

**Radboud Universiteit Nijmegen**

## **Spraaksynthese**

“Difoon- vs. unit selection synthese”

Johan Stortelder

s0355593

johanstortelder@student.ru.nl

Nijmeegs instituut voor informatica en informatiekunde

Radboud Universiteit Nijmegen

## Inhoud

Inleiding .....	2
Geschiedenis .....	2
Toepassingen .....	4
Toekomstperspectieven .....	5
Theorie .....	6
Synthese methoden.....	6
Concatenatie synthese .....	7
Foonsynthese .....	8
Difoonsynthese.....	8
Unit selection .....	9
Formantsynthese .....	10
Prosodie .....	11
Onderzoeksopzet .....	13
Onderzoeksvraag.....	13
Aanpak .....	13
Respondenten .....	13
Stimuli.....	13
Scoring.....	15
Systemen .....	15
Uitkomsten.....	16
Verdeeldheid .....	20
Kanttekeningen .....	20
Conclusies .....	22
Bronnen .....	23
Literatuur .....	23
Webpagina's .....	23
Appendix A: Stimuli.....	24
Appendix B: Gerandomiseerde lijst stimuli.....	26
Appendix C: Antwoordlijst .....	27

## Inleiding

Spraaksynthese is, zoals de naam al doet vermoeden, het kunstmatig produceren of samenstellen (synthetiseren) van spraak. Deze scriptie gaat over dit syntheseproces, en met name over de vraag welke synthesemethode spraak met de meest natuurgetrouwe prosodie oplevert.

Om een beeld te krijgen van de manieren waarop men geprobeerd heeft spraak te genereren, wordt in de eerste paragraaf van de inleiding kort de geschiedenis van de spraaksynthese behandeld. Er zijn verschillende toepassingen van spraaksynthese. In de tweede paragraaf van dit hoofdstuk wordt hier een overzicht van gegeven en komen een aantal voorbeelden aan de orde om het gebruik van spraaksynthese in het hedendaagse leven te illustreren. Tot slot worden in de derde paragraaf kort de toekomstperspectieven van spraaksynthese behandeld.

Moderne spraaksynthese wordt door middel van computers gerealiseerd. Hiervoor zijn verschillende methoden ontwikkeld. In het tweede hoofdstuk van deze scriptie, dat de theorie achter de spraaksynthese behandelt, zullen deze verschillende methoden besproken worden. Daarnaast behandelt het hoofdstuk de theorie over prosodie. Prosodie is, zoals Rietveld en van Heuven [1] beschrijven, het geheel van eigenschappen van een spraakuiting die niet herleid kunnen worden tot een opeenvolging van klinkers en medeklinkers. Hierbij moet aan eigenschappen zoals klemtonen, spreeknelheid, ritmiek en melodie gedacht worden.

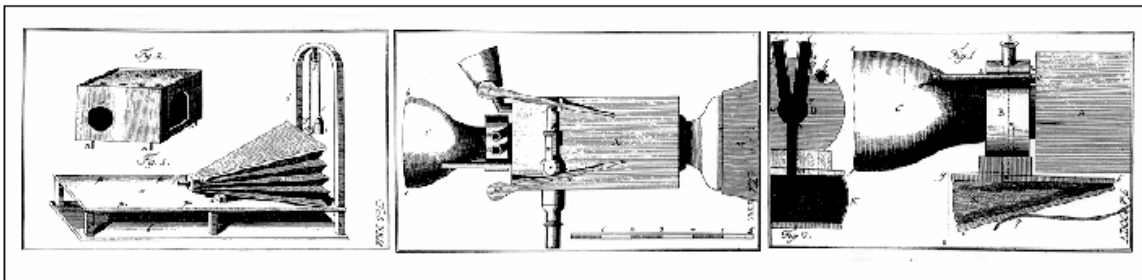
Om te bepalen welke methode als meest natuurgetrouw wordt ervaren, is er een onderzoek uitgevoerd. Hierin is een aantal respondenten gevraagd twee verschillende systemen die volgens een verschillende methode spraak genereren, te vergelijken. In het derde hoofdstuk worden de opzet voor dit onderzoek en de keuzes aangaande dit onderzoek beschreven.

Tot slot worden in het vierde hoofdstuk conclusies getrokken door de uitkomsten van het onderzoek en de theorie te combineren. Er wordt een antwoord gegeven op de vraag welke methode de meest natuurgetrouwe spraaksynthese oplevert.

## Geschiedenis

De meeste mensen denken bij spraaksynthese aan elektronische spraakproducerende systemen. Toch zijn er al lange tijd mensen bezig met het

namaken van menselijke spraak. Bijvoorbeeld Gerbert of Aurillac, Albertus Magnus en Roger Bacon die zich in de elfde en twaalfde eeuw al bezig hielden met het produceren van een soort sprekende hoofden. In de 18<sup>e</sup> eeuw zijn mechanische machines ontwikkeld om het menselijke spraakorgaan te simuleren. Wolfgang von Kempelen ontwikkelde destijds, gebruik makend van alledaagse materialen zoals hout, leer en metaal, een mechanisme dat in staat was spraakgeluiden en zelfs woorden en kleine zinnestjes te produceren. In figuur 1 is een afbeelding van deze machine opgenomen. Enkele overeenkomsten met het menselijk spraaksysteem zijn goed te zien. Ten eerste een gedeelte om de luchtstroom op te wekken; in de machine een blaasbalg, bij mensen de longen. Daarnaast is ook een gedeelte herkenbaar waarmee de juiste klank gegenereerd moet worden. Bij de mens gebeurt dit door middel van de stembanden en de vorm van de mondkeelholte, in de machine van Von Kempelen door het megafoonachtige gedeelte.



Figuur 1: Von Kempelen's Spraakmachine

Naast Von Kempelen zijn er ook anderen geweest die soortgelijke mechanismen gemaakt hebben. Ch. G. Kratzenstein, professor fysiologie in Kopenhagen, probeerde eveneens in de 18<sup>e</sup> eeuw klinkers na te bootsen met behulp van orgelpijpen.

Met de opkomst van de elektronica is het mechanische grotendeels losgelaten. De menselijke manier van spraak produceren diende vanaf toen niet meer als voorbeeld. Verschillende elektronische speech synthesizers zijn ontwikkeld, waaronder de VOCODER van Bell Labs (1930). Ook de telefoon, die aan het eind van de 18<sup>e</sup> eeuw is uitgevonden door Meucci en gepatenteerd door Bell, kan gezien worden als spraaksynthese. Het omzetten van gesproken tekst naar elektrische signalen die vervolgens weer naar spraak worden omgezet.

Tegenwoordig is telefonie niet meer weg te denken en micro-elektronica maakt het mogelijk om ook andere manieren van synthese toe te passen. Ook als er geen gesproken origineel is, zoals bij telefonie wel het geval is, wordt productie

van spraak nagestreefd. Text-to-speech (TTS) en Concept-to-speech (CTS) systemen zijn hier voorbeelden van. TTS systemen hebben een stuk geschreven tekst als input en de gesproken versie van die tekst als output. CTS systemen produceren een gesproken uiting op basis van een toestandsbeschrijving. Aan de hand van een aantal parameters heeft het systeem een beeld van de situatie en genereert het desgevraagd de tekst die bij die situatie hoort. Doordat deze systemen kennis hebben van de context is rekening houden met prosodie bij dit type systemen eenvoudiger dan bij TTS systemen. Deze scriptie richt zich vooral op de TTS systemen.

