functie tussen de ruis door de daadwerkelijke beweging weergeeft. Om deze afweging concreet te maken wordt een zekere ruwheidstrafmaat λ meegenomen bij het fitten van de functie. Hiermee kan formule 9 worden uitgebreid: naast het verminderen van het verschil tussen de datapunten en de functie wordt er geprobeerd een zo ‘glad’ mogelijke functie te vinden. Hoe deze twee zaken precies worden afgewogen hangt af van de gekozen λ in formule 11 (Eilers & Marx, 1996).

$$\hat{f}_i(t) = \arg \min_a \left(\sum_{t=0}^T \left[y_i(t) - \sum_{j=1}^n a_j B_j(t, q) \right]^2 + \lambda \int_{t'=0}^{t'=T} \left[\sum_{j=1}^n a_j B''_j(t', q) \right]^2 dx \right) \quad (11)$$

Voor ieder van de n segmenten wordt naar een B-splinefunctie B_j gezocht van graad q . Hierbij zijn a_j de controlepunten voor de B-spline-segmentfuncties B_j , is het domein $t = [0, T]$ en het aantal segmenten n . Hier is te zien dat een minimalisatie van de afwijking van de B-splines tegen het minimum wordt beperkt, maar dat deze geanneerd wordt door de smoothingsvariabele λ die gebruik maakt van de tweede afgeleide van B_j .

Methode

Voor het bestuderen van de intonatie van opsommingen is een script geschreven om een aantal opsommingen te genereren. Met behulp van dat script kregen participanten een aantal opsommingen om in te spreken. Deze opsommingen zijn vervolgens in Praat (Boersma & Weenink, 2013) geannoteerd, waarbij ieder woord van de opsomming is gemarkeerd. Deze annotatie werd vervolgens tezamen met de opnames gebruikt om de intonatie- en amplitudecontour



Figuur 4. Een schematische weergave van het verkrijgen en bewerken van de verzamelde data. Het begint bij de opsommingengeneratie (aangegeven in het zwart) en de data wordt vanaf daar uiteindelijk omgevormd voor verdere analyse (ook aangegeven in het zwart). De vlakken stellen de verschillende processen voor in de dataverwerking, de pijlen de verschillende vormen van in- en uitvoer data, de stippellijnen geven weer waar data ook in een later proces worden gebruikt. Bij de verdere analyse waren ook de overige invoergegevens beschikbaar (zoals bijvoorbeeld de plaatsnamen), dit is niet meegenomen in deze weergave.

per opgenomen woord te verkrijgen. De contouren worden genormeerd en kunnen vervolgens statistisch met elkaar worden vergeleken.

In figuur 4 is een overzicht te zien van dit proces. In de volgende paragrafen en in *Procedure* wordt het proces verder uitgewerkt.

Materialen

De woorden (zie woordenbank in appendix) die in de opsommingen worden gebruikt, zijn plaatsnamen. Hiervoor is gekozen omdat deze in het AUTONOMATA-corpus (Van den Heuvel et al., 2008) voorkomen, welke gebruikt zou kunnen worden voor het manipuleren met behulp van de verkregen/gesynthetiseerde intonatiecontouren. Hiernaast is het met plaatsnamen eenvoudig mogelijk geloofwaardige opsommingen te maken in willekeurige volgorde.

Bij de keuze voor *welke* plaatsnamen is gekeken naar plekken met een zekere naamsbekendheid. Dit is bepaald op basis van de aanwezigheid van een treinstation. Dit om er zeker van te zijn dat de respondenten er geen problemen mee hebben met het uitspreken van de plaatsnaam. Zeer grote steden zijn uitgesloten om afleiding te voorkomen. Daarnaast zijn plaatsnamen waarin epenthesis mogelijk is (zoals Alkmaar) of klemtoon niet voor iedereen vaststaat (zoals

Nijverdal) uitgesloten om er zeker van te zijn dat een plaatsnaam altijd met hetzelfde aantal syllaben en klemtoon wordt uitgesproken. Tot slot zijn plaatsnamen die uit meerdere woorden bestaan (zoals Den Haag) uitgesloten. Dit omdat het niet zeker is of de respondent expres een pauze zou kunnen laten vallen tussen de woorden.

De plaatsnamen zijn onderverdeeld op basis van *woordkenmerk* (het aantal syllaben en klemtoonpositie, aangegeven met (L, k)). Er waren geen gangbare plaatsnamen van 3 syllaben met de klemtoon op de laatste syllabe, daarom werd er alleen gekeken naar woorden met de klemtoon op de eerste of laatste syllabe. Dat geldt uiteraard niet voor woorden met 1 syllabe, daar speelt klemtoonpositie geen rol. Dit betekent dat er vijf mogelijke woordkenmerken zijn:

- 1 syllabe. Bijvoorbeeld *Goes*.
- 2 syllaben, klemtoon eerste syllabe. Bijvoorbeeld *Haarlem*.
- 2 syllaben, klemtoon tweede syllabe. Bijvoorbeeld *Roermond*.
- 3 syllaben, klemtoon eerste syllabe. Bijvoorbeeld *Deventer*.
- 3 syllaben, klemtoon tweede syllabe. Bijvoorbeeld *Abcoude*.

Voor elk van deze woordkenmerken zijn er 5 plaatsnamen gekozen, deze staan vermeld in de woordenbank in de appendix. Er zijn verschillende woordkenmerken gekozen om meer gevarieerde zinnen te kunnen genereren en om iets te kunnen zeggen over de eventuele generaliseerbaarheid van de intonatiepatronen. De opsommingen (stimuli) die gegenereerd werden hadden een lengte van 2 tot maximaal 5 woorden. Dit betekent dat er een totaal van 4 *stimuluszinslengtes* zijn.

Wanneer er vanuit wordt gegaan dat voor ieder mogelijk *positiekenmerk* (de lengte van een opsomming en de plek van een woordrealisatie daarin, aangegeven met $x_i \in S(n)$), ieder mogelijk woordkenmerk minstens één keer voorkomt, vereist dit minimaal 20 zinnen (4 zinslengtes x 5 woordkenmerken). Deze zinnen eenvoudigweg zonder randomisatie genereren resulteert in zinnen die bestaan uit woorden met hetzelfde woordkenmerk, bijvoorbeeld een zin met alleen woorden van 1 syllabe.

$$\begin{bmatrix} \text{Goes}_{(1,1)} & \text{Tiel}_{(1,1)} & \cdots & \text{Soest}_{(1,1)} \\ \text{Haarlem}_{(2,1)} & \text{Venlo}_{(2,1)} & \cdots & \text{Tilburg}_{(2,1)} \\ \text{Maassluis}_{(2,2)} & \text{Uitgeest}_{(2,2)} & \cdots & \text{Roermond}_{(2,2)} \end{bmatrix} \quad (12)$$

Een oplossing zou zijn om woorden met hetzelfde positiekenmerk te husselen: alle woorden die bijvoorbeeld op de derde plek staan in een opsomming van lengte 5 worden in een andere zin op dezelfde plek neergezet.

$$\begin{bmatrix} \text{Haarlem}_{(2,1)} & \text{Tiel}_{(1,1)} & \cdots & \text{Roermond}_{(2,2)} \\ \text{Goes}_{(1,1)} & \text{Uitgeest}_{(2,2)} & \cdots & \text{Soest}_{(1,1)} \\ \text{Maassluis}_{(2,2)} & \text{Venlo}_{(2,1)} & \cdots & \text{Tilburg}_{(2,1)} \end{bmatrix} \quad (13)$$

Hier ontstaat een probleem: woorden worden willekeurig op basis van woordkenmerk geplaatst en gehusseld op basis van positiekenmerk. Er wordt daarbij niet gekeken naar de context waardoor het mogelijk is dat hetzelfde woord meerdere malen in een zin voorkomt, zo is door het husselen van voorbeeld 14 de opsommingen 15 ontstaan waarin het woord *Goes* tweemaal voorkomt.

$$\begin{bmatrix} \text{Goes}_{(1,1)} & \text{Tiel}_{(1,1)} & \cdots & \text{Soest}_{(1,1)} \\ \text{Vught}_{(1,1)} & \text{Sneek}_{(1,1)} & \cdots & \text{Goes}_{(1,1)} \\ \text{Haarlem}_{(2,1)} & \text{Venlo}_{(2,1)} & \cdots & \text{Tilburg}_{(2,1)} \\ \text{Maassluis}_{(2,2)} & \text{Uitgeest}_{(2,2)} & \cdots & \text{Roermond}_{(2,2)} \end{bmatrix} \quad (14)$$

