

Foneemduren binnen RDS-TMC

Citation for published version (APA):

Rutten, J., de Pijper, J. R., & Boves, L. (1995). *Foneemduren binnen RDS-TMC*. (IPO rapport; Vol. 1061). Instituut voor Perceptie Onderzoek (IPO).

Document status and date:

Gepubliceerd: 17/07/1995

Document Version:

Uitgevers PDF, ook bekend als Version of Record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Rapport no. 1061

Foneemduren binnen RDS-TMC

Joost Rutten

Foneemduren binnen RDS-TMC

Door: Joost Rutten

Juli 1995

(c) 1995 IPO Instituut voor Perceptie Onderzoek Eindhoven

Begeleiders: IPO: Dr. Jan Roelof de Pijper

KUN: Prof. dr. L. Boves

Voorwoord

Het onderzoek voor deze scriptie heb ik uitgevoerd aan het Instituut voor Perceptieonderzoek in Eindhoven. Hierbij wil ik mijn begeleider Jan Roelof de Pijper bedanken voor zijn regelmatige steun en kritiek.

Dank ook aan mijn kamergenoot Remco Teunen, voor zijn hulp met de noodzakelijke programmaatjes welke nodig waren voor de experimenten en het verwerken van de resultaten. Daarnaast bedank ik ook mijn begeleider aan de Katholieke Universiteit Nijmegen, Dhr. L. Boves.

Tenslotte bedank ik alle andere medewerkers, afstudeerders en stagiaires op het IPO voor hun medewerking aan de experimenten en de leuke tijd.

Samenvatting

Het hier beschreven onderzoek is verricht in het kader van het RDS-TMC-project. Met het RDS-TMC-systeem in de autoradio is verkeersinformatie altijd opvraagbaar. Verkeersinformatie wordt gecodeerd overgezonden naar de autoradio. Deze code wordt gebruikt om een gesproken boodschap te genereren waarbinnen de betreffende plaatsnamen gezet worden. De plaatsnamen worden omgezet in synthetische spraak. Hierbij moeten de duren van de fonemen en de prosodie beregeld worden.

In dit onderzoek heb ik me gericht op de duren van de fonemen. Naar de duurberegeling van de Nederlandse versie was tot nog toe nog nooit serieus gekeken. Daarom ben ik dit onderzoek begonnen met te kijken wat er zoal voorhanden was op het gebied van duurberegeling.

In het vooronderzoek heb de in mijn ogen belangrijkste duurregels getest en gekeken of er enige regelmaat te vinden is in foutief beoordeelde duren.

Met deze achtergrond en aanpassingen op grond van beluistering, is een duurregelset ontworpen welke in het eindexperiment onderzocht is. Geconcludeerd kan worden dat de beregeling van de duren van fonemen belangrijker is naarmate de gesynthetiseerde uiting langer wordt. Bovendien is gebleken dat de uiteindelijk verkregen duurregelset beter is dan een regelset met alleen de belangrijkste duurregels erin opgenomen.

Inhoudsopgave

Hoofdstuk 1 Inleiding	1
1.1 Inleiding.....	1
1.2 RDS-TMC	1
1.3 PTS	5
1.3.1 Inleiding.....	5
1.3.2 Het duurberegelingformalisme	7
1.4 Doel van dit afstudeerproject.....	8
Hoofdstuk 2 De duurregels	9
2.1 Het ontwerpen van een duurregelset.	9
2.2 Bekende duurregels	10
2.3 De duurregels in het vooronderzoek.....	11
Hoofdstuk 3 Het vooronderzoek	12
3.1 Doel	12
3.2 Opzet.....	12
3.3 Uitvoering.....	15
3.4 Verwerking van de resultaten	15
3.5 Resultaten	17
3.5.1 Inleiding.....	17
3.5.2 De klemtoonregel.....	17
3.5.3 Woordfinale.....	18
3.5.4 De ruimtes tussen de gesynthetiseerde uitingen	19
3.5.5 Het aantal syllabes en fonemen in een woord	19
3.5.6 De prevocale consonant in prepausale positie	20
3.6 Conclusies.....	20
Hoofdstuk 4 De duurregels in het eindonderzoek	23
Hoofdstuk 5 Het eindonderzoek	26
5.1 Doel	26
5.2 Opzet.....	26
5.3 Uitvoering.....	29
5.4 Verwerking van de resultaten	30
Hoofdstuk 6 Conclusies en aanbevelingen	34
Referenties	36
Appendix A Foneem-notaties	37
Appendix B Luisterexperiment vooronderzoek	38
Appendix C De initiële duren	39
Appendix D De 26 testzinnen in het eindexperiment	40
Appendix E Luisterexperiment	41

Hoofdstuk 1 Inleiding

1.1 Inleiding

Spraak kan op vele manieren gezien worden. Bijvoorbeeld fysisch, als continue stroom luchtdrukverschillen, of linguïstisch/fonetisch, als een serie klanken. In de spraaksynthese wordt spraak meestal behandeld als opeenvolgende klanken. Deze klanken beïnvloeden elkaar. Daarbij zijn de overgangen tussen klanken vaak typerender voor de betreffende klanken dan het stationaire deel van een klank zelf. Daarom worden ook deze overgangen gebruikt als eenheden voor synthetische spraak. Een voorbeeld hiervan zijn difonen. Deze omvatten klankovergangen uit natuurlijke spraak. Deze overgangen zijn geknipt van het stationaire deel van de ene klank tot het stationaire deel van de tweede klank. Waarbij het stationaire deel van een klank dát deel is waarin de klank nagenoeg niet veranderd binnen korte tijd.

Als we bijvoorbeeld de uiting 'bot' willen maken met difonen, dan zullen we de volgende difonen achter elkaar moeten plakken:

si-b, b-o, o-t, t-si

Waarbij si voor silence staat, de stilte voor en na een los uitgesproken woord.

Maar met het aan elkaar plakken van deze eenheden heb je nog geen acceptabele spraak. Spraak bevat ook toonhoogte- en duurvariatie. Een zin bevat ritme en variatie in toonhoogte en luidheid (van Os, 1988). Daarnaast kan een zin snel of langzaam uitgesproken worden (tempo van de zin). Deze factoren in de spraak heten prosodische of suprasegmentele factoren en zijn gerelateerd aan één of meer akoestische grootheden in het spraakgeluid. Zo staat intonatie (toonhoogteverloop) in verband met de stembandfrequentie, tempo met duur en luidheid met intensiteit. Ritme staat met alle drie deze factoren in verband. Bovendien draagt een goede duurberegeling bij aan de natuurlijkheid en betekenis van synthetische spraak. Vergelijk voor dit laatste de uitingen 'beet' en 'bit', welke voornamelijk verschillen in de duur van de vocaal.

In dit onderzoek zal ik mij bezig houden met de duurberegeling van klanken in synthetische spraak oftewel met ritme en tempo in spraak. De duurberegeling beperkt zich tot de duren van fonemen in Nederlandse plaatsnamen in de RDS-TMC-applicatie, welke in de volgende paragraaf besproken zal worden.

1.2 RDS-TMC

RDS-TMC staat voor Radio Data System, Traffic Message Channel. Het Radio Data System omvat een aantal kanalen op de radio waarover codes uitgezonden kunnen worden. Deze codes worden gewoon met de rest van het signaal meegezonden door zendstations. RDS bestaat al langer en wordt bijvoorbeeld al gebruikt om namen van radiostations mee te zenden, zodat een luisteraar weet naar welk station hij of zij luistert.

Een nieuw kanaal op dit systeem is het Traffic Message Channel. Via dit kanaal kunnen verkeersboodschappen gecodeerd uitgezonden worden en opgeslagen in de autoradio. De gebruiker kan dan op elk gewenst moment de meest recente verkeersinformatie

opvragen. Het systeem werkt als volgt. Door detecteersystemen onder het asfalt (welke nog aangelegd moeten worden), worden files automatisch gedetecteerd. Deze automatische file-meldingen tezamen met meldingen van andere bronnen (bijvoorbeeld de ANWB) worden verzameld en gecodeerd. Deze gecodeerde informatie wordt verstuurd naar een radiozender met RDS-TMC service. Deze zender zendt de gecodeerde verkeersinformatie naar autoradio's. Autoradio's met het ingebouwde RDS-TMC systeem slaan deze informatie op. De gebruiker kan nu op elk moment de verkeersinformatie opvragen. Hij of zij kan zelf bepalen over welk gebied de informatie moet gaan.

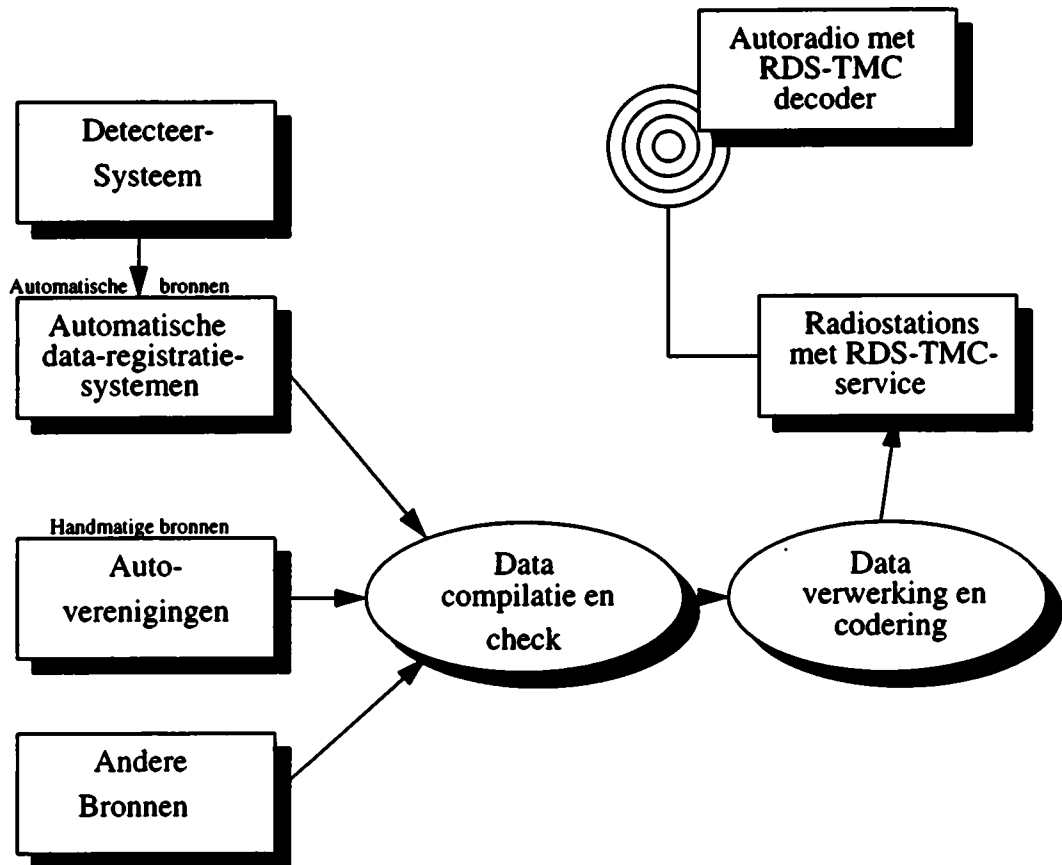
Het RDS-TMC-systeem heeft grote voordelen ten opzichte van de huidige overdracht van verkeersinformatie:

- De verkeersinformatie is op elk gewenst moment op te vragen.
- De verkeersinformatie is altijd actueel.
- Alleen relevante verkeersinformatie zal vermeld worden.
- De verkeersinformatie kan in verschillende talen opgevraagd worden.
- De verkeersinformatie is door bovengenoemde voordelen uitstekend geschikt voor toekomstige navigatie-systemen.

Het idee van een 'Traffic Message Channel' is al eerder gebruikt in het *Rhein-Corridor* project. Dit is een sub-project van het internationale systeem, welke alleen in Nederland bruikbaar was. In Frankrijk heeft men als sub-project *Cities*, welke bruikbaar is rond Parijs. In Duitsland onderzoeken ze de mogelijkheden van het Traffic Message Channel aan de hand van het *Bevei*-project, welke draait om de verkeersinformatie rond Stuttgart. Deze drie pilot-projecten vormen de basis van het RDS-TMC-systeem.

In het Rhein-Corridor project is niet echt veel aandacht geschonken aan de duurregels. De inherente duren zijn afhankelijk van de (toevallige) lengtes van de difonen, waarmee wel de invloed van direct omliggende klanken meegenomen is.

Schematisch kan het RDS-TMC systeem als volgt weergegeven worden:



De verkeersinformatie zit als volgt in elkaar:

Elke draagzin heeft zijn eigen code. Een draagzin bestaat uit:

- De weg waar de gebeurtenis plaatsvindt,
- De twee plaatsnamen waar deze weg tussen ligt met de richting,
- De specifieke plaats,
- De gebeurtenis zelf.