### **Toepassingen**

In het dagelijks leven zijn er tal van toepassingen van spraaksynthese, waar iedereen wel eens mee te maken heeft gehad. Iedereen heeft wel eens aan de lijn gehangen met een systeem dat automatisch telefoontjes afhandelt. In het openbaar vervoer worden omroepberichten gesynthetiseerd uit van te voren opgenomen uitspraken. Deze uitspraken worden afhankelijk van tijdstip, buslijn of treindienst, perron en de opgelopen vertraging samengesteld. Het voordeel van deze systemen is dat bedrijven geen werknemers meer nodig hebben om hun klanten van informatie te voorzien. Een ander voorbeeld van dergelijke spraaksynthese is de routebeschrijving die door GPS systemen gegeven wordt aan automobilisten. Deze systemen heten spraakassemblage systemen. Uitingen worden samengesteld door in een vaste contextzinnen variabele elementen in te voegen. "U gaat hier (*rechts* | *links*) af". Soms worden meerdere uitingen gecombineerd, aan elkaar geplakt. Dan gaat het om de zogenaamde concatenatiesystemen.

Ook andere soorten van synthese worden steeds meer gebruikt. Deze systemen gebruiken geavanceerdere methoden om van een tekst een gesproken uiting te maken. In hoofdstuk twee wordt hier dieper op in gegaan. Het gaat hier om toepassingen zoals het voorlezen van websites voor blinden of slechtzienden. Vooral bij websites van overheidsorganisaties en hulpverlenende instanties is dit een nuttige toepassing. Zo wordt de informatie voor een groter publiek toegankelijk. Ook voor mensen die niet kunnen praten kan het een uitkomst bieden. Door spraaksynthese worden ze in staat gesteld vrij complexe uitingen te realiseren. Denk bijvoorbeeld aan Stephen Hawking. Niet in staat om te spreken, maar wel in staat om de wereld deelgenoot te maken van zijn ideeën over het universum. Andere voorbeelden zijn digitale woordenboeken waarin de uitspraak van een woord hoorbaar gemaakt kan worden.

**Toekomstperspectieven**

Het gebied van de taalsynthese is nog steeds in ontwikkeling. Het ultieme streven is het nabootsen van het menselijk spraakgeluid zonder dat te horen is dat de uiting enkel en alleen door een computer gegenereerd is. Op het moment is er al een heel goed verstaanbare synthese mogelijk. Rekening houden met prosodie is iets dat nog maar in beperkte mate lukt. Hier is in de toekomst dan ook nog winst te behalen.

Een ander gebied waar in de toekomst nog aan gewerkt kan worden is het synthetiseren van een specifiek persoon. De gesynthetiseerde uiting moet dan de persoonlijke kenmerken dragen van een spreker, zodanig dat deze niet te onderscheiden is van een uitspraak die door de spreker zelf gerealiseerd is.

Als een dergelijke techniek gecombineerd wordt met andere spraak- en taaltechnologische technieken zou in de toekomst kunnen leiden tot wat nu alleen in theorie nog het boegbeeld van *Language Engineering* is n.l.: Vertaaltelefonie. Onder deze toepassing wordt (Rietveld en van Heuven [1]) een systeem verstaan dat een bepaalde taal herkent, automatisch vertaald naar bijvoorbeeld het engels, de vertaalde tekst via telefoonsatelliet verstuurt naar een andere plaats op aarde, waar de vertaling correct uitgesproken tot klinken wordt gebracht met de oorspronkelijke stemkarakteristiek van de oorspronkelijke spreker. Iemand met een zakenpartner in Japan kan hem dan in zijn eigen taal, telefonisch en real-time te woord staan, waarbij de spreker zijn eigen stemgeluid houdt.

## Theorie

Spraaksynthese is een gebied waar verschillende takken van wetenschap samen komen. Fonetiek, linguïstiek en informatiekunde/informatica zijn enkele voorbeelden. Er is dan ook een enorme verzameling theorie die achter spraaksynthese schuil gaat. In dit hoofdstuk wordt geprobeerd een compact overzicht te geven van de theorie die voor het in hoofdstuk drie beschreven onderzoek noodzakelijk is. De eerste paragraaf van dit hoofdstuk behandelt verschillende synthese methoden. Daarna zal ook nog worden ingegaan op de rol van prosodie in spraak en de invloed hiervan op spraaksynthese.

## Synthese methoden

In de spraaksynthese methoden zijn eigenlijk twee manieren van synthese het meest gangbaar. Concatenatie synthese en formantsynthese. Concatenatie synthese systemen plakken stukken spraak aan elkaar. Waarbij er verschillen bestaan in de grootte van de aan elkaar te plakken bouwstenen. Formantsynthese maakt gebruik van informatie over de energie die er per frequentiehoogte in het spraaksignaal zit, ook wel formanten genoemd. Deze formanten worden opgeslagen als parameters. Concatenatie systemen combineren stukken spraak van verschillende grootte die opgeslagen zijn in golven, echte geluidsfragmenten. Bij formantsynthese systemen is de informatie opgeslagen in parameters.

Een andere mogelijkheid is articulatorische synthese. Hierbij wordt een model gemaakt van de articulatoren, stembanden, keel, kaak en lippen (Marsi [4]). Op basis van dit model wordt dan spraak nagebootst. Hierbij moeten de vorm van het aanzetstuk (mondkeelholte), dwarsdoorsneden van strot tot lippen en resonanties berekend worden. Het is daarmee een erg bewerkelijke methode en er zijn weinig tot geen modellen van die geavanceerde of efficiënt genoeg zijn om voor synthese te gebruiken. Bij deze methode is het wel weer de menselijke manier van spraak produceren die als uitgangspunt voor de modellen wordt genomen.

Omdat articulatorische synthese een niet erg gangbare methode is, wordt er in deze scriptie verder geen aandacht aan besteed. De andere methoden, en hun verschillende vormen, worden in de rest van deze paragraaf behandeld.

*Concatenatie synthese*

Concatenatie synthese is een methode om kunstmatige spraak te produceren waarbij verschillende losse stukken spraak aan elkaar geplakt worden. Er zijn verschillende soorten concatenatie synthese, waarbij de grootte van de aan elkaar te plakken stukken varieert van hele zinnen (domein specifieke, assemblage en unit selection systemen) tot (groepen) fonemen (foon-, difoonsynthese).

Bij domein specifieke systemen moet gedacht worden aan een context met een beperkt aantal mogelijke spraakuitingen waarin het systeem moet voorzien. Voorbeelden hiervan zijn navigatie systemen in auto's en omroep berichten bij het openbaar vervoer. Hier is sprake van een aantal standaard zinnen die, op de juiste manier aan elkaar geplakt, steeds een op de situatie afgestemde boodschap op kunnen leveren. Ook door delen van de uitspraken variabel te maken, kan het aantal mogelijke uitdrukkingen sterk vergroot worden.

“De trein van (8|9|10|...) uur heeft (5|10|15|...) minuten vertraging, en vertrekt vanaf spoor (1|2|3|...)”.

Om een dergelijk systeem te ontwikkelen zal eerst met alle stake holders die het systeem moeten gaan gebruiken, geïnventariseerd moeten worden welke uitspraken er allemaal gedaan moeten kunnen worden. Vervolgens dienen de uitdrukkingen ingesproken te worden te worden door een professionele spreker. Als er gebruik gemaakt moet gaan worden van variabele delen in zinnen, is het verstandig een zo goed mogelijke afweging te maken tussen kwaliteit en het aantal zinnen. Iedere mogelijke zin apart inspreken levert natuurlijk een betere kwaliteit op aangezien iedere “lasnaad” kwaliteitsverlies oplevert. Alles apart inspreken is meestal ook geen optie. Denk maar aan de enorme hoeveelheid zinnen die dat op zou leveren in het geval van het omroepsysteem voor de spoorwegen.

(#lijnen) x (#vertrektijden) x (#vertragingen) x (#perrons) x ...

