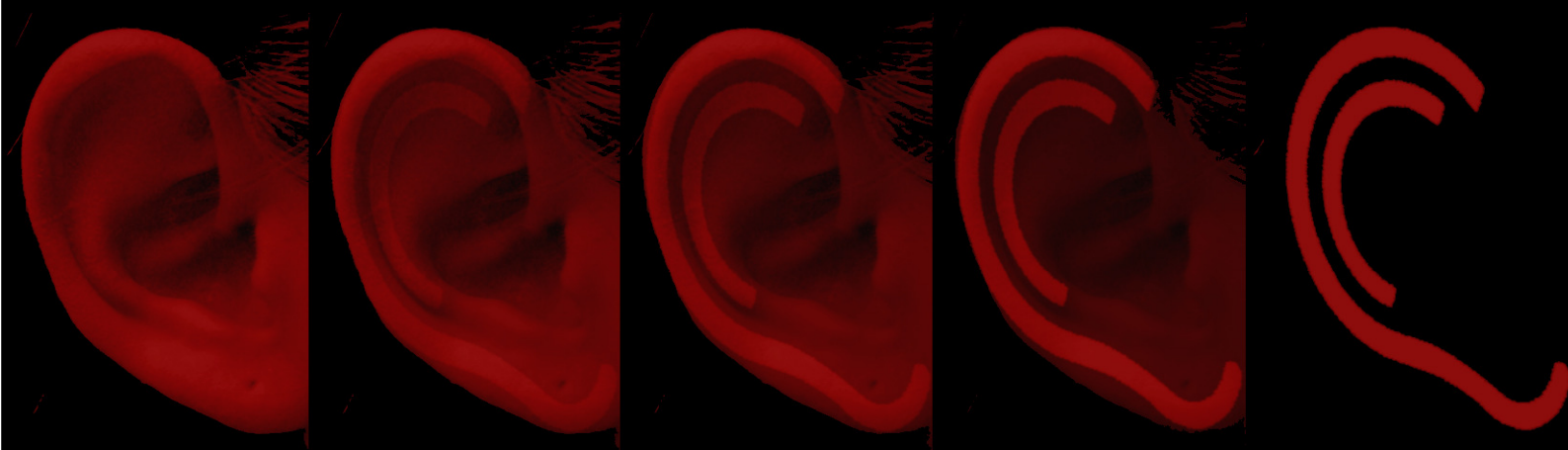


Oorcheck en lawaaislechthorendheid

M.C.J. Leensen en W.A. Dreschler



De resultaten van dit project zijn mede gebaseerd op metingen in het
Leids Universitair Medisch Centrum (dr.ir. J.A.P.M. de Laat) en het
Universitair Medisch Centrum Nijmegen (prof.dr. A.F.M. Snik).

Niets uit dit rapport mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of wijze, hetzij elektronisch, mechanisch, door fotokopieën, opname of enige andere manier, zonder voorafgaande schriftelijke toestemming van de afdeling Klinische & Experimentele Audiologie van het AMC te Amsterdam.

Eindrapport project

Oorcheck en lawaaislechthorendheid



document nr AMC-CEA-117

datum	23 december 2009
versie	V4.0
status	Eindrapportage
auteur(s)	Drs. M.C.J. Leensen en prof.dr.ir. W.A. Dreschler
trefwoorden	Lawaaislechthorendheid, screening, internettest, spraak-in-lawaai
opmerkingen	Een studie naar het gebruik van spraak-in-ruistesten voor internetscreening op lawaaislechthorendheid (NIHL); test (on)gevoeligheid en mogelijkheden ter verbetering.

INHOUDSOPGAVE

Samenvatting.....	7
1 Inleiding.....	11
1.1 Doel van het project.....	12
2 Materiaal en methoden	13
2.1 Toonaudiogram	13
2.2 Gouden standaard voor spraakverstaan.....	13
2.3 Populatie.....	14
2.4 Meetopstelling.....	16
2.5 Kalibratie.....	17
3 Evaluatie van bestaande internettesten.....	19
3.1 Experiment 1: vergelijking van testen	19
Nationale hoortest (NHT).....	19
Oorcheck.....	20
Bedrijfsoorcheck (BOC).....	20
3.2 Meetprotocol.....	21
3.3 Resultaten van experiment 1	21
Gevoeligheid van de testen	21
Oorcheck in thuissituatie versus experimentele testomgeving.....	23
Verdere evaluatie van bestaande testen	24
3.4 Conclusies t.a.v. de evaluatie van de bestaande testen	26
4 Effecten van spraakmodificatie op de test gevoeligheid	27
4.1 Experiment 2A: optimalisatie van spraakmateriaal.....	27
4.2 Resultaten van experiment 2A.....	29
4.3 Normering van de nieuwe oorcheck.....	33
4.4 Conclusies t.a.v. optimalisatie spraakmateriaal.....	34

5	Effecten van ruismodificaties op de test gevoeligheid	35
5.1	Experiment 2B: optimalisatie van het ruismateriaal	35
	Samenstellen van het ruismateriaal	36
	SII voorspellingen.....	36
	Filtering van het spraaksignaal	37
	Pilotmetingen.....	37
5.2	Meetprotocol.....	39
5.3	Resultaten van experiment 2B.....	40
5.4	Conclusie t.a.v. de ruismodificaties.....	44
5.5	Normering van de experimentele Oorchecks.....	45
6	Discussie	47
	Bestaande testen.....	47
	Verbeteringen van de bestaande Oorcheck.....	47
	Experimentele ruisconditie.....	48
	De invloed van het presentatieniveau	49
7	Conclusies.....	51
8	Dankwoord.....	52
9	Referenties	53
	Bijlagen	54
	Bijlage I: Geluidsniveau in meetruimte LUMC.....	55
	Bijlage II: Berekenen van woordspecifieke discriminatiecurven.....	57
	Bijlage III: SII voorspellingen	61
	Bijlage IV: Resultaten pilotmetingen.....	67

SAMENVATTING

INLEIDING

De Nationale Hoorstichting heeft diverse online hoortesten ontwikkeld om de gehoorfunctie te testen en daarmee bewustwording van gehoorverlies onder de bevolking te vergroten. Dit betreft spraak-in-ruistesten, waarbij woorden in continue achtergrondruis aangeboden worden en de testuitslag gebaseerd is op de spraakverstaanvaardigheidsdrempel (SRT).

De testen zijn ontwikkeld voor verschillende doelgroepen; de Nationale hoortest (NHT) voor (oudere) volwassenen, de Bedrijfsoorcheck (BOC) voor werknemers die werkzaam zijn in lawaai en de Oorcheck (OC) voor jongeren in de leeftijd 12 tot en met 24 jaar. Deze Oorcheck is bedoeld om de gevolgen van blootstelling aan recreatief lawaai, zoals luide muziek, bij deze jongeren onder de aandacht te brengen en richt zich daarom voornamelijk op het ontdekken van lawaaislechthorendheid.

Het is echter de vraag of een spraak-in-ruistest met continue breedbandige ruis voldoende gevoelig is om lawaaislechthorendheid, in een vroeg stadium, te ontdekken.

DOEL

Het doel van dit onderzoek is dan ook om bij Oorcheck en de andere online spraak-in-ruistesten de gevoeligheid voor lawaaislechthorendheid te onderzoeken.

Omdat we verwachten dat deze gevoeligheid beperkt is proberen we in deze studie tevens manieren te vinden om deze gevoeligheid te verbeteren. Dit doen we enerzijds door het optimaliseren van het spraakmateriaal, en anderzijds door het modificeren van de maskeerruis. Uiteindelijk willen we Oorcheck zo aanpassen dat er een screeningstest ontstaat die gevoelig genoeg is voor het ontdekken van lawaaislechthorendheid.

METHODEN

Hiertoe wordt multi-center studie uitgevoerd, in het LUMC Leiden, UMCN Nijmegen en AMC Amsterdam. In totaal hebben 98 personen aan het onderzoek deelgenomen, verdeeld over drie groepen; 49 normaalhorenden, 25 slechthorenden met een smalle lawaaidip in het audiogram en 24 slechthorenden met een brede lawaaidip. Bij deze personen wordt eerst een toonaudiogram afgenomen en vervolgens een aantal spraak-in-ruistesten met Plompzinnen (gouden standaard). Daarna worden de drie bestaande internettesten en zeven verschillende aangepaste varianten van Oorcheck uitgevoerd. Al deze testen worden uitgevoerd in een stille ruimte en op een vast afspeelniveau van 20 dB boven de SRT in stilte met een minimum van 65 dB(A). De testen werden twee keer, in gebalanceerde volgorde, afgenomen aan één oor.

RESULTATEN

GEVOELIGHEID VAN DE BESTAANDE INTERNETTESTEN

De evaluatie van de verschillende online screeningstesten laat zien dat de testbetrouwbaarheid van de drie spraak-in-ruistesten goed is en onderling vergelijkbaar. Echter, Oorcheck vindt geen significant verschil tussen de normaalhorende groep en de slechthorenden met een smalle dip.

Bovendien is de gevoeligheid van de huidige testen voor lawaaislechthorendheid relatief laag; de NHT en de OC classificeren ongeveer de helft van de totale groep slechthorenden als normaalhorend. De BOC doet het wat dat betreft beter, maar deze test geeft aan dat alle normaalhorenden een 'minder goed' gehoor hebben.

Wanneer we de resultaten van Oorcheck in de thuissituatie bekijken is de sensitiviteit van de test hoger, maar worden er nog steeds drie op de tien personen verkeerd geclassificeerd. De gemiddelde SRTs zijn in de thuissituatie hoger dan wanneer de test onder gecontroleerde experimentele omstandigheden afgenomen is. Dit verschil is het grootst, gemiddeld 2.0 dB, wanneer men de test thuis uitvoert via luidsprekers.

Geen van de testen blijkt gevoelig genoeg te zijn voor het ontdekken van lawaaislechthorendheid. Om te trachten de testgevoeligheid voor lawaaislechthorendheid te verbeteren, wordt allereerst het gebruikte spraakmateriaal geoptimaliseerd.

OORCHECK MET GEOPTIMALISEERD SPRAAKMATERIAAL

Het spraakmateriaal van Oorcheck (OC) bestaat uit negen verschillende woorden. Sommige van deze woorden zijn makkelijker te verstaan dan andere. Op basis van resultaten van ongeveer 100.000 testen die in 2004 t/m 2007 via internet zijn afgenomen, zijn woordspecifieke discriminatiecurven berekend. De gemiddelde verschuiving van deze curven wordt gebruikt om de intensiteit van ieder woord aan te passen, zodat het materiaal homogener wordt. Hierna wordt ook het spectrum van de ruis aangepast, zodat dit precies overeenkomt met het gemiddelde spectrum van de spraak.

Deze nieuwe Oorcheck wordt bij dezelfde onderzoekspopulatie afgenomen en resulteert in significante verschillen in SRT tussen de drie onderzoeksgroepen, welke bovendien groter zijn dan voor de originele versie. Ook correleert deze nieuwe test beter dan de originele Oorcheck met zowel de gouden standaard (SRT Plomp) als met het toonaudiogram.

Het gevonden leereffect is echter groter voor de nieuwe OC (1.6 dB) dan voor de originele OC (0.8 dB). De verwachte verbetering in test-retest betrouwbaarheid door het homogeniseren van de spraak wordt niet duidelijk aangetoond. De test-retest correlatie is zoals verwacht wel hoger voor de nieuwe Oorcheck, maar de sd_{intra} is niet beter dan die van de originele versie.

De vernieuwde Oorcheck vertoont echter wel een verbeterde sensitiviteit, van 87.8%, met een specificiteit van 67.3%, wat een sterke verbetering is van de sensitiviteit van de originele Oorcheck. De Oorcheck met gehomogeniseerd spraakmateriaal lijkt dus een verbetering te zijn van de originele versie. Echter, aanpassingen in de maskeerruis kunnen tot verdere verbetering leiden en worden onderzocht.

OORCHECK MET VERSCHILLENDE RUISMODIFICATIES

De maskeerruis wordt aangepast door deze spectraal en/of temporeel te filteren. Hiervoor wordt de breedbandige stationaire ruis gefilterd met laag- of hoogdoorlaat filters, met een afkapfrequentie van 1.4 kHz, en worden delen van de maskeerruis soms in de tijd gemoduleerd met een 16 Hz blokmodulatie. Aan alle ruizen wordt een ruisvloer van -15 dB toegevoegd, om de invloed van afspeelniveau en achtergrondlawaai te beperken. In totaal zijn er zes experimentele maskeerruizen gedefinieerd; Fluctuerende ruis (16 Hz), Low-pass gefilterde ruis (LP), High-pass gefilterde ruis (HP), een combinatie van deze ruizen met de fluctuerende ruis (LP 16 en HP 16) en een ruis die alleen uit de ruisvloer bestond (-15).

Ook deze testen worden bij dezelfde onderzoekspopulatie afgenomen.

Voor de meeste testen worden er significante verschillen in SRT gevonden tussen normaalhorenden en beide groepen slechthorenden, behalve voor de testen met High-pass gefilterde ruiscondities. De Oorcheck-varianten met Low-pass gefilterde ruis vertonen de grootste verschillen tussen de drie onderzoekspopulaties. Ook voor de fluctuerende ruis worden significante verschillen tussen de groepen gevonden, maar de spreiding binnen de groepen is hier groter dan in de testresultaten van de Low-pass gefilterde ruizen.

Alle experimentele Oorcheck-varianten, met uitzondering van de HP 16, vertonen een significant leereffect (<1 dB), wat overigens kleiner is dan het leereffect dat werd gevonden voor de nieuwe Oorcheck. De testen met experimentele ruiscondities hebben een onderling vergelijkbare test-retest betrouwbaarheid die ook overeenkomt met die van de oorspronkelijke Oorcheck. De beste correlatie tussen de testuitslag en de gemeten gehoordrempels wordt gevonden voor de Oorcheck met Low-pass gefilterde ruis, juist voor de frequenties die relevant zijn voor NIHL.

Deze testen met Low-pass gefilterde ruis maken een beter onderscheid tussen normaalhorenden en de groepen met een verschillende mate van gehoorverlies en correleren sterk met het gehoorverlies op de hogere frequenties. Ze lijken daarom gevoeliger voor NIHL en dus beter bruikbaar bij het screenen op lawaaislechthorendheid.

CONCLUSIES

Hoewel de online hoortest 'Oorcheck' veel gebruikt wordt en een belangrijke rol speelt bij het waarschuwen voor en ontdekken van gehoorverlies door lawaai, is de test in zijn huidige hoedanigheid niet gevoelig genoeg om beginnende lawaaislechthorendheid te ontdekken.

Het homogeniseren van het spraakmateriaal resulteert in een test met een iets groter discriminerend vermogen tussen normaal- en slechthorenden en sterkere correlaties met het toonaudiogram en de Plomptest.

De grootste winst wordt echter behaald wanneer er gebruik gemaakt wordt van slim gekozen ruismateriaal. Bij een Low-pass gefilterde maskeerruis worden er grote verschillen in de SRT gemeten tussen normaalhorenden en personen met een verschillende mate van lawaaislechthorendheid. De betrouwbaarheid en het leereffect van deze test zijn vergelijkbaar met die van de oorspronkelijke Oorcheck. Er is echter sprake van een sterke verbetering van de sensitiviteit (95%), en een verdere toename van de specificiteit (98%).

De Oorcheck met Low-pass gefilterde continue maskeerruis is daarom bij uitstek het alternatief voor de huidige Oorcheck, met de hoogste gevoeligheid voor beginnende lawaaislechthorendheid binnen de in dit onderzoek geteste Oorcheck-varianten.

1 INLEIDING

De huidige muziekconsumptie van jongeren, met veelvuldig gebruik van de populaire mp3-speler en disco- en concertbezoek, leidt tot een forse blootstelling aan luide muziek in deze populatie. Frequente blootstelling aan hoge geluidsintensiteiten kan op den duur leiden tot slechthorendheid door lawaai; noise-induced hearing loss (NIHL).

Om de gevolgen van lawaai-blootstelling voor het gehoor bij jongeren onder de aandacht te brengen heeft de Nationale Hoorstichting een website voor jongeren ontwikkeld, met daarop informatie over gehoor, geluid, lawaai-blootstelling en gehoorbescherming. Een belangrijk onderdeel van deze website wordt gevormd door de interactieve hoortest: de Oorcheck.