↓

$$\left[\begin{array}{cccc} \text{Vught}_{(1,1)} & \text{Uitgeest}_{(2,2)} & \cdots & \text{Tilburg}_{(2,1)} \\ \mathbf{Goes}_{(1,1)} & \text{Venlo}_{(2,1)} & \cdots & \mathbf{Goes}_{(1,1)} \\ \text{Haarlem}_{(2,1)} & \text{Sneek}_{(1,1)} & \cdots & \text{Soest}_{(1,1)} \\ \text{Maassluis}_{(2,2)} & \text{Tiel}_{(1,1)} & \cdots & \text{Roermond}_{(2,2)} \end{array} \right] \quad (15)$$

Om dubbele woorden te voorkomen wordt er gebruik gemaakt van placeholders $*(L,k)$. Deze placeholders worden eerst op woordkenmerk gehusseld, pas hierna worden woorden uit de woordenbank Ω geplaatst. Hierbij wordt voor iedere placeholder alleen woorden gekozen met een gelijkend kenmerk. Dit wil zegen dat bijvoorbeeld bij een placeholder $*(1,1)$ alleen een woord van één syllabe kan worden geplaatst uit de deelverzameling $\Omega_{(1,1)} \in \Omega$. Hiernaast wordt bij het plaatsen van een woord gecontroleerd of deze niet al eerder in de opsomming is gebruikt. Het plaatsen gebeurt asefect met als enige beperking dat een woord niet meer dan één keer in dezelfde opsomming mag worden gebruikt. Het is niet praktisch realiseerbaar om alle mogelijke combinaties van woordvormen te genereren. Zelfs het ogenschijnlijke beperkte aantal van 4 zinslengtes en 5 woordvormen resulteert in 3900 mogelijke permutaties.

$$\sum_{n=2}^5 5^n = 3900 \quad (16)$$

Om in het gehele corpus zoveel mogelijk vergelijkbare data te verkrijgen, is er als referentiepunt het controlewoord *Venlo* gekozen. Het controlewoord komt bij alle respondenten minimaal tweemaal voor bij ieder mogelijk positiekenmerk. Dit specifieke woord is gekozen omdat het geen plosieven bevat, wat het verkrijgen van intonatiepatronen vereenvoudigt. Het plaatsen van het controlewoord gebeurt wanneer een placeholder $*$ met woordkenmerk $(2,1)$ wordt gevuld, in plaats van een willekeurig woord uit $\Omega_{(2,1)}$ wordt gekeken of het controlewoord kan worden geplaatst. Hierbij wordt ook gecontroleerd dat dit woord niet al voorkomt in de opsomming, in welk geval een willekeurig woord uit $\Omega_{(2,1)}$ wordt geplaatst. Als het controlewoord tweemaal is geplaatst voor een positiekenmerk $x_i \in S(n)$ voor deze respondent, wordt asefect een woord met het woordkenmerk $(2,1)$ gekozen.

Bij het maken van de opsommingen worden er 4 sets van 20 placeholders-opsommingen gegenereerd voor een totaal van 80 opsommingen. Hierdoor kan het controlewoord voor iedere respondent tweemaal per positiekenmerk worden bevraagd en kan even vaak een ander woord met hetzelfde kenmerk worden geplaatst. Om te voorkomen dat achtereenvolgens blokken van dezelfde opsomminglengte worden gevraagd, worden aan het eind van de generatie alle opsommingen in zijn geheel gehusseld. Hierna worden asefect draagzinnen (zie de draagzinnen in de appendix) voor de opsomming geplaatst.

Bijvoorbeeld:

$$\left[\begin{array}{cccc} *(1,1) & *(1,1) & \cdots & *(1,1) \\ *(1,1) & *(1,1) & \cdots & *(1,1) \\ \vdots & \vdots & \ddots & \vdots \\ *(2,1) & *(2,1) & \cdots & *(2,1) \\ \vdots & \vdots & \ddots & \vdots \\ *(3,1) & *(3,1) & \cdots & *(3,1) \end{array} \right] \quad (17)$$

↓

$$\begin{bmatrix} *(1,1) & *(3,1) & \cdots & *(2,1) \\ *(2,1) & *(2,1) & \cdots & *(1,1) \\ \vdots & \vdots & \ddots & \vdots \\ *(1,1) & *(1,1) & \cdots & *(3,1) \\ \vdots & \vdots & \ddots & \vdots \\ *(3,1) & *(1,1) & \cdots & *(1,1) \end{bmatrix} \quad (18)$$

↓

$$\begin{bmatrix} \text{Vught}_{(1,1)} & \text{Deventer}_{(3,1)} & \cdots & \text{Tilburg}_{(2,1)} \\ \text{Haarlem}_{(2,1)} & \text{Venlo}_{(2,1)} & \cdots & \text{Tiel}_{(1,1)} \\ \vdots & \vdots & \ddots & \vdots \\ \text{Goes}_{(1,1)} & \text{Sneek}_{(1,1)} & \cdots & \text{Vlissingen}_{(3,1)} \\ \vdots & \vdots & \ddots & \vdots \\ \text{Hilversum}_{(3,1)} & \text{Tiel}_{(1,1)} & \cdots & \text{Soest}_{(1,1)} \end{bmatrix} \quad (19)$$

Voor iedere respondent wordt er aan het begin van het onderzoek een nieuwe verzameling stimuli aangemaakt. Dit wordt gedaan met behulp van een Python-script (Van Rossum, 2013) die op de bovenstaande manier zinnen aanmaakt en wordt automatisch aangeroepen voor het begin van de opnames. De bevroegde stimuli worden opgeslagen voor verdere analyse.

Respondenten

Aan vrienden en kennissen werd gevraagd om deel te nemen aan het onderzoek. Uiteindelijk waren er 10 respondenten, waarvan de leeftijd varieerde van 20 tot 32 jaar ($\mu = 24,7; \sigma = 3,37$). Van de respondenten zijn er 6 vrouw en 4 man. De meeste van de respondenten hebben hoger onderwijs gevolgd (Universiteit of HBO = 8, VMBO = 1, MBO = 1). Een van de respondent is momenteel woonachtig in provincie Utrecht, de anderen zijn allemaal woonachtig in Noord-Holland. De meeste van de respondenten zijn minimaal een aantal jaar woonachtig in de omgeving Amsterdam (Amsterdam = 6, Utrecht = 1, Haarlem = 2) en één in Den Helder. De meeste respondenten zijn opgegroeid met een Hollands accent (Noord Holland en Utrecht), 2 respondenten zijn opgegroeid in Brabant.

Procedure

Het onderzoek werd afgenomen bij de onderzoeker thuis onder vrienden en bekenden. Bij iedere respondent werden het raam en de deur van de kamer gesloten. Hiernaast werd het gordijn dichtgedaan om afleiding en geluid van buiten tegen te gaan.

De instructies en doelzinnen werden aan de respondent getoond door middel van een computerscherm. De uitgesproken zinnen werden opgenomen met behulp van een microfoon die aangesloten is op de computer waarmee het experiment werd afgenomen. Naast deze opname werd op de achtergrond een respondentnummer met een overzicht van de doelzinnen weggeschreven.

De 80 doelzinnen (zie *Materialen*) verschenen achtereenvolgend op het computerscherm. De respondent kon door middel van het toetsenbord aangeven dat de zin opnieuw moest worden opgenomen of dat naar de volgende zin kon worden gegaan. Halverwege was er een pauze en kreeg de respondent de gelegenheid iets te eten en te drinken. Gemiddeld deden de respondenten er 5:07 minuten ($\sigma = 61s$) over voor het eerst deel van het onderzoek, 7:47 minuten ($\sigma = 370s$) voor de pauze en 4:39 minuten ($\sigma = 45s$) voor het tweede deel.

