

Invloed van open quotiënt, spectrale helling, skewness en amplitude op de perceptie van de relatieve toonhoogte

Citation for published version (APA):

Prinsen, H., & Collier, R. P. G. (1996). *Invloed van open quotiënt, spectrale helling, skewness en amplitude op de perceptie van de relatieve toonhoogte*. (IPO rapport; Vol. 1139). Instituut voor Perceptie Onderzoek (IPO).

Document status and date:

Gepubliceerd: 13/12/1996

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. 1139

**Invloed van open quotiënt,
spectrale helling, skewness en
amplitude op de perceptie van
de relatieve toonhoogte**

Harrie Prinsen

Voor akkoord: Prof.dr. R. Collier



INVLOED VAN OPEN QUOTIENT, SPECTRALE
HELLING, SKEWNESS EN AMPLITUDE OP DE
PERCEPTIE VAN DE RELATIEVE TOONHOOGTE.

Onderzoek aan het Instituut voor Perceptie Onderzoek te Eindhoven als afstudeerproject voor de Faculteit Educatieve Opleidingen, het Mollerinstituut te Tilburg.

Harrie Prinsen

1996

INVLOED VAN OPEN QUOTIENT, SPECTRALE HELLING, SKEWNESS EN AMPLITUDE OP DE PERCEPTIE VAN DE RELATIEVE TOONHOOGTE.

SAMENVATTING

Dit rapport doet verslag van twee experimenten. In het eerste experiment wordt d.m.v. spraakanalyse onderzoek gedaan aan spraakparameters. Spraakparameters worden hier gedefinieerd als fysische grootheden die bepalen op welke manier de luchtstroom door de stemspleet gaat. Daardoor bepalen zij direct de stemkwaliteit. Verondersteld wordt dat enkele van deze parameters beïnvloed worden door de toonhoogte waarop gesproken wordt. In het eerste experiment zullen we onderzoeken of dergelijke relaties tussen toonhoogte en spraakparameters inderdaad bestaan en zoja, hoe deze relaties eruitzien (meer specifiek: de relatie tussen relatieve toonhoogte enerzijds en open quotient, skewness, spectrale helling en amplitude anderzijds). In het tweede experiment onderzoeken we of parametervariatiaties in kunstmatige spraakuitingen de toonhoogteperceptie beïnvloeden.

PROBLEEMSTELLING

INLEIDING

Tot nu toe klinkt kunstmatig gesynthetiseerde spraak, zeker volledige zinnen of teksten, tamelijk onnatuurlijk. Vaak is deze spraak monotoon, egaal of robotachtig, kortom ze vertoont niet de afwisseling die natuurlijke spraak zo aangenaam maakt om naar te luisteren.

Op vele plaatsen in de wereld en op vele manieren probeert men de kwaliteit van kunstmatige spraak te verbeteren. Getracht wordt de wetten te ontdekken die ten grondslag liggen aan de vele variaties in de menselijke spraak. Als deze variaties op een beregelde manier kunnen worden ingebouwd in de synthese, dan is het misschien mogelijk om in de toekomst kunstmatige spraak te genereren die niet of nauwelijks te onderscheiden is van natuurlijke spraak.

We kunnen ruwweg twee soorten variaties onderscheiden in de menselijke spraak:

- prosodische variaties
- stemvariatie

Een deel van het wetenschappelijk onderzoek richt zich op de prosodie en dus op het ontdekken en toepassen van regels die gelden voor zinsmelodie, ritme, pauzering en accentuering. Ondertussen weet men al zeer veel over de systematiek van de intonatie.

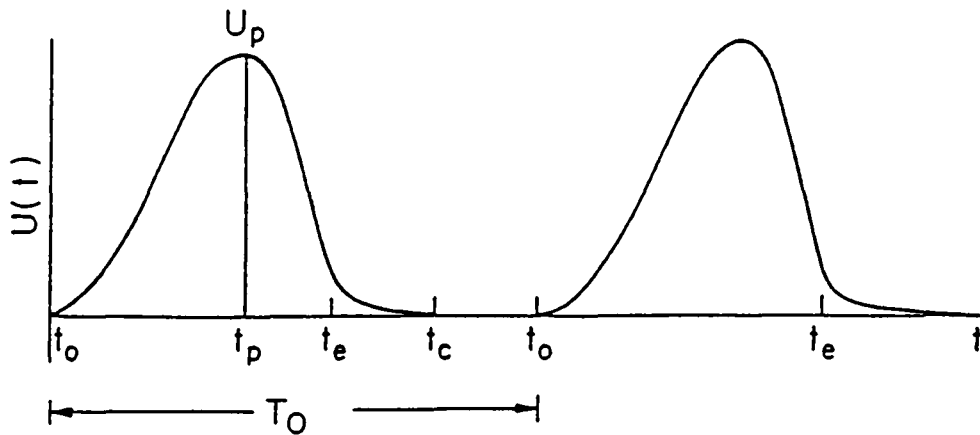
Een ander deel van het wetenschappelijk onderzoek richt zich op de stemkarakteristieken. Van verschillende parameters is nog niet eenduidig vastgesteld hoe ze zich gedragen in de natuurlijke spraak. Zo vermoedt men relaties tussen toonhoogte enerzijds en andere spraakparameters anderzijds. Juist het onderzoek naar de relatie tussen toonhoogte en andere spraakparameters is zo belangrijk omdat de resultaten hiervan direct gekoppeld kunnen worden aan de inzichten die ondertussen verkregen zijn over toonhoogtevariatie door intonatie. Er is echter nog veel research te doen naar de relaties tussen de verschillende spraakparameters. Dit in kaart brengen en later toepassen kan een grote bijdrage leveren aan de verbetering van kunstmatige spraak.