De geluidsbestanden worden opgeslagen en wanneer ze nodig zijn in de gewenste volgorde aan elkaar geplakt en afgespeeld. Vaak moeten de geluidsbestanden voor opslag nog bewerkt worden (Marsi [4]). De volumes van de fragmenten moeten op elkaar afgestemd worden en de overgangen tussen de losse stukken moten gladgestreken worden (smoothing, zie voor meer informatie over smoothing: Martin J. H., D. Jurafski [2]).



Het voordeel van deze methode is dat het natuurlijke spraak oplevert waar af en toe nog een plakrandje in te ontdekken valt. Nadeel is wel dat het slechts voor een beperkt domein geldt. Een ander nadeel dat door Marsi [4] wordt genoemd is de moeilijkheid om de set met uitspraken uit te breiden. Als een bepaalde spreker niet meer beschikbaar is moet alles opnieuw ingesproken worden, of dienen twee sprekers gecombineerd te worden. Denk hierbij aan de NS waar voor de standaard berichten een vrouwenstem gebruikt is (domein synthese), en voor de afwijkende berichten zoals werkzaamheden een mannenstem (apart opgenomen boodschappen).

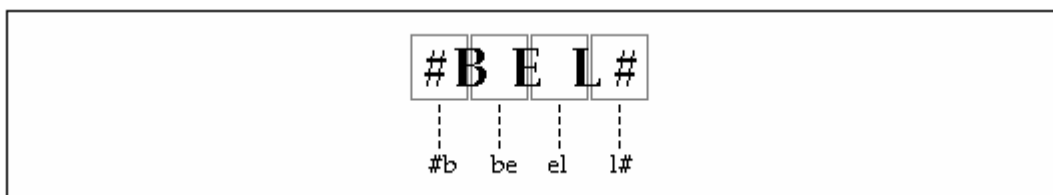
### *Foonsynthese*

Concatenatie systemen maken gebruik van verschillende bouwstenen om de boodschap op te bouwen. Door nu de bouwstenen zo klein mogelijk te maken, zo groot als een foneem, kan iedere uitdrukking gemaakt worden door de losse fonemen te combineren. Het Nederlands kent slechts een bekend aantal fonemen, zo'n veertig. Zelfs als je van alle verschillende fonemen uitspraakvarianten meeneemt, blijft een overzichtelijk aantal bouwstenen over.

Deze methode blijkt echter niet of nauwelijks te werken. Op deze manier gegenereerde spraak is nagenoeg onverstaanbaar. Dit komt doordat de verschillende fonemen zich aanpassen aan de fonemen in hun omgeving. Bij het combineren van losse fonemen wordt dit effect niet meegenomen.

### *Difoonsynthese*

Een methode die ook van kleine bouwstenen uitgaat is de difoonsynthese. Deze methode is een stuk slimmer, omdat hier gebruik gemaakt wordt van de overgangen tussen fonemen. Deze overgangen worden vervolgens aan elkaar geplakt. Het woord "bel" wordt dan samengesteld uit vier verschillende difonen. De eerste is de overgang van stilte naar b, de tweede is de overgang van b naar e, de derde de overgang van e naar l en tot slot die van l naar stilte. Figuur 2 geeft deze opdeling in difonen weer.



Figuur 2: Opdeling van "bel" in difonen

Om op deze manier alle woorden samen te kunnen stellen moet er voor iedere overgang een difoon opgenomen worden. De veertig verschillende difonen die het Nederlands rijk is leveren een eerste schatting op van  $40 \times 40 = 160$  difonen. Rekening houdend met accenten kunnen alle Nederlandse teksten met een kleine 3000 difonen tot klinken worden gebracht (Rietveld en van Heuven [1]). Deze difonen zijn opgeslagen geluiden, die gedestilleerd worden uit een gesproken tekst.

Ook deze methode heeft zijn nadelen. Ten eerste is er geen een op een relatie tussen de geschreven tekst en gewenste geluiden. Er moet een vertaling plaatsvinden tussen de karakters van een woord naar de difonen die daarbij horen. Dit heet grafeem-foneem omzetting. Voorkomens van dezelfde letterparen kunnen verschillende uitspraken hebben. Daarnaast moeten ook afwijkende tekstvormen, zoals acroniemen, afkortingen, getallen, symbolen zoals het euro teken, etc. verwerkt kunnen worden. Dit vereist voorbewerking van de tekst, waarin dergelijke situaties uitgeschreven worden of een difoon synthese systeem dat regels heeft meegekregen hoe om te gaan met dit soort situaties. Tot slot moeten ook prosodieregeling en signaalverwerking toegepast worden om de gewenste uitspraak te krijgen.

Marsi [4] geeft een duidelijk stapsgewijs overzicht van het proces waarmee taal naar spraak kan worden omgezet. Om te beginnen dient er een tekstvoorbewerking plaats te vinden. Hierbij moet gedacht worden aan het isoleren van afzonderlijke zinnen d.m.v. zinseindedetectie, en het uitschrijven van getallen en symbolen en dergelijke. Vervolgens dient de tekst aan een taalkundige analyse onderworpen te worden. Hierbij wordt informatie over de opbouw van de tekst verkregen en van ieder woord bepaald wat de woordsoort is (Part Of Speech Tagging). Dit is soms handig om de ambiguïteit op te lossen die het gevolg is van homografen, hetzelfde woord met meerdere betekenissen. Daarna dient de prosodie toegekend te worden, gevolgd door het omzetten van het woord naar de uitspraak. Voordat de difoonsynthese daadwerkelijk kan plaatsvinden moet ook nog aan de fonetische implementatie gesleuteld worden. Hieronder wordt het regelen van de duur en de grondtoon verstaan. Als dit allemaal gebeurd is kan de spraak gegenereerd worden.

#### *Unit selection*

Unit selection is een methode waarbij gebruik gemaakt wordt van zo groot mogelijke stukken bestaande spraakuitingen. Deze systemen gebruiken een database waarin een grote hoeveelheid gesproken tekst opgeslagen wordt (> 1

uur). Alle mogelijke deelverzamelingen van deze uitgesproken tekst worden ook apart opgeslagen. Dat wil zeggen dat ieder difoon, lettergreep, morfeem (kleinst mogelijke betekenisdragende deel van een woord), woord, zinsdeel en zin apart opgeslagen wordt. Als er spraak gegenereerd moet worden dan wordt volgens deze methode een zo groot mogelijk stuk uit de database gepakt dat overlapt met het stuk uit te spreken tekst. Deze stukken worden vervolgens aan elkaar geplakt (concatenatie). Voor woorden die niet in de database voorkomen, wordt dan in het ergste geval difoonsynthese gebruikt. Het is dus als het ware een combinatie van beide methoden, concatenatie en difoonsynthese, waarbij de grote van de bouwsteen varieert.

Het voordeel van deze methode boven concatenatie synthese is dat een unit synthese systeem niet domein gebonden is. Het voordeel boven difoonsynthese is dat er geen voorbewerking nodig is. Het is dan ook niet verwonderlijk dat unit synthese de meest gangbare methode is om begrijpelijke, natuurlijke, kunstmatige spraak op te leveren.

Toch kleven ook aan deze methode nadelen. Ten eerste is het erg duur en vereist het veel geheugen. Daarnaast zijn er ook kwalitatieve beperkingen. De prosodie laat vaak te wensen over. Dit komt omdat van een gesproken uiting voor een bepaald stuk tekst uit de database nog niet gezegd is dat dit de representatie met de juiste beklemtoning is. Een ander nadeel is dat de kwaliteit variabel is. Wanneer het systeem kan kiezen uit volledige gesproken zinnen uit de database om de tekst te representeren dan klinkt een unit synthese perfect. Komt er echter een term voor die niet in de database staat, dan moet deze met behulp van difonen opgebouwd worden, zonder bewerkingen. Midden in een goedlopende zin kan dan een hele slechte weergave van een woord voorkomen. Vaak zijn deze woorden ook nog belangrijke woorden in die zinnen. Het gaat dan vaak om specifieke woorden, jargon of eigennamen.