Elke plaatsnaam heeft ook zijn eigen code, welke samen met de event-code overgezonden wordt naar een autoradio met het RDS-TMC-systeem erin. Op deze manier kunnen gebeurtenissen op de weg als:

A12, Arnhem richting Utrecht, tussen afrit Ede/Wageningen en Veenendaal, 2 kilometer file

uitgesproken worden door de radio, door slechts een beperkt aantal codes over te zenden.

Het aantal gebeurtenissen dat plaats kan vinden is een relatief kleine eindige verzameling. Daardoor hoeven de duren en het intonatiepatroon daarvan in het toekomstige systeem niet beregeld te worden, maar kunnen in plaats van deze beregelingen, de waardes die gebaseerd zijn op metingen in natuurlijke spraak opgeslagen worden in het systeem ('copy prosody'). In de eerste versie(s) van het systeem zullen de draagzinnen bestaan uit natuurlijke spraak.

Het aantal plaatsnamen is echter veel groter. Bovendien kan het zijn dat er later plaatsnamen toegevoegd moeten worden. Hierdoor kunnen de plaatsnamen niet opgeslagen worden met hun duren en intonatiepatroon. Er is gekozen voor een chipkaartje met daarop de verrijkte foneemrepresentaties van de plaatsnamen.

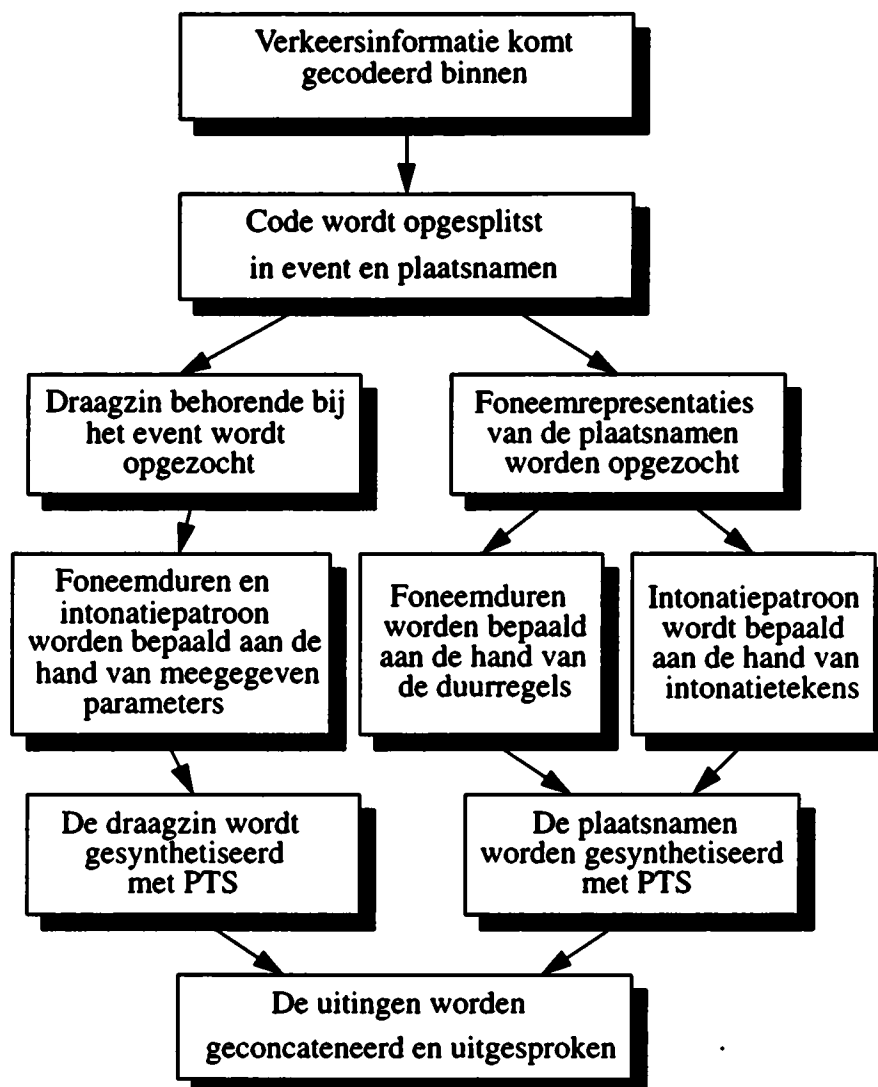
Dat de plaatsnamen los op een chipkaartje staan heeft twee voordelen:

1. Nieuwe plaatsnamen kunnen eenvoudig toegevoegd worden op het chipkaartje.
2. Het chipkaartje kan vervangen worden door een chipkaartje met buitenlandse plaatsnamen, zodat verkeersinformatie in het buitenland op te vragen is in de eigen taal. Dit is mede mogelijk doordat internationaal afspraken gemaakt zijn over de codes die bij een bepaalde gebeurtenis horen.

De plaatsnamen staan dus in verrijkte foneemrepresentatie op een chipkaartje en moeten om verklankt te worden helemaal worden gesynthetiseerd. Om deze reden zullen de duren van de fonemen en het intonatiepatroon van de plaatsnamen beregeld moeten worden.

Naast het verschil in verzenden van verkeersinformatie, is de autoradio zelf ook aangepast. Er is een systeem in geplaatst dat iets met de overgezonden codes kan doen.

Dit (toekomstige) systeem kan als volgt geschetst worden (waarbij de draagzin ook gesynthetiseerd is):



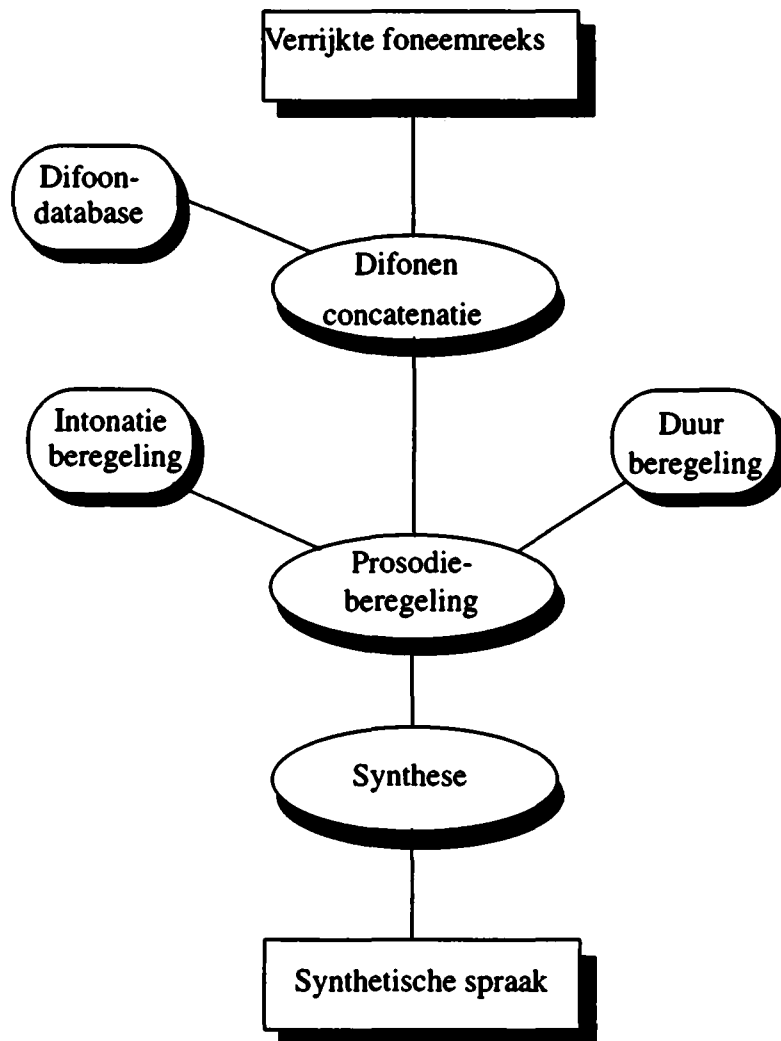
In dit schema is te zien dat de uitingen gesynthetiseerd worden door het PTS-systeem. Dit systeem wordt beschreven in de volgende paragraaf.

1.3 PTS

1.3.1 Inleiding

Het PTS-systeem is een programma dat foneemreeksen omzet in een spraak-file. Tijdens dit eindonderzoek was PTS nog in ontwikkeling, waardoor het systeem soms aangepast moest worden.

Het systeem bestaat uit een aantal modules, welke schematisch als volgt voorgesteld kunnen worden:



De invoerstring

De foneemreeks die uitgesproken moet worden, staat in een losse file. Door deze file te veranderen kunnen dus eenvoudig nieuwe uitingen gegenereerd worden. De foneemreeks is verrijkt met tekens voor het intonatiepatroon en de duurberegeling.

De prosodie

De prosodie wordt beregeld door twee losse modules: de duurmodule en de intonatie-module. Het intonatiepatroon van een uiting wordt beregeld met parameters die in een losse file staan. De duren van de fonemen worden beregeld door duurregels die in een losse file staan. Over de duurberegeling later meer.

Concatenatie en synthese

Difonen worden geconcateneerd en beregeld aan de hand van de gevonden duren en het intonatiepatroon. Dit gebeurt met PSOLA.

1.3.2 Het duurberegelingformalisme

In een bestand staat beschreven hoe de duren beregeld worden. PTS gebruikt deze regels om in de aangeboden foneemstring de afzonderlijke klanken de juiste duur te geven. Het bestand waarin de duurregels gezet moeten worden bestaat uit twee gedeeltes.

1. In het eerste gedeelte worden de initiële duren gegeven. Dit zijn de beginwaarden van de lengtes van de fonemen. Deze initiële duren vormen de basis voor de regels die toegepast gaan worden, en zijn dus erg belangrijk.
2. In het tweede gedeelte van het bestand staan de regels die fonemen verlengen of verkorten. Deze regels zien er als volgt uit:

$\langle F1, \dots, Fn; L1, \dots, Ln; R1, \dots, Rn; s; f \rangle +/=/\% x$

- Hierin zijn F1 tot en met Fn de fonemen waarop een regel betrekking heeft. Dit zijn dus de fonemen die eventueel verkort of verlengd worden.
- Met L1 tot Ln wordt de linkercontext aangegeven van het te behandelen foneem.
- De rechtercontext wordt gegeven door R1 tot Rn. De contexten kunnen ook leeg zijn. Hiermee kan aangegeven worden dat de context niet van belang is. In plaats van in de context alle fonemen op te sommen, kan dus een spatie ingevuld worden.
- Met s wordt aangegeven of het foneem in een beklemtoond woord ($s=1$) of juist een onbeklemtoond ($s=0$) woord moet staan. Als voor s geen waarde gegeven wordt, dan geeft dit aan dat het niet uitmaakt of het te behandelen foneem beklemtoond is of niet.
- f geeft aan of het foneem in finale positie staat of niet (0 of 1). Ook deze waarde kan weer onbepaald blijven, waarmee aangeduid wordt dat het wel of niet finaal zijn van het foneem niet van belang is. Een foneem wordt beschouwd als finaal als deze in de laatste lettergreep van het laatste woord van de uiting staat. In de uiting 'Rotterdam Terneuzen' zijn alleen de fonemen 'z', '@' en 'n' uit de lettergreep 'zen' finaal.
- Met x kunnen 4 mogelijkheden aangegeven worden;
 1. +x wil zeggen: verleng het foneem met x ms.
 2. -x wil zeggen: verkort het foneem met x ms.
 3. =x wil zeggen: maak het foneem x ms lang.
 4. %x wil zeggen: maak het foneem x% van de huidige waarde lang.

Voorbeeld:

De regel:

$\langle t; ;@; ;1 \rangle \%120$

zegt dat de 't' met 20% verlengd moet worden, als deze gevolgd wordt door het foneem '@' en in prepausale positie staat. De linkercontext is daarbij niet van belang.

Deze regel verlengt dus de 't' in 'rotte'.

Dit formalisme is beperkt in zijn mogelijkheden:

- Er kan slechts gekeken worden naar één foneem tegelijkertijd, naar de aangrenzende fonemen, naar het wel of niet beklemtoond zijn en naar het wel of niet finaal zijn.
- In de context kunnen alleen de direct aangrenzende fonemen gespecificeerd worden. Hierdoor is de clusterberegeling beperkt. Fonemen in een consonantcluster van twee consonanten zullen hetzelfde beregeld worden als fonemen in een consonantcluster bestaande uit drie consonanten.
- Beregeling op grond van het aantal syllabes, of het aantal fonemen per syllabe, kan in de huidige situatie niet gerealiseerd worden.
- Ook werken de regels niet met zogenaamde features, wat inhoudt dat steeds een hele reeks fonemen opgesomd moet worden. Dit zal de leesbaarheid van de duurregels niet ten goede komen, maar is geen echte beperking.