De Nationale Hoorstichting heeft verschillende spraak-in-ruis screeningstesten waarmee men op een laagdrempelige manier het gehoor kan testen. Dit betreft de Nationale hoortest (NHT), de eerste screeningstest op gehoor, ontwikkeld door het VUmc (Smits 2004). Deze test is bedoeld voor volwassenen vanaf 25 jaar. Gebaseerd op de principes van deze test heeft het LUMC vervolgens meerdere varianten ontwikkeld; de Kinderhoortest voor kinderen van 4 tot 12 jaar, Oorcheck (OC) voor jongeren van 12 tot 25 jaar en de Bedrijfsoorcheck (BOC) speciaal voor werknemers werkzaam in lawaai. In tegenstelling tot de NHT, die gebruik maakt van cijfers in ruis met een beperkte bandbreedte, wordt het signaal in deze andere testen op volle bandbreedte aangeboden en er zijn steeds negen testwoorden gekozen die een betere spreiding over de Nederlandse spraakklanken vertegenwoordigen dan de getallen in de NHT.

Deze spraak-in-ruistesten geven snel een objectief oordeel over het gehoor en zijn eenvoudig thuis uit te voeren. Het zijn adaptieve testen met spraak in continue achtergrondruis. De frequentiespectra van zowel het spraak- als het ruismateriaal zijn exact gelijk, waardoor invloeden van systeeminstellingen en de kwaliteit van de luidspreker of hoofdtelefoon grotendeels worden ondervangen. De testen zoeken adaptief de signaal-ruis verhouding waarbij 50% van het spraakmateriaal herkend wordt (speech reception threshold; SRT). Deze signaal-ruis verhouding is, mits het signaal voldoende boven de gehoordrempel wordt aangeboden, min of meer onafhankelijk van het exacte afspeelniveau.

Oorcheck meet op eenvoudige wijze in een paar minuten de gehoorfunctie van de deelnemer. Meteen na afronding van de test wordt de uitslag gegeven: je hoort goed, iets minder goed, onvoldoende of slecht. Deze uitslag wordt vergezeld van een advies om het luistergedrag aan te passen, of audiologische hulp te zoeken.

Het doel van deze test is niet alleen de controle van de individuele gehoorscherpthe, maar ook het vergroten van de bewustwording van de risico's van lawaai, in dit geval voornamelijk (te) harde muziek, op het gehoor bij de betreffende deelnemer. Dat er een grote belangstelling voor Oorcheck bestaat blijkt uit het feit dat de test jaarlijks meer dan 100.000 keer wordt uitgevoerd (Jongmans 2009).

Het is echter de vraag of een spraak-in-ruistest met continue ruis voldoende gevoelig is om (beginnende) lawaaislechthorendheid te ontdekken. In een vroeg stadium van lawaai-beschadiging zijn de gehoordrempels voor de lagere en middenfrequenties, die van belang zijn voor het spraakverstaan, nog (redelijk) normaal. Bovendien bestaat het testmateriaal

uit een gesloten set met unieke klinkers, waardoor slechts een deel van de spraakinformatie voldoende is om met het aanwezige restgehoor het correcte antwoord te 'raden'.

Ook eerdere onderzoeken vinden weinig verschil in woordscore tussen normaalhorenden en personen met lawaaislechthorendheid (Philips 1994, Bosman 1995) of tonen slechts een slechts matige correlatie van toonaudiogram met resultaten van spraak-in-ruistesten (Smooenburg 1992).

1.1 DOEL VAN HET PROJECT

Omdat de hoortesten via het internet een grote rol kunnen spelen bij het waarschuwen voor en ontdekken van (beginnende) lawaaislechthorendheid, zowel bij jongeren (t.g.v. mp3-gebruik en discotheekbezoek) als bij werkenden (t.g.v. lawaai op het werk) is een valide screeningstest van belang. Om de gevoeligheid van een spraak-in-ruis test voor lawaaischade in een vroegtijdig stadium te onderzoeken, is een multi-center evaluatie opgezet.

Vanwege de onzekerheid over de waarde van de bestaande internettesten voor het ontdekken van (beginnende) lawaaischade, is allereerst een vergelijkende studie nodig van de drie bestaande testen.

Deze testen worden geëvalueerd met behulp van specifieke populaties met een normaal gehoor en met een verschillende mate van lawaaislechthorendheid. De vraag hierbij is:

1. *Wat is de gevoeligheid van de bestaande spraak-in-ruistesten voor beginnende lawaaischade aan het gehoor?*

Verwacht wordt dat de waarde van de internettesten voor het ontdekken van (beginnende) gehoorschade door lawaai beperkt is. Daarom is het ontwikkelen van een alternatieve, dan wel aanvullende, test noodzakelijk. Hierbij richten we ons voornamelijk op het ontwikkelen van aangepaste versies van de bestaande Oorcheck, met de vraag:

2. *Welke aanpassingen aan de bestaande Oorcheck kunnen ervoor zorgen dat de deze test gevoeliger wordt voor lawaaislechthorendheid?*

Hierbij onderzoeken wij achtereenvolgens de effecten van:

- a. het homogeniseren van het spraakmateriaal
- b. specifieke filtering van het spraak- en/of ruissignaal

Het doel van dit project is uiteindelijk het ontwikkelen van een internet test die specifiek gevoelig is voor het detecteren van gehoorverlies in het frequentiegebied van 3000 – 6000 Hz, de zogenaamde lawaaidip.

2 MATERIAAL EN METHODEN

Het doel van de studie is tweeledig en daarom worden in dit rapport verschillende experimenten beschreven. Gestart wordt met een vergelijkend onderzoek van de bestaande testen in experiment 1, om antwoord op de eerste onderzoeksvraag te vinden.

In experiment 2 worden manieren bestudeerd om de Oorcheck zodanig aan te passen dat deze gevoeliger zal zijn voor het ontdekken van lawaaischade, teneinde antwoord te geven op de tweede onderzoeksvraag. In experiment 2a wordt daarom gekeken naar het effect van het homogeniseren van het woordmateriaal van de originele Oorcheck. In experiment 2b worden verschillende modificaties van het ruissignaal gedefinieerd en vervolgens getest. Uiteindelijk zal de test met het beste resultaat gevalideerd worden.

Alle experimenten zijn gedaan met dezelfde populatie en dezelfde opstelling, welke hieronder worden beschreven.

2.1 TOONAUDIOGRAM

Voordat er met de verschillende spraak-in-ruistesten wordt gestart, wordt er eerst een toonaudiogram afgenomen. Op basis van de gemeten gehoordrempels kan een proefpersoon geïncludeerd worden in een van de hieronder gedefinieerde onderzoeksgroepen.

Het toonaudiogram wordt afgenomen in een geluidsdichte audiocabine.

Luchtgeleidingsdrempels worden per oor voor de volgende frequenties bepaald; 125, 250, 500, 1000, 2000, 3000, 4000, 6000 en 8000 Hz. In de slechthorende groep wordt ook de beengeleidingsdrempel bepaald voor 500, 1000, 2000 en 4000 Hz.

2.2 GOUDEN STANDAARD VOOR SPRAAKVERSTAAN

Bij alle proefpersonen wordt de SRT in stilte en in ruis gemeten met behulp van zinnen ontwikkeld door Plomp en Mimpen (Plomp 1979). Dit is de gouden standaard op het gebied van spraak-in-ruistesten. De uitkomsten van de interactieve hoortesten kunnen hiermee vergeleken worden.

In dit onderzoek wordt gebruik gemaakt van het spraakmateriaal ingesproken door de vrouwelijke spreekster. Dit materiaal bestaat uit tien lijsten van 13 zinnen. De eerste zin wordt op een niveau aangeboden waarop de zin nog niet verstaan zal worden. Dit niveau wordt met 4 dB verhoogd totdat de zin correct herhaald wordt. Dan wordt de volgende zin aangeboden. Een respons geldt als correct wanneer de zin geheel juist gereproduceerd wordt. In dat geval wordt het afspeelniveau van de volgende zin met 2 dB verlaagd. Wordt de zin niet geheel correct herhaald dan gaat het niveau van de volgende zin met 2 dB omhoog. De gemiddelde signaal-ruis verhouding van zin 5 t/m 13, plus een virtuele 14e aanbidding, vormt de SRT (Plomp 1979). Deze procedure wordt gemeten in stilte en in stationaire ruis, met een vast ruisniveau.

Om de 'release of masking' als gevolg van een temporeel fluctuerende ruis te onderzoeken wordt nog een derde conditie aan toegevoegd, namelijk het spraakverstaan van zinnen in een ruis met 16 Hz blokmodulatie. Deze conditie is analoog aan een van de gekozen alternatieve Oorcheck-varianten met gemodificeerde maskeerruis, welke later aan bod zal komen.

Voor deze conditie wordt een oefenlijst afgenomen, omdat er voor deze ruis sprake kan zijn van een leereffect. Alle testcondities worden twee keer afgenomen, in test en retest. De testen in ruis worden afgenomen met een ruisniveau vastgesteld op een niveau 20 dB boven de SRT in stilte, met een minimum van 65 dB(A).

2.3 POPULATIE

In deze multicenter studie werden 50 volwassenen met een normaal gehoor en 50 slechthorenden geïnccludeerd. De metingen vonden plaats in drie verschillende audiologische centra; het Leids Universitair Medisch Centrum (LUMC), het St Radboud Universitair Medisch Centrum Nijmegen (UMCN) en het Academisch Medisch Centrum Amsterdam (AMC).

De normaalhorende groep bestond uit zowel collega's als studenten, gerekruteerd uit de studentenpopulaties van de universiteiten waaraan de drie audiologische centra zijn verbonden. Alle normaalhorende deelnemers hadden gehoordrempels van 15 dB HL of beter op de octaaffrequenties 125 tot 8000 Hz. Eén van de personen voldeed niet aan de opgestelde criteria en werd daarom geëxcludeerd.

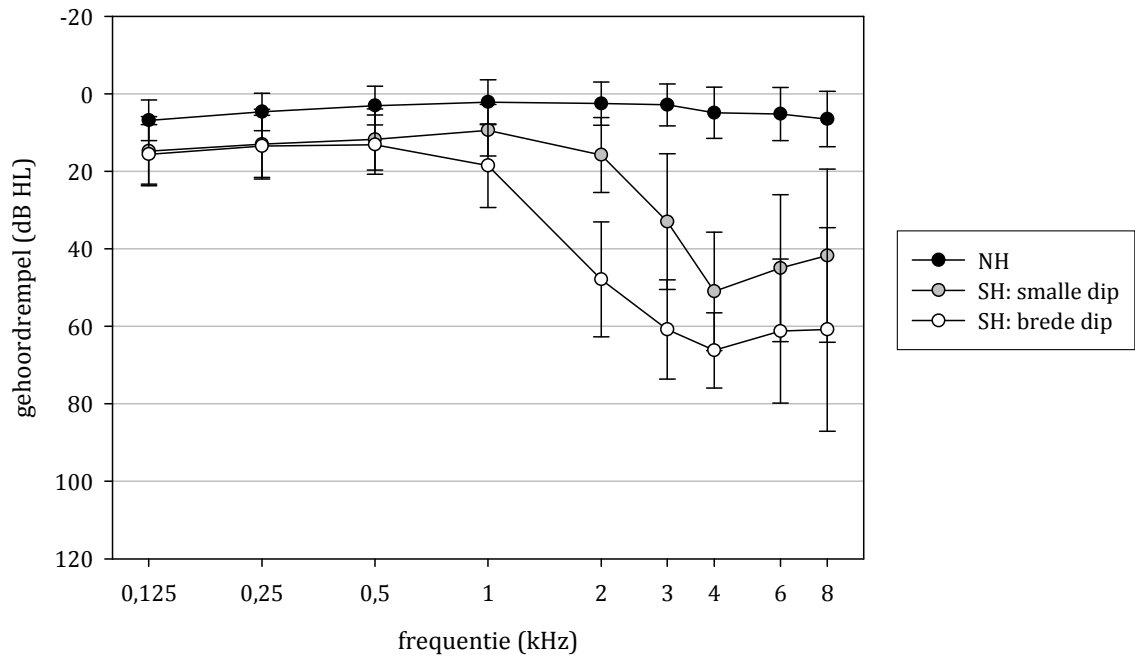
De slechthorende groep bestond uit patiënten van één van de drie KNO-afdelingen waarbij onlangs een toonaudiogram was afgenomen. Een klein deel van deze groep bestond uit werknemers die deelnamen aan eerder onderzoek naar beroepsmatige blootstelling aan lawaai en gehoorverlies van het AMC Amsterdam.

De slechthorende deelnemers hadden een perceptief gehoorverlies, gedefinieerd als volgt:

- afwezigheid van een air-bone gap > 15 dB voor de frequenties 500, 1000 en 2000 Hz.
- gehoordrempels op 250, 500 en 1000 Hz zijn ≤ 30 dB HL.
- gehoordrempels op 2, 3, 4 en/of 6 kHz zijn > 25 dB HL
- gehoordrempel op 4 of 6 kHz is tenminste 20 dB slechter dan PTA_{512}

Ook in deze groep voldeed één van de personen niet aan de opgestelde criteria en werd daarom geëxcludeerd. Binnen de groep slechthorenden is onderscheid gemaakt tussen personen met een matig gehoorverlies, gekenmerkt door een smalle dip (n=25), en personen met een ernstig gehoorverlies, gekenmerkt door een brede dip (n=24). Een brede dip werd gedefinieerd als een toonaudiogram waarbij de gehoordrempel op 2000 Hz tenminste 20 dB slechter is dan het gemiddelde verlies op 500 en 1000 Hz.

De gemiddelde gehoordrempels voor de normaal- en de beide slechthorende groepen (± 1 sd) zijn weergegeven in figuur 1. De deelnemers werden alleen geïnccludeerd als hun moedertaal Nederlands is.



Figuur 1: Gemiddelde audiogram voor de normaalhorende groep (zwarte symbolen) en de slechthorende groep met smalle dip (grijze symbolen) en slechthorende groep met brede dip (witte symbolen). De error bars geven één standaarddeviatie weer.

Het aantal deelnemers in elk centrum, hun demografische kenmerken en de SRT data gemeten met de Plompzinnen zijn weergegeven in tabel 1. Bijna één derde van de normaalhorende groep, met een gemiddelde leeftijd van 27.0 (\pm 8.5) jaar, was man. De slechthorende groep was ouder, gemiddeld 56.3 (\pm 9.4) jaar, en bestond bijna geheel uit mannen (96%). In de slechthorende groep zijn linker- en rechteroren ongeveer even vaak gemeten, afhankelijk van de audiogramconfiguratie. Bij tweederde van de normaalhorenden werd het rechteroor gemeten, aangezien het rechteroor werd gekozen indien er geen duidelijk verschil tussen beide oren was. Als dat wel het geval was, werd er voor het beste oor gekozen.

		n	leeftijd	seks (M)	oor (rechts)	SRT _{stille}	SRT _{cont ruis}	SRT _{fluc ruis}
			gem (sd)	n (%)	n (%)	gem (sd)	gem (sd)	gem (sd)
NH	AMC	22	26.5 (6.0)	8 (36%)	14 (64%)	26.8 (2.3)	-5.6 (1.2)	-17.6 (2.4)
	LUMC	12	21.5 (4.9)	2 (17%)	8 (67%)	31.2 (2.6)	-5.7 (1.0)	-14.1 (2.7)
	UMCN	15	32.0 (11.1)	6 (40%)	9 (60%)	34.6 (2.2)	-5.1 (1.6)	-14.2 (3.6)
	Totaal	49	27.0 (8.5)	16 (33%)	31 (63%)	30.3 (4.1)	-5.4 (1.3)	-15.7 (3.3)
HI	AMC	10	49.0 (7.8)	9 (90%)	4 (40%)	30.7 (3.1)	-3.6 (1.5)	-10.5 (3.4)
	LUMC	25	59.3 (9.5)	24 (96%)	12 (48%)	44.5 (8.1)	-0.4 (2.9)	-3.5 (4.9)
	UMCN	14	56.2 (6.5)	14 (100%)	8 (57%)	42.0 (5.4)	-1.0 (3.1)	-5.1 (5.9)
	Totaal	49	56.3 (9.4)	47 (96%)	25 (51%)	41.0 (8.4)	-1.2 (2.9)	-5.4 (5.6)

Tabel 1: Demografische kenmerken van de onderzoekspopulatie, per centrum en per groep. Gemiddelde SRT waarden zoals gemeten met Plompzinnen in stille, stationaire ruis en fluctuerende ruis (16 Hz) worden gegeven. Deze resultaten zijn gemiddeld over test en retest metingen.

Normaalhorende deelnemers die getest zijn in het LUMC waren significant jonger dan de deelnemers getest in het UMCN. De slechthorenden in het LUMC waren significant ouder dan de slechthorenden die in het AMC zijn getest. Er waren geen significante verschillen voor geslacht en gemeten oor tussen de drie centra.

Wat betreft de resultaten van de Plomptesten hebben de normaalhorenden gemeten in het AMC significant betere SRTs in stilte en in fluctuerende ruis. Dit kan verklaard worden doordat deze groep normaalhorenden ook betere gehoordrempels heeft dan de normaalhorenden gemeten in de andere centra. Voor de resultaten van de Plompzinnen in stationaire ruis worden geen significante verschillen tussen de centra gevonden.