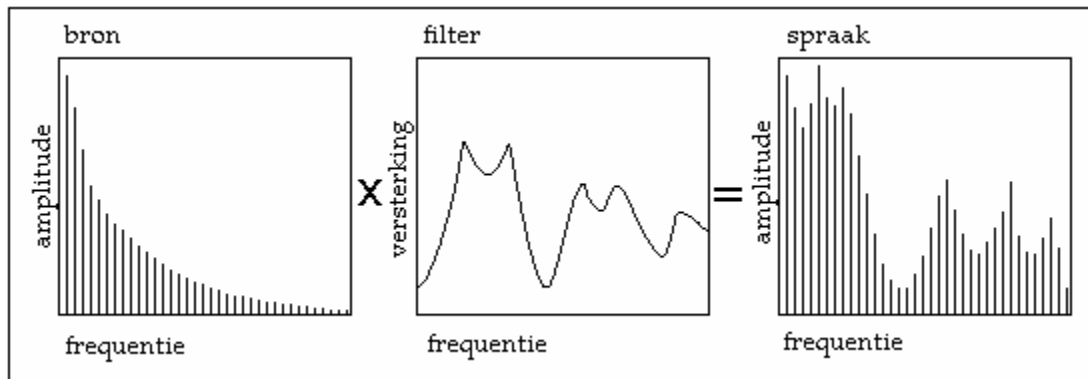
#### *Formantsynthese*

Formantsynthese is een heel andere manier van kunstmatige spraak genereren. Bij deze methode wordt niet real-time gebruik gemaakt van menselijke spraak. Er worden niet van te voren opgenomen golfvormen aan elkaar geplakt, maar op basis van parameters die onder andere formanten en prosodie beschrijven wordt een kunstmatig spraaksignaal gegenereerd.

Formanten zijn de maxima van de grafiek die verkregen wordt wanneer de amplitude van het signaal uitgezet wordt tegen de frequentie. Deze grafiek is

karacteristiek voor de fonemen. Dat is niet zo verwonderlijk omdat voor ieder foneem ook een aparte stand van de articulatoren nodig is.

Het bron-filter model beschrijft goed hoe dit werkt. De stembanden creëren een basis frequentie. De articulatoren leggen hier een voor het foneem karakteristieke formantfrequentie overheen die samen resulteren in de spraak. Figuur 3 geeft een schematische weergave van het bron-filter model.



Figuur 3: Bron-filter model

Bij formant synthese wordt de te synthetiseren uitspraak opgedeeld in frames van meestal 10 ms. Voor iedere frame worden de parameters bepaald op basis waarvan het signaal gegenereerd moet worden.

### Prosodie

De term prosodie verwijst naar de kenmerken van spraak die gerelateerd zijn aan hoorbare veranderingen in snelheid, luidheid en de lengte van lettergrepen (Dutoit [3]). In andere bronnen worden prosodische kenmerken ook wel beschreven als de suprasegmentale kenmerken, kenmerken van meerdere klanken samen (Rietveld en van Heuven [1], Marsi [4]). Prosodie heeft verschillende functies waarvan de belangrijkste het vestigen van aandacht op een bepaald deel van de zin is. Andere functies zijn bijvoorbeeld het structureren van informatie en het tonen van emoties. Prosodie kan de betekenis van woorden en zinnen veranderen. Zonder het aanbrengen van klemtonen kan een woord of zin voor tweërlei uitleg vatbaar zijn. De zin "Jan komt niet naar school omdat hij verliefd is op de juf" kan afhankelijk van de klemtoon betekenen dat a: Jan's verliefdheid reden is om niet naar school te komen of dat b: Jan's verliefdheid niet de reden is waarom hij op school komt. Foute prosodie kan daarom leiden tot een verkeerd begrip van hetgeen gezegd is. Daarnaast wordt een spraakuiting met foute prosodie als onnatuurlijk ervaren en is het moeilijker te verstaan.

Voorbeelden van prosodie waarbij de beklemtoning als onnatuurlijk ervaren kan worden zijn vraag-antwoord constructies waarbij een element in zowel de vraag als het antwoord voorkomt. Vergelijk hiertoe de volgende zinnen:

Q. "Hoe heet de conciërge?"

A. "De conciërge heet Jan."

Q. "Wie heet er Jan op deze school?"

A. "De conciërge heet Jan."

Het antwoord op de eerste vraag moet de klemtoon op "Jan" hebben liggen terwijl het antwoord op de tweede vraag de klemtoon op conciërge moet hebben. Is dit niet het geval, dan wordt dit als onnatuurlijk taalgebruik ervaren.

Systemen kunnen rekening houden met prosodie door naar verschillende indicatoren van prosodie te kijken. Is het woord al een keer genoemd in het voorgaande (geen klemtoon), wat is de woordsoort (zelfstandige naamwoorden vaker beklemtoond). Er zijn zelfs manieren (Formele grammatica's) om hele semantische boomstructuren te maken, zodat met de betekenis rekening gehouden kan worden als prosodie aan de spraak toegekend dient te worden (Martin J. H., D. Jurafski [2]). Dit is echter erg veel werk en garandeert nog steeds niet het ontbreken van fouten.

## Onderzoeksopzet

Uit het vorige hoofdstuk blijkt dat unit synthese een meer gebruikte methode is die een natuurgetrouwere spraak oplevert. Het deel over prosodie gaf echter aan dat fouten in de prosodie niet altijd voorkomen kunnen worden. Doordat dit bij difoon synthese handmatig geregeld kan worden, is deze methode beter op het prosodische vlak. Het natuurgetrouwe karakter van de spraakuitingen wordt hier echter weer ondermijnd doordat met losse difonen altijd hoorbaar is dat het om kunstmatige spraak gaat.

## Onderzoeksvraag

Het is dan ook interessant om te kijken waar de voorkeur van de luisteraar naar uitgaat, wat degene die de kunstmatige spraak hoort het dichtst bij een menselijke spreker in de buurt vind komen: Een spraakuiting met een vrij constante kwaliteit en redelijk goede prosodie (difoonsynthese), of een spraakuiting met een variabele kwaliteit waarvan delen niet van echte spraak te onderscheiden zijn maar waarvan andere delen ineens een stuk slechter zijn, zowel wat kwaliteit als prosodie betreft (unit synthese)?

## Aanpak

Om te kijken welke methode door de luisteraar geprefereerd wordt, krijgt deze steeds twee stimuli aangeboden. De stimuli zijn verklankingen van getypte zinnen. In elk paar van stimuli dat de respondent wordt aangeboden is er één gemaakt met behulp van een unit synthese systeem, de andere met behulp van een difoonsynthese systeem. De gebruiker wordt dan gevraagd om op een schaal van 1-7 aan te geven welke zin zijn voorkeur heeft.

### *Respondenten*

Het onderzoek wordt gehouden onder Nederlandse respondenten. Dit betekent wel dat de systemen ook gemaakt moeten zijn voor het genereren van Nederlandse spraak. Voor het onderzoek hebben twintig respondenten hun oordeel over de stimuli gegeven.