1.4 Doel van dit afstudeerproject

Het doel van dit onderzoek is het bepalen van een optimale duurregelset binnen RDS-TMC. Met optimaal bedoel ik de verhouding tussen de kwaliteit van de spraak en het aantal duurregels. De duurberegeling vindt plaats in het kader van het RDS-TMC-systeem. Dit systeem maakt (vooralsnog) voor de spraaksynthese gebruik van PTS, wat inhoudt dat de beregeling van de foneemduren binnen dit systeem plaatsvindt. Gekeken zal worden of er aanpassingen nodig zijn aan het formalisme ten behoeve van de duurberegeling. Dit doe ik voornamelijk omdat het aantal syllabes waaruit een plaatsnaam bestaat vermoedelijk een grote invloed heeft op de duur die de fonemen zouden moeten hebben. Mede daarom zal een vooronderzoek opgezet worden, met alleen enkele basale duurregels, welke uit de literatuur gehaald zullen worden. Onder basale duurregels versta ik duurregels die vaak toegepast worden.

Samengevat zal er in het vooronderzoek op drie dingen gelet worden:

1. Hoe goed of slecht werken de basale duurregels.
2. Zijn er regelmatigigheden te ontdekken op grond waarvan nieuwe duurregels gemaakt kunnen worden.
3. In hoeverre voldoet het formalisme.

Uiteindelijk zal worden bekeken of de set met alle gevonden duurregels (dus de aangepaste basale duurregelset met alle gevonden duurregels; literatuur en introspectie) daadwerkelijk beter is dan de op grond van het pilot-experiment aangepaste basale duurregelset.

Hoofdstuk 2 De duurregels

2.1 Het ontwerpen van een duurregelset.

Een duurregelset ontwerpen kan op verschillende manieren aangepakt worden.

Zo zou men in de literatuur kunnen duiken en alle regelmatigheden die gevonden zijn in de duren van klanken van natuurlijke spraak beschrijven in termen van het formalisme waarin je werkt. Echter als al deze bevindingen zouden leiden naar de perfecte duurregelset, dan zou het onnodig zijn nog een duurregelset te maken, dan moet alleen een goed formalisme gevormd worden, waarin al deze duurregels geïmplementeerd zouden kunnen worden. Kortom, er moet meer gebeuren om een betere duurregelset te maken dan er op dit moment al voorhanden is. Nieuwe data voor aanpassingen in de duurregels is op verschillende manieren te verkrijgen.

- Je zou in stukken natuurlijke spraak de duren kunnen meten en hierin zoeken naar nieuwe regelmatigheden.
- Een andere mogelijkheid, welke voor dit onderzoek gebruikt is in het RDS-TMC (PTS)-systeem, is door te luisteren naar de gesynthetiseerde spraak, die beregeld is met de meest bekende duurregels. Zoek in deze spraak naar fonemen die niet kloppen wat hun duur betreft, en probeer regels te formuleren welke die fouten corrigeren.

Bijkomend probleem bij de duurberegeling is het feit dat er een interactie is tussen het intonatieverloop van een uiting en de duren van de fonemen. Het lijkt dus voor de hand te liggen het intonatiepatroon uit natuurlijke spraak over te nemen. Dit patroon zou echter puur afhankelijk zijn van één uitspraak, waaruit de duren dan ook rechtstreeks overgenomen zouden moeten worden, omdat anders bepaalde stijgingen en dalingen in het intonatieverloop wellicht niet afgemaakt kunnen worden. En het is nu juist niet de bedoeling duren rechtstreeks over te nemen uit spraak, maar alleen de regelmatigheden te beschrijven in duurregels.

Een ander probleem is het formalisme. Hoe moet het formalisme aangrijpen op de fonemen en hoe moeten afhankelijkheden tussen fonemen onderling opgenomen worden? Moet het formalisme werken op foneem-, syllabe- of zinsniveau?

Daarnaast zijn er afhankelijkheden tussen intonatie- en duurberegeling (verlenging beklemtoonde vocaal). In hoeverre moeten deze afhankelijkheden worden opgenomen in het formalisme en moeten deze afhankelijkheden nog gestuurd kunnen worden door de gebruiker van het formalisme, of moet de afhankelijkheid hard in de code van het formalisme opgenomen worden.

Binnen het PTS-formalisme wordt alleen gekeken naar losse fonemen. Naar het aantal syllabes in een woord, frase of zin wordt niet gekeken. Afhankelijkheden tussen intonatie en duren van fonemen worden indirect beregeld, doordat zowel het duurformalisme als het intonatieformalisme gebruikt maakt van bepaalde symbolen in de aangeboden foneemreeks.

2.2 Bekende duurregels

Als achtergrond voor dit onderzoek heb ik gezocht naar overeenkomsten in eerder ontworpen duurregels, te weten, de duurregels van Klatt (Klatt, 1979, 1987), Van Coile (van Coile, 1990, pagina 3.52-3.62) en uit NIROS. Klatt is in mijn ogen de grondlegger van het onderzoek naar duren van fonemen. Van Coile en NIROS zijn in het kader van dit onderzoek belangrijk, omdat het onderzoek betreft naar duurregels voor het Nederlands.

- Klatt is één van de eersten geweest die zich heeft beziggehouden met de regelmatigheden van verkorting en verlenging van fonemen, op basis van hun context. Hij heeft dit gedaan voor de Amerikaans/Engelse taal, maar duurregels blijken veel overeenkomsten te vertonen voor verschillende talen. Hij heeft een model gemaakt waarmee duren van fonemen te beschrijven zijn en heeft daar een aantal aannames voor gemaakt:

1. Elk fonetisch segment heeft een inherente duur.
2. Elke regel verandert de duur van een segment een bepaald percentage.
3. Segmenten kunnen nooit korter worden dan een bepaalde minimumduur.

Het model is samengevat in de volgende formule:

$$DUR = MINDUR + \frac{(INHDUR - MINDUR) \times PRCNT}{100}$$

Waarbij MINDUR de minimumduur is van een foneem, INHDUR de inherente duur en PRCNT het percentage waarmee een foneem verlengd of verkort wordt. Klatt heeft de duurregels opgesteld aan de hand van duurregels die gebruikt zijn in verschillende spraaksystemen. Voorbeelden van systemen met duurregels van Klatt zijn Dectalk en Klattalk (Klatt, 1979, 1987).

- Van Coile heeft op basis van de lengtes van fonemen in een gesproken tekst van acht en halve minuut een model gemaakt. Hij heeft hierbij gelet op de plaats van een foneem, of een foneem beklemtoond was of niet, en op het aantal syllabes dat het woord had waar een foneem zich in bevond. De gevonden regelmatigheden zijn verwerkt in het LEM-tekst-naar-spraak-systeem. (van Coile, 1990).
- Hiernaast is naar de set duurregels gekeken die in Nijmegen ontwikkeld is; NIROS. NIROS is de afkorting van Nijmegen Interactive Rule Oriented Speech Synthesis. Het is een beschrijving van een verzameling duurregels die in de loop der tijd onderzocht zijn. Bij elke duurregel is vermeld wat de bron van de regel is, en wat de regel precies doet. De regels van NIROS worden gebruikt in het Polyglot-spraakstelsel.

De belangrijkste duurregels zijn de volgende: (de fonemen staan in de aangepaste sampanotatie, Appendix A):

1. De /i/, /u/ en /y/ worden verlengd voor de /r/.
2. Beklemtoonde vocalen worden verlengd, beklemtoonde vocalen in een geaccentueerd woord worden nog sterker verlengd.
3. Prepausale verlenging: Een vocaal of consonant in een syllabe die voor een pauze staat wordt verlengd.
4. Clusterverkorting: Opeenvolgende consonanten worden verkort.
5. Polysyllabische verkorting: Fonemen in een woord bestaande uit meerdere syllabes worden verkort.

Uiteraard zijn er meer regelmatigheden bekend wat betreft de duren van fonemen, maar dit lijken mij de meest belangrijke omdat ze in elk onderzoek naar duren weer naar voren komen. Bovendien zijn het regels die vrijwel in elke uiting toepasbaar zijn.

Regels 1 tot en met 4 zijn te realiseren binnen het PTS-formalisme. Polysyllabische verkorting is niet te beregelen in het huidige systeem. Zou uit het vooronderzoek blijken dat polysyllabische verkorting voor significante verbeteringen kan zorgen, dat zal het systeem aangepast moeten worden.

2.3 De duurregels in het vooronderzoek

De volgende regels worden getoetst in het vooronderzoek:

1. Allereerst natuurlijk de initiële duren van de fonemen.
2. De /u/ (boer), /i/ (bier) en /y/ (buur) worden verlengd vóór de /r/.
3. Beklemtoonde vocalen worden verlengd, behalve de schwa (@).
4. Een syllabe in prepausale positie wordt verlengd.

Alle beklemtoonde vocalen in de plaatsnamen hebben naast klemtoon ook zinsaccent, omdat de plaatsnamen in geaccentueerde positie staan. Vocalen in geaccentueerde positie worden meestal ook iets verlengd. Maar omdat in zijn algemeenheid niet geldt dat beklemtoonde vocalen ook in geaccentueerde positie staan en in PTS alleen klemtoon aan te geven is, is hier niets mee gedaan. Vandaar dat ik verder spreek over de klemtoonregel en niet over de accentsregel.

De vraag is nu hoeveel de duren verlengd moeten worden. Omdat het systeem van Van Coile is gebaseerd op één spreker, lijkt me dit geen goede basis voor algemene duurregels. De regels van NIROS heb ik uiteindelijk gebruikt als basis van dit experiment, omdat deze toch in hoge mate overeenkomen met de regels van Klatt, maar actueler en uiteraard enigszins gebaseerd op de regels van Klatt.

Hoofdstuk 3 Het vooronderzoek

3.1 Doel

Het vooronderzoek is een verkennend onderzoek. Ik zal hierin onderzoeken of de twee belangrijke regels, te weten de klemtoonregel en de prepausaalregel, de foneemduren goed beregelen. Bovendien zal ik onderzoeken of er regelmaat te vinden is in foneemduren die niet goed beregeld worden en of het formalisme misschien aangepast moet worden.

3.2 Opzet

De in paragraaf 2.3 genoemde duurregels zijn van de NIROS-duurregelset omgezet naar het formalisme zoals dat opgenomen is in het PTS-systeem. De initiële duren zijn ook overgenomen uit NIROS. Concreet ziet de duurberegeling er als volgt uit (de foneemen staan in de aangepaste sampa-notatie, appendix A):

```
1.<i,y,u; ;r; ; > +20
2.<2,3,4,5,9,A,E,I,O,Y,a,e,i,o,u,y; ; ;1; > +20
3.<t,k,p,s,f,x,N,n,m; ; ; ;1> %140
   <2,3,4,5,9,@,A,E,I,O,Y,a,e,i,o,u,y; ; ; ;1> +35
   <l,r,j,w; ;#; ;1> =120
```

Regel 1 verlengt de duren van de /i, y, u/ met 20 milliseconden als deze voor een /r/ staan (bier, buur, boer).

De tweede regel is de klemtoonregel. Deze verlengt een vocaal met lexicaal accent met 20 milliseconden.

De prepausaalregel bestaat uit 3 subregels:

- Stemloze plofklanken, stemloze fricatieven en nasalen, worden verlengd met 40% van hun huidige waarde, als ze in prepausale positie staan.
- Vocalen in prepausale positie worden met 35 milliseconden verlengd.
- De glides (/l, r, j, w/) worden 120 milliseconden lang als ze in de allerlaatste positie staan van een woord in prepausale positie.

Regel 1 zal altijd werken. Omdat regels 2 en 3 de regels zijn die getest worden, zullen deze regels de experimentele variabelen zijn. Deze regels worden soms wel en soms niet toegepast, onafhankelijk van de positie en/of accentuering van een vocaal. Met deze twee variabelen zijn vier varianten mogelijk van de duurregelset die hierboven beschreven is:

1. [-klem, -prep]
2. [+klem, -prep]
3. [-klem, +prep]
4. [+klem, +prep]

Waarbij 'klem' voor de klemtoonregel en 'prep' voor de prepausaalregel staat. Een minnetje wil zeggen dat de betreffende regel niet werkzaam is in de variant en een plusje wil zeggen dat de regel wel werkzaam is in de variant.

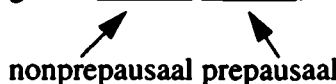
Met deze vier varianten zijn plaatsnamen gesynthetiseerd. De plaatsnamen die gesynthetiseerd zijn, zijn gekozen uit de plaatsnamen die in het Rhein-Korridor project gebruikt zijn. Dit zijn de volgende plaatsnamen:

Amsterdam, Ede, Eindhoven, Den Haag, Doetinchem, Driebergen, Heerlen, Hellevoetsluis, Hengelo en Tiel.

Getracht is zoveel mogelijk variatie in de plaatsnamen mee te nemen. Zo zijn er lange (Hellevoetsluis) en korte (Ede, Tiel), vooraan beklemtoonde (Ede, Hengelo) en achteraan beklemtoonde (Amsterdam, Den Haag). Hiernaast heb ik ook gelet op de bekendheid van de plaatsnamen, omdat het aannemelijk is dat een proefpersoon de duren van de fonemen van deze plaatsnamen beter kan beoordelen als hij weet hoe de plaatsnaam normaal uitgesproken wordt.