2.4 MEETOPSTELLING

Voorafgaand aan het bezoek aan het audiologisch centrum voeren de deelnemers de originele Oorcheck in hun thuissituatie uit. Ze gebruiken daarvoor hun eigen computersysteem en doen de test op het door hen gewenste geluidsniveau. Aangeraden wordt om de test uit te voeren met een hoofdtelefoon, maar dat is niet in alle gevallen mogelijk. Bij het bezoek aan het audiologisch centrum wordt nagegaan of men de test uitgevoerd heeft met behulp van luidsprekers of een hoofdtelefoon.

Tijdens het bezoek aan het audiologisch centrum worden toonaudiogram, Plomptesten en internettesten allemaal afgenomen in een geluidsdichte audiocabine. Voor de meetopstelling is een computer nodig met internetverbinding. De verschillende internettesten staan klaar op een aantal tijdelijke internetpagina's, die alleen voor dit experiment gebruikt werden.

In het LUMC is het niet mogelijk om de internettesten in een geluidsdichte cabine af te nemen. Daarom wordt daar gebruik gemaakt van een geluidsarme ruimte; een stille kamer zonder afleiding. Om de eventuele aanwezigheid van omgevingslawaai te controleren is het niveau van het achtergrondgeluid gemeten tijdens het uitvoeren van de metingen. De details staan beschreven in bijlage I.

Deze metingen zijn uitgevoerd op drie tijdstippen, ter hoogte van de proefpersoon. Deze geluidsniveaus worden vergeleken met de grenswaarden genoemd in ISO 6189 (1983), met inachtneming van de demping van de hoofdtelefoon. De grenswaarden geven geluidsniveaus weer die niet overschreden mogen worden wanneer luchtgeleidingsdrempels tot 0 dB HL gemeten worden. Deze grenswaarden worden in de testruimte niet overschreden, behalve op 1000 Hz, met minder dan 1 dB SPL, en op 500 Hz. Op deze frequentie is het niveau van het achtergrondgeluid 5,5 dB hoger dan de grenswaarde. Echter, er wordt in dit onderzoek niet tot niveaus van 0 dB gemeten, aangezien de internettesten bovendrempelig worden afgenomen op een startniveau van 65 dB(A), wat met maximaal 30 dB verzwakt kan worden tot minimaal 35 dB(A). Bovendien wordt er in de test zelf een ruisvloer toegevoegd die in elke test minimaal 50 dB(A) bedraagt. Daarom gaan we er in dit onderzoek vanuit dat er in tijdens het uitvoeren van de testen in het LUMC geen sprake is van invloed van omgevingsgeluid op de resultaten en dat de resultaten van de internettesten gemeten in de verschillende centra dus samengenomen kunnen worden voor analyse. De afwezigheid van systematische verschillen in SRT data in continue ruis tussen de centra bevestigt dit.

Alle spraak-in-ruistesten, behalve de Oorcheck in de thuissituatie, worden afgenomen op een vast geluidsniveau van 65 dB(A), dus de output van de test moet regelbaar en controleerbaar zijn. Hiervoor wordt er gebruik gemaakt van een GINA geluidskaart. Het signaal hiervan wordt via een TDT-PA5 verzwakt en vervolgens naar een Sennheiser HDA 200 hoofdtelefoon gestuurd. In het UMCN St Radboud wordt voor de verzwakking van het signaal een AC 40 gebruikt.

De SRT testen met het zinsmateriaal van Plomp en Mimpen worden afgenomen met een op het AMC ontwikkeld SRT programma. Het SRT programma kan echter niet met een Sennheiser HDA 200 worden afgenomen omdat deze gekalibreerd is voor een TDH-39 hoofdtelefoon. Daarom worden Plomp testen met een TDH-39 hoofdtelefoon, in verder dezelfde opstelling, afgenomen. Het is niet mogelijk dit SRT programma binnen het LUMC te gebruiken, daarom worden in dit centrum alle Plomptesten via Decos, met TDH-39 hoofdtelefoon, afgenomen.

Alle spraak-in-ruistesten worden aangeboden op één oor.

2.5 KALIBRATIE

Met behulp van een kunstoor (B&K 4153) wordt de output van de hoofdtelefoon gemeten. Op deze manier kan het afspeelniveau bepaald worden. Voor zowel de Plomptest als de internettesten kan er een ijkruis afgespeeld worden. De ijkruis kan dan met behulp van de verzwakker op het juiste niveau ingesteld worden.

Voor de internettesten geldt dat ze worden afgenomen op een afspeelniveau van 65 dB(A). Dit niveau wordt ingesteld voor de stationaire breedbandige Oorcheck ruis, gebruikt in de herkalibreerde Oorcheck. Alle andere testen dienen afgenomen te worden met dezelfde volume-instellingen. Het geluidsniveau van de verschillende internettesten met deze instellingen is via geluidmetingen met het kunstoor geverifieerd.

Ook is er bij elke test sprake van een button waarmee het spraakmateriaal afgespeeld kan worden. Deze button dient ervoor om de juiste verhouding tussen spraak en ruis te kunnen controleren en hiervoor eventueel later te corrigeren. Deze ijking en correctie kunnen alleen plaatsvinden binnen het AMC.

3 EVALUATIE VAN BESTAANDE INTERNETTESTEN

3.1 EXPERIMENT 1: VERGELIJKING VAN TESTEN

Alle in deze studie te vergelijken online hoortesten zijn interactieve spraak-in-ruistesten; er wordt spraak aangeboden in stoorruis, en de luisteraar kan via toetsen op het beeldscherm zijn/haar respons aanklikken. Na afronding van de test wordt direct de uitslag gegeven. Echter de testen verschillen onderling in het aangeboden materiaal en in de gevolgde procedure. Daarom worden ze hieronder afzonderlijk nader beschreven.

NATIONALE HOORTEST (NHT)

De Nationale Hoortest (NHT) is gebaseerd op de automatische spraak-in-ruis screening test met telefoon van Smits (2004) en is gemaakt in samenwerking met de Nationale Hoorstichting (www.nationalehoortest.nl). Omdat de internetversie gebaseerd is op de telefoontest heeft men bewust de beperkte bandbreedte (300-3400 Hz) van de telefonische test behouden. Op deze manier konden beide testen beter met elkaar vergeleken worden (Smits 2006).

In de test krijgen de respondenten achtereenvolgens 23 keer at random drie cijfers (tussen nul en negen) te horen, waarbij de signaal-ruis verhouding varieert. Op het scherm krijgt de respondent 10 antwoordmogelijkheden: de cijfers van 0 tot 9, weergegeven in een schema zoals de toetsen op een telefoon. Het is een adaptieve test, dus het geluidsniveau van de ruis blijft constant, maar het niveau van de aangeboden cijfers wordt elke keer met 2 dB verlaagd als men het juiste antwoord geeft en met 2 dB verhoogd na een foutieve respons. Een respons is juist als men alle drie de cijfers in de juiste volgorde heeft aangeklikt of ingetoetst. Op deze manier wordt de SNR berekend waarop 50% van de cijfercombinaties correct kan worden verstaan. Dat is de zogenaamde spraak-in-ruis drempel (SRT). De eerste vier aanbiedingen worden beschouwd als aanloop naar de drempel. De laatste 19 aanbiedingen, plus een virtuele 20e aanbieding, vormen de daadwerkelijke meting. Op basis van deze aanbiedingen worden de SRT-drempel en bijbehorende standaarddeviatie bepaald. De testuitslag wordt meteen na het invullen van de laatste cijfers aan de respondent meegedeeld. De mogelijkheden zijn: uw gehoor is goed, uw gehoor is onvoldoende of uw gehoor is slecht. De categorie 'goed' correspondeert met een SRT-waarde van $-5,5$ dB of lager ten opzichte van het aangeboden ruisniveau. De uitslag 'onvoldoende' correspondeert met een SRT-waarde tussen $-5,5$ dB en $-2,8$ dB. De categorie 'slecht' ten slotte correspondeert met een SRT-waarde van $-2,8$ dB of hoger, zie tabel 2.

SRT waarden per categorie (dB)	Nationale Hoortest	Oorcheck	Bedrijfsoorcheck
goed	≤ -5.5	≤ -10	≤ -12
minder goed	-	$-10 < \text{SRT} \leq -7$	$-12 < \text{SRT} \leq -9$
onvoldoende	$-5.5 < \text{SRT} \leq -2.8$	$-7 < \text{SRT} \leq -4$	$-9 < \text{SRT} \leq -6$
slecht	> -2.8	> -4	$-6 < \text{SRT} \leq -3$
zeer slecht	-	-	> -3

Tabel 2: SRT-waarden per categorie van de Nationale Hoortest, Oorcheck en Bedrijfsoorcheck

OORCHECK

De Oorcheck (OC) is een gehoorscreeningstest speciaal ontwikkeld voor jongeren tussen 12 en 24 jaar. Deze test is beschikbaar op de website www.oorcheck.nl. Op het scherm verschijnt een volumebalk waarmee het volume op het gewenste niveau ingesteld kan worden. Daarna verschijnen er negen plaatjes die corresponderen met de stimuli en één plaatje met 'niet verstaan'. Er zijn dus zoals bij de NHT tien keuzemogelijkheden. Een ander verschil met de NHT is dat de stimuli nu op volle bandbreedte worden aangeboden in een achtergrondruis met het gemiddelde spectrum van spraak. Bovendien wordt er gebruik gemaakt van woordmateriaal uit de NVA-lijsten met een goede spreiding over de Nederlandse spraakklanken. De gebruikte woorden zijn zaag, leeuw, duim, vuur, rat, wiel, kip, geit en poes.

De respondenten krijgen achtereenvolgens de negen woorden ieder drie keer at random in ruis te horen, waarbij de SNR varieert. De Oorcheck start met een signaal-ruis verhouding van 0 dB en werkt net zoals de Nationale Hoortest volgens een adaptieve methode. De eerste zeven aanbiedingen worden beschouwd als aanloop van de meting en de laatste 20 als eigenlijke meting. Van deze 20 aanbiedingen wordt de SRT-drempel en de standaardafwijking bepaald.

De testuitslag wordt meteen na het aanklikken van het laatste woord aan de respondent meegedeeld. De mogelijkheden zijn: het gehoor is goed, het gehoor is minder goed, het gehoor is onvoldoende of het gehoor is slecht. Zie tabel 2 voor de corresponderende SRT-waarden.

BEDRIJFSOORCHECK (BOC)

De Bedrijfsoorcheck (BOC) is net als de NHT en de OC een screeningstest van het gehoor¹. Maar deze versie is speciaal voor werknemers die in lawaai werken. De BOC lijkt op de OC, maar er zijn toch een aantal verschillen. Ten eerste werden er andere woorden gebruikt; sok, zon, pan, tas, bed, mes, boek, poes en pijl. Deze woorden hebben per paar dezelfde vocalen. Bovendien worden er meer hoogfrequente consonanten gebruikt.

Een tweede verschil is dat de respondenten nu verplicht met koptelefoon moeten luisteren en dat beide oren afzonderlijk worden getest. De BOC duurt dus langer, dit komt ook doordat er meer stimuli worden gebruikt: 35 woorden per oor. Een derde verschil is dat er pas met de daadwerkelijke meting begonnen wordt wanneer het eerste onjuiste antwoord gegeven is. Op deze manier wordt voor elke persoon afzonderlijk een startwaarde bepaald. Tenslotte zijn er in de BOC meer uitkomstcategorieën gedefinieerd (tabel 2).

Net zoals de NHT en de OC werkt deze test volgens de adaptieve methode en wordt de testuitslag onmiddellijk na het aanklikken van het laatste woord weergegeven. Deze uitslag wordt gebaseerd op de laatste 30 antwoorden per oor.

¹ Naast de drie bovenbeschreven testen heeft de Nationale Hoorstichting ook een test voor kinderen ontwikkeld; de Kinderhoortest. Deze test is bedoeld voor kinderen van 4 tot 12 jaar. De test is qua procedure vergelijkbaar met de Oorcheck, maar maakt gebruik van een aantal andere woorden, omdat deze beter bij de belevingswereld van kinderen aansluiten. De testuitslag wordt onderverdeeld in drie categorieën; goed (<-8 dB), minder (-8 < SRT < -4 dB) en slecht (> -4dB). De kinderhoortest wordt hier slechts genoemd om het overzicht van gehoortesten te completeren en zal in dit onderzoek buiten beschouwing gelaten worden.

3.2 MEETPROTOCOL

Voorafgaand aan het experiment doen de deelnemers eenmaal de Oorcheck in hun thuissituatie, op het door hen gewenste volume en met beide oren.

De metingen op het AC worden gestart met het afnemen van een toonaudiogram. Op basis van deze uitslag wordt het beste oor, of meest geschikte oor, bepaald waaraan de overige testen afgenomen worden.

Daarna wordt gestart met de spraak-in-ruistesten, die allemaal twee keer afgenomen worden (test en retest). Hierbij wordt begonnen met de lijst 1 en lijst 2 van de Plompzinnen, aangeboden in stilte. Indien de SRT in stilte hoger is dan 45 dB(A), wordt het afspeelniveau van alle overige testen aangepast aan de verhoogde drempel, naar $SRT_{\text{stilte}} + 20 \text{ dB(A)}$.

Vervolgens wordt lijst 4 in fluctuerende ruis aangeboden, als oefenlijst voor deze conditie. Dan volgen twee lijsten in stationaire ruis en twee lijsten in fluctuerende ruis waarvan de volgorde van de condities bepaald wordt door een 4x4 Latijns vierkant (Wagenaar 1969). De lijsten worden in vaste volgorde van lijst 5 t/m lijst 8 afgenomen.

Als laatste onderdeel worden de te vergelijken internettesten afgenomen. Dit betreft de drie bestaande testen, plus de geoptimaliseerde Oorcheck (zie experiment 2a). De testen worden monoraal aangeboden door middel van een Sennheiser HDA 200. Het ruisniveau wordt vastgesteld op 65 dB(A), of op een niveau van 20 dB boven de SRT in stilte indien deze drempel voor de proefpersoon hoger ligt dan 45 dB(A). De proefpersoon bevindt zich bij het uitvoeren van de testen in een geluidsarme ruimte en doorloopt elke test tweemaal, gescheiden door een korte pauze. De volgorde van de vier internettesten is gebalanceerde volgens een 4x4 Latijns vierkant.

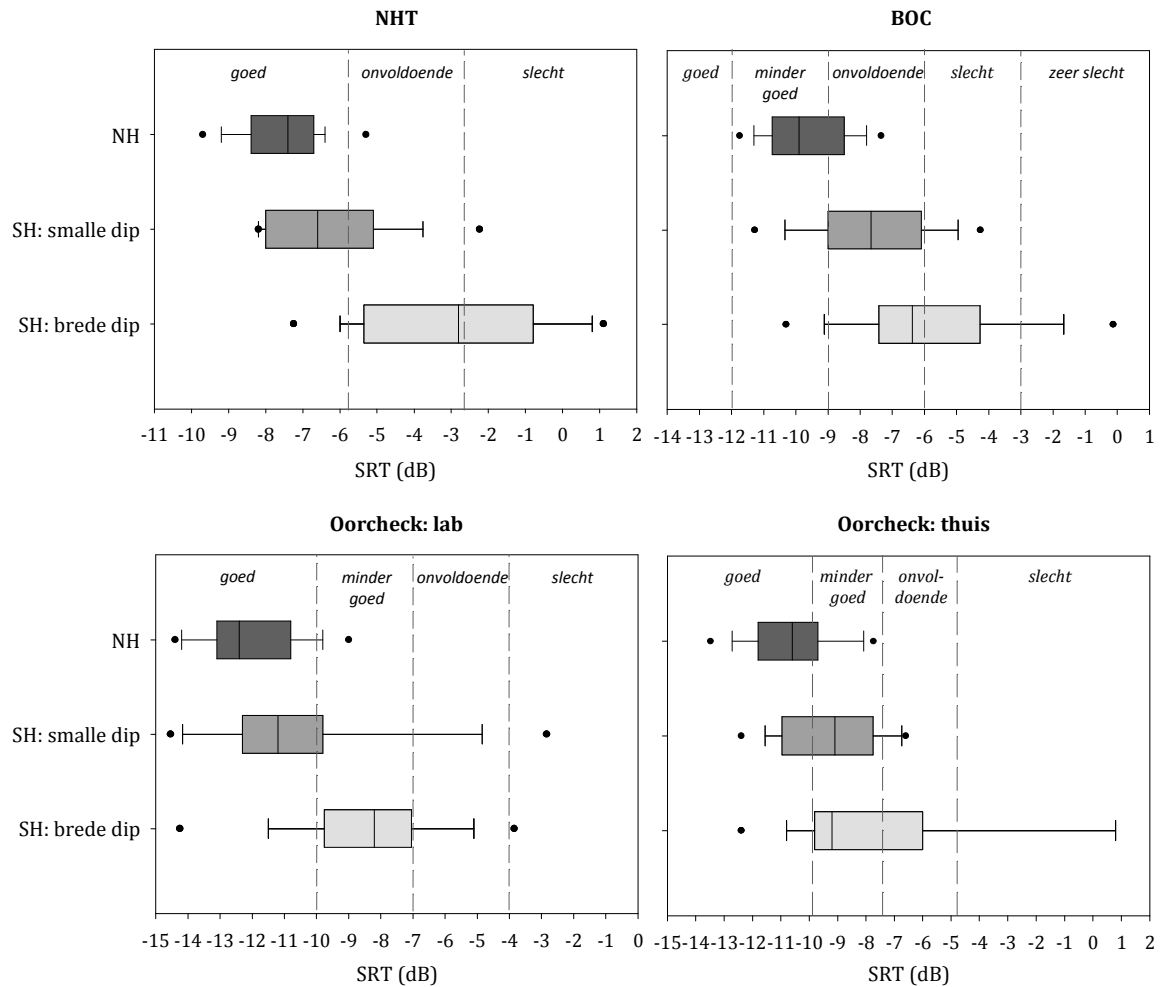
3.3 RESULTATEN VAN EXPERIMENT 1

GEVOELIGHEID VAN DE TESTEN

De drie bestaande interactieve hoortesten worden in dit onderzoek geëvalueerd, waarbij antwoord wordt gegeven op de vraag of deze testen voldoende gevoelig zijn om lawaaislechthorendheid te ontdekken. In figuur 2 worden boxplots weergegeven van de resultaten van de verschillende testen gesplitst voor de groep met normaal gehoor en beide groepen slechthorenden.