In dit rapport worden eerst twee modellen besproken die veel gebruikt worden bij spraaksynthese en spraakanalyse. Met behulp van deze modellen zijn we in staat verschillende spraakparameters kwantitatief en kwalitatief te definiëren. Vervolgens wordt een samenvatting gegeven van wat tot nu toe bekend is over bepaalde correlaties tussen enkele van deze parameters. In 'Experiment 1' wordt analyseonderzoek gedaan aan spraakuitingen op verschillende toonhoogtes. We onderzoeken hierbij de mogelijke relatie tussen de toonhoogte en drie bronparameters: open quotient(OQ), spectrale helling(tilt) en amplitude(AV). 'Experiment 2' bestaat uit een serie perceptieproeven. Stimuli met een vaste f_0 maar met verschillende OQ-, tilt- en AV-waarden worden door proefpersonen beoordeeld op relatieve toonhoogte. Aan het slot van dit rapport doen we enkele suggesties voor vervolgonderzoek.

MODELLEN

Veelal wordt in de spraakanalyse en spraaksynthese gebruik gemaakt van het bron-filtermodel. Dat model gaat ervanuit dat het spraaksignaal het gecombineerde resultaat is van een bron en een filter. De stembanden, die door passerende lucht in trilling zijn gebracht (de bron), produceren een signaal dat door de keel-, neus- en mondholte wordt bewerkt (het filter). In de meest simpele vorm worden bron en filter beschouwd als losse zelfstandige mechanismen, die elkaar niet beïnvloeden tijdens de spraakproductie.

In de bron ontstaat door het open- en dichtgaan van de stembanden een tijdafhankelijke luchtstroom : de glottale puls. De glottale puls kan beschreven worden als de stroomsterkte van de lucht (U) door de stemspleet (glottis) afhankelijk van de tijd.(zie figuur 1)



Figuur 1 : Stroomsterkte van de lucht (U) door de stemspleet afhankelijk van de tijd (t).

Relevante tijdstippen in de glottale puls zijn:

t_0 = tijdstip waarop stemspleet opent.

t_p = tijdstip waarop luchtstroomsterkte maximaal is.

t_e = tijdstip waarop het sluiten van de stemspleet abrupt vertraagt. (afgeleide maximaal negatief)

t_c = tijdstip waarop stemspleet volledig gesloten is.

In de grafiek zijn vier perioden te onderscheiden:

$t_0 < t < t_p$ = periode waarin de luchtstroom door de stemspleet groter wordt.

$t_p < t < t_e$ = periode waarin de luchtstroom door de stemspleet kleiner wordt.

$t_e < t < t_0$ = periode waarin de stemspleet gesloten is tot deze weer open gaat voor de volgende puls.

T_0 = pulsperiode, tijdsduur tussen twee opeenvolgende t_0 .

De toonhoogte van de spraak (f_0) wordt bepaald door T_0 volgens:

$$f_0 = 1/T_0 \quad (\text{eenheid: Hertz})$$

De fysische parameters open quotient, spectrale helling, skewness en amplitude zijn alle direct of indirect af te leiden uit de glottale puls. Deze parameters kunnen worden gedefinieerd volgens twee modellen die de glottale puls (of de afgeleide hiervan) beschrijven. Dit zijn het Rosenbergmodel en het LF-model.

Het Rosenbergmodel

(Rosenberg, 1971)

De stroomsterkte $U(t)$ wordt volgens het Rosenberg-model beschreven door een derde-graadsvergelijking:

$$U(t) = at^2 - bt^3$$

a en b zijn constanten waarvan de waarde afhangt van de amplitude en de duur van de glottale puls.

De volgende parameters kunnen met behulp van dit model worden gedefinieerd:

- * Amplitude (AV) = maximale stroomsterkte door de stemspleet.
 $U(t) = a(t^2 - b/a * t^3)$
 a is een maat voor AV
- * Open quotient (OQ) = verhouding tussen periode waarin de stemspleet volledig geopend is en de trillingstijd (periode) van f_0
 $U(t) = at^2(1 - b/a * t)$
 $t_e = a/b$
 $OQ = a/b * 1/T_0$
 a/b is een maat voor OQ
- * Skewness (Rk) = maat voor de symmetrie van de glottale puls
 $Rk = (t_e - t_p)/(t_p - t_0)$
 $t_e = a/b$
 $t_p = 2a/3b$
 $Rk = 0,5$ (constante waarde)

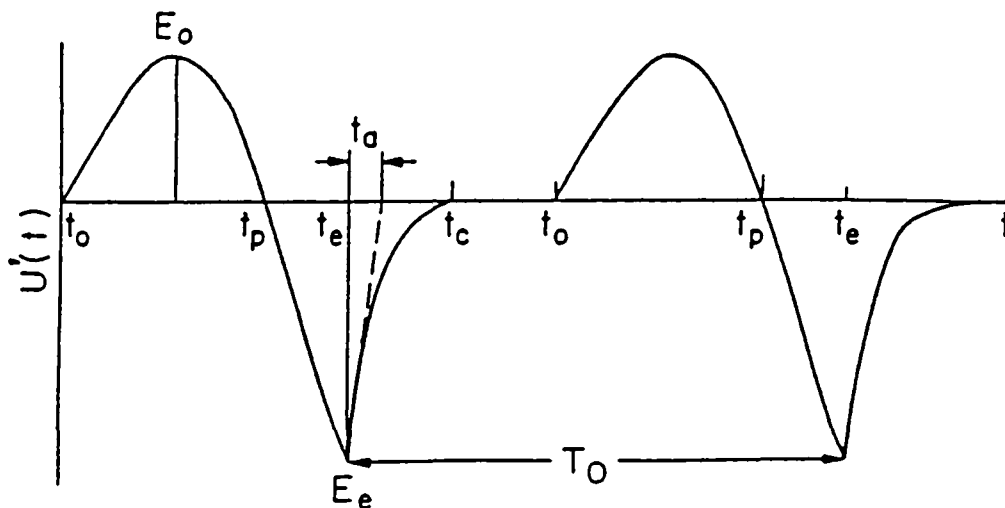
Het LF- model

(Fant, Liljencrants en Lin, 1985).