Deze gesynthetiseerde plaatsnamen worden in de volgende draagzin geplaatst:

Snelweg A10 _____, over een lengte van 22 kilometer file.

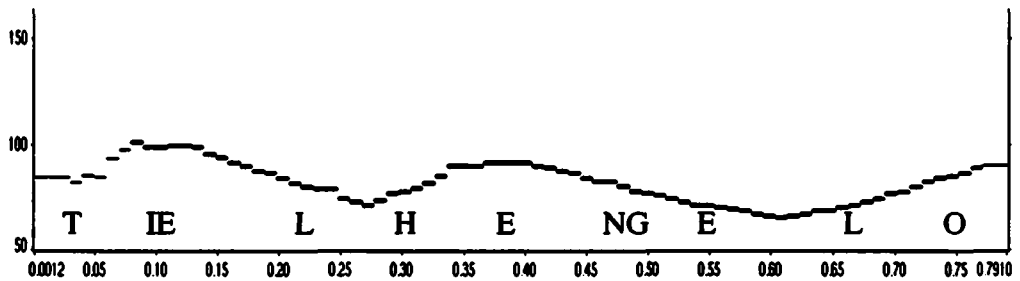


Deze draagzin biedt plaats aan twee plaatsnamen, waarvan de eerste in niet-prepausale positie staat en de twee in prepausale positie.

De twee gesynthetiseerde plaatsnamen die in één zin voorkomen, zullen als één geheel gesynthetiseerd worden. Op deze manier worden ook de overgangen tussen de twee plaatsnamen beregeld en getest. Daarnaast gaat het beregelen van de intonatiecontour dan ook goed. Zou je de twee gesynthetiseerde plaatsnamen namelijk apart synthetiseren en concateneren, dan zou je twee frases krijgen, met een dalend intonatiepatroon, een reset en nog eens hetzelfde patroon. Dit is echter niet juist, de twee plaatsnamen samen zijn één frase, met pieken in het intonatiepatroon op de beklemtoonde vocalen en een eindstijging welke de zware grens na de tweede plaatsnaam aangeeft. Tussen de plaatsnamen in ligt een lichte grens, welke niet prosodisch gerealiseerd is.

Het intonatiecontour dat gebruikt is, is het zogenaamde zaagtandmodel. Dit is een contour met een snelle stijging voor de beklemtoonde vocaal, en een daling tot aan de volgende beklemtoonde vocaal (Terken, 1993) (zie figuur 1).

Het intonatiepatroon van de gesynthetiseerde plaatsnamen 'Tiel Hengelo' ziet er als volgt uit:



FIGUUR 1.

Normaal gesproken verlengt de klemtoonregel een vocaal als deze beklemtoond is en de prepausaalregels werken op fonemen in de laatste syllabe voor een pauze. Dit is nu echter niet de bedoeling; de klemtoonregel en de prepausaalregels moeten onafhankelijk van de context werken. De klemtoonregel kan niet uitgeschakeld worden door het klemtoonsymbool (/') weg te halen, omdat de intonatieberegeling ook gebruik maakt van dit symbool in de invoerstring. Het toepassen van de klemtoonregel en de prepausaalregel moet dus losgekoppeld worden van daadwerkelijke fonologische context.

Dit probleem heb ik opgelost door de betreffende regels te koppelen aan een zelfgedefinieerde linkercontext. De klemtoonregel wordt nu toegepast als het betreffende foneem voorafgegaan wordt door een /+/. Voor fonemen die verlengd moeten worden door de prepausaalregels geldt dat ze voorafgegaan moeten worden door een /-/. Moet een foneem verlengd worden door beide regels, dan moet er een /=/ voor staan.

Op deze manier kan in de foneemreeks bepaald worden welke regels er moeten werken op welk foneem. De aangepaste duurberegeling ziet er als volgt uit:

1. <i,y,u; ;r; ; > +20
2. <2,3,4,5,9,A,E,I,O,Y,a,e,i,o,u,y;+; ; ; > +20
3. <t,k,p,s,f,x,N,n,m;-; ; ; > %140
 <2,3,4,5,9,@,A,E,I,O,Y,a,e,i,o,u,y;-; ; ; > +35
 <1,r,j,w;-; ; ; > =120
- 2+3.<2,3,4,5,9,A,E,I,O,Y,a,e,i,o,u,y;=; ; ; > +55
- 2+3.<@;=; ; ; > +35

Als nu bijvoorbeeld de uiting 'Amsterdam Hengelo' gesynthetiseerd moet worden, met voor 'Amsterdam' de klemtoonregel en de prepausaalregels en voor 'Hengelo' geen regels, dan moet de volgende foneemreeks meegegeven worden:

/Amst@rd"=A-m#h"EN@lo/

Op de tweede /A/ moeten zowel de klemtoonregel als de prepausaalregels werken, zodat de /A/ in totaal 55 milliseconden verlengd wordt. Deze verlenging wordt nu in één keer gedaan en wel door de /=/ voor de /A/.

Stimuli

Nu kan dus begonnen worden met het maken van de testzinnen. Er zijn tien plaatsnamen, vier varianten van de duurregelset en twee in te vullen posities in de draagzin. Hiermee zouden 400 verschillende testzinnen gemaakt kunnen worden. Dit zijn er veel te veel om te testen. Bovendien vind ik het niet wenselijk dat in één zin, twee keer dezelfde plaatsnaam voorkomt.

Om het aantal testzinnen terug te brengen naar een aantal dat in een redelijke tijd te beoordelen is door de proefpersonen, heb ik besloten dat een plaatsnaam niet in combinatie hoeft te staan met elke andere plaatsnaam. Elke plaatsnaam moet alleen met alle vier de varianten van de regelset op allebei de posities voorkomen. De eis dat een plaatsnaam niet in één testzin op allebei de posities mag staan, blijft gehandhaafd.

Op deze wijze zijn 40 testzinnen gemaakt; elke plaatsnaam is vier keer beregeld en kan op twee posities staan. Dit zijn 80 mogelijkheden, maar omdat er in elke zin twee plaatsnamen staan, zijn er dus 40 testzinnen.

3.3 Uitvoering

De op bovenstaande manier verkregen zinnen werden zowel voorgelegd aan ervaren luisteraars (IPO-medewerkers) als niet-ervaren luisteraars. Hierdoor kon worden vergeleken of ervaren luisteraars de zinnen hetzelfde beoordelen als de niet-ervaren luisteraars. Het zou namelijk zo kunnen zijn dat de lengtes zeer moeilijk te beoordelen zijn voor niet-ervaren proefpersonen, in welk geval je niets met de resultaten zou kunnen doen. Het zou ook zo kunnen zijn dat getrainde proefpersonen niet objectief beoordelen. Dat ze weten waar op te letten, zonder echt op de natuurlijkheid te letten. In dat geval zouden de resultaten van de ervaren proefpersonen niet bruikbaar zijn.

Aan dit experiment deden twaalf proefpersonen mee, waarvan zes getrainde luisteraars (medewerkers van het IPO) en zes naïeve luisteraars. Het experiment is individueel afgenomen in de studio van het IPO met behulp van een computer. De zinnen werden aangeboden door een hoge-kwaliteits hoofdtelefoon en de luisteraars konden zelf bepalen wanneer ze de zin startten en hoe vaak ze een zin wilden beluisteren. Zie appendix B voor de instructies van dit experiment. De beoordeling moesten ze invullen op een scoreformulier. Daarop stonden de veertig zinnen met onder elk foneem een ruimte om een plusje, minnetje of nulletje in te vullen. Een plusje moest worden ingevuld als ze vonden dat het foneem langer moest, een nulletje als ze het foneem goed van lengte vonden en een minnetje moest ingevuld worden als ze vonden dat het foneem korter zou moeten.

3.4 Verwerking van de resultaten

Eerst heb ik voor elk foneem het aantal gescoorde minnetjes, nulletjes en plusjes geturfd. Tussen de twee groepen (ervaren en niet-ervaren) bleken geen extreme verschillen te zitten, zodat ik de twee groepen bij elkaar genomen heb. Op deze manier is

één tabel verkregen met voor elk foneem het aantal gescoorde plussen, nullen en minnen. Dit zag er dan bijvoorbeeld als volgt uit (tabel 1):

TABEL 1. Aantal minnen, nullen en plussen per foneem

Zin l	T	ie	l	-	A	m	s	t	e	r	d	a	m
-	2	1	1		6	4	2		1	2	7	3	1
0	9	11	11	12	5	8	1	12	11	10	5	1	2
+	1				1		9					8	9

Hierna heb ik een nieuwe variabele genaamd 'score' gedefinieerd, omdat er behoefte is aan één getal dat een weergave is van de oordelen per foneem.

Deze variabele is per foneem als volgt berekend:

$$Score = \left(\frac{Nplus - Nmin}{Nmin + Nnul + Nplus} \right)$$

Waarbij Nplus het aantal gescoorde plussen voor het foneem is, Nmin het aantal minnen en Nnul het aantal nullen. Het aantal minnen is afgetrokken van het aantal plussen: Dit getal geeft aan hoeveel plussen er zijn in verhouding tot het aantal minnen. Dit getal is vervolgens gedeeld door het totaal aantal beoordelingen, zodat een getal tussen -1 (iedereen heeft een min ingevuld) en +1 (iedereen heeft een plus ingevuld) verkregen wordt.

Als de score nul is vinden evenveel mensen het foneem te kort als te lang. Deze nulwaarde is echter afhankelijk van het spreektempo van de draagzin, waarmee geen rekening gehouden is bij de synthese. Er zal een referentiewaarde berekend moeten worden, welke rekening houdt met het spreektempo van de draagzin. Hoe sneller de draagzin uitgesproken wordt, hoe sneller de synthetische uiting uitgesproken zal moeten worden, dus hoe lager het referentiepunt. Hiervoor zijn de scores van de fonemen met een bepaalde eigenschap (bijvoorbeeld: vocaal in beklemtoonde positie) genormaliseerd, en is de referentiewaarde het gemiddelde van deze scores. De fonemen waarnaar gekeken wordt zijn nu opgedeeld in groepen met een bepaalde positie en een bepaalde beregeling (bijvoorbeeld: vocalen in beklemtoonde positie die daadwerkelijk verlengd zijn door de klemtoonregel) is nu de gemiddelde z-waarde genomen ten opzichte van het referentiepunt. Deze z-waarde is de afstand in standaarddeviaties ten opzichte van het referentiepunt, met welke verder gerekend zal worden. Deze variabele zal ik in het verdere vooronderzoek eindoordeel noemen. De variabele eindoordeel loopt grof gezegd van -3 tot +3 en geeft aan hoeveel standaarddeviaties de beoordeling van het referentiepunt af ligt. Hoe dichter het eindoordeel van een bepaalde groep fonemen bij nul ligt, hoe beter de luisteraars het foneem vonden. Ligt het eindoordeel onder nul, dan zou dit foneem ten opzichte van de lengtes van andere fonemen iets korter moeten. Ligt het eindoordeel boven nul, dan zou de lengte iets langer moeten zijn.

3.5 Resultaten

3.5.1 Inleiding

In dit vooronderzoek heb ik me gericht op de volgende items, dit naar aanleiding van het doel van het vooronderzoek, en opvallend gescoorde fonemen:

1. De beklemtoonde vocalen. Op deze vocalen werkt de klemtoonregel.
2. De laatste fonemen in de gesynthetiseerde uitingen. Hiermee wordt de prepausaalregel getest.
3. De ruimtes tussen de twee gesynthetiseerde uitingen. Dit is gedaan omdat uit de tabel met resultaten blijkt dat deze ruimte vaak te kort bevonden wordt.
4. Het aantal fonemen en syllabes in een plaatsnaam. Dit is gedaan om te kijken of het opvalt dat verkorting of verlenging op grond van het aantal syllabes in een woord weggelaten is.
5. De prevocale consonant in prepausale positie. Dit is gedaan omdat in de literatuur onenigheid is over het beregelen van de prevocale consonant in prepausale positie.

3.5.2 De klemtoonregel

Om te onderzoeken in hoeverre het nodig is dat de beklemtoonde vocalen verlengd worden door de klemtoonregel, heb ik gekeken naar de eindoordelen van deze vocalen. Daaruit is tabel 2 komen rollen:

TABEL 2. Eindoordelen van de beklemtoonde vocaal

	PREPAUSAALREGEL			
	niet verlengd		wel verlengd	
	KLEMTOONREGEL		KLEMTOONREGEL	
	niet verlengd	wel verlengd	niet verlengd	wel verlengd
VOCAAL NIET IN LAATSTE SYLLABE POSITIE				
Niet-prepausaal	0.02 (n=6)	-1.0 (n=6)	-0.23 (n=6)	-0.23 (n=6)
Prepausaal	0.32 (n=6)	0.12 (n=6)	0.76 (n=6)	-0.18 (n=6)
VOCAAL WEL IN LAATSTE SYLLABE POSITIE				
Niet-prepausaal	0.66 (n=4)	0.22 (n=4)	-0.38 (n=4)	-0.83 (n=4)
Prepausaal	0.66 (n=4)	0.37 (n=4)	-0.16 (n=4)	0.07 (n=4)

Deze tabel verdient wellicht enige toelichting. Bovenaan staan de beregelingen die toegepast zijn. Dit zijn er twee (klemtoonberegeling en prepausaalberegeling), waardoor er vier mogelijkheden (kolommen) zijn. Links in de tabel is onderscheid gemaakt in de positie van de beklemtoonde vocalen. Er is onderscheid gemaakt in de positie in de zin (prepausaal of nonprepausaal) en positie in een woord (wel in laatste syllabe of niet in laatste syllabe). Hierbij dient opgemerkt te worden dat vocalen die in een prepausaal

woord staan, maar niet in de laatste syllabe van dat woord, nooit verlengd worden door de prepausaalregels.