De meeste normaalhorenden krijgen de uitslag 'goed' na het uitvoeren van de NHT of de OC. Na het uitvoeren van de BOC behaalt echter geen van de normaalhorende deelnemers de uitslag 'goed', maar alle deelnemers scoren 'minder goed' of lager.

De slechthorenden laten een grotere spreiding binnen de verschillende testen zien. De meeste slechthorende personen behalen een uitslag die slechter is dan 'goed', met name degene met een brede dip. Maar er zijn toch ook een groot aantal slechthorenden dat voor de NHT en OC een goede uitslag krijgen. Dit zijn met name de deelnemers in de groep met een smalle dip.



Figuur 2: Boxplots van SRTs van de eerste meting voor normaalhoorenden (donkergrijs) slechthorenden met een smalle dip (grijs) en slechthorenden met een brede dip retest (lichtgrijs) voor de drie verschillende screeningstesten. De verticale stippellijnen geven per test de uitslagcategorieën aan, die variëren van goed (links) tot zeer slecht (rechts).

Om het onderscheidend vermogen van de testen beter in kaart te brengen worden de sensitiviteit en specificiteit van de drie testen berekend, op basis van de resultaten van de eerste meting bij elke proefpersoon.

Zoals eerder genoemd kreeg niemand van de proefpersonen bij de BOC een uitslag in de categorie 'goed'. Daarom wordt er voor de berekening van de sensitiviteit en specificiteit uitgegaan van de grenswaarde -9, de grens tussen 'minder goed' en 'onvoldoende'. De sensitiviteit en specificiteit van de drie bestaande testen wordt gegeven in tabel 3. De sensitiviteit van een test geeft aan welk percentage van de slechthorenden door de test (terecht) als afwijkend (dus niet normaalhoorend) wordt geclassificeerd. De specificiteit geeft het percentage normaalhoorenden weer, dat door de test (terecht) als niet-afwijkend (en dus normaalhoorend) wordt aangeduid.

In het geval van Oorcheck wordt dus slechts de helft van de slechthorenden als zodanig geclassificeerd.

	sensitiviteit	specificiteit
NHT	55%	94%
BOC	78%	69%
OC	51%	90%
OC - thuis	71%	70%

Tabel 3: Testspecifieke kenmerken van de drie bestaande internettesten; sensitiviteit en specificiteit. Deze kenmerken zijn ook gegeven voor de OC in de thuissituatie.

OORCHECK IN THUISSITUATIE VERSUS EXPERIMENTELE TESTOMGEVING

De Oorcheck gemeten in de thuissituatie geeft het beste beeld van de huidige prestaties van deze spraak-in-ruistest. Daarnaast worden de resultaten van de originele Oorcheck uitgevoerd in de thuissituatie vergeleken met de resultaten verkregen onder de gecontroleerde experimentele omstandigheden in het lab. Een dergelijke vergelijking is nuttig, omdat het inzicht geeft in de invloed van de verschillen in de testsituatie, zoals het gebruikte computersysteem, achtergrondlawaai en afspeelniveau. Bovendien wordt de Oorcheck in het experiment monoraal getest, in tegenstelling tot het testen via internet, zoals in de thuissituatie, waarbij binauraal wordt gemeten. Gevonden verschillen tussen beide situaties kunnen in een later stadium eventueel gebruikt worden bij de vertaling van de resultaten gevonden in het experiment naar de verwachte resultaten in de praktijk. In totaal zijn de resultaten van Oorcheck in de thuissituatie van 69 deelnemers geanalyseerd^{2,3}. De verdeling van de SRTs is in figuur 2d weergegeven.

Gemiddeld genomen komen de drempels in de thuissituatie iets ongunstiger uit. De gemiddelde SRTs zijn hoger, waarschijnlijk ten gevolge van het gebruikte systeem en aanwezig achtergrondgeluid. Dit heeft ook gevolgen voor de sensitiviteit en de specificiteit zoals deze gelden voor de toepassing van OC in de praktijk (tabel 3). De sensitiviteit van de Oorcheck in de thuissituatie is iets hoger (71%) dan wanneer de test uitgevoerd wordt onder de gecontroleerde experimentele omstandigheden. Dit gaat wel ten koste van de specificiteit die lager is in de praktijk, namelijk 70%. Dit betekent voor de gevoeligheid van Oorcheck dat in de praktijk drie op de tien personen verkeerd geclassificeerd worden.

Verschillen tussen testresultaten verkregen met hoofdtelefoon en met luidsprekers

Voordat de resultaten van Oorcheck in de beide testsituaties met elkaar vergeleken worden, wordt eerst de data verkregen in de thuissituatie nader bekeken. Niet alle deelnemers konden namelijk de test uitvoeren met een hoofdtelefoon. Analyse van de NHT laat zien dat uitvoeren

² Van één van de normaalhorende personen konden de resultaten van de Oorcheck thuis niet teruggevonden worden.

³ Het UMCN heeft de originele Oorcheck uitgevoerd voorafgaand aan het experiment op het audiologisch centrum, met dezelfde opstelling als de overige testen. Deze testomgeving is echter niet vergelijkbaar met een thuissituatie. Daarom worden deze resultaten, hoewel ze niet significant afwijken van de uitkomsten in de andere centra, niet meegenomen in de volgende analyses.

van de test met gebruik van een hoofdtelefoon resulteert in een lagere SRT dan wanneer luidsprekers gebruikt worden. Voor het totaal aantal uitgevoerde testen in 2007 bedroeg dat verschil 1.8 dB (Jongmans 2008). Ook Smits vindt in zijn onderzoek (2006), na correctie voor leeftijd en geslacht, een voordeel van testen via hoofdtelefoon van 1.1 dB.

Van drie personen was niet bekend via welke transducer ze de test uitgevoerd hadden. Van de overige personen wordt de uitslag van de Oorcheck-thuis verkregen met beide transducers vergeleken. Ongeveer één derde van de personen heeft de test met behulp van luidsprekers gedaan, zij hebben een gemiddelde SRT van -8.9 dB. De personen die bij uitvoer van de test een hoofdtelefoon gebruikt hebben scoorden gemiddeld iets beter: -9.6 dB (zie tabel 4). Dit verschil van 0.7 dB is niet significant, ook niet wanneer er gecorrigeerd wordt voor onderzoeksgroep ($F=2.887$, $p=0.094$).

Verschillen tussen testresultaten verkregen in verschillende testsituaties

Ondanks het feit dat er geen significant verschil wordt gevonden tussen de resultaten verkregen met luidsprekers en verkregen met hoofdtelefoon, wordt de vergelijking van de resultaten van Oorcheck gemeten onder gecontroleerde omstandigheden en gemeten in de thuissituatie gesplitst naar gebruikte transducer.

Bij deze vergelijking moet opgemerkt worden dat de experimentele testen monoraal afgenomen worden, terwijl via internet binauraal getest wordt. Smits rapporteert binaurale resultaten van de NHT die 1.4 dB lager, en dus beter, zijn dan resultaten gemeten met één oor (2006). Een gepaarde vergelijking van de resultaten van Oorcheck gemeten in de praktijk en in het experiment laat zien dat de SRT gemeten onder gecontroleerde omstandigheden 1.2 dB significant lager is dan wanneer de test thuis via een hoofdtelefoon gedaan wordt ($p=0.002$), zie tabel 4. De SRT gemeten onder gecontroleerde omstandigheden is zelfs 2.0 dB lager dan wanneer men de test thuis uitvoert via luidsprekers ($p=0.004$). Men haalt in de thuissituatie, ondanks het gebruik van beide oren, dus hogere SRT's dan onder gecontroleerde omstandigheden met één oor.

	n	Thuis-situatie gem (sd)	Lab-situatie gem (sd)
luidsprekers	21	-8.9 (3.6)	-10.9 (2.0)
hoofdtelefoon	45	-9.6 (2.7)	-10.8 (2.8)

Tabel 4. Resultaten van de totale groep gemeten met Oorcheck in de thuissituatie, gesplitst naar gebruikte transducers en in de experimentele testomgeving

VERDERE EVALUATIE VAN BESTAANDE TESTEN

De drie bestaande testen, uitgevoerd onder gecontroleerde omstandigheden, zijn nader geëvalueerd. Naast de verschillen tussen de drie groepen in de gemiddelde SRT data zijn ook andere testkarakteristieken berekend, namelijk de eventuele aanwezigheid van een leereffect en de test-retest betrouwbaarheid.

Verschillen tussen de drie subgroepen

In tabel 5 zijn de SRT resultaten van de drie bestaande testen vergeleken tussen de groepen, waarbij de gemiddelde verschillen zijn berekend tussen de groep normaalhorenden en de twee groepen slechthorenden, en deze laatste groepen onderling, voor elk van de drie bestaande testen.

De normaalhorenden hebben op elke test een significant lagere SRT dan de slechthorenden met brede dip. Voor de NHT en BOC hebben normaalhorenden ook een significant lagere SRT dan de slechthorenden met een smalle dip, maar voor OC is dit verschil niet significant. De SRTs van de slechthorenden met een brede dip zijn voor alle testen significant lager dan de SRTs van de slechthorenden met een smalle dip. De verschillen tussen de normaal- en slechthorende groepen zijn ongeveer even groot voor de verschillende testen.

Test	ANOVA		Δ NH - SH _{SD}		Δ NH - SH _{BD}		Δ SH _{SD} - SH _{BD}	
	F	p	gem	p	gem	p	gem	p
NHT	60,01	0.000	-1.2	0.015	-4.6	0.000	-3.4	0.000
BOC	32,04	0.000	-2.0	0.000	-3.7	0.000	-1.7	0.008
OC	20,56	0.000	-1.2	0.079	-3.6	0.000	-2.4	0.001

Tabel 5: Uitkomst van ANOVA met groep als factor en de SRT van meting 1 als uitkomst. Per test is het gemiddelde verschil tussen de groepen, met bijbehorende p-waarden gegeven (na Bonferroni correctie).

Leereffecten in de bestaande testen

Het gemiddelde verschil tussen test en retest zegt iets over de aanwezigheid en de grootte van een leereffect binnen de test. De gemiddelde waarden voor de beide metingen en voor de drie groepen worden in tabel 6 weergegeven. Daarnaast is het gemiddelde verschil tussen beide metingen weergegeven, met bijbehorend significantieniveau. Voor de NHT is er geen sprake van een significant leereffect. Voor de BOC en OC is dit wel het geval, waarbij het leereffect voor OC het grootst is. Dit effect is echter kleiner dan 1 dB.

Het kritische verschil wordt bepaald door de bovengrens van het 95% betrouwbaarheidsinterval van het verschil tussen test en retest, en geeft aan hoe groot het verschil tussen twee testen minimaal moet zijn om als significant verschil te worden beschouwd bij een individuele test. Ook deze waarde is het grootst voor OC en ligt net boven 1 dB.

test	groep	meting 1 gem (sd)	meting 2 gem (sd)	gem. verschil	p verschil	kritisch verschil	sd intra	correlatie test-retest
NHT	NH	-7.5 (1.2)	-7.9(0.8)	0.29	0.097	0.63	1.21	0.79
	SH-S	-6.3 (1.7)	-6.7 (2.3)					
	SH-B	-2.9 (2.5)	-2.8 (2.0)					
BOC	NH	-9.6 (1.4)	-9.7 (1.7)	0.45	0.015	0.81	1.33	0.71
	SH-S	-7.6 (2.0)	-7.9 (2.0)					
	SH-B	-5.9 (2.7)	-7.2 (1.7)					
OC	NH	-12.0 (1.7)	-12.6 (1.5)	0.82	0.000	1.18	1.37	0.76
	SH-S	-10.7 (3.0)	-11.4 (2.3)					
	SH-B	-8.4 (2.4)	-9.6 (2.3)					

Tabel 6: Gemiddelde SRT-waarden en test-retest karakteristieken voor de bestaande testen.

Test-retest betrouwbaarheid van de bestaande testen

De test-retest betrouwbaarheid kan weergegeven worden door de intra-individuele standaarddeviatie van het verschil (sd_{intra}). Deze maat wordt als volgt berekend:

$$\sigma_{\text{intra}} = \frac{1}{\sqrt{2}} \cdot \sqrt{\sum_n \frac{1}{n} \cdot (x_{n,1} - x_{n,2})^2}$$

De sd_{intra} is vergelijkbaar voor de drie testen.

Ook de mate van correlatie tussen de SRT gemeten met meting 1 en met meting 2 is een maat voor test-retest betrouwbaarheid. Ook deze correlatiecoëfficiënt is vergelijkbaar voor de drie testen.

3.4 CONCLUSIES T.A.V. DE EVALUATIE VAN DE BESTAANDE TESTEN

De sensitiviteit van NHT en OC is in deze populatie relatief laag: ongeveer de helft van de totale groep slechthorenden, voornamelijk de deelnemers met een smalle dip in het audiogram, worden door deze testen als normaalhorend geclassificeerd. Wanneer we de resultaten van de Oorcheck in de praktijk bekijken is de sensitiviteit iets hoger, maar de specificiteit iets lager. Echter, drie op de tien personen worden in de praktijk verkeerd geclassificeerd.

De sensitiviteit en specificiteit van de BOC zijn vergelijkbaar met die van de thuis - Oorcheck, maar alleen als de hoogste grenswaarde buiten beschouwing wordt gelaten. Wanneer de afkappunten van de BOC juist gehanteerd worden zou niemand van de normaalhorenden als zodanig door de test aangemerkt worden. Dit komt overeen met een specificiteit van 0%.

De resultaten van de Oorcheck zijn gemiddeld hoger wanneer deze test thuis uitgevoerd wordt dan wanneer deze gedaan wordt onder gecontroleerde experimentele omstandigheden. De verschillen tussen de resultaten verkregen in de thuissituatie met verschillende transducers zijn binnen de kleine groepen waarvoor dit kon worden vergeleken niet significant.

Voor de NHT wordt geen leereffect gevonden, voor de BOC en OC wel, van respectievelijk 0.5 en 0.8 dB. Dit is lager dan de individuele testonnauwkeurigheid (sd_{intra}). De normaalhorenden scoren (zoals verwacht) significant beter dan de slechthorenden op alle testen, maar gemeten met de Oorcheck is de SRT van de normaalhorenden niet significant verschillend van die van de slechthorende deelnemers met een smalle dip.

De grootte van de verschillen tussen de groepen zijn vergelijkbaar voor de drie testen. Verder hebben de drie bestaande testen een goede test-retest betrouwbaarheid, die vergelijkbaar is voor de drie testen.

Uit experiment 1 blijkt dat geen van de vergeleken testen gevoelig genoeg is voor het ontdekken van lawaaislechthorendheid. Omdat in de doelgroep van de Oorcheck naar verwachting lawaaislechthorendheid zal voorkomen, wordt er in het tweede deel van dit onderzoek gezocht naar een manier om de gevoeligheid van de Oorcheck voor (beginnende) lawaaislechthorendheid te verbeteren.