Dit model beschrijft de afgeleide van $U(t)$ en houdt rekening met de vrij abrupte vertraging in de afname van de stroomsterkte die optreedt aan het einde van de glottale puls. (Het Rosenbergmodel doet dat niet) Vlak voor de definitieve sluiting van de stemspleet raken de stembanden elkaar aan ventrale zijde. Vanaf dat moment (t_e) gaat de stemspleet verder dicht als een ritssluiting (Sluijter, 1995). Daardoor neemt de luchtstroom geleidelijk af naar nul. Deze vertraagde adductie, die ook wel 'returnphase' wordt genoemd, begint op t_e en eindigt op t_c en wordt beschreven door de functie 2. De afgeleide van de glottale puls van t_0 tot t_e wordt beschreven door functie 1.

$$t_0 < t < t_e \quad U'(t) = E(t) = E_0 \cdot e^{at} \cdot \sin(\omega_g t) \quad (1)$$

$$t_e < t < t_c \quad U'(t) = E_e / (e \cdot t_a) \cdot [e^{-e(tc-te)} - e^{-e(t-te)}] \quad (2)$$



Figuur 2 : Afgeleide van de glottale puls $U'(t)$ zoals beschreven door functie 1 en functie 2 van het LF-model.

De volgende parameters kunnen met behulp van dit model worden gedefinieerd:

- * Amplitude
 E_0 is een maat voor de amplitude van de glottispuls.
- * Open quotient
 $OQ = t_e/T_0$
- * Spectrale helling
 $t_a = \text{returntime} = \text{tijdinterval tussen } t_e \text{ en de projectie van de returnphase-helling in } t_e$. Hoe groter t_a , hoe trager de returnphase, hoe groter de reductie van de hogere frequenties, hoe negatiever de spectrale helling. (omdat snelle veranderingen gepaard gaan met hogere frequenties (Gobl))
- * Skewness
 $Rk = (t_e - t_p)/(t_p - t_0)$

Beide modellen hebben hun voor- en nadelen. Het Rosenbergmodel is relatief eenvoudig met slechts twee onafhankelijke parameters. Dat maakt het makkelijk passend bij analyse. Bij synthese wordt de vorm van de glottale puls

echter minder nauwkeurig benaderd dan in het LF-model omdat de 'returnphase' niet beschreven wordt. Dit valt in de praktijk te ondervangen door achter de puls een laagdoorlaatfilter te plaatsen. (zie KLatt & KLatt, 1989) Een ander nadeel is dat binnen het Rosenbergmodel de skewness een vaste waarde heeft ($R_k = 0,5$). Deze tekortkoming kan ondervangen worden door de Rosenbergformule uit te breiden met een polynoom.

Voordeel van het LF-model is de nauwkeurigheid en grote variatie in pulsvormen die mogelijk is dankzij de vier verschillende parameters. Dit maakt de implementatie echter minder makkelijk. Het LF-model heeft een randvoorwaarde waaraan altijd moet worden voldaan: de integraal van functie 1 en 2 moet samen nul zijn. Daardoor zijn de vier parameters dus niet geheel onafhankelijk van elkaar te variëren en heeft het dus minder variatiemogelijkheden dan je op grond van vier parameters zou vermoeden. Omdat dit model de afgeleide van de glottale puls voorstelt, beschrijft het impliciet het uitstralingseffect van de geluidsgolf bij de lippen. Voor sommige toepassingen kan dat gezien worden als een ander voordeel van het LF-model.

Zowel het Rosenberg- als het LF-model houden geen rekening met het bestaan van de zogenaamde 'constant flow leakage', een voortdurende kleine luchtstroom door de stemspleet ten gevolge van het niet volledig sluiten van de stembanden bij sommige personen. Deze luchtstroom veroorzaakt een soort ruis die de perceptieve natuurlijkheid van de spraak ten goede komt. Beide modellen gaan ook voorbij aan het feit dat er wel degelijk interactie bestaat tussen het brongedeelte en het filtergedeelte tijdens de spraakproductie.

PARAMETERCORRELATIES, LITERATUUROVERZICHT

Onderzoek naar de relaties tussen bronparameters is relatief jong. De hoeveelheid literatuur zoals we die kennen over prosodie-onderzoek, treffen we dan ook niet aan. De resultaten en conclusies van eerder parameteronderzoek blijken niet altijd eenduidig, soms zelfs in tegenspraak met elkaar te zijn. Onderstaand literatuuroverzicht is een samenvatting van wat tot nu toe bekend is over de relatie tussen de toonhoogte en de open quotiënt, spectrale helling, skewness en amplitude. Andere relaties of andere parameters zijn niet vermeld omdat ze niet relevant zijn voor onze experimenten.

* Relatie f_0 - OQ:

Bij het verhogen van f_0 vinden Cleceland & Sunberg(1983) een toename van OQ, Pierrehumbert (1989) vindt de schijnbaar tegenstrijdige resultaten. Binnen een zin hebben door toonhoogte beklemtoonde vocalen een grotere OQ. Als de hele zin wordt uitgesproken met een relatief luidere stem ziet zij

een gemiddeld kleinere OQ. Het verhogen van f_0 kan dus verschillende consequenties hebben voor de OQ, afhankelijk van de situatie waarin dit gebeurt. Sluijter (1995) kwantificeert de OQ-toename ten gevolge van beklemtoning door verhoogde f_0 op 8%. Fant (1959) en Monsen & Engebretson (1977) vinden juist een afname van OQ. Fant verklaart dit door de verhoogde pressie op de stembanden die optreedt bij f_0 -verhoging. Daardoor gaat de stemspleet minder ver open en zal de OQ en returnphase korter zijn. Monsen & Engebretson meten bij vrouwen een gemiddeld lagere OQ, Sluijter (1995) vindt een groter OQ. Klatt & Klatt (1988) spreken over een OQ-waarde van 0,5 - 0.6 bij een modale stem. Door stempessie wordt de glottale puls smaller, de OQ dus kleiner en de f_0 substantieel lager! Sundberg & Gauffin (1979) en Karlsson (1985) concluderen dat f_0 nauwelijks invloed heeft op de OQ en onder normale omstandigheden nagenoeg constant blijft.