Aan het begin van het onderzoek werd de respondent geïnstrueerd over deze procedure, daarnaast kreeg hij of zij de instructie de gepresenteerde zinnen op een natuurlijke, niet-formele

manier rustig uit te spreken. Zodra de doelzinnen waren opgenomen werd het geslacht, leeftijd, opleiding, dialect, thuisstad en eventuele opmerkingen van de respondent genoteerd. Bij het experiment is 1 keer het opslaan van het stimuloverzicht mislukt, dit overzicht is toen handmatig uitgezocht aan de hand van de ingesproken gegevens.

De opgenomen doelzinnen werden vervolgens in Praat (Boersma & Weenink, 2013) handmatig op woordniveau geannoteerd; hierbij werden alleen de woorden van de uitgesproken opsomming zelf, het voegwoord ‘en’ en het laatste woord van de draagzin geannoteerd. Dit was gedaan om van alle woorden in de opsomming de eventuele tijdsduur van stiltes ervoor en daarna te kunnen analyseren. Na de annotatie is een gescripte controle uitgevoerd om te controleren of alle geannoteerde woorden overeenkwamen met de doelzinnen. Er bleken geen zinnen te zijn waar de uitgesproken woorden niet overeenkwamen met de doelzinnen.

Bij het annoteren waren er praktische problemen met de enorme hoeveelheid van te annoteren gegevens (800 zinnen), het annoteren alleen heeft uiteindelijk ruim een werkweek gekost. Bij het annoteren waren er praktische problemen met de klankbotsingen en assimilatie. Het was vaak niet direct de precieze ‘grens’ duidelijk te zien tussen de woorden, zoals bij bijvoorbeeld “Enkhuizen, Enschede en Goes” (respondent 4, opsomming #25), “en Maassluis” (respondent 7, #6), “en Haarlem” (#7). Wanneer dit überhaupt niet mogelijk was werd het midden gekozen van de klank. Bij een degeminatie van een plosief (bijvoorbeeld “uit Tilburg”) werd het hoorbare begin van de plosieflank gekozen als afbreekpunt. Voor de analyse van de stiltes tussen woorden kunnen deze plosieven nog voor een kanttekening zorgen: het is niet altijd geheel zeker of de stilte tussen twee woorden komt door de plosief of een pauze.

Respondenten 1 t/m 4 kregen door een fout in de programmering soms dezelfde draagzin twee keer achter elkaar, de opsommingen zelf waren wel altijd willekeurig. Er waren een aantal onverwachte uitspraken van de respondenten, zo realiseerde de eerste respondent consequent Delfzijl als Delftzijl. Ook respondent 10 realiseerde Delfzijl incidenteel op deze manier. De resultaten van deze realisaties zijn niet uitgesloten omdat deze verder geen invloed hadden op de klemtoonstructuur en rest van de opsomming. Andere uitingen waarbij de klemtoon niet overeenkwam met de verwachte woordvorm zijn volledig uitgesloten van verdere analyse, dit om het effect van klemtoon zuiver te houden in de verdere analyse. Dit trad bijvoorbeeld op bij respondent 5 en 6 bij de uitspraak van resp. Roermond en Maassluis, waar de klemtoon altijd op de eerste syllabe werd geplaatst in plaats van de tweede.

Een enkele maal werd het geluid voortijdig beëindigd, dit kwam doordat respondent 3 voor het afronden van de uiting op ‘verder’ had gedrukt. Deze uiting #30 is daarom volledig uitgesloten van analyse. Bij respondent 10 zijn opnames #15 en #23 uitgesloten vanwege aarzeling bij het uitspreken van Hoogkarspel, hiernaast was bij #41 van deze respondent ‘Venlo’ uitgesloten vanwege een versproken klemtoon.

Van de 797 (2800 – 11 woordrealisaties) resterende geannoteerde opnames zijn vervolgens, uitgesplitst per woord, de intonatie F_0 en amplitude opgevraagd vanuit Praat. Voor de intonatie-analyse in Praat werd de autocorrelatie-methode gebruikt (Boersma, 1993) met standaardinstellingen en een ondergrens van 70 en 150 Hz voor respectievelijk mannen en vrouwen en een bovengrens van resp. 300 Hz en 500 Hz. Op een vergelijkbare manier als bij Gubian et al. (2011) en Zellers et al. (2010) werd vervolgens voor ieder woord in R deze F_0 opgehaald en omgezet naar semitonen. De contouren zijn omgezet naar semitonen om uitschieters te verminderen en de verschillen tussen waardes meer overeen te laten komen met perceptie (Nolan, 2003). De woordrealisaties met een klemtoon die afweek van de verwachte woordvorm werden uitgesloten, waarna er $2789 - 36 = 2753$ woordrealisaties overbleven.

Alle intonatie- en amplitudewaardes zijn per woord op twee B-spline-functies gefit (één voor intonatie, één voor amplitude) met behulp van de FDA-toolkit (Ramsay et al., 2009). Hiervoor is een smoothingvariabele $\lambda = 1 \times 10^{-7}$ gebruikt, een B-spline van orde 4 en 40 knopen

(met 2 knopen aan de uiteinden). Dit zijn dezelfde instellingen als bij Turco et al. (2011, tutorial Gubian). Na een visuele inspectie van de contouren zijn er geen redenen gezien om van deze waarden af te wijken; ook al gaat het hier om intonatiecontouren van woorden in plaats van gehele zinnen.

In plaats van de semitonen direct op een B-spline te fitten werd een Z-normering $Z_{i,p}(t) \equiv (f_i(t) - \mu) / \sigma$ (Rose, 1991; 1987) toegepast om verschillen tussen sprekers te verminderen. Hierbij werd voor iedere spreker de μ en σ gebruikt van de ongenormeerde functies van de opnames van de respondent zelf. Omdat FPCA wordt toegepast op functies in hetzelfde tijdsinterval zijn deze functies ook genormeerd naar $\tau = t \rightarrow [0, 1]$ seconden. Omdat het meestal niet mogelijk is om over de gehele tijdsduur van een woordrealisatie de intonatiehoogtes te verkrijgen is er voor gezorgd dat aan het begin en eind van de woordrealisatie ($\tau = 0$, $\tau = 1$) de laatst gevonden en direct volgende intonatiewaarden worden gebruikt.

Tijdens de Z-normering zijn uitschieters waarbij $|Z(t)| > 4$ uitgesloten om uitschieters te verminderen, uiteindelijk zijn er hiermee 166 woordrealisaties uitgesloten. In de appendix is een overzicht te zien van alle uitgesloten woordrealisaties. De genormeerde amplitudefuncties zijn vervolgens geautomatiseerd gelandmarkt, dit is gedaan door de woorden uit te lijnen op de grootste beginnende piek in de amplitude. Hiervoor is een heuristisch algoritme geschreven en toegepast waarna de resultaten zijn gecontroleerd met visuele inspectie. Vervolgens is er op deze gegevens een continue registratie (Ramsay et al., 2009) toegepast om de timewarp verder te verfijnen.

De timewarp van deze amplitudefuncties zijn vervolgens toegepast op de intonatiefuncties om deze op dezelfde manier te vervormen, waarbij de amplitude als ‘invariant’ wordt beschouwd. Op deze genormeerde en geregistreerde intonatiefuncties is vervolgens een FPCA toegepast, waarna een Varimax-rotatie (Kaiser, 1958) is uitgevoerd voor een vereenvoudigde interpretatie.

Net zoals bij Gubian et al. (2010) worden alle normeringsgegevens en harmonics bewaard zodat deze kunnen worden toegepast voor het reconstrueren van intonatiecontouren en het manipuleren van opnames.