### *Stimuli*

De stimuli bestaan uit paren spraakgeluiden. Eén gemaakt met difoonsynthese, één gemaakt door unit synthese. In totaal zijn aan elke respondent twintig paren stimuli voorgelegd. Deze twintig stimuli zijn weer opgedeeld in vier groepen

van vijf, die verschillende factoren onderzoeken die van invloed zijn op het spraaksignaal. De eerste groep bevat zinnen met woorden die niet als volledig woord voorkomen in de unit synthese database, zoals vakjargon of eigen namenuitdrukkingen en buitenlandse woorden. De tweede groep bevat zinnen waarin frasering een rol speelt. De derde groep bestaat uit zinnen waarin beklemtoning een rol speelt. Tot slot zijn er ook nog vijf teststimuli gemaakt, die bestaan uit zeer algemene woorden, die allemaal als geheel voorkomen in de unit synthese en waarvan dus aangenomen mag worden dat deze als prettiger ervaren worden door de luisteraar. Deze teststimuli zijn opgenomen om de resultaten van de respondenten met een gehoorbeschadiging of respondenten die het niet serieus nemen buiten beschouwing te laten. Voorbeelden van zinnen uit iedere categorie:

- Onbekende woordsoorten/eigen namen/uitdrukkingen/buitenlandse woorden:  
*Hillary Clinton heeft naar verwachting nog wel een appeltje te schillen met Monica Lewinsky.*
- Frasering:  
*Zij zochten de schaar met grote ogen*
- Beklemtoning:  
*Is André van Duin een aangever of een afmaker? André van Duin is een afmaker.*
- Test items:  
*Wij zijn gisteren met de auto naar de winkel gegaan.*

In appendix A is een per categorie gesorteerd overzicht van de stimuli opgenomen. In de eerste zin heeft de unit synthese moeite met eigen namen en de Engelse woorden. De tweede zin kan op twee manieren gelezen worden: Hebben de zoekers of heeft de schaar grote ogen? In de derde zin moet door de vraag de klemtoon op afmaker liggen. In de appendix zijn bij deze stimuli ook de varianten gegeven waar de klemtoon op een ander zinsdeel ligt. De laatste zin is duidelijk een test item met algemene woorden. Om respondenten niet te veel inzicht te laten krijgen in het onderzoek zijn de stimuli niet per categorie afgewerkt maar gerandomiseerd, en zonder uitleg aan de respondent voorgelegd. De gerandomiseerde lijst is opgenomen in appendix B.

*Scoring*

Om iets meer te kunnen zeggen dan alleen “unit synthese wordt als beter ervaren dan difoon synthese” of andersom, is niet gevraagd of de respondenten aan wilden geven welke stimulus uit het paar beter op hen over kwam. In plaats daarvan is gevraagd of ze op een schaal van goed tot slecht, opgedeeld in zeven stukken, aan wilden geven waar ze de beide spraakuitingen vonden staan. Deze manier van scoring meet impliciet welke uiting als beter wordt ervaren, maar zegt daarnaast ook nog iets over hoever de scores uit elkaar liggen. En daarmee welke factor de grootste invloed heeft op de ervaring van de spraak. Ook kan er op deze manier aangegeven worden dat de stimuli van dezelfde kwaliteit zijn.

Het is moeilijk te geloven dat Erica Terpstra in vroeger tijden excelleerde in borst crawl en domineerde op menig zwemtoernooi.

6A	slecht	1. <input type="radio"/>	2. <input type="radio"/>	3. <input type="radio"/>	4. <input type="radio"/>	5. <input type="radio"/>	6. <input type="radio"/>	7. <input type="radio"/>	goed
6B	slecht	1. <input type="radio"/>	2. <input type="radio"/>	3. <input type="radio"/>	4. <input type="radio"/>	5. <input type="radio"/>	6. <input type="radio"/>	7. <input type="radio"/>	goed

In Appendix C is de complete antwoordlijst opgenomen.

*Systemen*

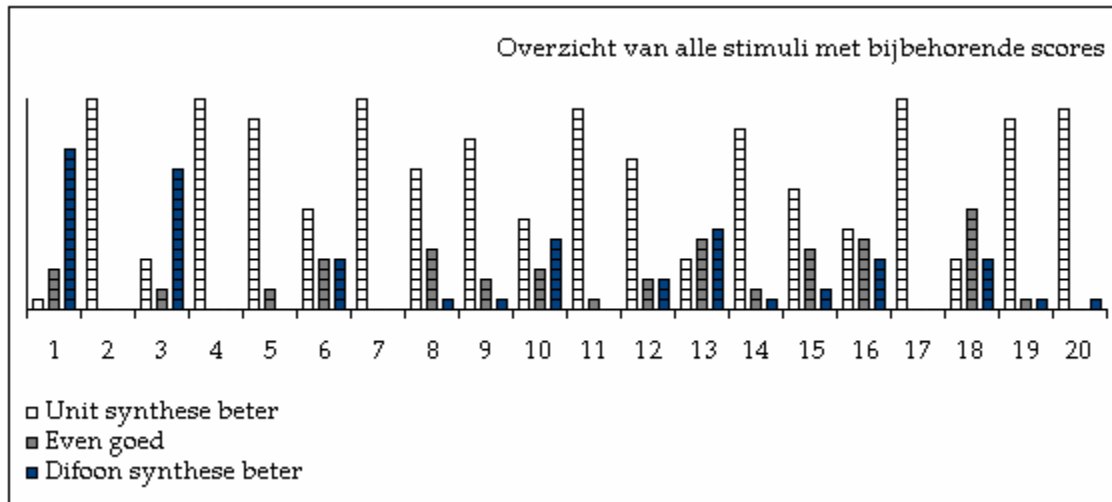
Het difoonsynthese systeem dat voor dit onderzoek gebruikt is, is Nextens. Nextens staat voor Nederlandse Extensie voor Tekst naar Spraak [7], waarvan een on-line demo beschikbaar is en een gratis te downloaden versie voor de pc.

Het unit synthese systeem dat gebruikt is voor het onderzoek is Acapela. Hiervan is eveneens een on-line demo beschikbaar [8] waarmee zinnen tot 250 karakters gegenereerd kunnen worden. Voor dit onderzoek is echter gebruik gemaakt van een demoversie die dertig dagen gratis gebruikt kan worden [9] en die op de pc te installeren is.



## Uitkomsten

De resultaten van de twintig respondenten staan hieronder weergegeven in figuur 4. Een snelle blik leert dat het unit synthese systeem in het merendeel van de gevallen beter uit de verf komt. In deze grafiek staan de verschillende soorten stimuli echter nog door elkaar heen omdat ze in een random volgorde aangeboden zijn aan de respondenten.



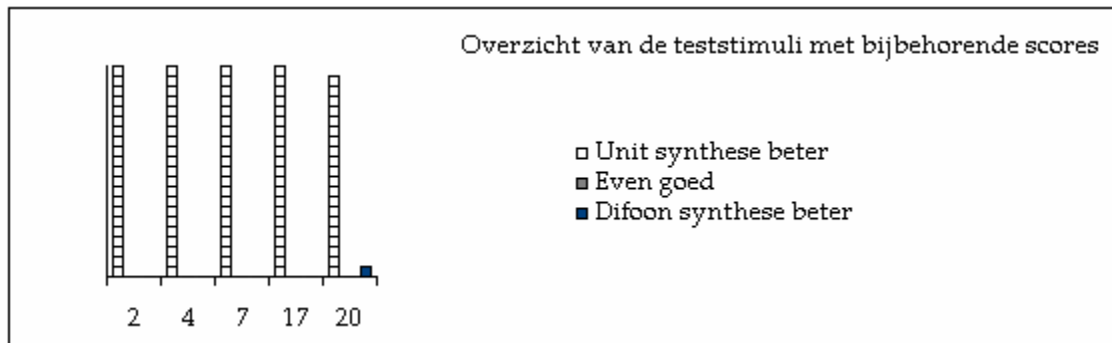
Figuur 4: Resultaten

Als de uitkomsten per soort, dus per groep stimuli, bekeken worden dan is er een beter inzicht mogelijk in wat de effecten op de perceptie waren.

Als eerste kijken we naar de teststimuli. Dit zijn de stimuli waarvan de woorden allemaal zo goed als zeker als geheel voorkomen in de database van het unit synthese systeem.