4 EFFECTEN VAN SPRAAKMODIFICATIE OP DE TEST GEVOELIGHEID

De eerste verbetering van de huidige Oorcheck wordt gezocht in het homogeniseren van het spraakmateriaal. Sommige testwoorden blijken relatief moeilijk te zijn en andere testwoorden relatief makkelijk. Door het aanbiedingsniveau van de woorden aan te passen kan de homogeniteit van het testmateriaal worden verbeterd en de test-retest betrouwbaarheid worden geoptimaliseerd.

Dit homogeniseren wordt gedaan op basis van ongeveer 100.000 testen die in 2004 t/m 2007 via internet zijn afgenomen. Na het aanpassen van de intensiteit van de negen woorden, wordt ook de ruis opnieuw gemaakt, zodat het spectrum goed overeenkomt met het gemiddelde spraakspectrum van de negen gehomogeniseerde woorden. Onze hypothese is dat de nieuwe test betrouwbaarder is dan de originele.

4.1 EXPERIMENT 2A: OPTIMALISATIE VAN SPRAAKMATERIAAL

Het spraakmateriaal van de Oorcheck bestaat uit negen verschillende woorden. Uit een analyse van de resultaten van 2004 (Martens 2005) blijkt dat het woord “wiel” een moeilijk woord is. Deze analyse is uitgevoerd bij een grote groep deelnemers aan de Oorcheck. Er werd gekeken naar de aangeboden stimulus, naar de signaal-ruis verhouding ten opzichte van de individuele SRT drempel waarop de stimulus aangeboden werd, naar de bijbehorende respons en of deze juist was. Voor elk woord werden de discriminatiecurven bepaald op de manier zoals beschreven in bijlage II. Het 50%-punt van de verschillende curven varieert tussen -1.2 tot 2.1 dB.

Wij hebben deze analyses opnieuw uitgevoerd, en wel voor de jaren 2004 (geheel), 2005, 2006 en 2007 (afstudeeronderzoek Celine Jacobs, 2008). Bij alle analyses is een selectie gemaakt van de deelnemers in de beoogde leeftijdsgroep (12 t/m 24 jaar). Bovendien zijn alleen die resultaten in de analyse betrokken, waarvan wij op basis van de standaarddeviatie binnen de adaptieve procedure konden vast stellen dat de meting (zeer) betrouwbaar is uitgevoerd. Als grens hiervoor is een standaarddeviatie kleiner dan of gelijk aan 2 dB gekozen (Martens 2005).

Woord	2004		2005		2006		2007		Totaal	
	Shift (dB)	Slope (%/dB)	Shift (dB)	Slope (%/dB)	Shift (dB)	Slope (%/dB)	Shift (dB)	Slope (%/dB)	Shift (dB)	Slope (%/dB)
Kip	-0.77	22.5	-0.89	22.2	-0.93	22.8	-0.82	22.7	-0.84	22.5
Duim	-1.28	23.0	-1.18	23.0	-1.38	23.0	-1.32	22.9	-1.29	23.3
Geit	-0.80	24.9	-0.68	24.4	-0.78	25.8	-0.81	25.1	-0.78	25.1
Leeuw	1.22	24.6	1.14	23.8	1.09	24.6	1.20	24.2	1.70	24.4
Poes	0.79	18.8	0.62	18.0	0.76	18.7	0.73	18.7	0.74	18.7
Rat	-0.35	21.8	-0.34	21.8	-0.33	21.6	-0.95	21.8	-0.39	21.7
Vuur	-0.11	21.4	-0.15	20.9	0.17	21.2	0.08	21.0	-0.10	21.2
Wiel	2.22	14.8	2.37	14.1	2.60	13.5	2.44	14.3	2.38	14.3
Zaag	-0.29	23.6	-0.46	22.9	-0.48	23.4	-0.29	23.6	-0.37	23.4

Tabel 7: De verschuiving en helling van de woordspecifieke discriminatiecurven, per jaar en voor alle jaren

Aan de hand van de resultaten van 2004 tot 2007 kunnen we per woord een discriminatiefunctie bepalen, waarin het percentage correct als functie van het aangeboden niveau t.o.v. de individuele drempel weergegeven wordt. De resultaten van 105.783 deelnemers zijn via een logistische regressie vertaald in een gemiddelde verschuiving (shift) en een gemiddelde helling (slope). De verschuiving is een maat voor de moeilijkheidsgraad van een testwoord. De helling van de curve hangt samen met het discriminerend vermogen van het betreffende woord. Deze resultaten zijn, per jaar en voor het totaal, voor elk woord weergegeven in tabel 7.

Uit tabel 7 blijkt dat de berekende verschuivingen en hellingen per jaar goed overeenkomen. Dit is nader geverifieerd door middel van een correlatieanalyse. De resultaten hiervan staan vermeld in tabel 8. De overeenkomst blijkt groot en alle correlatiecoëfficiënten zijn significant en groter dan 0,98 ($p < 0.001$).

	2004	2005	2006	2007	Totaal
2004	*	0.994 $p < 0.001$	0.990 $p < 0.001$	0.984 $p < 0.001$	0.999 $p < 0.001$
2005	0.996 $p < 0.001$	*	0.994 $p < 0.001$	0.980 $p < 0.001$	0.998 $p < 0.001$
2006	0.996 $p < 0.001$	0.992 $p < 0.001$	*	0.981 $p < 0.001$	0.995 $p < 0.001$
2007	0.998 $p < 0.001$	0.995 $p < 0.001$	0.998 $p < 0.001$	*	0.986 $p < 0.001$
Totaal	0.999 $p < 0.001$	0.997 $p < 0.001$	0.997 $p < 0.001$	0.998 $p < 0.001$	*

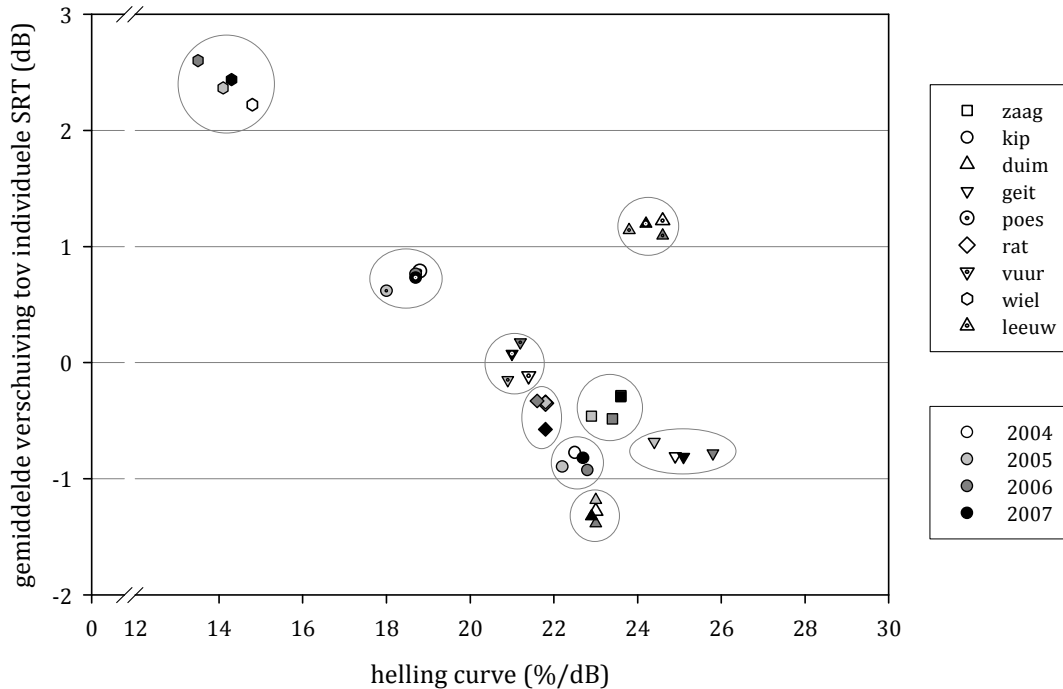
Tabel 8: *Correlatiecoëfficiënten met bijbehorende p-waarden van correlaties tussen de jaren. De witte cellen geven de correlatie aan tussen de verschuiving van elke woordspecifieke curve per jaar en de grijze cellen geven de correlatie aan tussen de helling van elke curve per jaar.*

De resultaten voor de verschillende jaren zijn tevens weergegeven in een scatterplot (figuur 3). Bij alle woorden treedt een duidelijke clustering op. Deze figuur rechtvaardigt, net als de hoge mate van correlatie gegeven in tabel 8, het samennemen van de resultaten, waardoor de resultaten van de totale populatie gebruikt zullen worden.

Nu we voor elk woord deze parameters hebben kunnen wij proberen het materiaal consistent te maken. Aan het verschil in de helling is niet veel te doen, tenzij wij nieuwe woorden zouden includeren. Aan het verschil in presentatieniveaus valt echter wel eer te behalen. Als wij de verschuivingen van de kolom 'totaal' uit tabel 7 gebruiken om de RMS van de woorden aan te passen, wordt het spraakmateriaal (perceptief) homogener en mag derhalve een nauwkeuriger resultaat worden verwacht.

Omdat de RMS waarden van de verschillende woorden naar aanleiding van bovenstaande berekeningen aangepast worden, wordt ook de ruis aan het vernieuwde spraakmateriaal aangepast. Het spectrum van de ruis moet exact overeenkomen met het gemiddelde spectrum van het gehomogeniseerde spraakmateriaal. Deze ruis wordt gemaakt door witte ruis te filteren met een filter gebaseerd op het long term average speech spectrum (LTASS) van het herkalibreerde spraakmateriaal. Hiertoe is een running speech gemaakt van de negen verschillende woordjes door deze in Cool Edit 2000 achter elkaar te plakken.

Van deze running speech is het spectrum bepaald en op basis daarvan is een FFT-filter gemaakt om de witte ruis te filteren [Versfeld 2000]. De RMS waarde van deze ruis is vervolgens gelijk gesteld aan de RMS waarde van de running speech, voordat de ruis aan de test toegevoegd wordt.



Figuur 3: De verschuiving en helling van de woordspecifieke discriminatiecurven, weergegeven per jaar.

4.2 RESULTATEN VAN EXPERIMENT 2A

De eerste verbetering van de huidige Oorcheck wordt dus gezocht in het verbeteren van de homogeniteit van het spraakmateriaal. Na het aanpassen van de intensiteit van de negen woorden en de ruis wordt verwacht dat de nieuwe test betrouwbaarder is dan de originele Oorcheck.

De resultaten van de proefpersonen verkregen met de nieuwe Oorcheck worden hieronder vergeleken met de resultaten van de originele versie van de Oorcheck.

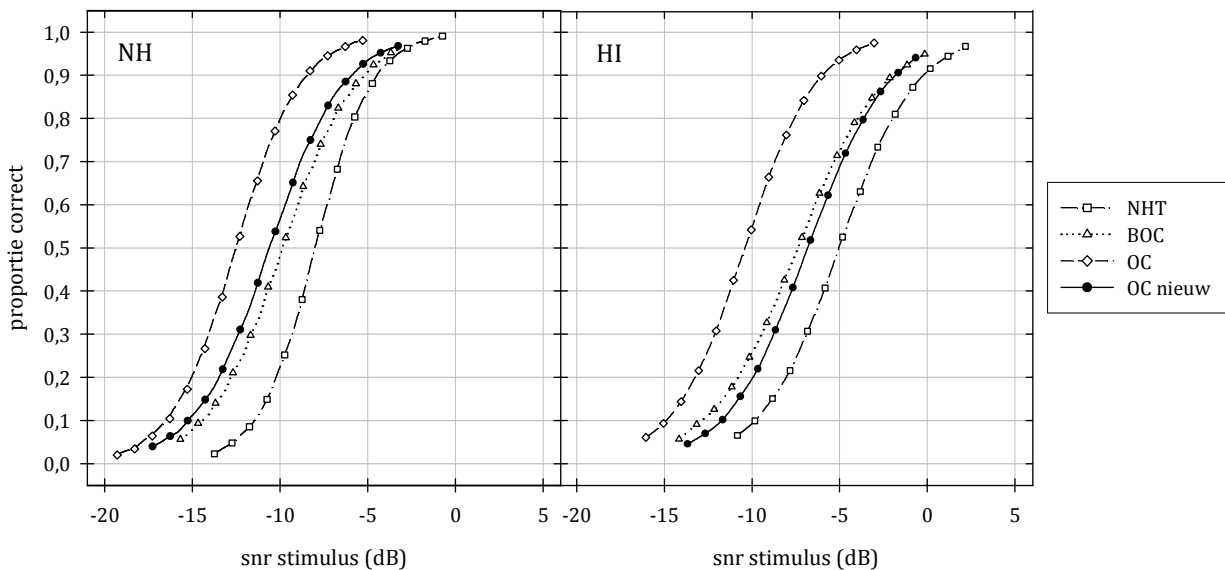
De resultaten van de nieuwe Oorcheck in vergelijking tot die van de originele Oorcheck

Allereerst worden de verschillen in bovenstaande testuitslagen tussen de drie onderzoeksgroepen bekeken (tabel 9). Deze resultaten laten zien er met de nieuwe test een groter verschil wordt gevonden tussen normaalhorenden en de twee groepen slechthorenden met verschillende maten van NIHL, en tussen de beide groepen slechthorenden onderling. Met deze vernieuwde test wordt er wel een significant verschil gevonden in resultaat tussen normaalhorenden en slechthorenden met een smalle dip.

Test	ANOVA		$\Delta NH - SH_{SD}$		$\Delta NH - SH_{BD}$		$\Delta SH_{SD} - SH_{BD}$	
	F	p	gem	p	gem	p	gem	p
OC	20.6	0.000	-1.2	0.079	-3.6	0.000	-2.4	0.001
OC nieuw	32.7	0.000	-2.6	0.000	-4.2	0.000	-1.6	0.035

Tabel 9: Uitkomst van de uitgevoerde ANOVA met groep als factor en de SRT van meting 1 als uitkomst. Per test is het gemiddelde verschil tussen de verschillende groepen, met bijbehorende p-waarden, na Bonferroni correctie, gegeven.

Van beide testen, en de twee andere bestaande testen, worden performance-intensiteit (PI) curven gemaakt, weergegeven in figuur 4. Om te voorkomen dat de aantallen per test te klein zouden worden om de psychometrische curve betrouwbaar te presenteren, zijn beide groepen slechthorenden als één groep weergegeven. De curve van de nieuwe OC ligt iets meer naar rechts dan de originele OC, omdat deze test resulteert in een hogere gemiddelde SRT. De helling van de test hangt, net als bij de woordspecifieke discriminatiecurven, samen met het discriminerend vermogen van de test; hoe steiler hoe nauwkeuriger is de test. De steilheid van de curve van beide Oorchecks is vergelijkbaar, zowel voor de normaal- als slechthorende groep. De curve van de NHT is het steilst.



Figuur 4: PI curven van de vier testen met proportie correct als functie van de signaal-ruis verhouding. De snr is weergegeven relatief t.o.v. individuele SRT en vervolgens gecorrigeerd naar gemiddelde SRT van de test.

Leereffecten voor de vernieuwde Oorcheck

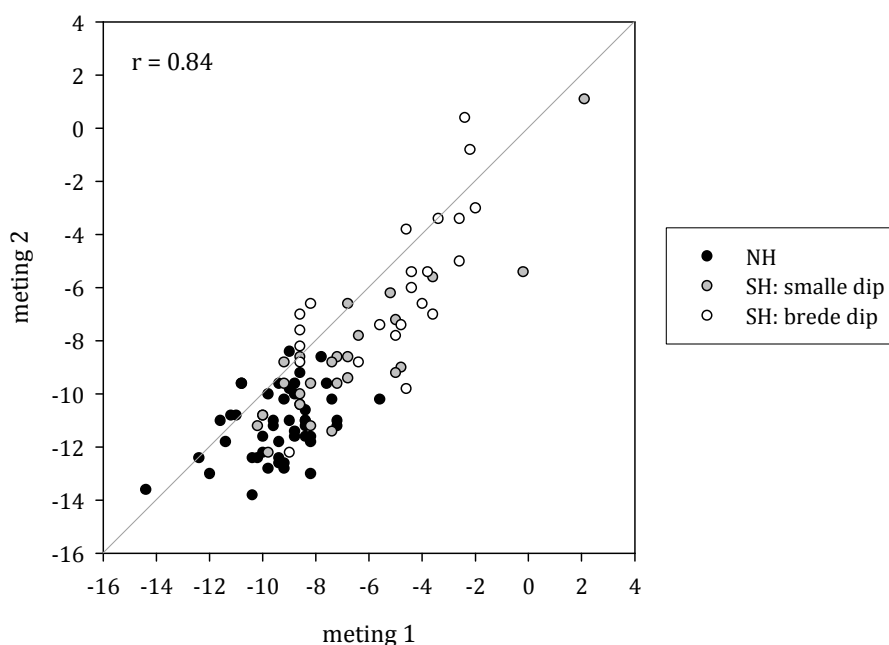
Evenals voor de originele Oorcheck zijn ook voor de vernieuwde de gemiddelde SRTs en sd's van de drie groepen berekend, met daarbij het gemiddelde test-retest verschil (tabel 10).