De vetgedrukte getallen zijn de voorkeursberegelingen volgens de proefpersonen (eindoordeel dichtst bij nul).

Op deze manier is te zien wat het perceptieve effect van de duurregels is als functie van de positie van de vocaal. Zo valt uit de eerste rij af te lezen dat een beklemtoonde vocaal in een niet-finale syllabe van een nonprepauaal woord (bijvoorbeeld de 'E' uit hEN@lo#til), niet verlengd moet worden door de klemtoonregel. De prepausaalregel zou hier eigenlijk geen invloed moeten hebben, daar deze nooit aan kan grijpen op deze vocaal. Als de prepausaalregel echter de vocaal in de laatste syllabe wel verlengt, dan zou de beklemtoonde vocaal korter moeten (-0.23). Dit zou te verklaren zijn door aan te nemen dat er naar gestreefd wordt bepaalde spraakeenheden (fonemen, syllabes, woorden) een vaste lengte te laten zijn, zodat wanneer de finale vocaal gerekt wordt, een ander gekort wordt.

Voor de prepauale beklemtoonde vocaal die niet in de laatste syllabe staat (en dus ook niet daadwerkelijk verlengd wordt door de prepausaalregel!), geldt dat deze wel verlengd moet worden door de klemtoonregel (0.12). Bovendien is er een lichte voorkeur voor het niet verlengen van de laatste syllabe in die plaatsnaam door de prepausaalregel (0.12 vs. -0.18). Ook hier doet zich weer het verschijnsel voor, dat wanneer de laatste syllabe gerekt wordt door de prepausaalregel, een andere syllabe korter zou moeten.

Als de vocaal wel in de laatste syllabe staat is er in niet-prepauale positie een voorkeur voor het wel verlengen door de klemtoonregel en niet verlengen door de prepausaalregel (0.22). In prepauale positie is de duur het best als de vocaal zowel door de klemtoon- als door de prepausaalregel verlengd wordt (0.07).

In tabel 3 heb ik de vocalen opgesplitst in lange (/a, e, o/) en korte (alle andere) vocalen. Hieruit valt af te lezen dat de korte vocaal beter niet verlengd kan worden door de klemtoonregel in beklemtoonde positie.

TABEL 3. De eindoordelen van de beklemtoonde vocalen, opgesplitst in lange en korte vocalen

	KLEMTTOONREGEL	
	niet verlengd	wel verlengd
LANGE VOCAAL	0.37 (n=12)	-0.11 (n=12)
KORTE VOCAAL	0.05 (n=28)	-0.31 (n=28)

3.5.3 Woordfinale

De prepausaalregel heb ik ook onderzocht aan de hand van het allerlaatste foneem in een uiting. Dit is bij 'Tiel' de /L/ en bij 'Hengelo' de /O/. Dit is onderzocht omdat de allerlaatste foneem in prepauale positie volgens mij in een bijzondere positie staat. Deze foneem sluit een frase af, zonder over te gaan in een ander foneem.

De resultaten hiervan staan in tabel 4:

TABEL 4. Eindoordelen van het laatste foneem in de uiting

POSITIE	PREPAUSAALREGEL	
	niet verlengd	wel verlengd
niet-prepauzaal	0.25 (n=20)	-0.63 (n=20)
prepauzaal	0.46 (n=20)	-0.08 (n=20)

In niet-prepauzale positie ligt het eindoordeel het dichtst bij nul als de prepauzaalregel niet toegepast is. In prepauzale positie is het eindoordeel het best als het foneem wel verlengd wordt door de prepauzaalregel.

3.5.4 De ruimtes tussen de gesynthetiseerde uitingen

Omdat opvallend veel mensen de ruimtes tussen de twee gesynthetiseerde plaatsnamen te kort vonden, leek het me interessant om te kijken of de omliggende fonemen van invloed zijn op de duurbeoordeling. In dit verband is het belangrijk te bedenken dat beide plaatsnamen in geaccentueerde positie staan, omdat beide plaatsnamen kernwoorden zijn in de uiting. Daarnaast moet bedacht worden dat na de tweede plaatsnaam een zware grens ligt, wat uitgedrukt wordt in een reset van de toonhoogtecontour en een pauze. Tussen de plaatsnamen in ligt echter ook een grens, zij het een lichte, juist omdat beide plaatsnamen kernwoorden zijn.

Deze lichte grens is in dit materiaal niet prosodisch gerealiseerd. Je zou je dus voor kunnen stellen dat het ontbreken van de pauze minder acceptabel gevonden wordt. Alleen als een woord begint met een vocaal, moet een glottal stop geplaatst worden, welke gezien kan worden als een woordscheider (vgl. Tiel-Ede met ledemaat). Tabel 5 geeft aan dat de er een grotere stilte tussen de uitingen zou moeten als het tweede woord begint met een vocaal. Met andere woorden, de glottal stop moet langer.

TABEL 5. De eindoordelen van de woordscheiders

RECHTERCONTEXT	
vocaal	0.69 (n=12)
consonant	-0.30 (n=28)

Als de tweede plaatsnaam begint met een consonant, dan zou de (niet aanwezige) pauze in de overgang korter moeten. Met andere woorden, iets wat er niet is, zou korter moeten. Dit zou er op kunnen duiden dat er geen pauze hoeft te zijn in de overgang; Mocht er wel een pauze gerealiseerd worden, dan moet deze wel lang genoeg zijn.

3.5.5 Het aantal syllabes en fonemen in een woord

Om te onderzoeken of de duurbeoordelingen afhankelijk zijn van het aantal fonemen en/of syllabes in een woord, heb ik de eindoordelen vergeleken met het aantal syllabes en fonemen. Hieruit zijn de tabellen 6 en 7 ontstaan, waarbij elk eindoordeel het gemiddelde is van de eindoordelen per foneem uit een plaatsnaam. Dus bij één syllabe

is het gemiddelde eindoordeel bepaald van alle eindoordelen van fonemen uit elk woord bestaande uit één syllabe (alleen Tiel);

TABEL 6. Gemiddelde eindoordelen als functie van het aantal syllabes in een woord

AANTAL SYLLABES	Eindoordeel
1	1.27 (n=24)
2	0.39 (n=120)
3	-0.33 (n=320)
4	-0.76 (n=88)

Uit tabel 6 valt af te lezen dat woorden die bestaan uit één syllabe duidelijk te kort bevonden worden (het eindoordeel ligt 1.27 standaarddeviatie boven nul). Woorden bestaande uit 4 syllabes worden te lang bevonden (-0.76). In deze tabel is duidelijk te zien dat de plaatsnamen te lang bevonden worden naarmate het aantal syllabes toeneemt.

3.5.6 De prevocale consonant in prepausale positie

In de regelset die in dit experiment gebruikt is, wordt de duur van de prevocale consonant in de laatste syllabe van een plaatsnaam (bijvoorbeeld de /d/ in Amsterdam) niet beïnvloed door de prepausaalregels. Om te kijken hoe de duren van deze fonemen waren beoordeeld heb ik de volgende tabel gemaakt:

TABEL 7. Eindoordelen van de prevocale consonant in woordfinale syllabe

POSITIE	PREPAUSAALREGEL	
	niet verlengd	wel verlengd
niet-prepausaal	0.20 (n=20)	-0.41 (n=20)
prepausaal	-0.01 (n=20)	0.22 (n=20)

Uit deze tabel blijkt dat de betreffende consonant in beide gevallen iets verlengd zou moeten worden (bedenk dat in prepausale positie de prepausaalregel wel moet werken, zie 3.5.2 en 3.5.3). De eindoordelen zijn echter dermate klein, dat het onwaarschijnlijk is dat daadwerkelijke verlenging van de prevocale consonant in prepausale positie voor verbetering zorgt.

3.6 Conclusies

Als eerste valt het op dat de eindoordelen alle dicht bij nul liggen, verreweg de meeste eindoordelen liggen tussen de -0,5 en +0,5. Dit betekent dat de inherente duren in ieder geval vrij goed gekozen zijn. Daarnaast kan het ook betekenen dat de duurregels weinig effect hebben, of dat luisteraars niet heel erg goed in staat zijn duren te beoordelen, waardoor veel nullen ingevuld zijn. Uit de resultaten blijkt echter dat de duurregels wel degelijk effect hebben, en dat de proefpersonen zeker wel in staat waren de duren te beoordelen (zie bijvoorbeeld tabel 6), zodat de conclusie getrokken kan worden dat de inherente duren goed gekozen zijn. Daarnaast betekent dit ook dat de invloed van de duurregels niet zeer extreem is.

Verder is in dit vooronderzoek gekeken naar de klemtoonregel. In de meeste gevallen beregelt de klemtoonregel de duren van beklemtoonde vocalen goed. Als de beklemtoonde vocaal echter niet in de laatsie syllabe van een woord in niet-prepauzale positie staat (bijvoorbeeld de /i/ in 'Driebergen' als deze in niet-prepauzale positie staat), dan zou de klemtoonregel de vocaal niet moeten verlengen.

In dezelfde paragraaf is gekeken naar de klemtoonregel in combinatie met lange en korte vocalen. Proefpersonen blijken er de voorkeur aan te geven als korte beklemtoonde vocalen (alle vocalen behalve /a, e, o/) niet verlengd wordt door de klemtoonregel. Lange vocalen (/a, e, o/) worden wel goed beregeld. Als algemene conclusie kan dus gesteld worden dat er een verdere opsplitsing moet komen in plaats en lengte van de vocalen in de klemtoonregel.

In paragraaf 3.5.2 is ook gekeken naar de prepauzale regel die werkzaam is op de beklemtoonde vocalen, in combinatie met de plaats van deze vocalen. Hieruit bleek inderdaad wat verwacht kan worden, dat de prepauzale regel die de vocalen verlengd alleen werkzaam moet zijn als deze vocalen in de laatste syllabe staat van een woord dat in prepauzale positie staat. In paragraaf 3.5.3 is gekeken naar de laatste foneem in een plaatsnaam. Onderzocht is of de proefpersonen er inderdaad de voorkeur aan geven als deze fonemen alleen verlengd zouden worden als ze in prepauzale positie staan. Dit bleek ook het geval. Hieruit kan de conclusie getrokken worden dat de prepauzale regel zoals deze nu is, de juiste fonemen met de juiste duren verlengt.

De glottal stop die op het woordbegin staat, is te kort. In bijvoorbeeld de uiting 'Tiel Ede', begint 'Ede' met een glottal stop om aan te geven dat er een woordgrens is. Deze glottal stop moet in deze situatie (tussen twee plaatsnamen) langer, omdat deze twee plaatsnamen een belangrijke betekenis hebben in de zin. Er is een lichte grens tussen de twee plaatsnamen, welke raar genoeg alleen benadrukt dient te worden met een pauze, als de tweede plaatsnaam met een vocaal (glottal stop) begint. Als de tweede plaatsnaam met een consonant begint, zou de overgang, welke niet gerealiseerd is, korter moeten.

Hoe meer syllabes een woord bevat, hoe korter de fonemen van dat woord. Dit verschijnsel is al langer bekend. Ook hier lijkt het erop (zie ook 3.5.2) dat er naar gestreefd wordt woorden zoveel mogelijk dezelfde duren te laten hebben, ongeacht het aantal syllabes. De conclusie kan getrokken worden dat verlenging of verkorting op grond van het aantal syllabes in een woord daadwerkelijk tot verbetering van de duurregelset kunnen leiden. Dit betekent tevens dat het formalisme aangepast zal moeten worden.

Verlenging van de prevocale consonant in prepauzale positie zal de duur niet significant verbeteren. De lengtes van deze consonant werden zowel in een nonprepauzale woord, als in een prepauzale woord, iets te kort bevonden. Bovendien waren de scores in beide posities vrijwel gelijk. De lengtes van alle fonemen in dit experiment, werden over het algemeen te kort bevonden. Het lijkt dus alsof alle initiële duren iets langer zouden moeten, of dat de gesynthetiseerde uitingen in het geheel iets verlengd zouden moeten worden. Volgens mij is deze globale lengte echter afhankelijk van het spreektempo van de spreker door wie de draagzin uitgesproken is. Daarnaast viel mij al op dat de oorspronkelijke zin ('Snelweg A10 Rotterdam Terneuzen, over een lengte van 22 km file') langzamer uitgesproken werd dan normaal bij de

verkeersinformatie. Het lijkt dus weinig zinvol om aan de hand van dit experiment de initiële duren van de fonemen te vergroten.