- Wij zijn gisteren met de auto naar de winkel gegaan.  
(stimulus 2)
- De boom die in de tuin staat is hoog en groen.  
(stimulus 4)
- De sporters lopen met hun tassen in de richting van de uitgang.  
(stimulus 7)
- Ze zaten wel drie uur over de wedstrijd na te praten.  
(stimulus 17)
- Hier kun je lekker eten en drinken en het is niet duur.  
(stimulus 20)

De teststimuli zijn ook naast elkaar in een grafiek gezet. Deze is weergegeven in figuur 5.



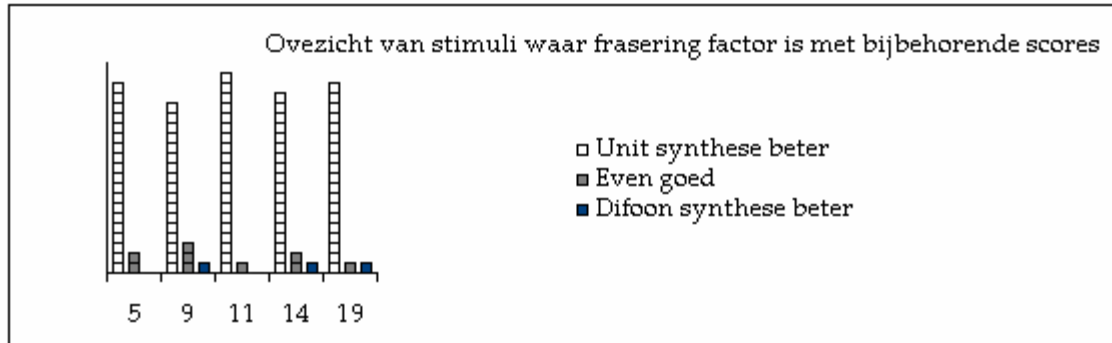
Figuur 5: Teststimuli

Een stimulus is door een respondent als beter aangegeven. Omdat de andere stimuli van deze respondent wel allemaal het unit synthese systeem als betere aanwezen, zijn ook de resultaten van deze respondent meegenomen in de evaluatie. Het was niet onoplettendheid of een slecht gehoor waardoor deze respondent tot deze keuze kwam, maar een weergave van zijn perceptie van de kwaliteit van de twee spraakuitingen voor stimulus 20.

Als tweede groep worden de zinnen waarin frasering een rol speelt apart genomen. Deze zijn hieronder weergegeven. Merk hierbij op dat iedere zin op twee manieren uitgelegd kan worden. Bijvoorbeeld in de zin "Zij zochten de schaar met grote ogen" kan het zo zijn dat een stel mensen grote ogen heeft of dat de schaar grote ogen heeft.

- De meester kijkt naar zijn leerlingen met een verrekijker.  
(stimulus 5)
- Zij zochten de schaar met grote ogen.  
(stimulus 9)
- Zij pakte de fles met een rode kleur.  
(stimulus 11)
- De pastoor pakte de kandelaar met een kwast aan de arm.  
(stimulus 14)
- Hij loopt naar het restaurant langs de weg.  
(stimulus 19)

Ook deze gegevens zijn in een tabel weergegeven op de volgende pagina in figuur 6.



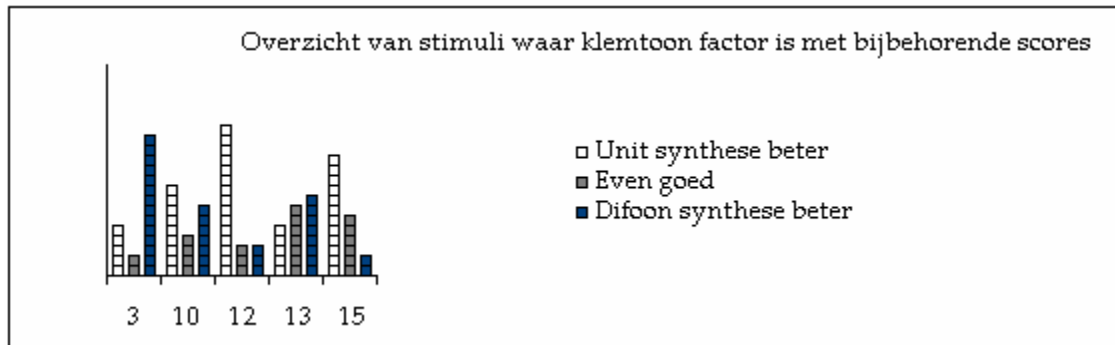
Figuur 6: Stimuli frasering

Ook in het geval van de frasering is het de unit synthese die als meest natuurgetrouw wordt ervaren. Dit is natuurlijk ook niet zo gek omdat het effect van frasering behoorlijk subtiel is. Daarnaast is het zo dat een gerealiseerde uitspraak waarin frasering een rol speelt niet als fout ervaren hoeft te worden. Alle vijf zinnen zijn op twee manieren op te vatten maar de kans dat iemand een uiting als onnatuurlijk bestempelt op basis hiervan is vrij klein. De keren dat het difoonsynthese systeem beter naar voren kwam dan de unit synthese zijn dan waarschijnlijk meer te wijten aan andere oorzaken dan frasering. Bijvoorbeeld aan één van de twee factoren die een rol spelen bij de overige tien stimuli. Deze hebben wel een duidelijkere invloed op de ervaren kwaliteit van de verklanking.

Als eerste worden de stimuli behandeld waarin klemtoon een rol speelt. Deze klemtoon is in de hieronder staande voorbeelden steeds bewerkstelligd door een vraag antwoord constructie. Doordat de vraag op een bepaalde manier gesteld is, is er maar n manier van beklemtonen mogelijk in het antwoord.

- Welke bus gaat naar het station? Bus 1 gaat naar het station.  
(stimulus 3)
- Is André van Duin een aangever of een afmaker? André van Duin is een afmaker.  
(stimulus 10)
- Wie komt er om half drie op visite? Oma komt om half drie.  
(stimulus 12)
- Wat is het symbool van het christelijk geloof? Het kruis is het symbool van het christelijk geloof.  
(stimulus 13)
- Is het konijn het favoriete dier van jou of van je broertje? Het konijn is mijn favoriete dier.  
(stimulus 15)

De scores van de respondenten voor deze stimuli staan in figuur 7.



Figuur 7: Stimuli klemtoon

Het is te zien dat de klemtoon een grotere invloed op het verschil in perceptie heeft. Niet alleen wordt de difoonsynthese regelmatig als de betere verklanking ervaren, voor sommige stimuli vindt zelfs het merendeel van de respondenten de difoonsynthese het beste.

Tot slot kijken we naar de resultaten van de stimuli waarin woorden voorkomen die niet in de unit synthese database voorkomen. Hieronder vallen eigen namen, uitdrukkingen, woorden die weinig voor komen, vakjargon en woorden uit andere talen.

De Hubble telescoop schiet schitterende plaatjes van het universum.

(stimulus 1)

Het is moeilijk te geloven dat Erica Terpstra in vroeger tijden excelleerde in borst crawl en domineerde op menig zwemtoernooi.

(stimulus 6)

Een cocker spaniel is een leuk beest, mopshondjes en pekineesjes zijn echter veel te truttig.

(stimulus 8)

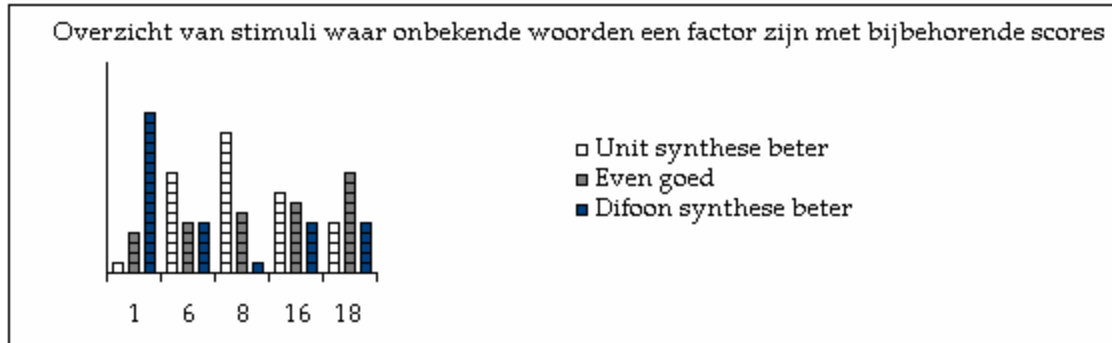
Hillary Clinton heeft naar verwachting nog wel een appeltje te schillen met Monica Lewinsky.