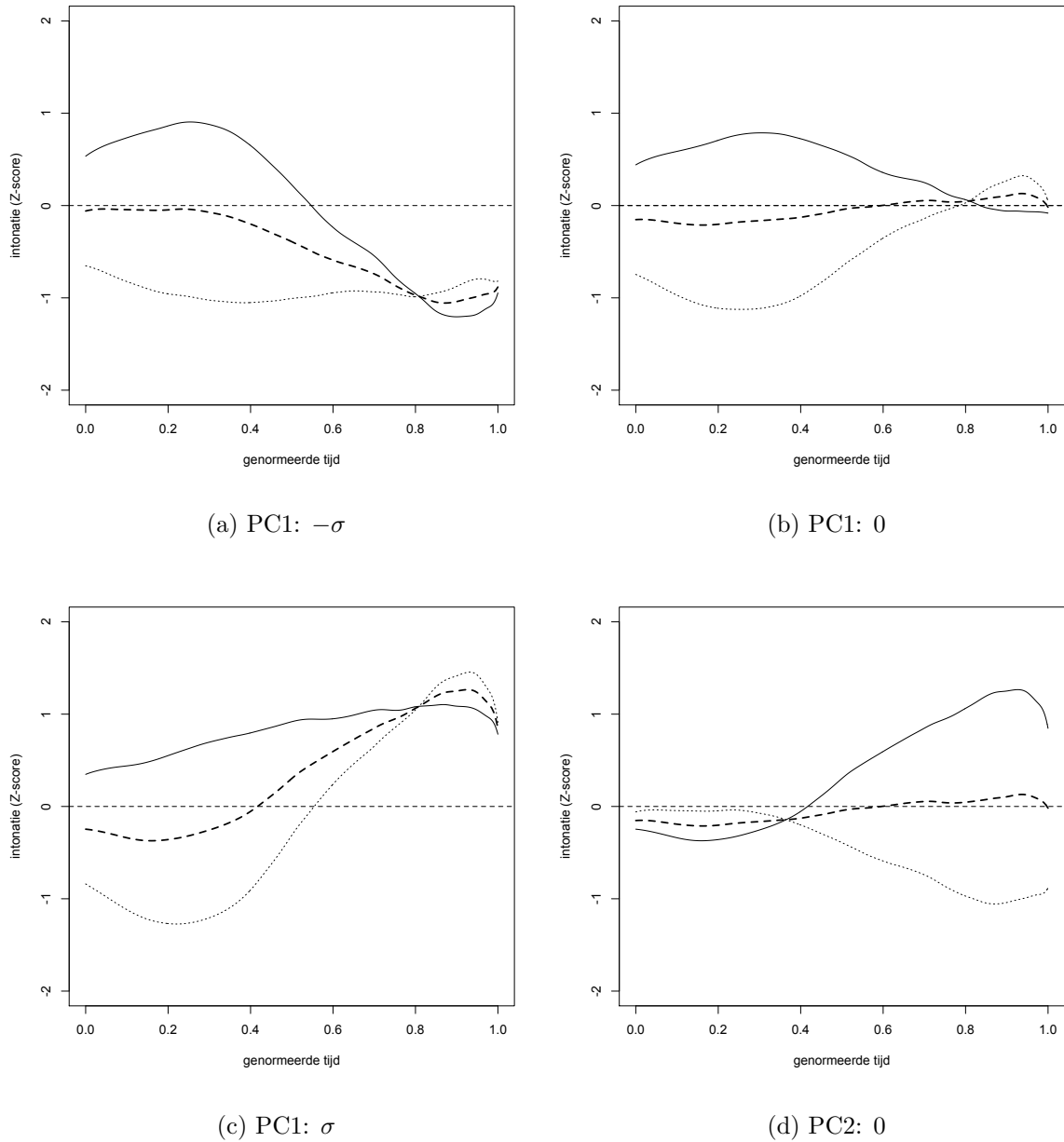
Resultaten

Na iedere afname was de respondent gevraagd over de aanwezigheid van een controlewoord. Er werden meerdere woorden geraden, maar het daadwerkelijke controlewoord *Venlo* werd door geen één van de respondenten genoemd. Van de 2576¹ overgebleven woordrealisaties zijn de eerste twee componentscores PC1 en PC2 van de intonatiecurves verkregen. Deze twee scores geven tezamen aan wat de vorm is van ieder van deze intonatiecurves. Deze scores worden met behulp van conventionele statistiek – een General Linear Model (GLM) – in verband gebracht met verschillende factoren zoals woordvorm en positie.

Tabel 1: Algemene statistiek over de FPCA-scores en tijdsduur, $N = 2576$.

PC1 intonatie		PC2 intonatie		Tijdsduur	
μ	σ	μ	σ	μ	σ
0,00	0,64	0,00	0,62	0,59s	0,14s

¹92% van de 2800 woordrealisaties, zie de appendix voor de uitgesloten woordrealisaties.



Figuur 5. De intonatiecontouren bij verschillende PC1- en PC2-scores. Bij 5a heeft PC1 een score van $-\sigma$, bij 5b een score van 0 en bij 5c een score van σ . De PC2-scores is in de grafieken 5a, 5b en 5c σ voor de bovenste zwarte lijn, 0 voor de middelste gestreepte lijn en $-\sigma$ voor de onderste gestipte lijn. In grafiek 5d heeft PC2 een score van 0 en PC1 is tussen $-\sigma$ (gestipte lijn) en σ (zwarte lijn).

De tijdsduur (voor normering) per woordrealisatie was gemiddeld $\mu = 0,59s$; $\sigma = 0,14s$. De gemiddelde PC1-score voor intonatie was gecentreerd op $\mu = 0,00$ met $\sigma = 0,64$ en de PC2-score was gecentreerd op $\mu = 0,00$ met $\sigma = 0,62$ (zie tabel 1). De verklaarde variantie van de eerste score was 38,5% en 36,6% voor de tweede score. Hiermee is 75,1% van de variantie qua intonatie in de dataset verklaard door deze componenten. In figuur 5 zijn de componentfuncties van de FPCA over de intonatiefuncties te zien. Hierin is het effect van een positieve dan wel

negatieve score te zien: een positieve score op het eerste component wordt geassocieerd met een stijgende lijn (figuur 5c), een negatieve met een dalende (figuur 5a). Bij het tweede component is het juist andersom: positief is een dalende lijn (figuur 5b). Waar deze lijnen stijgen of dalen verschilt echter: bij het eerste component vindt de stijging plaatst rond $0,4s$ terwijl bij het tweede component er juist eerst nog een stijging is voordat hij na $\pm 0,3s$ gaat dalen of stijgen. Het is ook te zien dat de PC2-score samenhangt met de beginwaarde op $t = 0$, bij een hoge score begint de intonatie boven het genormeerde gemiddelde, bij een lage score juist daaronder. De PC1-score hangt samen met de intonatiehoogte rond $t = 0,8s$ daar wordt deze ‘hoog’ bij een positieve PC1-score of ‘laag’ bij een negatieve PC1-score. Dit is vooral duidelijk te zien bij figuur 5b, de drie grafieken convergeren naar 0, of dit een daling of stijging is hangt af van de beginwaarde op $t = 0$ die wordt bepaald door PC2.

Deze FPCA-scores van de intonatie en amplitude zijn vervolgens in SPSS² gemodelleerd met een General Linear Model (GLM) waarbij de FPCA-scores en tijdsduren van de woordrealisaties als afhankelijke variabelen werden gebruikt, woordkenmerk, positiekenmerk, geslacht en dialect werden gebruikt als onafhankelijke variabelen, en opsommingspositie (de hoeveelste opsomming van de respondent) als covariabele. De uitkomsten hiervan zijn te zien in tabel 2.

Er wordt uitgegaan van een toetsingsniveau van $\alpha = 0,05$. Hierbij is te zien dat er een significant effect optreedt op basis van de factor Geslacht op de PC2 amplitude ($p < 0,001$). De effectgrootte is echter klein ($\eta_p^2 = 0,013$). De effectgrootte bij de factor Geslacht op tijdsduur is groter ($p < 0,001$; $\eta_p^2 = 0,15$) waarbij het gemiddelde bij mannen $0,54s$ ($\sigma = 0,12$) is en bij vrouwen $0,62s$ ($\sigma = 0,13$) is. Bij Dialect traden ook significante effecten op bij de PC-scores voor intonatie en tijdsduur, maar deze hadden allemaal een minieme effectgrootte $\eta_p^2 < 0,1$. Er werden op basis van de opsommingspositie geen resultaten gevonden die significant waren én een wezenlijke effectgrootte hadden ($\eta_p^2 > 0,1$). Bij het significante effect van de factor Woordvorm op PC2 intonatie ($p < 0,001$; $\eta_p^2 = 0,20$) is in figuur 6a te zien dat de PC2 score bij een klemtoon op de eerste syllabe hoger ligt dan bij de tweede.