Hoofdstuk 4 De duurregels in het eindonderzoek

Zoals gezegd zijn de duurregels voor het RDS-TMC systeem verkregen door bestaande duurregels om te zetten in het PTS-formalisme en vervolgens te luisteren wat er nog verbeterd kon worden. Daarnaast zijn bestaande regels aangepast naar aanleiding van resultaten uit het vooronderzoek.

Het uiteindelijk verkregen regelsysteem kan opgedeeld worden in vier stukken, te weten:

1. De initiële duren. In dit gedeelte worden de beginduren van elk foneem gedefinieerd.
2. De fonologische regels. Dit zijn de prepausaalregels en de klemtoonregel.
3. Regels die afhankelijk zijn van de segmentele context van een foneem.
4. Regels afhankelijk van het aantal syllabes waaruit een woord bestaat.

De in mijn ogen twee belangrijkste regels zijn in het vooronderzoek getest op verbetering ten opzichte van geen beregeling. Dit waren de prepausaalregels en de klemtoonregel. Daarna zijn regels toegevoegd die de consonant-clusters beregelen. De regels die na beluistering nog voor verbeteringen zorgden, zijn alle contextgevoelige regels eventueel rekening houdend met positie of beklemtoning.

Hieronder zal ik de regels per categorie beschrijven (tussen haakjes staat het overeenkomstige regelnummer en de bron van de regels staat tussen vierkante haken):

1. De initiële duren.

Hiermee worden de beginduren van elk foneem bepaald. Het is puur een opsomming van alle fonemen met hun bijbehorende duur (zie appendix C).

2. De fonologische regels.

De klemtoonregels:

- (1) Lange beklemtoonde vocalen, de /2, 9, a, e, o/, worden met 60 procent verlengd [NIROS aangepast door Rutten (vooronderzoek)]. Voorbeeld: 'Den Haag'.
- (2) De halflange vocalen (/3, 4, 5, i, u, y/) worden met 20 procent verlengd als ze in beklemtoonde positie staan [NIROS aangepast door Rutten (vooronderzoek)]. Voorbeeld: 'Tiel'.
- (12) Plofklanken in een beklemtoonde syllabe worden met 15 ms verlengd [NIROS]. Voorbeeld: 'Tiel'.

De prepausaalregels:

- (3) Alle stemloze plofklanken, stemloze fricatieven en alle nasalen in finale positie worden verlengd met 60 procent [NIROS aangepast door Rutten (vooronderzoek)]. Voorbeeld: 'Utrecht'.
- (4) De vocalen, op de schwa na, worden met 35 ms verlengd in finale positie [NIROS (vooronderzoek)]. Voorbeeld: 'Utrecht'.
- (5) In zinsfinale positie en op wordeinde worden de glides en lateralen 120 ms lang [NIROS (vooronderzoek)]. Voorbeeld: 'Tiel'.

3. De regels die afhankelijk zijn van de segmentele context van een foneem.

De consonant-cluster regels:

- (6) De fricatieven die voorafgaan aan een consonant met 15 ms worden verkort [NIROS]. Voorbeeld: 'Amsterdam'
 - (7+8) Fricatieven die voorafgegaan worden door een consonant en gevolgd door een vocaal worden met 15 ms verkort [NIROS]. Voorbeeld: 'Amersfoort'.
 - (9) Plosieven die voorafgaan aan een consonant worden verkort met 40 procent [NIROS aangepast door Rutten]. Voorbeeld: 'Geertruidenberg'.
 - (10+11) Plosieven volgend op een consonant en voorafgaan aan een vocaal worden met 15 ms verkort [NIROS]. Voorbeeld: 'Harderwijk'.
 - (13) De /N/ (lang) wordt met 40 procent verkort als deze gevolgd wordt door een consonant [Rutten]. Voorbeeld: 'Doetinchem'.
- Daarnaast zijn er nog enkele contextgevoelige regels waarvan sommige uit NIROS komen en regels die ik zelf opgesteld heb, na beluistering.
- (14) De /i, y, u/ worden verlengd met 20 ms als ze voorafgaan aan een /r/ [NIROS]. Voorbeeld: 'Byren'.
 - (15) De schwa wordt 50 ms langer als deze onbeklemtoond is en op het wordeinde staat [Rutten]. Voorbeeld: 'Oude Pekela'.
 - (16) De /n/ wordt op wordeinde en indien deze voorafgegaan wordt door een schwa, met 20 procent verkort (een soort n-deletie) [Rutten]. Voorbeeld: 'Eindhovenn'.
 - (17) De glottal stop wordt op woordbegin met 15 ms verlengd [Rutten]. Voorbeeld: 'ʔEde'.
 - (18) De /h/ wordt met 10 ms verkort als deze voorafgaat aan een korte klinker [NIROS]. Voorbeeld 'Hengelo'.
 - (19) Van de combinatie /t/ in finale positie wordt de /t/ met 40 procent verlengd [Rutten]. Voorbeeld: 'Belfet'.
 - (20) De /w/ in finale positie, voorafgegaan door een klinker en gevolgd door de schwa, wordt met 30 procent verlengd [Rutten]. Voorbeeld: 'Hoge Veluw'.
 - (21) Van de combinatie /t@/ in finale positie wordt de /t/ met 20 procent verlengd [Rutten]. Voorbeeld: 'Raalte'.
 - (22) De beklemtoonde /A/ die voorafgaat aan de /N/ wordt met 20 procent verkort [Rutten]. Voorbeeld 'Angeren'.
4. **Regels afhankelijk van het aantal syllabes waaruit een woord bestaat** (deze regels staan niet in de duurregelset, maar zijn hard in het systeem opgenomen).
- Een monosyllabisch woord wordt met 20 procent verlengd.
 - Een woord bestaande uit twee syllabes verandert niet van duur door deze regels.
 - Een tri-syllabisch woord wordt met 5 procent verkort.
 - een quattro- of meersyllabisch woord wordt met 10 procent verkort.

Deze woorden worden verkort door elk foneem uit dat woord procentueel evenveel te verkorten.

In het PTS-formalisme zien bovengenoemde regels er als volgt uit:

- 1.<2,9,a,e,o; ; ;1; > %160
- 2.<3,4,5,i,u,y; ; ;1; > %120
- 3.<t,k,p,s,f,x,N,n,m; ; ;1> %160
- 4.<2,3,4,5,9,A,E,I,O,Y,a,e,i,o,u,y; ; ; ;1> +35
- 5.<l,j,w,r; ;#; ;1> =120
- 6.<S,Z,f,h,s,v,x,z;
;g,N,S,Z,b,d,f,h,j,k,l,m,n,p,r,s,t,v,w,x,z; ; > -15
- 7.<S,Z,f,h,s,v,x,z;g,N,S,Z,b,d,f,h,j,k,l,m,n,p,r,s,t,v,w,
x,z;a,e,E,i,I,o,O,9,u,y,Y; ; > -15
- 8.<S,Z,f,h,s,v,x,z;g,N,S,Z,b,d,f,h,j,k,l,m,n,p,r,s,t,v,w,
x,z;@,A,2,3,4,5; ; > -15
- 9.<p,t,k,b,d,g;
;g,N,S,Z,b,d,f,h,j,k,l,m,n,p,r,s,t,v,w,x,z; ; > %60
- 10.<p,t,k,b,d,g;g,N,S,Z,b,d,f,h,j,k,l,m,n,p,r,s,t,v,w,x,z
;a,e,E,i,I,o,O,9,u,y,Y; ; > -15
- 11.<p,t,k,b,d,g;g,N,S,Z,b,d,f,h,j,k,l,m,n,p,r,s,t,v,w,x,z
;@,A,2,3,4,5; ; > -15
- 12.<p,t,k,b,d,g; ; ;1; > +15
- 13.<N; ;gS,Z,b,d,f,h,j,k,l,m,n,p,r,s,t,v,w,x,z; ; > %60
- 14.<i,y,u; r; ; > +20
- 15.<@; ;#;0; > +50
- 16.<n;@;#; ; > %80
- 17.<?;#; ; ; > +15
- 18.<h; ;@,A,E,I,O,i,u,y; ; > -10
- 19.<l; ;t; ;1> %140
- 20.<w;2,3,4,5,9,@,A,E,I,O,i,u,y;@; ;1> %130
- 21.<t; ;@; ;1> %120
- 22.<A; ;N;1; > %80

Hoofdstuk 5 Het eindonderzoek

5.1 Doel

Het doel van dit eindonderzoek is te controleren of de uitgebreide duurregelset (hoofdstuk 4) daadwerkelijk beter is dan de duurregelset met alleen de essentiële duurregels (paragraaf 2.3, aangepast na het vooronderzoek).

5.2 Opzet

De beide duurregelsets zullen met elkaar vergeleken worden, door plaatsnamen te synthetiseren met de verschillende sets. Deze plaatsnamen zullen in draagzinnen gezet worden. Dit gaat tenslotte in het RDS-TMC-systeem ook gebeuren. Twee dezelfde testzinnen, die met een andere duurregelset gemaakt zijn, worden voorgelegd aan proefpersonen. Deze moeten bepalen welke ze het beste vinden.

Bij het maken van het stimulus-materiaal spelen verschillende factoren een rol. Er moeten beslissingen genomen worden over de te gebruiken draagzinnen en plaatsnamen. Bovendien moet bepaald worden welk intonatie-patroon gebruikt gaat worden.

Bij het kiezen van de draagzinnen speelt één factor een belangrijke rol: Hoeveel plaatsnamen moeten worden geplaatst in een slot. Er zijn zinnen denkbaar waarin één plaatsnaam geplaatst moet worden in een slot en er zijn ook zinnen waarin twee plaatsnamen geplaatst moeten worden in een slot. Om verschillen in de resultaten op grond van het aantal plaatsnamen in een slot uit te sluiten, zullen beide mogelijkheden meegenomen worden in dit experiment. Verder is in het Pilot-experiment gebleken dat het erg vermoeiend voor de luisteraar is, elke keer weer de hele zin te moeten beluisteren. Daarom is het raadzaam korte draagzinnen te nemen.

De volgende draagzinnen voldoen aan de eisen en zullen in het eindexperiment gebruikt worden:

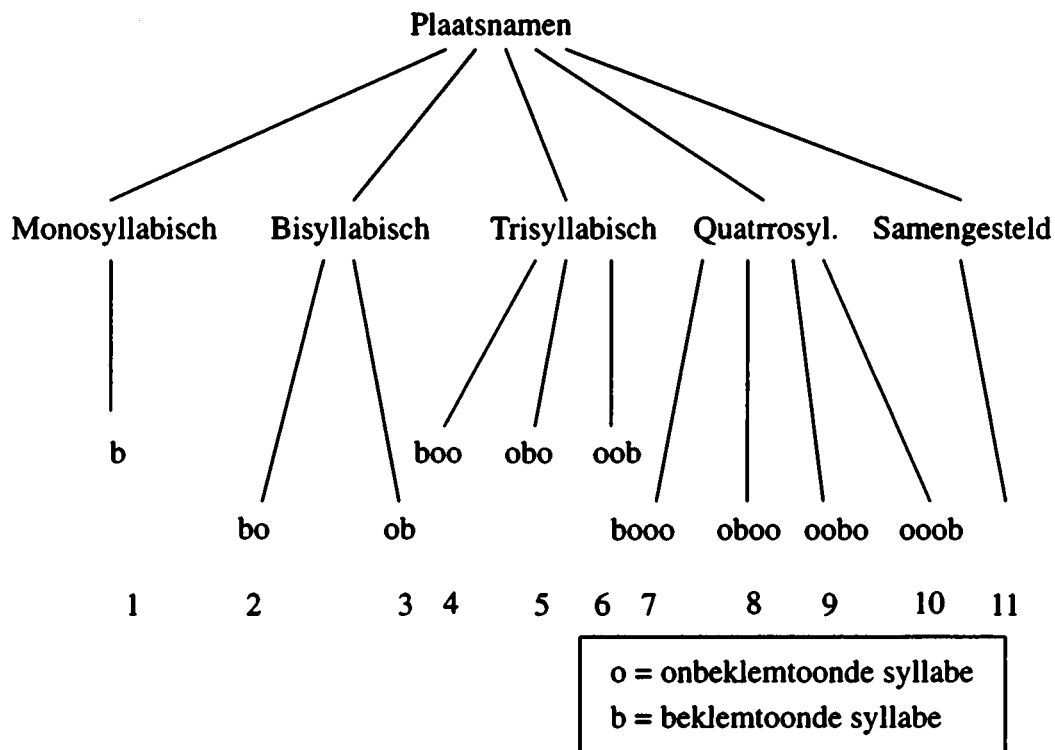
1. A27 __ __, 22 kilometer file.
2. P en R terrein _____ gesloten.