Het verschil tussen test en retest is ook voor de nieuwe Oorcheck significant en zelfs een factor 2 groter dan het leereffect van de originele OC. Gemiddeld genomen doet men het 1,6 dB beter wanneer men de test voor de tweede keer uitvoert.

test	groep	meting 1 gem (sd)	meting 2 gem (sd)	gem. verschil	p verschil	kritisch verschil	sd intra	correlatie test-retest
OC	NH	-12.0 (1.7)	-12.6 (1.5)	0.82	0.000	1.18	1.37	0.76
	SH-S	-10.7 (3.0)	-11.4 (2.3)					
	SH-B	-8.4 (2.4)	-9.6 (2.3)					
OC nieuw	NH	-9.4 (1.5)	-11.2 (1.3)	1.62	0.000	1.95	1.63	0.84
	SH-S	-6.7 (2.9)	-8.6 (2.7)					
	SH-B	-5.1 (2.4)	-6.1 (2.8)					

Tabel 10: Gemiddelde SRT-waarden en test-retest karakteristieken voor de nieuwe Oorcheck, vergeleken met de originele Oorcheck

Dit effect is ook te zien in figuur 5, een scatterplot waarbij de SRT van meting 2 is uitgezet tegen de SRT van meting 1. De meeste punten liggen onder de diagonale lijn, wat aangeeft dat de SRT van meting 2 lager is dan de SRT van meting 1. Door dit grotere leereffect is ook het kritische verschil groter voor de nieuwe OC; een verschil tussen twee opeenvolgende testen van 2 dB of meer kan worden beschouwd als een significant verschil.



Figuur 5: Scatterplot voor de SRTs gemeten in test en retest voor de nieuwe OC, gesplitst naar groep.

In een poging dit leereffect, dat voor de nieuwe Oorcheck ongeveer de twee keer groter is dan het leereffect voor de originele OC, te verklaren wordt er gekeken naar de meest betrouwbare testuitslagen. Per test is naast de SRT ook de standaarddeviatie van de adaptieve procedure van de test berekend. Dit zegt iets over de schommeling van de snr rond de SRT. Hoe kleiner deze sd, des te consistentere de uitvoering van de test en des te betrouwbaarder is de meting.

Om te bekijken of het grotere leereffect veroorzaakt wordt door onbetrouwbaardere testresultaten wordt deze interne sd gebruikt als criterium om de minst betrouwbare resultaten buiten de analyse te laten. Hiervoor wordt het 90^e percentiel van de sd's als grenswaarde genomen, dat is 3.0 dB. Dit betekent dat 10% van de testen en interne sd had die hoger was dan

3.0 dB. Het verschil tussen test en retest wordt opnieuw berekend voor personen die voor beide testen een sd hadden ≤ 3.0 dB. Dit verschilt echter weinig met de vorige bevinding, het gemiddelde leereffect is voor deze populatie 1.56 dB. De grootte van het leereffect is dus niet te wijten aan minder betrouwbare testresultaten.

Test-retest betrouwbaarheid van de nieuwe Oorcheck

In figuur 5 is de scatterplot gegeven voor de correlatie tussen de resultaten van test en retest; de correlatie tussen de resultaten van meting 1 en 2 is iets groter dan voor de originele OC. De sd_{intra} is echter weer iets groter voor de nieuwe Oorcheck dan voor de originele OC. De parameters voor test retest betrouwbaarheid, zoals weergegeven in tabel 10, geven een wisselend beeld. Er is dus geen eenduidig bewijs dat de nieuwe test betrouwbaarder is dan de originele versie.

Correlatieanalyse met toonaudiogram en SRT gemeten met Plompzinnen.

De SRT-data gemeten met beide varianten van Oorcheck, en de twee andere bestaande testen, worden vergeleken met de gouden standaard voor spraakverstaan in ruis; de SRT in stationaire ruis gemeten m.b.v. Plompzinnen. Dit gebeurt door middel van een correlatieanalyse. Ook wordt de correlatie van de SRTs van de bestaande testen met het toonaudiogram bekeken, per frequentie en voor een aantal toondrempel-gemiddelden (PTA's). In figuur 6 staan de gevonden correlatiecoëfficiënten van de verschillende testen als functie van audiogramfrequentie. In deze figuur zijn ook de correlatiecoëfficiënten van de SRT gemeten met Plompzinnen met de gehoordrempels weergegeven. De overige correlatiecoëfficiënten zijn in tabel 11 weergegeven.

	Plomp SRT	PTA₅₁₂	PTA₂₄	PTA₃₄₆
NHT	0.77	0.72	0.74	0.69
BOC	0.76	0.67	0.67	0.66
OC	0.65	0.67	0.64	0.62
OC nieuw	0.72	0.72	0.75	0.76

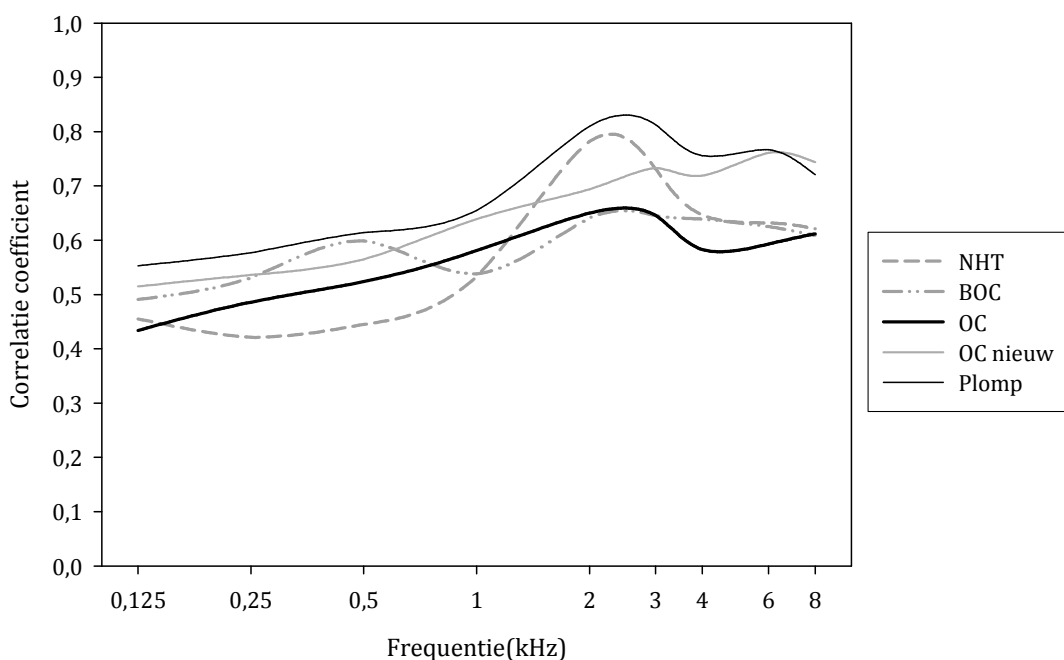
Tabel 11: *Correlatiecoëfficiënten voor de bestaande testen met de SRT in stationaire ruis gemeten met Plompzinnen en gemiddelde gehoordrempels, op basis van de eerste meting.*

Uit deze correlatieanalyse blijkt dat de bestaande testen redelijk goed correleren met SRT in stationaire ruis gemeten met Plompzinnen en met de verschillende PTA-waarden. De correlaties met de verschillende parameters zijn het laagst voor de originele Oorcheck. Bovendien vertonen de bestaande testen de minst sterke correlatie met het gemiddelde van de hoogfrequente, lawaaigevoelige, gehoordrempels van 3, 4 en 6 kHz.

De nieuwe Oorcheck vertoont hogere correlaties met de SRT gemeten met Plompzinnen en met de verschillende PTA-waarden, dan de originele OC. Bovendien vertonen deze testresultaten de beste relatie met de PTA₃₄₆ van alle testen.

Dat blijkt ook uit figuur 6, de resultaten van de herkalibreerde Oorcheck correleren over het hele frequentiegebied genomen beter met de gehoordrempels dan de resultaten van de originele versie van de test. Deze figuur laat ook zien dat de resultaten van de BOC op vergelijkbare wijze correleren met de audiogramdata als de resultaten van de originele Oorcheck. De resultaten van

de NHT vertonen een minder sterke correlatie met de gehoordrempels op de lage frequenties dan de resultaten van de andere bestaande testen. De mate van correlatie neemt echter toe met toenemende frequentie, en is het hoogst van alle testen voor de gehoordrempels gemeten op 2 en 3 kHz. Op 4 en 6 kHz zijn de correlatiecoëfficiënten van de NHT-resultaten met de toondrempels vergelijkbaar met de overige originele testen. De beste correlatie van SRT-data en gehoordrempels wordt echter verkregen met Plompzinnen in stationaire ruis.



Figuur 6: Correlatiecoëfficiënten van de bestaande spraak-in-ruistesten met de gehoordrempels gemeten op de verschillende audiogramfrequenties.

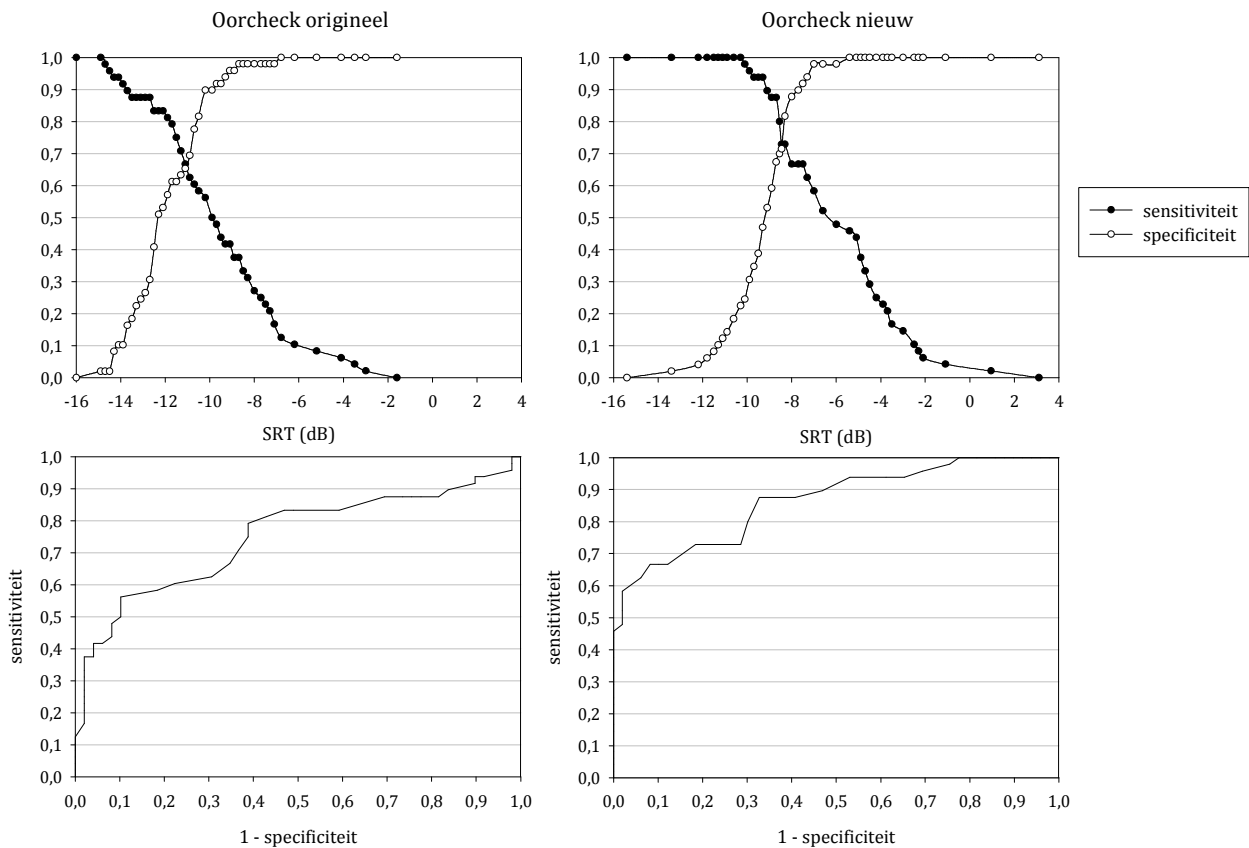
4.3 NORMERING VAN DE NIEUWE OORCHECK

De aanpassingen van de vernieuwde Oorcheck leveren ook andere referentiewaarden op. Deze kunnen teruggevonden worden in de psychometrische functie van normaalhorenden, gegeven in figuur 4. Doordat deze nieuwe OC resulteert in andere, hogere, gemiddelde SRTs voor de verschillende groepen dan de originele versie, zullen er ook andere afkappunten gekozen moeten worden. Deze kunnen worden gevonden door middel van een ROC analyse. Deze analyse kan voor elke uitslag de sensitiviteit en de specificiteit van de testen berekenen. Op basis hiervan kan een afkappunt gekozen worden voor de nieuwe Oorcheck om onderscheid te maken in normaal- en slechthorenden.

De uitkomst van deze analyse is weergegeven in figuur 7, waarbij zowel de verschillende waarden voor sensitiviteit en specificiteit voor de verschillende SRT-uitkomsten als de daadwerkelijke ROC curve zijn weergegeven. Ter vergelijking zijn deze figuren ook weergegeven voor de oorspronkelijke Oorcheck.

De beste waarde voor het afkappunt is -8.7 dB. Met dit afkappunt heeft de nieuwe Oorcheck, zoals gemeten in deze populatie, een sensitiviteit van 87.8% en een specificiteit van 67.3%. Dat is een sterke verbetering van de sensitiviteit van de originele Oorcheck, waarvoor het afkappunt,

-10 dB, op basis van de in figuur 7 gepresenteerde analyse (linksboven) iets te laag gekozen lijkt. De specificiteit is echter iets lager voor de nieuwe test dan voor de originele Oorcheck.



Figuur 7: Boven: Sensitiviteit en specificiteit tegen elkaar uitgezet. Onder: ROC curven. Het linkerpaneel geeft deze gegevens weer voor de originele Oorcheck, het rechterpaneel voor de nieuwe Oorcheck

4.4 CONCLUSIES T.A.V. OPTIMALISATIE SPRAAKMATERIAAL

De nieuwe Oorcheck vindt - na optimalisatie van het woordmateriaal - een groter verschil tussen normaal- en slechthorenden. Bovendien correleert deze nieuwe test beter met zowel de gouden standaard (SRT Plomp) als het toonaudiogram dan de originele versie.

Echter het gevonden leereffect voor de nieuwe Oorcheck is groter dan voor de originele test. Ook is de test-retest betrouwbaarheid vergeleken met de originele test niet duidelijk verbeterd, maar zijn beide mate voor de betrouwbaarheid vergelijkbaar. Het zou kunnen zijn dat de combinatie met de geoptimaliseerde ruis ervoor zorgt dat ook de verstaanbaarheid van de afzonderlijke woorden weer iets veranderd is. Dit kan nagegaan worden door voor de verkregen resultaten opnieuw woordspecifieke discriminatiefuncties te berekenen, zoals in beschreven in §4.1, en op basis daarvan de RMS van de woorden aan te passen. De sensitiviteit van de vernieuwde Oorcheck is wel duidelijk verbeterd ten opzichte van de oorspronkelijke test. Door de sterke correlatie met de gehoordrempels en de gouden standaard en de sterk verbeterde sensitiviteit, lijkt de Oorcheck met het gehomogeniseerde spraakmateriaal, ondanks het grotere leereffect, een verbetering te zijn van de originele versie.

5 EFFECTEN VAN RUISMODIFICATIES OP DE TEST GEVOELIGHEID

In het vorige hoofdstuk hebben we gekeken naar het effect van de optimalisatie van het spraakmateriaal op de testgevoeligheid voor NIHL. Een andere methode om de Oorcheck gevoeliger te maken voor beginnende lawaaislechthorendheid is het filteren van de gebruikte stimuli.