Een ander GLM is opgesteld waarbij de factoren Geslacht en Dialect vervangen zijn door de factor Participant, hierin werd een significant effect gevonden bij de factor Participant op tijdsduur van $p < 0,001$; $\eta_p^2 = 0,42$, de overige effecten hadden bij deze factor een minieme effectgrootte. Hierna is nog een GLM opgesteld op basis van dezelfde tabel 2 waarbij Geslacht en Dialect vervangen zijn door de factor Controle die aangeeft of het woord wel of niet het controlewoord *Venlo* is. Dit model is uitgevoerd op woordrealisaties met hetzelfde woordkenmerk hieruit volgde significante effecten $p \leq 0,02$ van de factor Controle op alle afhankelijke variabelen: PC1 amplitude ($\eta_p^2 = 0,18$), PC2 amplitude ($\eta_p^2 = 0,02$), PC1 intonatie ($\eta_p^2 = 0,02$), PC2 intonatie ($\eta_p^2 = 0,07$) en tijdsduur ($\eta_p^2 = 0,10$). Het totale aantal (niet uitgesloten) realisaties van het controlewoord bedroeg $N = 322$, van dezelfde woordvorm (2, 1) zijn er $N = 211$ anderen gerealiseerd, dit aantal is te klein om op woordniveau vergelijkingen te doen waarbij positie- en participantkenmerken worden meegenomen.

De componentscores van de intonatiecontouren zijn uitgesplitst op positiekenmerk in tabel 3 en 4. De intonatiescores worden ook weergegeven in figuur 6c en 6d, hier is duidelijk te zien dat de scores bij de woorden aan het begin $x_1 \in S(n)$ positief zijn, terwijl die aan het einde $x_n \in S(n)$ juist negatief zijn. Er is hier echter nog meer te zien: bij een opsomming van twee woorden is het PC1-intonatiescoreverschil tussen het eerste en het laatste woord minder groot dan bij de langere opsommingen. Dit is ook een significant effect (independent samples t-test tussen $x_1 \in S(2)$ en $x_1 \in S(n > 2)$, $t = 4,98$; $p < 0,001$; $r = 0,20$). In figuur 6b zijn de PC1-scores van de eerste woorden te zien.

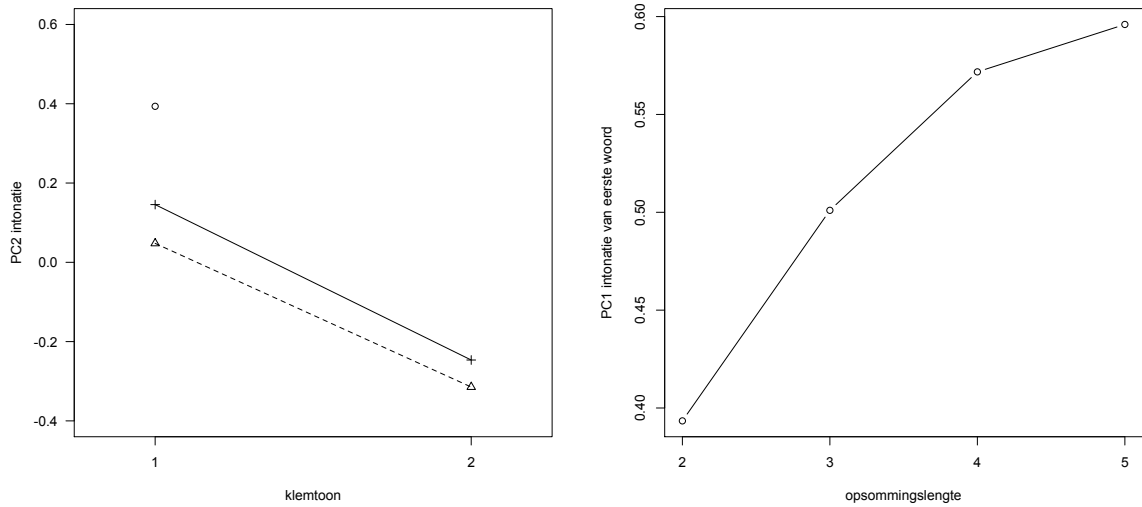
Hiernaast is in figuur 6c te zien dat op de PC1-score van de intonatie van een woord in

²SPSS versie 20 voor Mac

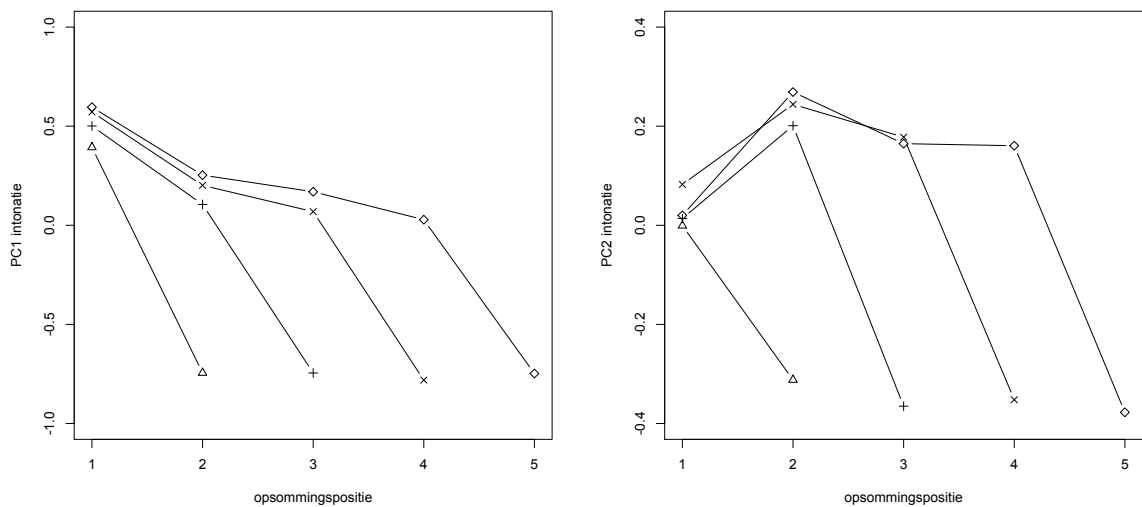
Tabel 2: Uitkomsten van de GLM over verschillende effecten op de FPCA-scores en tijdsduur, $N = 2576$.

Factor	Afhankelijke variabele	df	MS	F	p	η_p^2
Positiekennmerk	PC1 amplitude	13	0,14	0,76	0,70	0,004
	PC2 amplitude	13	9,70	66,25	< 0,001*	0,26
	PC1 intonatie	13	48,25	326,07	< 0,001*	0,62
	PC2 intonatie	13	10,22	38,34	< 0,001*	0,16
	Tijdsduur	13	0,20	19,94	< 0,001*	0,09
Woordvorm	PC1 amplitude	4	48,42	258,87	< 0,001*	0,29
	PC2 amplitude	4	19,23	135,20	< 0,001*	0,18
	PC1 intonatie	4	7,38	49,83	< 0,001*	0,07
	PC2 intonatie	4	41,18	154,50	< 0,001*	0,20
	Tijdsduur	4	3,04	311,21	< 0,001*	0,33
Geslacht	PC1 amplitude	1	0,39	2,06	0,15	0,001
	PC2 amplitude	1	2,62	18,42	< 0,001*	0,007
	PC1 intonatie	1	0,005	0,03	0,85	< 0,001
	PC2 intonatie	1	0,89	3,35	0,07	0,001
	Tijdsduur	1	4,51	462,26	< 0,001*	0,15
Dialect	PC1 amplitude	1	1,89	10,11	0,001*	0,004
	PC2 amplitude	1	0,001	0,005	0,95	< 0,001
	PC1 intonatie	1	0,47	3,17	0,08	0,001
	PC2 intonatie	1	6,63	24,86	< 0,001	0,01
	Tijdsduur	1	2,01	205,59	< 0,001	0,07
Opsommingspositie	PC1 amplitude	1	0,85	4,52	0,03	0,002
	PC2 amplitude	1	0,19	1,34	0,25	0,001
	PC1 intonatie	1	0,005	0,03	0,86	< 0,001
	PC2 intonatie	1	0,30	1,11	0,29	< 0,001
	Tijdsduur	1	0,87	89,29	< 0,001*	0,034

het midden van een opsomming ($x_2, \dots, x_{n-1} \in S(n)$), een lichte daling plaatsvindt die gevolgd wordt door een scherpere daling naar ongeveer $-0,75$. Hierbij wordt de PC1 voor x_2 hoger naarmate de opsomming langer wordt. Op het gebied van de PC2-scores is er vooral te zien dat



(a) De PC2-scores intonatie tegen woordkenmerk. (b) De PC1-scores intonatie van de eerste woorden. De stippellijn geeft de woorden aan met 2 syllaben, ($x_1 \in S(n)$) tegen de opsommingslengte n . de doorlopende lijn de woorden met 3 syllaben, en de stip geeft de woorden aan met 1 syllabe.