Draagzin 1 is gemaakt door uit de natuurlijk uitgesproken zinnen 'A27 Groningen Amsterdam carpoolpunt Lunetten, 10 parkeerplaatsen vrij' en 'Snelweg A10 Rotterdam Terneuzen, over een lengte van 22 kilometer file' de onderstreepte delen te knippen. Draagzin 2 is verkregen door van de natuurlijk uitgesproken zin 'P en R terrein Nijmegen gesloten' het eerste en het laatste stuk te knippen. Alle uitingen zijn uitgesproken door E. Jeukens. De difoonset die gebruikt is bij de synthese is ook met de stem van deze man gemaakt.

De keuze van de plaatsnamen is uiteraard nog belangrijker. Om tot een keuze te komen van de te gebruiken plaatsnamen zijn een aantal criteria gehanteerd:

- De plaatsnamen moeten variëren in aantal syllabes.
- De plaats van de klemtoon van de plaatsnaam moet variëren.
- Plaatsnamen bestaande uit meerdere woorden moeten ook gebruikt worden.
- De plaatsnamen moeten bekend zijn.
- Plaatsnamen waarmee eerder getest is, zijn uitgesloten.

Aan de hand van de eerste drie criteria kunnen 11 groepen onderscheiden worden:

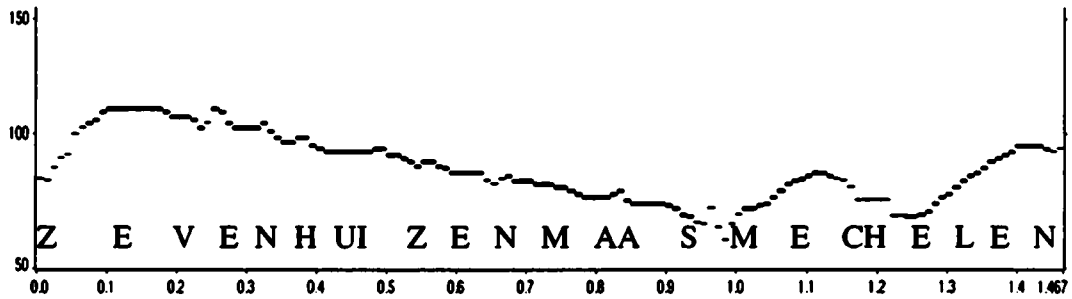


Aan de hand van dit model is voor elke groep een plaatsnaam gezocht uit de lijst met plaatsnamen die voor het Rhein-Corridor project (paragraaf 1.2) gebruikt is. Daaruit is voor elke groep de in mijn ogen meest bekende genomen. De volgende lijst is hieruit gekomen:

1. Sneek, Goes
2. Gouda
3. Maastricht
4. Vlissingen
5. Zaltbommel
6. Nieuwegein
7. Zevenhuizen
8. Maasmechelen
9. Leenderheide
10. Hilvarenbeek
11. Wijk bij duurstede, Krimpen aan de ijssel

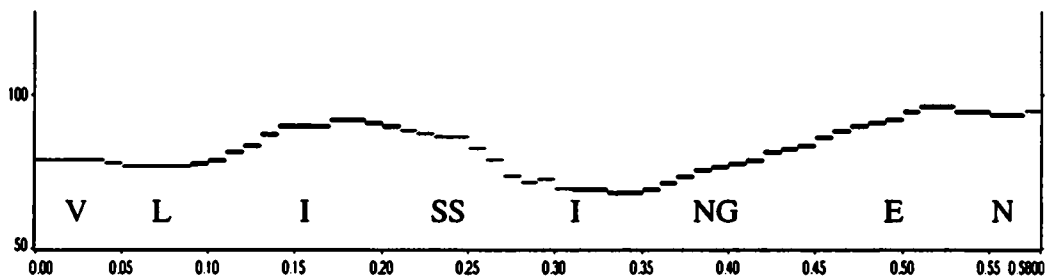
Eén van de aspecten die bekeken wordt in dit onderzoek is het verschijnsel dat in woorden met meer syllabes de fonemen korter worden. In de basale set wordt deze beregeling niet meegenomen, in de uitgebreide set wel. Om nu het effect van deze beregeling optimaal te kunnen analyseren, is het verstandig om in zinnen waarin twee plaatsnamen ingevuld moeten worden, twee plaatsnamen met evenveel syllabes te plaatsen. Op deze manier is er geen variatie wat betreft de beregeling op grond van het aantal syllabes. Om in een draagzin twee verschillende plaatsnamen in te kunnen vullen is aan set één en elf een plaatsnaam toegevoegd, waarbij opgemerkt kan worden dat de toevoeging bij set elf eigenlijk niet nodig is, omdat de verkorting evenveel is als de quattrosyllabische groep (4 of meer syllabes in een frase).

Bij het maken van de stimuli is het van belang dat het intonatiepatroon gelijk blijft. In ieder geval bij de twee zinnen die vergeleken gaan worden, maar beter nog over alle gesynthetiseerde plaatsnamen. Na het bekijken van het intonatiepatroon bij de natuurlijk uitgesproken plaatsnamen en vergelijken van verschillende intonatiepatronen, is gekozen voor een zogenaamde zaagtandcontour (Terken, 1993). Deze contour laat een snelle stijging zien bij de vowel-onset van een beklemtoonde syllabe. Als er nog een beklemtoonde syllabe is, dan volgt een geleidelijke daling tot deze syllabe, anders volgt een snelle daling. Aan het eind van de gesynthetiseerde uiting laat het intonatiepatroon een continuation-rise zien. Zie figuur 2 voor het intonatiepatroon van de synthetische uiting 'Zevenhuizen Maasmechelen':



FIGUUR 2.

Het intonatiepatroon van de synthetische uiting 'Vlissingen' ziet er als volgt uit (figuur 3):



FIGUUR 3.

Hierna zijn testzinnen gemaakt met de natuurlijk uitgesproken draagzinnen en de gesynthetiseerde plaatsnamen. Met de twee draagzinnen, de 13 plaatsnamen en de eis in de draagzin met twee plaatsnamen, twee plaatsnamen met evenveel syllables te plaatsen, kunnen 37 zinnen gemaakt worden. Dit is gereduceerd tot 26 door niet elke plaatsnaam met een bepaald aantal syllables te combineren met elke andere plaatsnaam met datzelfde aantal syllables (zie appendix D voor de 26 testzinnen).

Deze 26 testzinnen zijn willekeurig door elkaar gezet. Elke testzin is met beide duurregels gesynthetiseerd zodat er 26 testzinparen zijn. Een testset bestaat uit drie proefzinparen, de 26 paren testzinnen in willekeurige volgorde en nog eens de 26 paren testzinnen in willekeurige volgorde, waarbij ook de volgorde van duurberegeling willekeurig is. Dit is gedaan om te kunnen controleren of een proefpersoon wel consistent is in zijn of haar beoordeling.

Op deze manier zijn vier testsets gemaakt om volgorde-effecten uit te sluiten.

5.3 Uitvoering

Aan dit experiment deden 30 proefpersonen mee. Het experiment is afgenomen in de IPO-studio. De zinnen waren te beluisteren door een hoge-kwaliteits hoofdtelefoon. Elke proefpersoon kreeg één van de vier testsets. Per zinspaar moest de proefpersoon bepalen welke hij of zij het best vond klinken. Zie voor de instructies van dit experiment appendix E.

5.4 Verwerking van de resultaten

Elke testzinpaar is twee keer beoordeeld door een proefpersoon. De proefpersoon kan per zin voorkeur hebben uitgesproken voor de basale duurregelset, de uitgebreide duurregelset, of inconsistent geoordeeld hebben.

Hieronder staat voor elke zin hoe vaak deze het best bevonden werd met de basale duurregelset, de uitgebreide duurregelset en hoe vaak deze inconsistent beoordeeld was:

TABEL 8. .Plaatsnamen en hun beoordelingen

zin	Basaal	Uitgebreid	Inconsistent
Goes	5	8	17
Sneek	13	5	12
Goes Sneek	7	10	13
Sneek Goes	8	7	15
Gouda	4	5	21
Maastricht	3	14	13
Gouda Maastricht	6	12	12
Maastricht Gouda	4	9	17
Nieuwegein	4	4	22
Vlissingen	3	9	18
Zaltbommel	9	9	12
Nieuwegein Zaltbommel	3	11	16
Vlissingen Nieuwegein	9	6	15
Zaltbommel Vlissingen	1	15	14
Hilvarenbeek	4	16	10
Krimpen aan de IJssel	0	17	13
Leenderheide	2	16	12
Maasmechelen	5	10	15
Wijk bij Duurstede	3	8	19
Zevenhuizen	7	11	12
Hilvarenbeek Leenderheide	0	25	5
Krimpen aan de IJssel Wijk bij Duurstede	3	15	12
Leenderheide Zevenhuizen	1	18	11
Maasmechelen Hilvarenbeek	4	15	11
Wijk bij Duurstede Krimpen aan de IJssel	1	14	15
Zevenhuizen Maasmechelen	2	11	17

Bij elkaar genomen levert dit tabel 9:

TABEL 9. Resultaten in aantallen

Resultaat	Aantal
Basaal	111
Uitgebreid	300
Inconsistent	369

Er zijn dus 411 consistente beoordelingen en 369 niet consistente beoordelingen. Er zijn niet significant meer consistente beoordelingen dan inconsistente beoordelingen. Dit wil zeggen dat als de proefpersonen de zinnen willekeurig hadden beoordeeld, er ongeveer hetzelfde aantal consistente beoordelingen waren geweest.

Kijken we echter alleen naar de consistente beoordelingen dan is eenvoudig na te gaan dat er significant vaker voor de uitgebreide duurregelset gekozen is.

Het vermoeden bestaat dat vooral de monosyllabische plaatsnamen erg inconsistent beoordeeld zijn. Daarom is de tabel 10 gemaakt:

TABEL 10. Significantie van beoordelingen als functie van het aantal syllabes

Aantal syllabes	Basaal	Expanded	Inconsistent	Significant
1	33	30	57	nee
2	17	40	63	nee
3	29	54	97	nee
4 of meer	32	176	152	ja

Hieruit blijkt echter dat alleen voor de quattro- of meersyllabische plaatsnamen geldt dat ze significant vaker consistent beoordeeld zijn.

TABEL 11. Significantie van beoordelingen als functie van het aantal plaatsnamen in een slot

Aantal plaatsnamen in een slot	Basaal	Expanded	Inconsistent	Significant
1	62	132	196	nee
2	49	168	173	ja

Uit tabel 11 blijkt dat proefpersonen consistentere gaan beoordelen als het aangeboden synthetische stuk langer wordt. Bovendien blijkt in de meeste gevallen de voorkeur uit te gaan naar de uitgebreide duurregelset.

Om deze resultaten om te zetten in getallen waarmee gerekend kan worden heb ik als eindoordeel het aantal keren dat voor de basale duurregelset gekozen is, afgetrokken van het aantal keren dat voor de uitgebreide duurregelset gekozen is. Door deze waarden door 30 te delen, wordt een gemiddeld oordeel verkregen dat ligt tussen -1 (iedereen heeft consistent voor de basale duurregelset gekozen) en +1 (iedereen koos consistent voor de uitgebreide duurregelset).

Dit gaf het volgende overzicht:

TABEL 12. Plaatsnamen en hun beoordelingen

zin	Eindoordeel	Gemiddelde
Goes	3	0,10
Sneek	-8	-0,27
Goes Sneek	3	0,10
Sneek Goes	-1	-0,03
Gouda	1	0,03
Maastricht	11	0,37
Gouda Maastricht	6	0,20
Maastricht Gouda	5	0,17
Nieuwegein	0	0,00
Vlissingen	6	0,20
Zaltbommel	0	0,00
Nieuwegein Zaltbommel	8	0,27
Vlissingen Nieuwegein	-3	-0,10
Zaltbommel Vlissingen	14	0,47
Hilvarenbeek	12	0,40
Krimpen aan de IJssel	17	0,57
Leenderheide	14	0,47
Maasmechelen	5	0,17
Wijk bij Duurstede	5	0,17
Zevenhuizen	4	0,13
Hilvarenbeek Leenderheide	25	0,83
Krimpen aan de IJssel Wijk bij Duurstede	12	0,40
Leenderheide Zevenhuizen	17	0,57
Maasmechelen Hilvarenbeek	11	0,37
Wijk bij Duurstede Krimpen aan de IJssel	13	0,43
Zevenhuizen Maasmechelen	9	0,30

Zoals te zien is, is er bij de meeste zinnen een waarde boven nul verkregen, wat aangeeft dat er een voorkeur is voor de uitgebreide duurregelset.

Het gemiddelde van alle resultaten is 0,24. Dit getal zegt echter niet zoveel, omdat veel beoordelingen inconsistent waren, zodat er veel nullen in de resultaten zitten.