* Relatie f_0 - spectrale helling:

Monsen & Engebretson (1977) vinden een negatievere (lees: grotere) spectrale helling bij vrouwen dan bij mannen. Fant (1959) verklaart waarom de spectrale helling minder steil wordt tijdens stempessie: door stempessie die gepaard gaat met een verhoogde f_0 is de returnphase korter en de amplitude van de hogere frequenties in het spectrum groter, daardoor wordt de spectrale helling minder steil. De resultaten van Pierrehumbert (1989) zijn hiermee in overeenstemming, evenals die van Klatt & Klatt (1988). Sluijter (1995) vindt een minder steile spectrale helling bij beklemtoonde vocalen. (en een grotere 'constant flow leakage,') Karlsson (1985) leidt theoretisch af dat een afname van t_a correleert met een toename van R_k .

* Relatie f_0 - skewness:

Pierrehumbert (1989) meet een kleinere R_k bij toename van toonhoogte, d.w.z. de glottale puls helt over naar rechts en de skewness neemt dus toe bij hogere f_0 . Ishizaka-Flanagan (1972) & Titze-Talkin (1979) vinden het tegenovergestelde: een toename van f_0 leidt juist tot een meer symmetrische pulsvorm.

* Relatie f_0 - amplitude:

Alleen Sluijter (1995) maakt melding van een AV-toename bij door toonhoogte beklemtoonde klinkers.

EXPERIMENT 1

ONDERZOEKSVRAAG EN HYPOTHESE

Alvorens variabele bronparameters kunnen worden toegepast in kunstmatige spraak ter verbetering van de natuurlijkheid, zullen de onderlinge relaties eenduidig gedefinieerd moeten zijn.

Onze eerste onderzoeksvraag luidt daarom:

* Bestaat er een relatie tussen relatieve toonhoogte enerzijds en open quotient, skewness, spectrale helling en pulsamplitude anderzijds?

Zoja, in welke richting gaat deze relatie?

Deze vraag zal onderzocht worden in experiment 1. We verwachten een lichte stijging van OQ t.g.v. een f_0 -toename omdat hiervan het meest melding wordt gemaakt in de literatuur. Tevens is de theoretische verklaring hiervoor (de verhoogde stempessie) zeer aannemelijk. Hiermee hangt de verwachting samen dat de spectrale helling afneemt (minder negatief). Ten aanzien van de amplitude verwachten we een toename bij stijging van de relatieve toonhoogte. Dit is niet zo zeer gebaseerd op literatuurgegevens maar eerder op de eigen perceptie: het lijkt alsof iemand die hoger gaat spreken ook harder gaat praten en dus meer volume produceert. Voor wat betreft de skewness hebben we geen verwachtingen vnl. omdat de literatuur hierover zo summier is. Naast deze hoofdvraag onderzoeken we tevens of bovengenoemde relaties nog afhankelijk zijn van het geslacht van de spreker en de vokaalsoort.

PROCEDURE

Een proefpersoon wordt verzocht vier maal dezelfde vocaal (klinker) uit te spreken op verschillende toonhoogtes. In feite krijgt de proefpersoon de opdracht een octaaf te maken in vier stappen: do - mi - sol - do. (met korte tussenpauzes) De toonhoogte van de eerste 'do' wordt door de proefpersoon vrij gekozen, maar wel zodanig dat het volledige oktaaf binnen het stembereik valt. Er worden vier verschillende vocalen uitgesproken die de centrale en uiterste plaatsen bezetten van de akoestische klinkerdriehoek: /a/ (baas), /i/ (biet), /u/ (boek) en /e/ (lopen: 'sjwa'). ('t Hart 1988) Van elke vocaal worden twee series digitaal opgenomen. De opnamen geschieden met een Bruel & Kjer microfoon (type 4003) en een Philips integrated digital amplifier (DFA 888) in een geluidsdichte ruimte. Als proefpersonen dienen vier mannen en vier vrouwen van volwassen leeftijd.

ANALYSE

Gipos is een software-pakket voor spraakanalyse ontwikkeld door het IPO te Eindhoven. Met behulp van dit pakket wordt uit elke uiting een stukje van 300 ms. geknipt met een zo constant mogelijke amplitude. Daarna volgt de analyse door hetzelfde programma. Het glottale-puls-algoritme binnen Gipos berekent door middel van Fourieranalyse het frequentiespectrum van de uiting en bepaalt hieruit de vorm van de glottale puls. Bij dit 'orgineel' wordt in een 'codebook' de bestpassende golf functie opgezocht (fitten). Uit de wiskundige representatie van deze golf functie worden de verschillende parameters berekend. Het algoritme binnen Gipos is gebaseerd op het Rosenbergmodel en voorzien van enkele uitbreidingen. Er blijkt nog geen tool voorhanden te zijn binnen dit gipos-programma om de skewness uit het wiskundig model te bepalen. Daarom wordt besloten alleen OQ, t_a en AV te onderzoeken.

RESULTATEN

Alle data die voortkomen uit het analyse-onderzoek zijn te vinden in een aparte bijlage aan het eind van dit rapport. Ten aanzien van de open quotient blijken significante verschillen op te treden tussen mannen en vrouwen, alsmede tussen de verschillende vokaalsoorten. Dit is een goede reden om de resultaten van deze variabelen apart te presenteren.

Tijdens het onderzoek blijken de vocalen: /i/, /u/ en /e/ minder goed te analyseren te zijn met het Gipos-softwarepakket. Deze klinkers bezitten namelijk een lage eerste formant die een zuivere schatting van de glottale puls via inverse filtering verstoort. De vocaal /a/ heeft geen last van dit verschijnsel omdat de eerste formant relatief hoog is. De data van de vocaal /a/ zijn dus betrouwbaarder en daarom zullen we ons bij het behandelen van de resultaten vooral concentreren op deze vocaal. Zowel bij mannen als bij vrouwen vinden we een aantoonbare correlatie tussen OQ en relatieve toonhoogte bij de vocaal /a/.