(stimulus 16)

Clint Eastwood speelt een stoere cowboy in de western "Two mules for sister sara".

(stimulus 18)

Figuur 8 toont de scores voor deze stimuli in een grafiek.



Figuur 8: Stimuli onbekende woorden

Ook hier is een aanzienlijk effect te zien van de inhaalslag die het difoonsynthese systeem maakt. De difoonsynthese wordt vaak als de betere ervaren. Maar slechts in één geval komt de difoonsynthese vaker dan de unit synthese als beste naar voren.

### Verdeeldheid

Niet alleen was het zo dat het unit synthese systeem bij de teststimuli en bij de meeste stimuli met frasering als beste werd ervaren, ook lagen de scores verder uit elkaar. Voor de scores van de stimuli met beklemtoning en die met woorden die niet in de database van het unit synthese systeem voorkomen, geldt juist het omgekeerde. Hier kwam de difoon synthese regelmatig als beste verklanking naar voren en daarnaast lagen de waarden ook vaak dicht bij elkaar.

### Kanttekeningen

Bij deze resultaten moet wel een aantal zaken opgemerkt worden. Ten eerste is het zo dat bij dit onderzoek het genereren van de difoonsynthese door een relatief onervaren persoon gedaan is. Een professional op het gebied van linguïstiek die al vaak met dergelijke programma's gewerkt heeft, zal wellicht een beter resultaat bewerkstelligen voor de difoonsynthese.

Daarnaast dient opgemerkt te worden dat luisteraars sterk waren gericht op de echtheid van het stemgeluid. En dat is bij unit synthese in het overgrote merendeel van de verklanking het geval. Vaak werd het echte stemgeluid van de unit synthese van acapela als zeer prettig ervaren. De verklanking die door het difoonsynthese systeem gemaakt zijn, werden vaak als "blikkerig" of "robotachtig" omschreven. Dit heeft echter niet zozeer met het beoordelen van prosodie te maken.

Ten derde speelden (lichte) emoties ook vaak een rol bij de beoordeling van de stimuli. Af en toe reageerden de respondenten op de opmerkingen, de stem van de unit synthese werd zelfs een keer als truttig omschreven.

Tot slot dient opgemerkt te worden dat in dit onderzoek niet alle factoren die een rol kunnen spelen, zijn meegenomen. Het effect van het gebruik van bijzondere tekens of van getallen, en hoe de verklanking hiervan wordt ervaren, is bijvoorbeeld niet meegenomen.

## Conclusies

Uit de verzamelde reacties van de respondenten blijkt duidelijk dat voor algemene zinnen, waarvan de woorden zo algemeen zijn dat ze als een heel woord in de database van het unit synthese systeem voorkomen, geldt dat de unit synthese de geprefereerde verklanking oplevert. Dit laatste kan ook gezegd worden voor zinnen waarin door frasering de betekenis van de zin bepaald moet worden. In verreweg de meeste gevallen genoot de unit synthese hier de voorkeur. De difoonsynthese kwam beter naar voren in zinnen waar beklemtoning een must is op bepaalde plaatsen. Dit geldt ook voor de zinnen met onbekende woorden. Het aangeven van prosodie en het constante kwaliteitsniveau zorgen ervoor dat in dit soort zinnen soms door difoonsynthese een betere verklanking gerealiseerd wordt.

We kunnen derhalve stellen dat de mogelijkheid tot het toevoegen van prosodie de difoonsynthese een geduchtere concurrent maakt voor de unit synthese. Toch is laatstgenoemde nog vaak de methode die de als prettigst ervaren spraak oplevert. De prosodie brengt het niveau van de difoonsynthese wel een stuk omhoog, maar niet altijd genoeg om de kwaliteit van de menselijke spraak te evenaren.

## Bronnen

### Literatuur

- [1] Rietveld A.C.M., V.J. van Heuven, *Algemene fonetiek*, Coutinho, bussum, 2001
- [2] Martin J. H., D. Jurafski, *An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice Hall, 2000
- [3] Dutoit T., A short introduction to Text-to-speech synthesis, *Journal of Electrical & Electronics Engineering*, vol. 17, p. 25-37
- [4] Marsi E., College Spraaksynthese, Universiteit van Tilburg (webpagina 7)
- [5] Pols, L. C. W. (1998b), "Speech synthesis evaluation", In: R. Cole (Ed.), *Survey of the state of the art in human language technology*, Pisa: Giardini Editori e Stampatori, 429-430.
- [6] Handbook of standards and resources for spoken language systems D Gibbon, R Moore, R Winski - 1997 - Berlin; New York: Mouton de Gruyter Cited by 115 - Web Search - Library Search

### Webpagina's

- [7] <http://nextens.uvt.nl/nextens-wiki/Links>
- [8] <http://demo.acapela-group.com/>
- [9] [http://www.advtechnics.com/Nederlands/Spraaksynthesizers/Infovox\\_De\\_sktop\\_Pro\\_NL/](http://www.advtechnics.com/Nederlands/Spraaksynthesizers/Infovox_De_sktop_Pro_NL/)



## Appendix A: Stimuli

---

De twintig stimuli met hun nummering en soort synthese, gegroepeerd per factor die van invloed verondersteld wordt op de perceptie van de luisteraar. Er worden drie factoren onderzocht, te weten: 1. Gebruik van woorden die niet als volledig woord voorkomen in de unit synthese database, zoals vakjargon of eigen namenuitdrukkingen en buitenlandse woorden, 2. Zinnen waarin frasering een rol speelt, 3. zinnen waarin beklemtoning een rol speelt. Tot slot is er nog een aantal teststimuli die moeten testen of respondenten wel geschikt zijn voor het onderzoek (gehoorcapaciteit) en of ze het wel serieus nemen. Deze stimuli worden geheel opgebouwd uit grote stukken uit de unit synthese database, en klinken derhalve natuurlijker.

---

### 1. Onbekende woordsoorten/eigen namen/uitdrukkingen/buitenlandse woorden

1. Hillary Clinton heeft naar verwachting nog wel een appeltje te schillen met Monica Lewinsky.
2. De Hubble telescoop schiet schitterende plaatjes van het universum.
3. Het is moeilijk te geloven dat Erica Terpstra in vroeger tijden excelleerde in borst crawl en domineerde op menig zwemtoernooi.
4. Een cocker spaniel is een leuk beest, mopshondjes en pekineesjes zijn echter veel te truttig.
5. Clint Eastwood speelt een stoere cowboy in de western "Two mules for sister sara"

### 2. Frasering

- De meester kijkt [naar zijn leerlingen met een verrekijker]
- De meester kijkt naar zijn leerlingen [met een verrekijker]
- Hij loopt [naar het restaurant langs de weg]
- Hij loopt naar het restaurant [langs de weg]
- Zij pakte [de fles met een rode kleur]
- Zij pakte de fles met [een rode kleur]
- De pastoor pakte [de kandelaar met een kwast aan de arm]
- De pastoor pakte de kandelaar [met een kwast aan de arm]
- Zij zochten [de schaar met grote ogen]
- Zij zochten de schaar [met grote ogen]

### 3. Beklemtuning

- Is André van Duin of Frans van Dusschoten de afmaker? André van Duin is een afmaker.
- Is André van Duin een aangever of een afmaker? André van Duin is een afmaker.
  