5.1 EXPERIMENT 2B: OPTIMALISATIE VAN HET RUISMATERIAAL

In de huidige versie van Oorcheck wordt breedbandige stationaire ruis gebruikt, maar men kan ook de ruis bewerken om meer onderscheid tussen normaal- en slechthorenden te verkrijgen. Dit kan op verschillende manieren gedaan worden, waarbij onderscheid gemaakt kan worden in twee aspecten:

- temporeel: het is bekend dat normaalhorenden een significant lagere SRT behalen in fluctuerende ruis dan in continue ruis. Dit verschil in SRT wordt bij slechthorenden niet in dezelfde mate gevonden. Slechthorenden kunnen door verbreding van auditieve filters en meer voorwaartse maskering minder goed gebruik maken van perioden zonder maskeerruis dan dat normaalhorenden dat kunnen. Smits toonde aan dat bij gebruik van fluctuerende ruis met een blokmodulatie van 16 Hz in de Nationale Hoortest de verschillen in SRT tussen normaalhorenden en slechthorenden groter waren dan bij continue ruis (Smits 2007).
- spectraal: door spectraal te filteren kan de proefpersoon gedwongen worden om de spraakinformatie uit een bepaald frequentiegebied te halen. In dit geval kan men de ruis zo filteren dat de spraakinformatie vooral uit het gebied van de hogere frequenties moet worden gehaald. Op die manier zou er een groter verschil moeten ontstaan tussen normaalhorenden en personen met NIHL.

Voor beide manieren van modificeren geldt dat er delen van de ruis verwijderd worden. Wordt de SRT gemeten in een breedbandige, continue ruis dan is er sprake van een één-op-één verhouding van spraak en ruis en is de test onafhankelijk van de intensiteit van de aanbieding. De uitkomsten van de testen met 'gaten' in de ruizen kunnen echter wel afhankelijk zijn van de intensiteit; de spraak kan lokaal harder worden zonder dat de maskering op dat punt in dezelfde mate toeneemt. Achtergrondruis zou dan een rol kunnen gaan spelen in het gedeelte van de spraak dat niet meer door ruis gemaskeerd wordt. Is het niveau van de achtergrondruis hoog, dan kan deze de gehoordrempel benaderen. Hierdoor kan de uitslag van de test beïnvloed worden. Daarnaast heeft de modulatie diepte ook invloed op de SRT gemeten in fluctuerende ruis. Om de invloed hiervan zo klein mogelijk te maken worden door middel van toevoeging van een breedbandige ruisvloer op een vast niveau de achtergrondruis en modulatie diepte voor iedereen op een zelfde drempelniveau gehouden.

Om te voorkomen dat het aantal condities dat getest wordt bij de onderzoekspopulatie te groot wordt, worden vooraf de juiste parameters voor de maskeerruizen gekozen, op basis van SII-voorspellingen en het uitvoeren van verschillende pilotmetingen. Dit zal uiteindelijk resulteren in een experimentele set van zes ruizen. Hiervoor moeten eerst de ruizen samengesteld worden.

SAMENSTELLEN VAN HET RUISMATERIAAL

De experimentele ruizen worden gemaakt door de ruis met hetzelfde gemiddelde spectrum als dat van het spraakmateriaal (LTASS) te modificeren (zie §4.1). Alle ruizen worden eerst ten opzichte van deze breedbandige stationaire LTASS-ruis van de vernieuwde Oorcheck geschaald, en daarna worden ze gefilterd en/of temporeel gemoduleerd.

De temporele fluctuatie van de ruis wordt bereikt door 16 Hz blokmodulatie. De andere ruizen worden gemaakt door spectrale filtering, met ofwel een laag- dan wel hoogdoorlaat 2000^e orde FIR filter. Deze filters kunnen gebruikt worden om zowel de continue als de fluctuerende ruis te filteren. Er kunnen dan bijvoorbeeld ook combinaties gemaakt worden van een Low-pass gefilterde continue ruis met een High-pass gefilterde fluctuerende ruis. Aan alle ruisvormen wordt een ruisvloer toegevoegd. Deze ruisvloer wordt gevormd door de breedbandige continue LTASS-ruis te verzwakken tot het benodigde niveau.

De op deze manier geconstrueerde ruizen kunnen vervolgens gebruikt worden voor de voorspellingen met behulp van het SII-model.

Voordat de ruizen daadwerkelijk aan de internetttest toegevoegd kunnen worden, wordt er een stuk van 2 seconde uit een langere file gekopieerd. In de eerste 0.5 seconde neemt vanaf 0 dB de intensiteit steeds meer toe (fade in) en in de laatste 0.5 seconde neemt de intensiteit weer tot 0 dB af (fade out).

Bij het ontwikkelen en testen van deze verschillende gemodificeerde ruiscondities, bleek dat met dit ruismateriaal een lagere SRT gemeten kon worden. In de testen wordt het spraaksignaal in stappen van 2 dB verzwakt. Echter, door beperkte dynamiek van de gebruikte flashapplicatie in de internettoepassing verliep deze verzwakking niet meer lineair voor signaal-ruis verhoudingen kleiner dan -20 dB. Om dit probleem te omzeilen wordt het spraakmateriaal voor de experimentele Oorcheck-versies vooraf softwarematig met 10 dB verzwakt. De dynamiek van de testen waarin de verzwakking van het spraakmateriaal nog lineair verloopt wordt hierdoor van 4 t/m -20 dB verschoven naar -6 tot -30 dB. Bij het berekenen van de drempel wordt deze verschuiving uiteraard weer gecorrigeerd.

Na toevoeging van de verschillende ruizen aan de internetttesten, worden er pilotmetingen uitgevoerd, om uiteindelijk tot een set van zes verschillende experimentele condities te komen, die vervolgens bij de onderzoekspopulatie afgenomen worden.

SII VOORSPELLINGEN

Het effect van de spectrale filteringen van de maskeerruis kunnen we voorspellen aan de hand van het SII-model. Dit kan ons helpen om vooraf de juiste criteria voor de maskeerruizen te bepalen. Deze voorspelling zijn weergegeven in bijlage III.

De SII voorspellingen worden gedaan voor audiogrammen met een verschillende mate van NIHL. Het model voorspelt weinig verschil tussen deze verschillende audiogrammen voor de breedbandige continue ruis. Voor een ruis met een notch rond 4 kHz van één octaafband zijn de voorspellingen niet veel anders, daarom wordt er van deze conditie afgezien.

Wanneer de gehele ruis Low-pass (LP) gefilterd zou worden geeft dit wel grote verschillen voor de verschillende audiogramdata. Het grootste verschil wordt gevonden voor een conditie met afsnijfrequentie van 1.4 kHz en een ruisvloer op -35 dB. Hierbij wordt echter een erg lage SRT

voorspeld voor degene met een (bijna) normaal audiogram, wat wel eens té laag zou kunnen zijn om met deze test betrouwbaar te meten bij een internettoepassing in een niet goed gecontroleerde testomgeving.

Wanneer de ruisvloer iets hoger gelegd wordt, op -20, wordt deze voorspelde SRT hoger, maar wordt ook het verschil tussen de personen kleiner. Hetzelfde geldt voor de voorspelde SRTs wanneer de afsnijfrequentie wordt verhoogd naar 1.6 of 1.8 kHz. Toch lijken dit condities te zijn die voldoende kunnen discrimineren tussen een verschillende mate van lawaaischade. Deze condities zullen door middel van pilotmetingen getest worden bij normaalhorende proefpersonen.

FILTERING VAN HET SPRAAKSIGNAAL

Een andere optie om de gevoeligheid van een spraak-in-ruis test voor NIHL te verhogen is om het spraakmateriaal zodanig te filteren dat de nadruk zal komen te liggen op de hogere frequenties, daar waar de gehoorfunctie het eerst verminderd wordt als gevolg van blootstelling aan lawaai. Ook hiervoor worden de resultaten voorspeld door middel van het SII model. Op basis van deze voorspellingen wordt er echter slechts een klein verschil tussen normaal- en slechthorenden verwacht.

Een andere mogelijkheid om de test aan te passen is met een plaatselijke versterking van spraak en ruis in het gebied 4-6 kHz. Een nadeel van deze methode is dat er nog steeds te veel spraakinformatie aanwezig is in de lage frequenties, vooral bij de vocalen. Bovendien gaat de spraak snel kunstmatig of onprettig klinken en daarom zal een dergelijke versterking ook niet echt een meerwaarde geven.

Omdat er slechts kleine effecten van spraakfiltering op het onderscheiden van normaal- en slechthorenden verwacht worden, wordt deze optie verder niet meer in het onderzoek meegenomen.

PILOTMETINGEN

Bij een kleine groep normaalhorenden (n=10) worden in het AMC verschillende Oorcheck-varianten met experimentele ruizen afgenomen, om de verschillende condities in de praktijk te testen (afstudeeronderzoek Celine Jacobs, 2008). De details staan beschreven in bijlage IV. Verwacht wordt dat een alternatieve test met gefilterde maskeerruis resulteert in een lagere SRT. Deze moet echter wel betrouwbaar via internettoepassing te testen zijn, en moet daarom ook niet resulteren in een te lage SRT.

Het doel van de pilottesten is om met verschillende parameters van ruisfiltering te variëren om de invloed van deze verschillende parameters op de testuitslag te onderzoeken. Op basis hiervan kan uiteindelijk de set van experimentele ruizen vastgesteld worden om in de onderzoekspopulatie te testen.

Het SII model voorspelt dat de meeste experimentele ruizen veel minder maskeren dan de breedbandige stationaire Oorcheck ruis. De uitslagen van de meeste testen met gefilterde ruizen zullen ongeveer 10-15 dB lager liggen dan de SRT gemeten met de OC met LTASS-ruis. Sommige varianten blijken echter zo veel minder maskerend te zijn dat de SRT-drempel tegen de begrenzing van de test aanliep. Hierdoor vormen de eigenschappen en mogelijkheden van de testen een beperkende factor. Ook gaan bij lagere snr beperkingen van de verschillende

computersystemen en gehoordrempels uiteindelijk een grotere rol spelen dan de daadwerkelijke spraakverstaanbaarheid van de respondent. Het is dus zaak dat de signaal-ruis verhoudingen binnen de test niet te laag worden, maar wel op een dusdanige manier dat de spreiding van de testresultaten groot genoeg is om voldoende onderscheid tussen normaal- en slechthorenden te maken.

Hiertoe wordt er in de pilot gevarieerd met de volgende parameters van de ruiscondities:

- duty cycle van de fluctuerende ruis. Vaak wordt er uitgegaan van een ruis met fluctuaties met een duty cycle van 50%, wat betekent dat er achtereenvolgens evenveel ruis als stilte is. Wanneer deze duty cycle verhoogd wordt tot 60%, bestaat elke fluctuatie voor 60% uit ruis en 40% uit stilte;
- afsnijfrequentie van de gefilterde ruis. De voorspellingen van het SII model laten zien dat er een hogere SRT verwacht wordt wanneer de afsnijfrequentie toeneemt tot 1.6 of 1.8 kHz.
- het niveau van de ruisvloer. De SII voorspellingen geven aan dat er een hogere SRT verwacht wordt wanneer de ruisvloer in de gefilterde ruisconditie verhoogd wordt.

De parameters die in verschillende ruiscondities met pilottesten bij normaalhorenden worden gemeten zijn in tabel 12 weergegeven.

De bevindingen van de pilotmetingen laten in de breedbandige fluctuerende ruizen slechts een klein verschil zien in SRT tussen de ruizen met duty cycle 50/50 en met 60/40. De conventionele en meest beschreven methode is die waarbij de duty cycle van de ruis 50/50 is en daar zullen we dan ook in deze studie vanuit gaan.

<i>temporeel</i>	<i>spectrum</i>	<i>afsnijfreq</i>	<i>duty cycle</i>	<i>ruisvloer</i>
continu	BB	-	-	-
	BB	-	-	-
	BB	-	-	-
	BB	-	-	-
	LP	1.4	-	-15
	LP	1.4	-	-20
	LP	1.4	-	-35
	LP	1.6	-	-35
	LP	1.8	-	-20
	LP	1.8	-	-35
	HP	1.4	-	-20
fluctuerend	BB	-	50/50	-35
	BB	-	60/40	-20
	BB	-	50/50	-15
	BB	-	50/50	-20
	LP	1.4	50/50	-15
	LP	1.4	50/50	-20
	LP	1.4	50/50	-35
	LP	1.4	60/40	-20

Tabel 12: Variaties in parameters van de experimentele ruiscondities. BB = breedbandig, LP = low-pass filtering, HP = high-pass filtering. Afsnijfrequentie wordt gegeven in kHz en voor de ruisvloer wordt de verzwakking in dB(A) gegeven.

Tussen de Low-pass ruis met afsnijfrequentie 1.4 kHz en die met afsnijfrequentie 1.6 en 1.8 kHz worden evenmin grote verschillen gevonden. Bij de filtering van de experimentele ruizen zal de afsnijfrequentie 1.4 kHz gehanteerd worden. Deze frequentie ligt in het midden van het spraakspectrum volgens de SPIN-weging en vindt volgens de SII voorspellingen de grootste verschillen tussen verschillende audiogrammen.

Tenslotte laten de pilot-metingen zien dat, met name bij de fluctuerende ruizen, bij een modulatie diepte van 15 dB de meest gunstige resultaten behaald worden. Met deze modulatie diepte spelen de beperkingen van de testen geen, of slechts een geringe rol. Bovendien komen de gemiddelde SRT-drempels iets hoger uit. Daarom zal er in het experiment uitgegaan worden van een modulatie diepte van 15 dB.

5.2 MEETPROTOCOL

In de proefopzet van experiment 2B worden, naast de bovengenoemde ruiscondities, extra condities toegevoegd om alle opties te combineren, namelijk een High-pass gefilterde continue ruis en een Low-pass gefilterde fluctuerende ruis gecombineerd met een High-pass gefilterde continue ruis. Deze ruizen zijn niet in de pilot meegenomen, omdat hierbij in eerste instantie weinig verschil wordt verwacht tussen normaalhorenden en personen met NIHL. Immers, beide groepen kunnen ongeveer even goed gebruik maken van de voor spraak essentiële informatie in de lagere frequenties. Voor screeningsdoeleinden zijn deze extra condities daarom iets minder van belang, maar voor het fundamentele inzicht zijn deze wel belangrijk.

De uiteindelijke set experimentele ruiscondities staat in tabel 13 weergegeven. De zes ruiscondities van experiment 2B onderscheiden zich van de herkalibreerde Oorcheck als volgt:

- 16 Hz: de ruis is gemoduleerd met een 16 Hz blokmodulatie (50% duty cycle)
- LP: de ruis is Low-pass gefilterd bij 1.4 kHz (2000^e orde FIR filter)
- LP 16: de ruis is LP-gefilterd en bevat HP een 16-Hz blokmodulatie
- HP: de ruis is High-pass gefilterd bij 1.4 kHz (2000^e orde FIR filter)
- HP 16: de ruis is HP-gefilterd en bevat LP een 16-Hz blokmodulatie
- Stil (-15): breedbandige continue ruis, met 15 dB verzwakking (gelijk aan ruisvloer)

Alle experimentele ruiscondities bevatten een continue breedbandige ruisvloer op 15 dB afstand van de actieve delen van de (primaire) ruis.

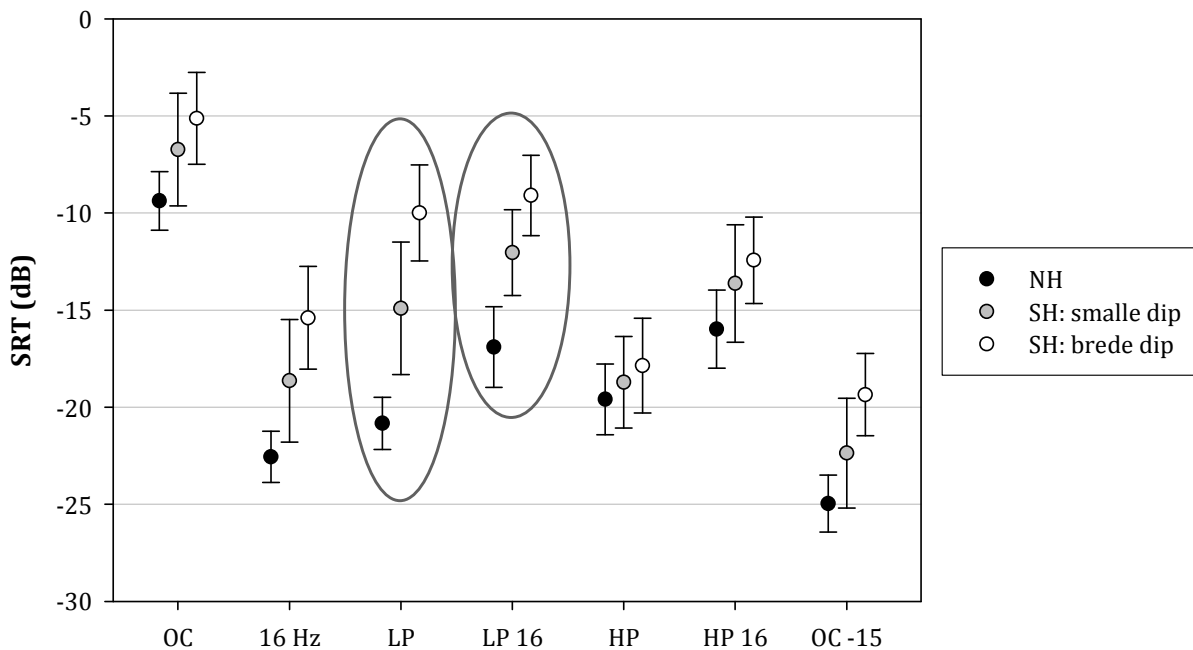
<i>test ruis</i>	<i>filtering</i>	<i>modulatie</i>	<i>ruisvloer</i>
Oorcheck	-	-	-
16 Hz	-	16 Hz	-15 dB
LP	LP 1.4 kHz	-	-15 dB
LP 16	LP 1.4 kHz	16 Hz	-15 dB
HP	HP 1.4 kHz	-	-15 dB
HP 16	HP 1.4 kHz	16 Hz	-15 dB
OC -15	-	-	-15 dB

Tabel 13: Overzicht van de specificaties van de definitieve set van experimentele ruizen.