Tabel 1 : Relatie OQ - sexe.

sexe	mean OQ	std dev
man	0,661	0,154
vrouw	0,724	0,117

Tabel 2 : Relatie OQ - vokaalsoort

vokaal- soort	mean OQ	std dev
a	0,642	0,126
e	0,688	0,129
i	0,746	0,132
u	0,688	0,155

Tabel 3 : Relatie OQ - rel. toonhoogte. (voor a-vokaal bij mannen)

toon hoogte	mean OQ	std dev
1	0,528	0,071
2	0,522	0,108
3	0,564	0,117
4	0,703	0,115

Tabel 4 : Relatie OQ - rel. toonhoogte. (voor a-vokaal bij vrouwen)

toon hoogte	mean OQ	std dev
1	0,637	0,092
2	0,660	0,124
3	0,679	0,158
4	0,747	0,054

Voor een volledig overzicht van alle resultaten, zie bijlage 1.

CONCLUSIE EN DISCUSSIE

* OQ - geslacht.

De OQ-waarden van mannen verschillen significant van die van vrouwen. $F_{(1,230)} = 6,70$ ($p < 0,05$), Bij mannen is deze gemiddeld 0,66 en bij vrouwen hoger nl. 0,72.

* OQ - vokaal.

Zowel bij mannen als bij vrouwen bestaat een significant verschil tussen de verschillende vokalen: $F_{(3,230)} = 5,45$ ($p < 0,05$). Post-hoc analyse geeft aan dat de a en de i significant van elkaar verschillen, andere combinaties niet.

* OQ - toonhoogte.

Bij de a-vokaal bestaat een correlatie tussen relatieve toonhoogte en OQ. [bij mannen: $r = 0,556$ ($p < 0,01$) bij vrouwen: $r = 0,367$ ($p < 0,05$)] De OQ-waarde neemt significant toe bij de hoogste relatieve toonhoogte. Dit stemt overeen met eerdere bevindingen (zie literatuur overzicht) Bij de andere vokalen (e, i en u) vinden we deze correlatie niet. De reden hiervan ligt wellicht in het feit dat deze vokalen zich minder goed lenen voor analyse volgens de gipos-methode. Bij deze vokalen valt de eerste formant ongeveer samen met de eerste of tweede harmonische toon. Gipos heeft dan moeite met het onderscheid tussen formant en boventoon: de analyse wordt minder betrouwbaar en de gegevens vertonen een grotere standaarddeviatie.

* Spectrale helling - toonhoogte.

De spectrale helling vertoont geen significante verschillen bij veranderende relatieve toonhoogten, $F_{(3,233)} = 2,19$ ($p > 0,05$).

Er bestaat echter wel een correlatie tussen de relatieve toonhoogte en de spectrale helling voor de vokaal a bij mannen.

* AV - toonhoogte

De AV vertoont geen significante verschillen bij een veranderende relatieve toonhoogte, $F_{(1,230)} = 1,40$ ($p > 0,05$).

Bovenstaande resultaten zijn gebaseerd op vocalen waarbij de fit kleiner is dan 0,1. De gemiddelde fit per vocaalsoort is ongeveer gelijk (0,04-0,05).

Opvallend in de oorspronkelijke data is de grote variatie in parameterwaarden van één persoon bij dezelfde vocaal. Mogelijk wordt dit veroorzaakt doordat de verschillende vocalen zijn uitgesproken als losse, opzichzelfstaande uitingen. De toonhoogte is immers trapsgewijs gevarieerd met na elke nieuwe uiting een korte pauze. Dit kan voor elke volgende vocaal leiden tot een nieuwe stand en spanning van de stembanden en een andere subglottale druk. Het is aanbevelenswaardig om de parameters opnieuw te bepalen maar nu als de toonhoogte van de vocalen vloeiend gevarieerd wordt.

CONCLUSIE

De OQ-waarden van de deelnemende mannen is gemiddeld lager dan die van vrouwen die deelnamen aan dit experiment.

De OQ vertoont een stijging bij uitingen van de a-klinker op een hogere relatieve toonhoogte.

Voor spectrale helling en amplitude vinden we geen duidelijke correlaties met F_0 .

EXPERIMENT 2

In het vorige experiment hebben we gezien dat er een correlatie bestaat tussen de relatieve toonhoogte en de OQ bij uitingen van de klinker /a/. In het tweede experiment wordt onderzocht of er perceptieve verschillen optreden t.a.v. de relatieve toonhoogte als gevolg van een veranderende OQ, AV en/of spectrale helling(tilt).

ONDERZOEKSVRAAG EBN HYPOTHESE

De tweede onderzoeksvraag in dit rapport luidt:

* Bestaat er een relatie tussen toonhoogteperceptie enerzijds en open quotient, skewness, spectrale helling en pulsamplitude anderzijds?

* Zoja, in welke richting gaat deze relatie?

Het is belangrijk om deze vragen beantwoord te krijgen omdat het geen zin heeft om de resultaten uit experiment 1 toe te passen in kunstmatige spraak als deze toch niet voor het menselijk oor waarneembaar zijn. Anderzijds is het interessant om te weten of parametervariatië perceptieve verschillen opleveren ongeacht of deze variatië in werkelijkheid in de gewone spraak voorkomen. Mogelijkerwijs beïnvloeden deze toch de natuurlijkheid van de kunstmatige spraak en kunnen dan direct toegepast worden. We verwachten dat de toonhoogte perceptief hoger beoordeeld wordt naarmate het volume van de spraak toeneemt. (om hierbovengenoemde reden) Ten aanzien van de andere bronparameters hebben we geen specifieke verwachtingen.