- Hoe laat komt oma op visite? Oma komt om half drie.
- Wie komt er om half drie op visite? Oma komt om half drie.
  
- Wat is het symbool van het christelijk geloof? Het kruis is het symbool van het christelijk geloof.
- Van welk geloof is het kruis het symbool? Het kruis is het symbool van het christelijk geloof.
  
- Wat is jouw favoriete dier? Het konijn is mijn favoriete dier.
- Is het konijn het favoriete dier van jou of van je broertje? Het konijn is mijn favoriete dier.
  
- Waarheen gaat bus 1? Bus 1 gaat naar het station.
- Welke bus gaat naar het station? Bus 1 gaat naar het station.

### 4. Test items

1. Wij zijn gisteren met de auto naar de winkel gegaan.
2. Ze zaten wel drie uur over de wedstrijd na te praten.
3. De boom die in de tuin staat is hoog en groen.
4. Hier kun je lekker eten en drinken en het is niet duur.
5. De sporters lopen met hun tassen in de richting van de uitgang.

## Appendix B: Gerandomiseerde lijst stimuli

---

Stimuli in de volgorde waarin ze zijn afgenomen.

---

1. De Hubble telescoop schiet schitterende plaatjes van het universum.
2. Wij zijn gisteren met de auto naar de winkel gegaan.
3. Welke bus gaat naar het station? Bus 1 gaat naar het station.
4. De boom die in de tuin staat is hoog en groen.
5. De meester kijkt naar zijn leerlingen met een verrekijker.
6. Het is moeilijk te geloven dat Erica Terpstra in vroeger tijden excelleerde in borst crawl en domineerde op menig zwemtoernooi.
7. De sporters lopen met hun tassen in de richting van de uitgang.
8. Een cocker spaniel is een leuk beest, mopshondjes en pekinneesjes zijn echter veel te truttig.
9. Zij zochten de schaar met grote ogen.
10. Is André van Duin een aangever of een afmaker? André van Duin is een afmaker.
11. Zij pakte de fles met een rode kleur.
12. Wie komt er om half drie op visite? Oma komt om half drie.
13. Wat is het symbool van het christelijk geloof? Het kruis is het symbool van het christelijk geloof.
14. De pastoor pakte de kandelaar met een kwast aan de arm.
15. Is het konijn het favoriete dier van jou of van je broertje? Het konijn is mijn favoriete dier.
16. Hillary Clinton heeft naar verwachting nog wel een appeltje te schillen met Monica Lewinsky.
17. Ze zaten wel drie uur over de wedstrijd na te praten.
18. Clint Eastwood speelt een stoere cowboy in de western "Two mules for sister sara".
19. Hij loopt naar het restaurant langs de weg.
20. Hier kun je lekker eten en drinken en het is niet duur.

## Appendix C: Antwoordlijst

---

Je krijgt telkens twee geluidsfragmenten te horen. Luister ze eerst allebei af. Vul vervolgens voor beide zinnen in, in hoeverre je de gehoorde fragmenten een goede benadering van echte spraak vindt (slecht tot goed). Let hierbij op de kwaliteit van het spraaksignaal en op de natuurlijkheid van klemtonen.

---

1. De Hubble telescoop schiet schitterende plaatjes van het universum.  
1A slecht 1.  2.  3.  4.  5.  6.  7.  goed  
1B slecht 1.  2.  3.  4.  5.  6.  7.  goed
2. Wij zijn gisteren met de auto naar de winkel gegaan.  
2A slecht 1.  2.  3.  4.  5.  6.  7.  goed  
2B slecht 1.  2.  3.  4.  5.  6.  7.  goed
3. Welke bus gaat naar het station? Bus 1 gaat naar het station.  
3A slecht 1.  2.  3.  4.  5.  6.  7.  goed  
3B slecht 1.  2.  3.  4.  5.  6.  7.  goed
4. De boom die in de tuin staat is hoog en groen.  
4A slecht 1.  2.  3.  4.  5.  6.  7.  goed  
4B slecht 1.  2.  3.  4.  5.  6.  7.  goed
5. De meester kijkt naar zijn leerlingen met een verre kijker.  
5A slecht 1.  2.  3.  4.  5.  6.  7.  goed  
5B slecht 1.  2.  3.  4.  5.  6.  7.  goed
6. Het is moeilijk te geloven dat Erica Terpstra in vroeger tijden excelleerde in borst crawl en domineerde op menig zwemtoernooi.  
6A slecht 1.  2.  3.  4.  5.  6.  7.  goed  
6B slecht 1.  2.  3.  4.  5.  6.  7.  goed
7. De sporters lopen met hun tassen in de richting van de uitgang.  
7A slecht 1.  2.  3.  4.  5.  6.  7.  goed  
7B slecht 1.  2.  3.  4.  5.  6.  7.  goed
8. Een cocker spaniel is een leuk beest, mopshondjes en pekineesjes zijn echter veel te truttig.  
8A slecht 1.  2.  3.  4.  5.  6.  7.  goed  
8B slecht 1.  2.  3.  4.  5.  6.  7.  goed
9. Zij zochten de schaar met grote ogen.  
9A slecht 1.  2.  3.  4.  5.  6.  7.  goed  
9B slecht 1.  2.  3.  4.  5.  6.  7.  goed
10. Is André van Duin een aangever of een afmaker? André van Duin is een afmaker.  
10A slecht 1.  2.  3.  4.  5.  6.  7.  goed  
10B slecht 1.  2.  3.  4.  5.  6.  7.  goed

11. Zij pakte de fles met een rode kleur.

11A slecht 1.  2.  3.  4.  5.  6.  7.  goed

11B slecht 1.  2.  3.  4.  5.  6.  7.  goed

12. Wie komt er om half drie op visite? Oma komt om half drie.

12A slecht 1.  2.  3.  4.  5.  6.  7.  goed

12B slecht 1.  2.  3.  4.  5.  6.  7.  goed

13. Wat is het symbool van het christelijk geloof? Het kruis is het symbool van het christelijk geloof.

13A slecht 1.  2.  3.  4.  5.  6.  7.  goed

13B slecht 1.  2.  3.  4.  5.  6.  7.  goed

14. De pastoor pakte de kandelaar met een kwast aan de arm.

14A slecht 1.  2.  3.  4.  5.  6.  7.  goed

14B slecht 1.  2.  3.  4.  5.  6.  7.  goed

15. Is het konijn het favoriete dier van jou of van je broertje? Het konijn is mijn favoriete dier.

15A slecht 1.  2.  3.  4.  5.  6.  7.  goed

15B slecht 1.  2.  3.  4.  5.  6.  7.  goed

16. Hillary Clinton heeft naar verwachting nog wel een appeltje te schillen met Monica Lewinsky.

16A slecht 1.  2.  3.  4.  5.  6.  7.  goed

16B slecht 1.  2.  3.  4.  5.  6.  7.  goed

17. Ze zaten wel drie uur over de wedstrijd na te praten.

17A slecht 1.  2.  3.  4.  5.  6.  7.  goed

17B slecht 1.  2.  3.  4.  5.  6.  7.  goed

18. Clint Eastwood speelt een stoere cowboy in de western "Two mules for sister sara".

18A slecht 1.  2.  3.  4.  5.  6.  7.  goed

18B slecht 1.  2.  3.  4.  5.  6.  7.  goed

19. Hij loopt naar het restaurant langs de weg.

19A slecht 1.  2.  3.  4.  5.  6.  7.  goed

19B slecht 1.  2.  3.  4.  5.  6.  7.  goed

20. Hier kun je lekker eten en drinken en het is niet duur.

20A slecht 1.  2.  3.  4.  5.  6.  7.  goed

20B slecht 1.  2.  3.  4.  5.  6.  7.  goed