(c) PC1-scores tegen positiekenmerk. De verschillende lijnen geven verschillende opsommingslengtes aan. Een opsomming is zo lang als de hoogste woordpositie. (d) PC2-scores tegen positiekenmerk. De verschillende lijnen geven verschillende opsommingslengtes aan. Een opsomming is zo lang als de hoogste woordpositie.

Figuur 6. De gemiddelde PC1- en PC2-scores van de intonatiecurves bij verschillende woord- en positiekenmerken.

deze op het einde negatief worden rond de $-0,40$ en daarvoor positief zijn rond de $0,20$.

Tot slot is gekeken naar een partiële correlatie tussen de PC-scores voor de amplitude- en intonatiecontouren. Er is gecontroleerd voor woordkenmerk, positiekenmerk en volgnummer. Hieruit kwam dat er een significante correlatie $p = 0,002$ is tussen de PC1-score van amplitude en

Tabel 3: PC1-scores intonatie tegen positiekenmerk

woordpositie x_i	opsommingslengte $S(n)$							
	2		3		4		5	
	μ	σ	μ	σ	μ	σ	μ	σ
1	0,39	0,40	0,50	0,38	0,57	0,33	0,60	0,42
2	-0,75	0,41	0,11	0,37	0,20	0,35	0,26	0,35
3			-0,75	0,47	0,07	0,34	0,17	0,39
4					-0,78	0,49	0,03	0,38
5							-0,75	0,49

Tabel 4: PC2-scores intonatie tegen positiekenmerk

woordpositie x_i	opsommingslengte $S(n)$							
	2		3		4		5	
	μ	σ	μ	σ	μ	σ	μ	σ
1	-0,00	0,69	0,01	0,62	0,08	0,70	0,02	0,64
2	-0,31	0,40	0,20	0,56	0,24	0,63	0,27	0,58
3			-0,37	0,46	0,18	0,52	0,16	0,59
4					-0,35	0,51	0,16	0,61
5							-0,38	0,45

de PC2-score van amplitude en de PC1-score van intonatie waarbij deze correlatie resp. $r = 0,10$ en $r = -0,06$ zijn. Tussen de PC2-score van amplitude en PC1- en PC2-scores van de intonatie zijn ook significante correlaties van resp. $r = 0,24$ en $r = 0,16$.

Conclusies

Op basis van woordpositie in een opsomming is een sterk effect ($\eta^2 = 0,62$) gezien op de eerste componentscore die samenhangt met de daling of stijging van de intonatie. Er is te zien dat aan het eind van de opsomming een daling plaatsvindt en aan het begin van de opsomming een stijging. Dit is een resultaat dat overeenkomt met de gangbare kennis over de intonatie van het Nederlands (b.v. Keijsper (1984); Van Heuven (2004)) en is ook duidelijk zelf te horen bij het uitspreken van een opsomming.

Hiernaast zijn er specifiekere effecten vastgesteld bij deze intonatiecontouren: de stijging bij het eerste woord in een opsomming van twee woorden is minder hoog dan het eerste woord bij opsommingen van meer woorden. Voor het laatste woord in de opsomming is de daling

echter ongeacht de lengte van opsomming gelijk. In het midden van een opsomming treedt een subtieler effect op: het lijkt erop dat het tweede woord in de opsomming meer stijgt naarmate de opsomming langer wordt, terwijl het op één na laatste woord juist steeds vlakker wordt naarmate de opsomming langer wordt. Dit effect komt overeen met de waarneming van Van Heuven (2004) dat de eerste downstep in een opsomming kleiner wordt, naarmate de opsomming langer wordt. Hiermee wordt duidelijk dat bij het uitspreken van de opsomming al meteen wordt ingespeeld op de voor de spreker bekende lengte van de opsomming.

De woordvorm is ook een belangrijke factor ($\eta_p^2 = 0,20$) gebleken op de PC2 van de intonatiecontouren, dit is te zien in de vorm van een hogere intonatie aan het begin van een woord bij een klemtoon op de eerste syllabe. Dit zou veroorzaakt kunnen worden wanneer bij het realiseren van de klemtoon een hogere intonatie wordt gebruikt. De eerste syllabe en dus het begin van het woord zal dan op een hoger intonatie-niveau beginnen.

Hier buiten is het verder niet duidelijk of verschillende intonaties binnen de verschillende woorden van dezelfde woordvorm mogelijk zijn. De vergelijking van het controlewoord *Venlo* met de overige woorden met hetzelfde woordkenmerk liet zien dat de verschillen qua intonatie niet groot zijn ($\eta_p^2 < 0,2$). Desondanks is dat een vertekend beeld, omdat hierbij wordt gekeken hoe één woord (*Venlo*) zich verhoudt ten opzichte van het gemiddelde van de andere 4 woorden. Hier zou verder onderzoek naar gedaan kunnen worden. Dit zijn ook subtielere effecten die minder goed te modelleren zijn met het autosegmenteelmodel: het globale ‘downstep’-effect treedt overduidelijk hier op, maar onderlinge verschillen in intonatie van de woorden onderling moet hier dan uit iets anders worden verklaard dan het intonatiepatroon van een downstep.

Bij de andere factoren is uiteindelijk, na de normering en registratie, een beperkte invloed op de intonatie overgebleven. De effecten van geslacht, dialect en participant zijn soms significant (zie de resultaten en tabel 2) op intonatie, maar hebben dan niet een grote effectgrootte ($\eta_p^2 < 0,1$). Bij dialect zou dit vooral verklaard kunnen worden uit de participantselectie: de participanten hebben geen erg uiteenlopende dialecten. Bij de overige aspecten is het gebrek aan grote verschillen vooral te verklaren door de normering en registratie.

Discussie en vervolgonderzoek

Uit dit onderzoek blijkt dat het met een functionele analyse goed mogelijk is om kwantitatief verschillen in intonatiecontouren te analyseren en hoe die contouren samenhangen met meerdere factoren. Het voordeel van deze analyse is dat er hiermee wordt gekeken naar gehele contouren in plaats van slechts een beperkt aantal gemarkeerde datapunten. De uiteindelijk gevonden scores komen overeen met de herkenbare elementen van de ‘gestileerde’ aanpak (Van Heuven, 2004), inclusief de subtielere effecten zoals de verschillen qua intonaties tussen het begin en het eind van de opsomming. Net zoals bij het vergelijken van gemarkeerde datapunten worden de curves uiteindelijk op basis van een beperkt aantal cijfers vergeleken: de scores. Het verschil is dat deze componentscores vervolgens wel eenvoudig herleidbaar zijn tot *realistische* intonatiecontouren. Dit is een voordeel ten opzichte van het werken met gemarkeerde punten zoals in het gesegmenteerde model.

Het niet hoeven markeren van punten bij een intonatiecontouranalyse met behulp van een functionele data-analyse zorgt ervoor dat de subjectiviteit en observer bias verminderd kan worden. Het uitgangspunt van een ‘gesegmenteerd model’ heeft geen invloed op de daadwerkelijk gemeten waarden en afgeleide intonatiefuncties. Hier zijn echter wel nog de nodig punten die kennis van de onderzoeker zelf vereisen om tot resultaten te komen. Zo kan bij de intonatie-extractie gekozen worden voor verschillende algoritmen en instellingen van het intonatie-algoritme. Het zou ook mogelijk kunnen zijn om dit handmatig te verfijnen op basis van kennis van de onderzoeker. In dit geval is er voor gekozen om zo dicht mogelijk bij de standaardinstellingen te

blijven en later de uitschieters uit te sluiten. Dit zal echter bij kleinere datasets of bij minder geautomatiseerde processen misschien minder haalbaar zijn. Omdat er hierdoor misschien te weinig data overblijft.