METHODE

Met behulp van een spraaksynthesizer worden 27 stimuli aangemaakt met steeds een andere combinatie van OQ-, AV- en tilt waarden. We gebruiken hiervoor de Klatt Cascade-Parallel Formant Synthesizer v 3.0 van Jon Iles en Nick Ing-Simons. Dit is een latere versie van de Klatt Cascade-Parallel Formant Speech Synthesizer van Dennis Klatt, nl. de KLGLOTT 88. Deze werkt standaard volgens het Rosenbergmodel. Met deze synthesizer wordt een /a/-vocaal gesynthetiseerd met de volgende toonhoogtes en formanten: $f_0=109$ Hz, $f_1=860$ Hz, $f_2=1315$ Hz, $f_3=501$ Hz, $f_4=3295$ Hz, $f_5=3751$ Hz. De tijdsduur van de vocaal is 600 ms. In deze vocaal is de f_0 nagenoeg constant (109 Hz) en worden de OQ-, AV- en tiltwaarden steeds gevarieerd. Voor elke bronparameter nemen we drie verschillende instelwaarden.

Criteria voor de instelwaarden zijn:

- verschillende waarden leveren waarneembare verschillen op.

- beperkingen van de synthesizer (deze kan bijvoorbeeld geen OQ > 60 % genereren)
 - stimuli moeten boven gehoordrempel blijven.
- Zo komen we tot de volgende instelwaarden:

Instelwaarden:				
	a	b	c	
open quotient (OQ)	: 10	35	60	%
spectrale helling	: 0	-15	-30	dB/oktaaf
pulsamplitude (AV)	: 48	54	60	dB

De mogelijke combinaties van deze instelwaarden leiden tot de volgende 27 stimuli:

code	OQ	spectrale helling	AV
aaa	10	0	48
aab	10	0	54
aac	10	0	60
aba	10	-15	48
abb	10	-15	54
abc	10	-15	60
aca	10	-30	48
acb	10	-30	54
acc	10	-30	60
baa	35	0	48
bab	35	0	54
bac	35	0	60
bba	35	-15	48
bbb	35	-15	54
bbc	35	-15	60
bca	35	-30	48
bcb	35	-30	54
bcc	35	-30	60
caa	60	0	48
cab	60	0	54
cac	60	0	60
cba	60	-15	48
cbb	60	-15	54
cbc	60	-15	60
cca	60	-30	48
ccb	60	-30	54
ccc	60	-30	60

Van deze stimuli wordt een serie gemaakt waarin elke stimulus tweemaal voorkomt in random volgorde. Deze serie wordt ter beoordeling aangeboden aan 16 proefpersonen (er zijn twee versies van de stimulusserie die elkaars inverse zijn). De

proefpersonen zijn volwassen mannen en vrouwen. Ze beoordelen de stimuli op relatieve f_0 . en maken hiervoor gebruik van een puntenschaal van -3 tot en met +3. Hoe lager het cijfer, hoe lager in het stembereik de stimulus wordt gelokaliseerd, hoe hoger het cijfer, hoe hoger in het stembereik de stimulus wordt gelokaliseerd terwijl 0 het midden van het stembereik representeert:

```

      0   0   0   0   0   0   0
     -3  -2  -1   0  +1  +2  +3
    lager in <----- <-----> hoger in
    stembereik                    stembereik

```

RESULTATEN

Elke stimulus is in totaal 32 maal beoordeeld met een cijfer dat de perceptieve relatieve toonhoogte voorstelt. Van al deze beoordelingen is per stimulus het gemiddelde bepaald. Deze gemiddelde waarden staan afgebeeld in onderstaande tabel.

Tabel 5 : Gemiddelde perceptieve beoordeling per stimulus.

tilt	A			B			C		
AV:	A	B	C	A	B	C	A	B	C
OQ:A	1,1	1,3	1,6	-0,1	0,9	1,1	-1,1	-0.5	-0.8
B	0,2	-0,1	0,1	-0,3	-0,2	-0,2	-1,2	-1,1	-0,9
C	-0,2	0,1	-0,1	-0,7	-0,6	-0,4	-1,2	-1,6	-0,9

Daarna is de gemiddelde perceptieve beoordeling berekend uit alle stimuli met één constante parameterwaarde. (Bijvoorbeeld het gemiddelde van alle stimuli met een OQ van 10 %) Deze gemiddeldes zijn per parameter weergegeven in tabel 6.

Tabel 6 : Gemiddelde perceptieve beoordeling per bronparameter.

open quotient (%)		
10	35	60
0,4	-0,4	-0,5

spectrale helling (dB/oct)		
0	-15	-30
0,4	-0,1	-1,0

pulsamplitude (dB)		
48	54	60
-0,4	-0,2	-0,1

DISCUSSIE

Uit correlaties tussen de eerste en de tweede beoordeling van één proefpersoon blijkt dat de proefpersonen consistent zijn in hun beoordeling van de stimuli op $p=0,05$ -niveau. De meeste zijn dit ook voor $p=0,01$ -niveau.

Een ANOVA toont aan dat er een overall significant effect is van OQ op de perceptie van de relatieve toonhoogte, $F_{(2,405)} = 20,53$ ($p < 0,01$). Post-hoc Scheffe-analyses geven aan dat de perceptie van stimuli met de laagste OQ-waarde (10%) significant verschilt van stimuli met de twee hogere OQ-waarden (35% en 60%). De stimuli met de lage OQ-waarde worden als relatief hogere toonhoogtes waargenomen. Tussen de twee OQ-waarden 35% en 60% vinden we geen significante verschillen. ($p > 0,05$).

De verschillende tiltwaarden zorgen voor significant verschillende percepties van de relatieve toonhoogte, $F_{(2,405)} = 38,32$ ($p < 0,01$). Hoe negatiever de spectrale helling, hoe lager de relatieve toonhoogte wordt beoordeeld.

De amplitude heeft geen effect op de relatieve toonhoogte-perceptie. ($p > 0,05$) Dat laatste lijkt vreemd omdat veel mensen de indruk hebben dat wanneer iemand hoger spreekt (of een lettergreep beklemtoont) ook het volume toeneemt. Wanneer iemand harder spreekt lijkt dat dus niet hoger.