Als we de gemiddelde beoordelingen en het aantal syllabes waaruit de plaatsnamen bestaan tegen elkaar uitzetten, krijgen we tabel 13 (bedenk dat bij zinnen met twee slots twee plaatsnamen staan met hetzelfde aantal syllabes):

TABEL 13. Gemiddelde beoordeling als functie van het aantal syllabes

Aantal syllabes	Gemiddelde beoordeling
1	-0,025 (n=120)
2	0,19 (n=120)
3	0,14 (n=180)
4 of meer	0,4 (n=360)

Uit deze tabel blijkt dat de voorkeur voor de uitgebreide duurregelset groter is naarmate de lengte van de plaatsnaam groter wordt. Voor plaatsnamen met één syllabe hebben de proefpersonen zelfs een lichte voorkeur voor de basale duurregelset. Dit zou kunnen betekenen dat de monosyllabische verlenging teveel is. Ook zou het zo kunnen zijn dat deze plaatsnamen door hun geringe lengte moeilijker te beoordelen waren, en dat deze daarom random beoordeeld zijn.

Het verschijnsel dat er meer voor de uitgebreide set gekozen wordt naarmate het synthetische deel groter wordt doet zich ook voor als we kijken naar de gemiddelde beoordelingen als functie van het aantal plaatsnamen in een slot (tabel 14):

TABEL 14. De gemiddelde beoordelingen als functie van het aantal slots in een zin

Aantal plaatsnamen in een slot	Gemiddelde beoordeling
1	0,15
2	0,31

Bedenk hierbij dat alle plaatsnamen bij allebei de mogelijkheden inbegrepen zijn, dus is het verschil in gemiddelde beoordeling puur te wijten aan het aantal plaatsnamen in een slot.

Samengevat kan gesteld worden dat proefpersonen niet zeer goed in staat waren consistent te oordelen. Naarmate de synthetische uitingen langer werden, werden de beoordelingen echter ook consistent en was er steeds meer voorkeur voor de uitgebreide duurregelset.

Hoofdstuk 6 Conclusies en aanbevelingen

De basis van dit onderzoek was de synthetische spraak binnen RDS-TMC van duurberegeling voorzien. Bij het begin van dit onderzoek was er nog geen enkele beregeling. Het PTS-formalisme moest verkend worden om te zien wat de mogelijkheden van duurberegeling waren.

Hierna is begonnen met het maken van een basale duurregelset, op basis van regels die in de literatuur gevonden zijn. De klemtoonregel en de prepausaalregel zijn hierin opgenomen. Deze basale duurregelset bleek verrassend goed te werken. De synthetische spraak was goed verstaanbaar en acceptabel. Bovendien bleken de regels een hoorbaar beter resultaat te geven dan de onberegelde synthetische spraak.

Vervolgens is de basale duurregelset aangepast en zijn regels toegevoegd op basis van introspectie en de literatuur. Hieruit is de uitgebreide duurregelset ontstaan. Deze uitgebreide duurregelset is significant beter dan de basale duurregelset, doch het verschil is erg klein. Geconcludeerd kan worden dat de basale duurregels voor een grote verbetering in de spraaksynthese zorgen, en dat het toevoegen van duurregels steeds minder invloed heeft.

De duurberegeling op grond van het aantal syllabes in een frase heeft naar alle waarschijnlijkheid binnen de uitbreiding van de basale duurregelset de meeste invloed. Dit is te horen aan de synthetische spraak en af te leiden als je bedenkt dat deze beregeling werkt op alle fonemen in de synthetische spraak. Deze beregeling zal van grote invloed geweest zijn op het eindresultaat. Andere toegevoegde regels geven slechts kleine verbeteringen, maar leveren toch een beter resultaat (zie 'Maastricht', waarbij de beregeling op grond van polysyllabiciteit geen rol speelt).

Binnen de regels die de duren op grond van het aantal syllabes beregelen, zal de regel die monosyllabische frases beregelt aangepast moeten worden. Deze regel verlengd de fonemen in monosyllabische frases met 20 procent. Deze verlenging zou iets minder moeten zijn. Het is verstandig om de beregeling op grond van het aantal syllabes in een frase op te nemen in het formalisme in plaats van in de programmacode. Dit zou het eenvoudiger maken te experimenteren met verschillende waardes.

Ook is in dit onderzoek naar voren gekomen dat de beregeling meer effect heeft naarmate de gesynthetiseerde uiting langer wordt. Bovendien geldt dat hoe langer de synthetische uiting wordt, hoe meer de proefpersonen de voorkeur hebben voor de uitgebreide duurregelset.

Wat de invloed van elke regel apart is, is niet onderzocht. Dit zou in een eventueel vervolgonderzoek kunnen gebeuren. Op deze manier kunnen de afzonderlijke regels fijntjes bijgesteld worden, iets wat bij lange synthetische uitingen zeker tot verbetering zou kunnen leiden. Als basis voor een vervolgonderzoek zou ik de uitgebreide duurregelset nemen, omdat gebleken is dat deze significant beter is dan de eenvoudige duurregelset.

Daarnaast moet een opmerking gemaakt worden over de gebruikte difoonset. Deze was dermate slecht dat het stimulusmateriaal niet optimaal was. Hoewel dit niet direct van invloed is op de resultaten, zou het wellicht eenvoudiger zijn voor proefpersonen om spraak te beoordelen, als de gebruikte difoonset beter was geweest.

Verder zou bij een vervolgonderzoek de intonatieberegeling tegelijkertijd onderzocht moeten worden. In dit onderzoek is gebleken dat de duren niet te beregelen zijn zonder naar het intonatiepatroon te kijken. Hoe de samenhang tussen intonatiepatroon en duurberegeling precies is zou dan ook meteen onderzocht kunnen worden.

Samenvattend kan gezegd worden dat de uitgebreide duurregelset de meest optimale duurregelset is tot nu toe. Uit dit onderzoek blijkt dat toevoegen van duurregels wel degelijk invloed heeft als de synthetische uitingen langer worden. In de toekomst is het de bedoeling dat ook de draagzinnen binnen het RDS-TMC-systeem gesynthetiseerd worden. Omdat dit vrij lange stukken spraak zullen zijn, is het zinvol steeds meer duurregels toe te voegen aan de uitgebreide duurregelset. Het fijntjes bijstellen van de waarden waarmee fonemen verlengd of verkort worden in de huidige regelset zal zeker ook voor verbetering zorgen.

Referenties

- Den Os, E (1988), Ritme en tempo in spraak, In: M.P.R. van den Broecke (ed.), *Ter sprake*, pagina 144-150, Dordrecht: Foris publications.
- Klatt, D.H. (1987), Review of text-to-speech conversion for English, *JASA*, Vol. 82, No 3, pagina 737-783.
- Klatt, D.H. (1979), Synthesis by rule of segmental durations in English sentences, *Ninth International Congress of Phonetic Sciences, Proceedings*, Vol. II, pagina 290-297.
- van Coile, B. (1990), *Tekst-naar-spraak omzetting: een taalkundig, fonetisch en akoestisch probleem*, Proefschrift, Gent: Rijksuniversiteit Gent.
- Terken, J. (1993), Synthesizing natural-sounding intonation for Dutch: rules and perceptual evaluation, *Computer Speech & Language*, 7, pagina 27-48.

Appendix A Foneem-notaties

De foneemrepresentatie die gebruikt wordt binnen RDS-TMC is afgeleid van de SAMPA-notatie. De SAMPA-notatie is een standaardfoneemrepresentatie welke voor meerdere talen ontwikkeld is.

TABEL 15. Foneem-notaties binnen RDS-TMC

Nederlandse grafemen	Sampa-notatie	Notatie binnen RDS-TMC
mAt	A	A
bEdoel	@	@
pUt	Y	Y
IEs	E	E
plI	I	I
rOt	O	O
rOEt	u	u
fUUt	y	y
IIeP	i	i
mAAAt	a:	a
IEEs	e:	e
kEUUs	2:	2
rOOd	o:	o
kOUd	Au	3
rEIs	Ei	4
mUIs	9y	5
frEUle	9:	9
Pas	p	p
Tas	t	t
Kas	k	k
Bas	b	b
Das	d	d
Goal	g	g
Lang	l	l
Rang	r	r
Jan	j	j
Wang	w	w
Fok	f	f
Sok	s	s
SJaak	S	S
Gok	x	x
Veer	v	v
Zeer	z	z
bagaGe	Z	Z
Hang	h	h
Meer	m	m
Neer	n	n
baNG	N	N
stilte	.	*
glottal stop	?	?

Appendix B Luisterexperiment vooronderzoek

Dit is het instructievel van het vooronderzoek:

U krijgt zometeen 40 zinnen te horen. Deze zinnen zien er als volgt uit:

Snelweg A10 'x' 'y', over een lengte van 22 kilometer file.

Het eerste stuk (snelweg A10) en het laatste stuk (over een lengte van 22 km file) zijn stukken natuurlijke spraak. 'x' en 'y' zijn synthetische uitingen, waarbij zowel 'x' als 'y' plaatsnamen zijn.

Het is nu de bedoeling dat u deze synthetische spraak beoordeelt op de duur van de klanken. Hierbij moet u niet de omliggende natuurlijke spraak gaan vergelijken met de synthetische spraak, maar puur op de duren van de fonemen van de plaatsnamen letten.

Op de bijgeleverde bladen (3) staan alle uitingen. Het is de bedoeling dat u op de stippe lijn onder elke foneem aangeeft welke korter, welke langer zouden moeten en welke fonemen goed zijn wat de duur betreft. Dit doet u door een '-' (klank zou korter moeten), '+' (klank zou langer moeten) of 'o' (lengte is oké) onder de betreffende klank te zetten. Verlenging of verkorting van de stiltes tussen de woorden kunt u aangeven onder de underscore (_).

Het experiment begint door 'start' in te typen. Bovenaan staat welke zin uitgesproken wordt. U kunt deze zin nog een keer beluisteren door op enter te drukken. De volgende uiting krijgt u door de 'v' <enter> in te drukken. Mocht u onverhoopt naar het volgende woord gegaan zijn zonder dat dit de bedoeling was, dan kunt u altijd nog terug door de 't' <enter> in te drukken.

Mocht u nog vragen hebben, dan kunt u deze nu stellen.

Alvast bedankt voor uw medewerking.

Appendix C De initiële duren

Hieronder staat een opsomming van de fonemen met hun initiële duren in milliseconden. De fonemen staan in de aangepaste sampa-notatie (appendix A).

/2/	120
/3/	136
/4/	135
/5/	125
/*/	200
/9/	130
/?/	15
/@/	45
/A/	81
/E/	80
/I/	80
/N/	80
/O/	80
/S/	85
/Y/	70
/Z/	85
/a/	120
/b/	60
/d/	60
/e/	120
/f/	85
/g/	60
/h/	85
/i/	80
/j/	60
/k/	80
/l/	85
/m/	80
/n/	70
/o/	120
/p/	80
/r/	70
/s/	85
/t/	80
/u/	80
/v/	85
/w/	60
/x/	85
/y/	80
/z/	85

Appendix D De 26 testzinnen in het eindexperiment

P en R terrein Goes gesloten.

P en R terrein Sneek gesloten.

A27 Goes Sneek 22 kilometer file.

A27 Sneek Goes 22 kilometer file.

P en R terrein Gouda gesloten.

P en R terrein Maastricht gesloten.

A27 Gouda Maastricht 22 kilometer file.

A27 Maastricht Gouda 22 kilometer file.

P en R terrein Nieuwegein gesloten.

P en R terrein Vlissingen gesloten.

P en R terrein Zaltbommel gesloten.

A27 Nieuwegein Zaltbommel 22 kilometer file.

A27 Vlissingen Nieuwegein 22 kilometer file.

A27 Zaltbommel Vlissingen 22 kilometer file.

P en R terrein Hilvarenbeek gesloten.

P en R terrein Krimpen aan de IJssel

P en R terrein Leenderheide gesloten.

P en R terrein Maasmechelen gesloten.

P en R terrein Wijk bij Duurstede gesloten.

P en R terrein Zevenhuizen gesloten.

A27 Hilvarenbeek Leenderheide 22 kilometer file.

A27 Krimpen aan de IJssel Wijk bij Duurstede 22 kilometer file.

A27 Leenderheide Zevenhuizen 22 kilometer file.

A27 Maasmechelen Hilvarenbeek 22 kilometer file.

A27 Wijk bij Duurstede Krimpen aan de IJssel 22 kilometer file.

A27 Zevenhuizen Maasmechelen 22 kilometer file.

Appendix E Luisterexperiment

Dit is het instructievel van het eindexperiment:

U krijgt zodadelijk 55 paren zinnen te horen. Elk paar zinnen verschilt in de uitspraak van de plaatsnamen. Het is nu de bedoeling dat u op het antwoordvel aangeeft welke zin van een paar u het best vindt.

Vindt u de eerste zin (A) het best, zet dan een kruis onder A. Vindt u de tweede (B) beter zet dan een kruisje onder B.

Na het beluisteren van een paar zinnen kunt u het paar altijd herhalen (H), naar het volgende paar (V) of terug naar het vorige paar (T).

Als u nog vragen hebt, kunt u deze nu stellen,

Succes!