Een ander punt waar sprake is van subjectiviteit, is bij de keuze voor een bepaalde intonatiecontourfunctie. Voor het omvormen van de datapunten naar een functie kan een bepaalde basisfunctie gekozen worden (in dit geval B-splines) welke eigenlijk al een voorafgesteld idee verwachten van de onderzoeker over hoe die functies er ongeveer eruit moeten zien. Bij het fitten aan deze B-splines moet vervolgens voor een bepaalde smoothingsvariabele gekozen worden, het aantal knopen en de orde van de B-splines. Dit zijn allemaal zaken die invloed hebben op de uiteindelijke vormen van de functies.

Bij de landmarking, registratie en analyse wordt er vanuit gegaan dat er een zekere ‘genormeerde’ functie bestaat die vergeleken kan worden. Kenmerken zouden moeten zijn uitgelijnd waarna hetgeen wat interessant is nog over blijft om te vergelijken. In dit onderzoek is hiervoor gebruik gemaakt van twee hulphypothesen. De eerste is dat de amplitudes van de woorden onderling niet veel verschillen, dat blijkt hier niet geheel op te gaan. De tweede is dat de samenhangende intonatie enkel op amplitude kan worden uitgelijnd. Het positiekenmerk heeft echter effect op de PC2-score van amplitude waardoor het duidelijk is dat amplitudecontouren van woorden niet invariant blijven. Ook was in de resultaten een effect van woordvorm op amplitude te zien. Het is duidelijk dat aan de manier van uitlijning bij vervolgonderzoek meer aandacht zou moeten worden besteed. Het is waarschijnlijk zuiverder om alleen woorden van een vergelijkbare vorm (of misschien zelfs precies hetzelfde woord) onderling te vergelijken.

Het niet langer uitlijnen aan één algemeen contour zou echter problematisch zijn: de resultaten binnen een groep (bijvoorbeeld dezelfde woordvorm) zijn dan niet meer één-op-één overzetbaar en vergelijkbaar met die van een andere groep. Dit komt omdat bij een (F)PCA de maximale variantie *binnen* de steekproef wordt verklaard. Die steekproef verschilt uiteraard, waardoor er voor het één-op-één kunnen gebruiken van deze scores in verder werk een normeringsslag dient plaatst te vinden. Eventueel zou een aangepaste vorm van ToDI (Gussenhoven et al., 2005) kunnen worden gebruikt waarbij saillante intonatiekenmerken gemarkeerd zijn – vergelijkbaar met hoe allofonen genoteerd kunnen worden. Hetzelfde onderscheid valt hier ook te maken: voor het simpelweg noteren van hoe een woord uitgesproken wordt, hoeft niet de exacte uitspraak te worden genoteerd maar volstaan doorgaans slechts de distinctieve klanken. Bij gangbaar onderzoek naar de rol van intonatie zullen deze FPCA-scores waarschijnlijk te specifiek en niet generaliseerbaar genoeg zijn.

Praktische zaken

Er zijn in de loop van dit onderzoek ook enkele praktische problemen aan het licht gevonden. Zo bleek bijvoorbeeld bij het starten van de afname van het onderzoek dat een trial met daarin enkele oefensequenties het voor de respondent begrijpelijker zou hebben gemaakt. Desondanks heeft dit uiteindelijk geen noemenswaardige problemen opgeleverd. Bij een vervolgonderzoek zou dit een goede toevoeging zijn, die de afnem zal vereenvoudigen.

Ten tweede bleek dat het handmatig annoteren van alle opnames zeer veel tijd in beslag nam. Het gebruik van een spraakherkenner (ASR) zoals bij Gubian et al. (2011) zou bij vervolgonderzoek aan te raden zijn.

Daarnaast bleek bij de uiteindelijke analyse ook dat de landmarkregistratie veel tijd in beslag nam. Tijdens dit proces traden een aantal randgevallen op die leiden tot een crash van de berekening, dit kwam bijvoorbeeld door verkeerd uitgesloten opnames of extreme waarden in de intonatiepatronen. Deze crashes leidde tot vertraging van het onderzoek. Het oplossen van dit soort zaken kan namelijk zeer veel tijd in beslag nemen, simpelweg omdat het opnieuw

uitvoeren van de mogelijk verbeterde code weer uren kan duren. Zo duurde het vinden van alle eerste pieken bij de landmarkregistratie bijna drie uur. Hiervoor werd een stuk heuristisch toegepast, deze oplossing was niet analytisch uitgevoerd wat waarschijnlijk wel had gekund. Een analytische oplossing zou in R waarschijnlijk een aanzienlijke tijds winst hebben opgeleverd. Het is daarom ook aan te raden om dit bij eventueel vervolgonderzoek mee te nemen.

Tot slot was een van die struikelblokken bij deze analyse dat het niet altijd mogelijk is om betrouwbare intonatiepatronen te verkrijgen. Het proces is erop gericht om eerst het intonatiepatroon van de gehele opname te verkrijgen vanuit Praat, om dit vervolgens op te knippen voor de verschillende woordrealisaties. In dit intonatiepatroon kunnen gaten zitten die bij de extrapolatie bij het omzetten naar B-splines tot extreme waarden kunnen leiden. Dit leidt dan tot uitschieters, die later weer moeten worden weggefilterd.

Toepassingen en verder onderzoek

Vergelijkbaar met Gubian et al. (2010) zou de data van dit onderzoek rechtstreeks kunnen worden ingezet om semi-geautomatiseerd manipulatie van intonatiecontouren uit te voeren. De vereiste intonatiepatronen zouden nu simpelweg kunnen worden herleid uit de gevonden intonatiecurves en scores. Dit model is, in ieder geval binnen dit dialect, generaliseerbaar over verschillende participanten en zou naast de positie in de opsomming slechts rekening hoeven te houden met de woordvorm. Bij deze manipulaties zou zowel gekeken kunnen worden naar het manipuleren van de opnames gemaakt in het kader van dit onderzoek, als naar andere corpora zoals het AUTONOMATA-corpus (Van den Heuvel et al., 2008). In alle gevallen zou gekeken kunnen worden naar de gevolgen voor perceptie en de kwaliteit van deze manipulaties.

Deze analyse zou ook kunnen worden gebruikt voor tekst-naar-spraaksystemen, waarbij deze scores gebruikt zouden kunnen worden voor het ‘on-the-fly’ genereren van een bruikbaar intonatiecontour. De vergelijking met systemen gebaseerd op gesegmenteerde intonatie is dan niet ver weg. Het voordeel van deze methode is echter dat het automatisch intonatiepatronen kan vaststellen zonder dat er vooraf categorieën moet zijn gedefinieerd. Het stijgen-dalenonderscheid was in de componentfuncties te vinden niet omdat het een uitgangspunt was in de theorie, maar omdat die vorm simpelweg de meeste variantie verklaarde in de data. Bij het toepassen van deze techniek op een groot corpus zou ‘automatisch’ een uitgebreid model van intonatie ontstaan. De toepassingen daarvan zouden eindeloos zijn.

Nawoord

Allereerst zou ik Paul Boersma willen bedanken voor zijn uitgebreide begeleiding tijdens mijn afstudeerscriptie. Hiernaast gaat mijn dank uit naar Gwenda Frederiks voor haar hulp en ondersteuning. Verder zou ik Sjanie Quist willen bedanken voor de uitleg over de stationsomroepen. Tot slot wil ik al mijn participanten bedanken voor hun deelname aan het onderzoek.

Referenties

- Aston, J. A., Chiou, J.-M., & Evans, J. P. (2010). Linguistic pitch analysis using functional principal component mixed effect models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(2), 297–317.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the institute of phonetic sciences* (Dl. 17, pp. 97–110).
- Boersma, P., & Kovacic, G. (2006). Spectral characteristics of three styles of Croatian folk singing. *The Journal of the Acoustical Society of America*, 119, 1805.