De instelwaarde 10 % voor OQ is een erg lage waarde die normaliter weinig voorkomt in natuurlijke spraak. In die zin vormt het een extreem effect in de stimulus. Voor de instelwaarden van de spectrale helling: 0 en -30 dB/oktaaf geldt min of meer hetzelfde. Waar deze parameterwaarden voor OQ en spectrale helling in stimuli gecombineerd worden, ontstaan zeer onnatuurlijk klinkende stimuli, bijvoorbeeld zeer hees of krakerig. Het is niet vreemd dat proefpersonen deze stimuli als uitersten beoordelen. We hebben deze stimuli toch aangeboden aan de proefpersonen, juist omdat we willen weten of er überhaupt effecten zijn.

CONCLUSIE

Een kunstmatig gesynthetiseerde a-klank met een OQ-waarde van 10 % wordt als relatief hoger in het stembereik beoordeeld dan een a-klank met een OQ-waarde van 35 of 60 %. Dit is in tegenspraak met nogal wat bevindingen uit eerder onderzoek.

Een kunstmatig gesynthetiseerde a-klank met een spectrale helling van -30 dB/oct wordt als relatief lager in het stembereik beoordeeld dan een a-klank met een spectrale helling van -15 dB/oct en deze weer als lager dan een a-klank met een spectrale helling van 0 dB/oct.

De amplitude van een kunstmatig gesynthetiseerde a-klank is niet van invloed op de perceptie t.a.v. relatieve toonhoogte.

ALGEMENE CONCLUSIE

Uit het analyse-experiment blijkt dat de OQ toeneemt als de toonhoogte waarop wij spreken toeneemt. We zien dit effect optreden bij de a-vocaal en in de hogere regionen van het stembereik. Het resultaat is bemoedigend genoeg om aan vervolgonderzoek te doen. Mogelijkerwijs zal in de nabije toekomst de analyse van andere vocalen dan de /a/ betrouwbaar kunnen geschieden. Dan is het zeker de moeite waard om te onderzoeken of bovengenoemd effect ook optreedt bij andere vocalen. Het is tevens aanbevelenswaardig om in een vervolgonderzoek de verhoging van de spreektoon vloeiend te laten verlopen inplaats van trapsgewijs zoals in ons onderzoek. Dit is een betere benadering van de natuurlijke spraak en geeft wellicht een beter inzicht in de correlatie tussen toonhoogte en bronparameters.

Uit het synthese-experiment blijken twee dingen: een a-klank met een extreem lage OQ-waarde wordt als relatief hoger beoordeeld en de toonhoogte van een a-klank wordt relatief steeds lager beoordeeld naarmate de spectrale helling negatiever wordt. Zeker de correlatie tussen spectrale helling en toonhoogteperceptie dient nader onderzocht te worden. Misschien is het mogelijk om door tiltvariaties de natuurlijkheid van

kunstmatige spraak te vergroten. Om die reden is het ook zinvol opnieuw de relatie OQ - toonhoogteperceptie te bestuderen maar dan met stimuli die minder extreme OQ-waarde hebben en met OQ-waarden die de natuurlijke waarden benaderen.

LITERATUURLIJST

- Fant, G., Liljencrants, J. and Lin, Q. (1985): A four-parameter model of the glottal flow.
STL-QPSR 4
- Gobl, C. (1988): Voice source dynamics in connected speech.
STL-QPSR 1, pp 123-159.
- 't Hart, J. (1988): Spraakgeluid.
Ter sprake, spraak als betekenisvol geluid., pp 40-72.
- Iles, J. en Ing-Simons, N. (datum onbekend)
Manuel Klatt Cascade-Parallel Formant Synthesizer v 3.0.
- Klatt, D.H. and Klatt, L.C.(1990): Analysis, synthesis and perception of voice quality variations among female and male speakers.
JASA 87 (2), pp 820-856.
- Pierrehumbert, J. (1989): A preliminary study of the consequences of the intonation for the voice source.
STL-QPSR 4, pp 23-36.
- Sluijter, A. (1995): Phonetic correlates of stress and accent.
Unpublished doctoral dissertation, Leiden University.

BEDANKT,

- * Marc Swerts,
voor je buitengewoon goede begeleiding. De deur van je kamer stond letterlijk en figuurlijk altijd open voor mijn vragen. Je hebt me op veel goede ideeën gebracht en soms ook heel wat tijdrovende karweien van me overgenomen. Door je voortdurende interesse en kundige supervisie kreeg ik het gevoel dat het niet mijn onderzoek was maar ons onderzoek. Dat stimuleerde enorm. Ook op het persoonlijke vlak klikte het goed. Daardoor hebben we het gelukkig over meer gehad dan glottale pulsen en waf-files alleen. Kortom: ik had me geen betere stagebegeleider kunnen wensen.
- * Raymond Veldhuis,
voor al je antwoorden op mijn fysische vragen. Je bereidheid was groot en je uitleg helder. Veel succes met je verdere ontwikkeling van de analyse-programmatuur.
- * Angeliën Sanderman,
voor je hulp bij de verwerking van alle resultaten. Je stond klaar met raad en daad betreffende data-analyse en statistiek.
- * kamer- en ganggenoten: Paul, Jeroen, Daniël, Emiel,
voor de gezelligheid en de melige of soms serieuze conversaties.
- * het IPO, in het bijzonder Jacques Terken,
voor het mogelijk maken van mijn afstudeeronderzoek. Waar had ik beter terecht gekund met mijn natuurkunde- en biologiebevoegdheid dan hier? Ik was aangenaam verrast door het gemak waarmee op het IPO een stageplaats georganiseerd werd. In het begin was het wel even wennen aan zo veel specialisten op zo weinig oppervlakte. Ik vond het maar wereldvreemde ambiance; al die mensen die acht uur per dag met vierkante ogen naar een computerscherm zitten te turen. Maar gaandeweg veranderde dat in ontzag en werd ik steeds meer gedreven door de professionele sfeer.
- * Piet Slooten van de FEO,
voor het blindelings vertrouwen in mijn afstudeeronderzoek.

Harrie Prinsen

17 oktober 1996

Bijlage 1

1 Relatieve toonhoogte - OQ

1A Relatieve toonhoogte - OQ

Allen

	OQ	OQ	OQ
	MEAN	STD	N
TOON			
1	0.69	0.13	61.00
2	0.69	0.14	60.00
3	0.67	0.17	58.00
4	0.73	0.11	52.00

1B Relatieve toonhoogte - OQ.

Man

	OQ	OQ	OQ
	MEAN	STD	N
TOON			
1	0.64	0.12	29.00
2	0.66	0.16	29.00
3	0.64	0.21	30.00
4	0.70	0.10	30.00

1C Relatieve toonhoogte - OQ.

Vrouw

	OQ	OQ	OQ
	MEAN	STD	N
TOON			
1	0.73	0.12	32.00
2	0.71	0.12	31.00
3	0.70	0.12	28.00
4	0.76	0.10	22.00

1D Relatieve toonhoogte - OQ.

klinker a

	OQ	OQ	OQ
	MEAN	STD	N
TOON			
1	0.60	0.10	13.00
2	0.61	0.13	13.00
3	0.62	0.15	12.00
4	0.72	0.09	16.00

1E Relatieve toonhoogte - OQ.

klinker e

	OQ	OQ	OQ
	MEAN	STD	N
TOON			
1	0.73	0.12	16.00
2	0.65	0.08	15.00
3	0.65	0.18	15.00
4	0.72	0.11	14.00

1F Relatieve toonhoogte - OQ.

klinker i

	OQ	OQ	OQ
	MEAN	STD	N
TOON			
1	0.73	0.13	16.00
2	0.76	0.12	16.00
3	0.77	0.15	15.00
4	0.71	0.13	10.00

1G Relatieve toonhoogte - OQ.

klinker u

	OQ	OQ	OQ
	MEAN	STD	N
TOON			
1	0.67	0.13	16.00
2	0.73	0.18	16.00
3	0.63	0.17	16.00
4	0.74	0.12	12.00

2A Relatieve toonhoogte - tilt.

2A Relatieve toonhoogte - tilt. Allen

	TILT	TILT	TILT
	MEAN	STD	N
TOON			
1	-8.82	1.60	61.00
2	-9.31	1.54	60.00
3	-9.34	1.49	58.00
4	-8.88	1.57	52.00

2B Relatieve toonhoogte - tilt. Man

	TILT	TILT	TILT
	MEAN	STD	N
TOON			
1	-8.55	1.74	29.00
2	-8.96	1.49	29.00
3	-9.18	1.55	30.00
4	-8.97	1.62	30.00

2C Relatieve toonhoogte - tilt. Vrouw

	TILT	TILT	TILT
	MEAN	STD	N
TOON			
1	-9.07	1.43	32.00
2	-9.63	1.54	31.00
3	-9.50	1.43	28.00
4	-8.75	1.52	22.00

2D Relatieve toonhoogte - tilt.

klinker a

	TILT	TILT	TILT
	MEAN	STD	N
TOON			
1	-8.30	1.60	13.00
2	-8.89	1.58	13.00
3	-8.46	1.36	12.00
4	-7.56	1.35	16.00

2E Relatieve toonhoogte - tilt.

klinker e

	TILT	TILT	TILT
	MEAN	STD	N
TOON			
1	-8.74	2.03	16.00
2	-8.93	1.77	15.00
3	-9.28	1.56	15.00
4	-9.10	1.29	14.00

2F Relatieve toonhoogte - tilt.

klinker i

	TILT	TILT	TILT
	MEAN	STD	N
TOON			
1	-8.90	1.66	16.00
2	-9.26	1.60	16.00
3	-9.21	1.48	15.00
4	-9.54	1.21	10.00

2G Relatieve toonhoogte - tilt.

klinker u

	TILT	TILT	TILT
	MEAN	STD	N
TOON			
1	-9.27	0.89	16.00
2	-10.05	0.99	16.00
3	-10.17	1.18	16.00
4	-9.81	1.33	12.00

3 Relatieve toonhoogte - AV.

3A Relatieve toonhoogte - AV.

Allen

	AV	AV	AV
	MEAN	STD	N
TOON			
1	19532.77	8334.94	61.00
2	24070.13	10160.02	60.00
3	23313.81	9435.97	58.00
4	23345.82	12567.18	52.00

3B Relatieve toonhoogte - AV.

Man

	AV	AV	AV
	MEAN	STD	N
TOON			
1	19096.45	5744.75	29.00
2	21956.48	6144.68	29.00
3	24972.77	9006.15	30.00
4	24487.57	12351.07	30.00

3C Relatieve toonhoogte - AV.

Vrouw

	AV	AV	AV
	MEAN	STD	N
TOON			
1	19928.18	10213.35	32.00
2	26047.40	12625.87	31.00
3	21536.36	9721.88	28.00
4	21788.89	12980.34	22.00

3D Relatieve toonhoogte - AV.

klinker a

	AV	AV	AV
	MEAN	STD	N
TOON			
1	23503.31	10762.21	13.00
2	23722.80	8889.96	13.00
3	24147.50	6113.15	12.00
4	18865.63	8403.84	16.00

3E Relatieve toonhoogte - AV.

klinker e

	AV	AV	AV
	MEAN	STD	N
TOON			
1	22021.27	8600.65	16.00
2	27516.61	12863.75	15.00
3	24320.33	11539.53	15.00
4	23160.88	14524.59	14.00

3F Relatieve toonhoogte - AV.

klinker i

	AV	AV	AV
	MEAN	STD	N
TOON			
1	16272.19	6399.56	16.00
2	22683.00	11112.93	16.00
3	21224.25	7953.48	15.00
4	22217.92	11897.66	10.00

3G Relatieve toonhoogte - AV.

klinker u

	AV	AV	AV
	MEAN	STD	N
TOON			
1	17078.78	5680.38	16.00
2	22508.38	6943.97	16.00
3	23703.91	11023.43	16.00
4	30475.08	13599.72	12.00