- Boersma, P., & Weenink, D. (2013). *Praat: doing phonetics by computer (version 5.3.52)*. Verkrijgbaar van <http://www.praat.org>
- Caspers, J. (2000). Experiments on the meaning of four types of single-accent intonation patterns in dutch. *Language and speech*, 43(2), 127–161.
- Dombrowski, E., & Niebuhr, O. (2010). Shaping phrase-final rising intonation in German. In *in Proceedings of the 5th International Conference on Speech Prosody, Chicago, Illinois, USA*.
- Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), pp. 89-102. Verkrijgbaar van <http://www.jstor.org/stable/2246049>
- Fodor, I. K. (2002). *A survey of dimension reduction techniques*. Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory.
- Foley, J. D., Van Dam, A., Feiner, S. K., & Hughes, J. F. (1996). *Introduction to computer graphics* (2e dr.). Addison-Wesley Reading. (Herdruk. 1997)
- Goldsmith, J. A. (1976). *Autosegmental phonology* (Dl. 159). Cambridge, MA: Massachusetts Institute of Technology.
- Gubian, M. (2011). Functional Data Analysis for Phonetic Research. *Very-Large-Scale Phonetics Research 2011*.
- Gubian, M., Boves, L., & Cangemi, F. (2011). Joint analysis of F_0 and speech rate with Functional Data Analysis. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4972–4975).
- Gubian, M., Cangemi, F., Boves, L., et al. (2010). Automatic and data driven pitch contour manipulation with functional data analysis. In *Proceedings of speech prosody*. Chicago, IL, USA:[sn].
- Gussenhoven, C. (1991). Tone segments in the intonation of dutch. In *The Berkley conference on Dutch linguistics, University Press of America, Lanham, MD* (pp. 139–155).
- Gussenhoven, C., et al. (2005). Transcription of dutch intonation. *Prosodic typology: The phonology of intonation and phrasing*, 118–145.
- Gussenhoven, C., Rietveld, T., Kerkhoff, J., & Terken, J. (2003). *Transcription of Dutch Intonation* (2e dr.). Verkrijgbaar van <http://todi.let.kun.nl/>
- Van 't Hart, J., & Collier, R. (1975). *Integrating different levels of intonation analysis*.
- Van den Heuvel, H., Martens, J.-P., D'hoore, B., D'hanens, K., & Konings, N. (2008). The AUTONOMATA spoken names corpus. In *LREC*. European Language Resources Association.
- Van Heuven, V. J. (2004). Planning in speech melody: production and perception of downstep in Dutch. *On speech and language: Studies for Sieb G. Nooteboom*, 83–93.
- D'Imperio, M. (2001). Focus and tonal structure in Neapolitan Italian. *Speech Communication*, 33(4), 339–356.
- Jacobi, I., Pols, L. C., & Stroop, J. (2005). Polder Dutch: Aspects of the /ei/-lowering in standard Dutch. In *Interspeech* (p. 2877-2880). ISCA.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187–200.
- Keijsper, C. E. (1984). Vorm en betekenis in Nederlandse toonhoogtecontouren I. In *Forum der Letteren* (Dl. 25, pp. 20–37).

- Van Leeuwen, H. C., & Te Lindert, E. (1993). Speech Maker: a flexible and general framework for text-to-speech synthesis, and its application to Dutch. *Computer Speech & Language*, 7(2), 149–167.
- Mertens, P., Beaugendre, F., & d’Alessandro, C. (1997). Comparing approaches to pitch contour stylization for speech synthesis. *Progress in Speech Synthesis*, 347–364.
- Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. In *Proceedings of the 15th international congress of phonetic sciences* (Dl. 771, p. 774).
- R Core Team. (2013). R: A Language and Environment for Statistical Computing [Handleiding van computersoftware]. Vienna, Austria. Verkrijgbaar van <http://www.R-project.org>
- Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and MATLAB* (1e dr.). Springer Publishing Company, Incorporated.
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2e dr.). Springer Publishing Company, Incorporated.
- Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication*, 6(4), 343–352.
- Rose, P. (1991). How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency? *Speech Communication*, 10(3), 229–247.
- Van Rossum, G. (2013). *Python v3.3.2 documentation*. Verkrijgbaar van <http://docs.python.org/3/>
- Terken, J. (1993). Synthesizing natural-sounding intonation for Dutch: rules and perceptual evaluation. *Computer Speech & Language*, 7(1), 27–48.
- Turco, G., Gubian, M., & Schertz, J. (2011). A quantitative investigation of the prosody of verum focus in italian. In *Interspeech* (p. 961-964). ISCA. Verkrijgbaar van <http://lands.let.ru.nl/FDA/FDA.Tutorials.htm>
- Van Dale. (2013). *Woordenboek Nederlands* (4e dr.). VBK Media.
- Zellers, M., Gubian, M., & Post, B. (2010). Redescribing intonational categories with functional data analysis. In *Proceedings of interspeech 2010*. Makuhari, Japan:[sn].

Appendix

Woordenbank

- | | |
|---------------------------------|---------------------------------|
| 1 syllabe | • Venlo |
| • Goes | • Tilburg |
| • Tiel | • Rijswijk |
| • Soest | • Hoofddorp |
| • Vught | 2 syllaben, accent op syllabe 2 |
| • Sneek | • Maassluis |
| 2 syllaben, accent op syllabe 1 | • Uitgeest |
| • Haarlem | • Roermond |

- Delfzijl
- Breda
3 syllaben, accent op syllabe 1
- Deventer
- Hilversum
- Vlissingen
- Doetinchem
- Enschede
3 syllaben, accent op syllabe 2
- Abcoude
- Enkhuizen
- Hoogkarspel
- Almere
- Rosmalen

Draagzinnen

- Ik heb gewoond in: ...
- Hij heeft vrienden uit: ...
- Zij werkte in: ...
- We gingen op vakantie naar: ...
- Ik ken mensen uit: ...
- De band speelt in: ...
- Zij hebben winkels in: ...

Begrippen

De begrippen en bijbehorende variabelenamen die gebruikt zijn.

Controlewoord *control word* Venlo

Draagzin *carrier*

Opsomming *enumeration*

Woordenbank *word index* Alle woorden die beschikbaar zijn om te kiezen. Deze woorden kunnen gefilterd zijn op woordvorm en in willekeurige volgorde worden getrokken.

Positiekenmerken *position features* De combinatie van de opsommingslengte (c.q. het aantal woorden in de opsomming) en positie binnen die opsomming. Wordt aangegeven door deze twee punten binnen haakjes. Bijvoorbeeld: (2, 1) voor het eerste woord in een opsomming van twee woorden.

Woordkenmerken *word features* De combinatie van het aantal syllaben en stresspositie. Wordt aangegeven door deze twee punten binnen haakjes. Bijvoorbeeld: (2, 1) voor een woord van twee syllaben met het accent op de eerste lettergreep.

Stimuluszin *sentence* Een enkele zin bestaande uit een draagzin en een aantal woorden.

Stimuluszinslengte *sentence length* Het aantal woorden uit de woordenbank dat voorkomt in de stimuluszin. De woorden in de draagzin en het voegwoord worden buiten beschouwing gelaten.

Uitgesloten woordrealisaties

In totaal: 224 (8% van alle woordrealisaties).

Tabel 5: Uitgesplitst per reden. Percentage ten opzichte van totaal (2800).

	Reden	<i>N</i>	%
	aarzeling	9	0,32%
	geen intonatie of amplitude verkregen	11	0,39%
	uitschieter, amplitude	33	1,18%
	uitschieter, amplitude & intonatie	1	0,04%
	uitschieter, intonatie	109	3,89%
	uitschieter, na landmarking	23	0,82%
	verkeerde klemtoon	36	1,29%
	voortijdig beëindigd	2	0,07%

Tabel 6: Uitgesplitst per respondent en opsommingspositie. *Per respondent*: Percentage ten opzichte van totaal respondent (280). *Per opsommingspositie*: Uitgesplitst op volgorde van inspreken door de respondent. Iedere respondent heeft 80 zinnen met ieder één opsomming ingesproken, deze zijn hieronder gegroepeerd op tientallen. Het percentage is ten opzichte van het totale aantal woordrealisaties binnen deze tien opsommingsposities (280).

Respondent	<i>N</i>	%	Opsommingspositie	<i>N</i>	%
1	59	21,07%	1 - 10	32	11,42%
2	9	3,21%	11 - 20	32	11,42%
3	17	6,07%	21 - 30	28	10,00%
4	7	2,50%	31 - 40	32	11,42%
5	25	8,92%	41 - 50	23	8,21%
6	27	9,64%	51 - 60	27	9,64%
7	33	11,79%	61 - 70	24	8,57%
8	14	5,00%	71 - 80	26	9,29%
9	10	3,57%			
10	23	8,21%			