De zes experimentele Oorcheck-varianten uit tabel 13 met gemodificeerde ruizen en het gehomogeniseerde spraakmateriaal worden allemaal monoraal getest, met behulp van de opstelling zoals eerder beschreven in §2.4. Alle testen worden twee keer afgenomen, gescheiden door een korte pauze. De volgorde van de testen is gebalanceerd volgens een 6x6 Latijns vierkant (Wagenaar 1969). De testen worden afgenomen met een vast ruisniveau dat voor de onbewerkte ruis wordt vastgesteld op 20 dB boven SRT gemeten met Plompzinnen in stilte, met een minimum van 65 dB(A). De andere testen worden afgespeeld met dezelfde instellingen, welke overeenkomen met 65 dB(A) voor de onbewerkte ruis, maar het eigenlijke ruisniveau kan - afhankelijk van de soort ruis - hier enigszins van afwijken. Hoe groot deze afwijking is, is met de kalibratie van de testen nagegaan. De uiteindelijke resultaten worden dan ook weergegeven als signaal-ruis verhouding, relatief ten opzichte van de onbewerkte Oorcheck.

5.3 RESULTATEN VAN EXPERIMENT 2B

Om de gevoeligheid van de Oorcheck voor lawaaislechthorendheid te verbeteren, is de maskeerruis van deze test dus op verschillende manieren aangepast. Deze testen zijn door de onderzoekspopulatie uitgevoerd in test en retest. De gemiddelde SRTs van de eerste meting van de verschillende testen, met bijbehorende spreiding, worden in figuur 8 weergegeven, voor de drie subgroepen: NH, SH-SD, and SH-BD.



Figuur 8: Gemiddelde SRT per test, voor de drie groepen. De error bars geven één standaarddeviatie weer.

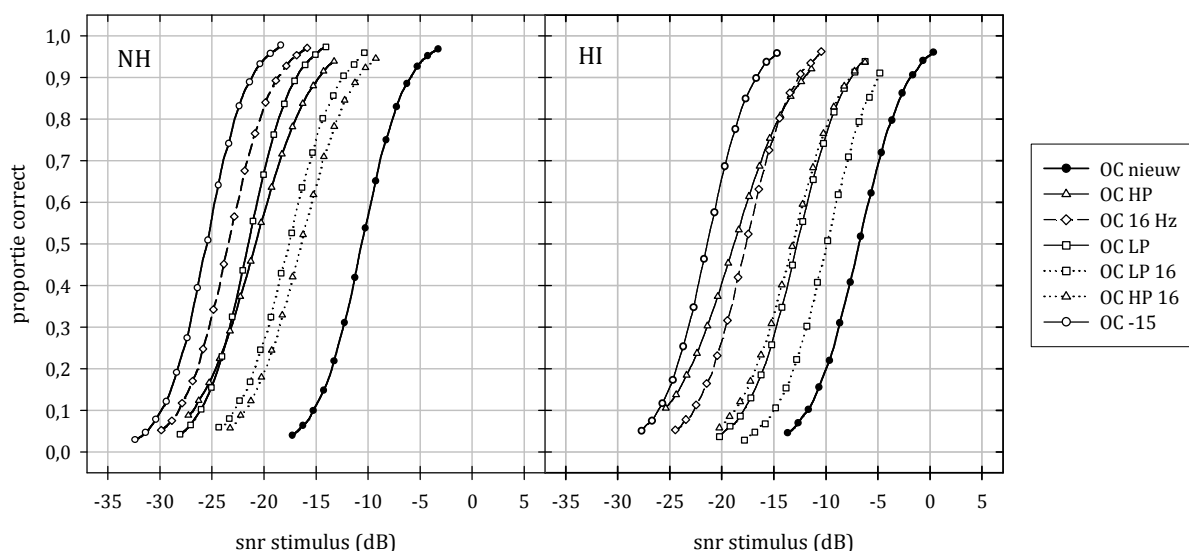
De verschillen tussen de resultaten van de drie subgroepen

Met behulp van een variantieanalyse worden de gemiddelde SRTs van de verschillende groepen met elkaar vergeleken. De gemiddelde verschillen, met bijbehorend significantieniveau worden in tabel 14, beschreven en getoetst. Alle testen vinden een significant verschil tussen normaalhorenden en beide groepen slechthorenden, behalve de Oorcheck met High-pass gefilterde ruis, waarvoor alleen een kleine afwijking wordt gevonden voor de gemiddelde SRT van de slechthorenden met een brede dip ten opzichte van de SRT van normaalhorenden.

Ook tussen beide groepen van slechthorenden worden met de meeste testen significante verschillen gevonden, waarbij de deelnemers met een brede dip hogere SRTs hebben dan deelnemers met een smalle dip. Deze verschillen zijn echter niet significant voor de condities met High-pass gefilterde ruis. De testen met Low-pass ruis en met fluctuerende ruis vertonen de grootste verschillen tussen alle groepen onderling.

Test	ANOVA		$\Delta NH - SH_{SD}$		$\Delta NH - SH_{BD}$		$\Delta SH_{SD} - SH_{BD}$	
	F	p	gem	p	gem	p	gem	p
OC nieuw	32.7	0.000	-2.6	0.000	-4.2	0.000	-1.6	0.035
OC 16 Hz	85.4	0.000	-3.9	0.000	-7.2	0.000	-3.3	0.000
OC LP	162.0	0.000	-5.9	0.000	-10.8	0.000	-4.9	0.000
OC LP - 16	93.5	0.000	-4.9	0.000	-7.8	0.000	-2.9	0.000
OC HP	5.6	0.000	0.9	0.285	-1.7	0.004	-0.9	0.477
OC HP - 16	20.5	0.000	-2.3	0.000	-3.5	0.000	-1.2	0.238
OC -15	61.8	0.000	-2.6	0.000	-5.6	0.000	-3.0	0.000

Tabel 14: Uitkomst van de uitgevoerde ANOVA met groep als factor en SRT van meting 1 als uitkomst. Per test is het gemiddelde verschil tussen de verschillende groepen gegeven, met bijbehorende p-waarden, na Bonferonni correctie



Figuur 9: PI curven van de zes testen met experimentele ruizen en OC nieuw, met proportie correct als functie van de signaal-ruis verhouding, relatief t.o.v. individuele SRT, gecorrigeerd naar gemiddelde SRT van de test. Om te voorkomen dat de aantallen per test te klein zouden worden om de psychometrische curve betrouwbaar te presenteren, zijn beide groepen slechthorenden als één groep weergegeven.

De bijbehorende PI-curve van de experimentele testen worden weergegeven in figuur 9. De curve voor de slechthorenden liggen iets meer naar rechts door de iets hogere gemiddelde SRT. Ook liggen de curve niet in dezelfde volgorde voor de slechthorende deelnemers vergeleken met de normaalhorenden, omdat zij niet in gelijke mate profijt hebben van een verminderde maskering van een aantal ruisen, zoals bij de Low-pass ruisen. De steilheid van de verschillende curve is echter vergelijkbaar, alleen de curve van Oorcheck met High-pass gefilterde ruis verloopt iets slapper dan de overige curve.

Leereffecten in de experimentele testen

Ook voor deze testen worden de verschillende testkarakteristieken, zoals het leereffect en de test-retestbetrouwbaarheid, berekend. De gemiddelde SRTs van de verschillende test en retest metingen en de andere testkarakteristieken zijn per groep weergegeven in tabel 15.

test	groep	meting 1 gem (sd)	meting 2 gem (sd)	gem. verschil	p verschil	kritisch verschil	sd intra	correlatie test-retest
OC 16	NH	-22.6 (1.3)	-23.2 (1.3)	0.70	0.001	1.12	1.45	0.86
	SH-S	-18.6 (3.2)	-19.6 (4.0)					
	SH-B	-15.4 (2.6)	-16.1 (2.8)					
OC LP	NH	-20.8 (1.3)	-21.3 (1.7)	0.69	0.000	1.06	1.25	0.93
	SH-S	-14.9 (3.4)	-17.2 (1.9)					
	SH-B	-10.0 (2.5)	-9.8 (2.8)					
OC LP 16	NH	-16.9 (2.1)	-17.8 (1.8)	0.48	0.028	0.90	1.39	0.87
	SH-S	-12.0 (2.2)	-12.2 (2.4)					
	SH-B	-9.1 (2.1)	-8.6 (1.7)					
OC HP	NH	-19.6 (1.8)	-21.0 (2.2)	0.73	0.009	1.28	1.92	0.42
	SH-S	-18.7 (2.4)	-18.9 (2.7)					
	SH-B	-17.9 (2.4)	-18.1 (2.7)					
OC HP 16	NH	-16.0 (2.0)	-16.5 (2.2)	0.46	0.073	0.96	1.77	0.60
	SH-S	-13.6 (3.0)	-13.9 (2.8)					
	SH-B	-12.4 (2.2)	-12.9 (2.1)					
OC -15	NH	-25.0 (1.5)	-25.8 (1.5)	0.73	0.000	1.09	1.28	0.83
	SH-S	-22.4 (2.8)	-22.7 (2.7)					
	SH-B	-19.4 (2.1)	-20.3 (2.9)					

Tabel 15: Gemiddelde SRT-waarden en test-retest karakteristieken voor de experimentele testen

Het gemiddelde verschil tussen test en retest laat zien dat er voor alle experimentele testen sprake is van een significant leereffect, behalve voor de High-pass gefilterde ruis gecombineerd met een Low-pass gefilterde fluctuerende ruiscombinatie (OC HP 16). De grootte van het leereffect is echter voor alle testen kleiner dan 1 dB en dus ook kleiner dan het leereffect van de nieuwe Oorcheck. De test met Low-pass gefilterde ruis heeft het kleinste leereffect, waardoor ook het kritische verschil klein is; kleiner dan 1 dB. Voor de overige testen is het kritische verschil vergelijkbaar.

Test-retest betrouwbaarheid van de experimentele testen

De test-retest betrouwbaarheid weergegeven door de sd_{intra} , is redelijk vergelijkbaar voor de testen. Voor de testen waarbij High-pass gefilterde ruis is gebruikt is deze maat iets hoger dan voor de andere testen. Hiervoor is deze maat ongeveer even groot, namelijk 1.3 tot 1.5 dB.

Ook de mate van correlatie tussen de resultaten van meting 1 en meting 2 is minder goed voor de testen met High-pass gefilterde ruis. Voor de overige experimentele testen is deze, significante, correlatie hoog. Voor de test met fluctuerende ruis en vooral voor de test met Low-pass gefilterde ruis is deze correlatie zelfs hoger dan voor de nieuwe Oorcheck.

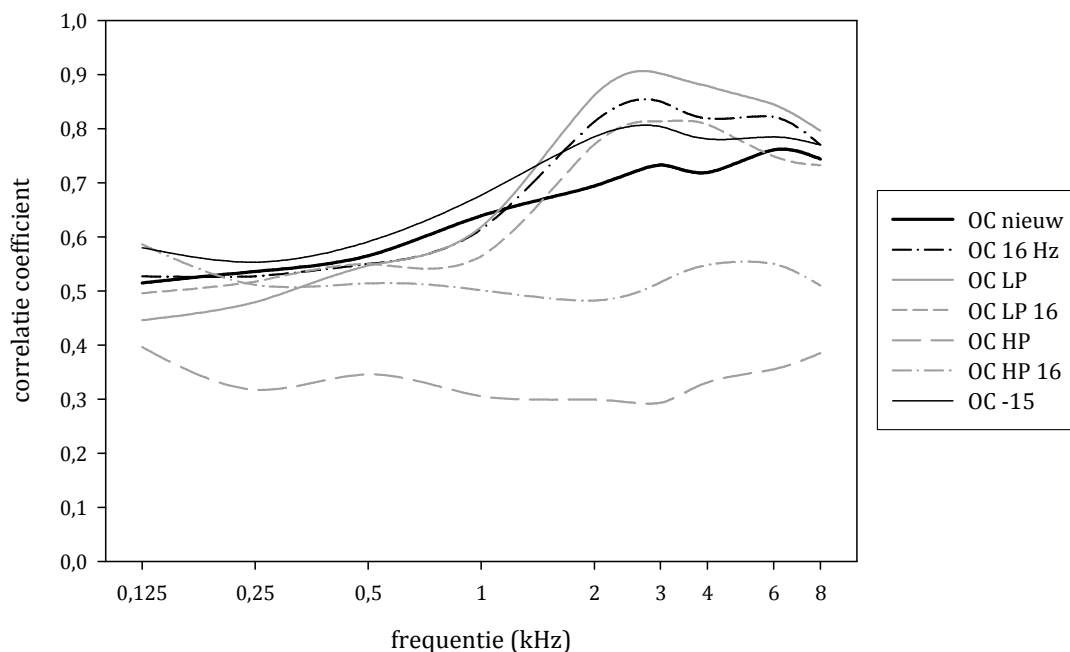
Correlatieanalyse met toonaudiogram en SRT Plompzinnen.

De resultaten van de Oorcheck met experimentele ruiscondities worden vergeleken met de SRT in stationaire ruis gemeten m.b.v. Plompzinnen (gouden standaard), door middel van correlatieanalyse. Ook wordt de correlatie van de SRTs van de verschillende internettesten met het toonaudiogram bekeken, voor een aantal PTA-waarden en per frequentie. De correlatiecoëfficiënten voor SRT met Plompzinnen en PTA-waarden zijn in tabel 16 weergegeven. In figuur 10 staan de gevonden correlatiecoëfficiënten van de verschillende testen als functie van audiogramfrequentie.

	Plomp SRT	PTA₅₁₂	PTA₂₄	PTA₃₄₆
OC 16 Hz	0.81	0.77	0.78	0.86
OC LP	0.78	0.71	0.81	0.92
OC LP 16	0.72	0.67	0.75	0.84
OC HP	0.40	0.39	0.34	0.34
OC HP 16	0.60	0.62	0.55	0.56
OC -15	0.79	0.77	0.79	0.82

Tabel 16: *Correlatiecoëfficiënten voor de SRT van de eerste meting de SRT Plomp en PTA, voor de verschillende Oorchecks met experimentele ruizen.*

Ook uit deze correlatieanalyse blijkt dat vooral de testen met Low-pass gefilterde ruizen en de test met fluctuerende ruis goed correleren met SRT in stationaire ruis gemeten met Plompzinnen, en met de verschillende maten voor gehoorverlies. Uit figuur 10 blijkt dat deze testen ook beter correleren met de lawaaigevoelige frequenties dan de herkalibreerde Oorcheck met breedbandige stationaire ruis.



Figuur 10: Correlatiecoëfficiënten van de SRT met de gehoordrempels gemeten op de verschillende audiogramfrequenties voor de verschillende Oorchecks met experimentele ruizen

5.4 CONCLUSIE T.A.V. DE RUISMODIFICATIES

Voor de meeste testen worden er significante verschillen in SRT gevonden tussen normaalhorenden en beide groepen slechthorenden, behalve voor de testen met High-pass gefilterde ruiscondities. De Oorcheck-varianten met Low-pass gefilterde ruis vertonen de grootste verschillen tussen de drie onderzoekspopulaties. Ook voor de fluctuerende ruis worden significante verschillen tussen de groepen gevonden, maar de spreiding binnen de groepen is hier groter dan in de testresultaten van de Low-pass gefilterde ruizen.

Alle experimentele Oorcheck-varianten, met uitzondering van de HP 16, vertonen een significant maar klein leereffect (< 1 dB). Dit leereffect is ook kleiner dan de in hoofdstuk 4 gevonden waarden voor de nieuwe Oorcheck. De test-retest betrouwbaarheid van de testen met experimentele ruisconditie is vergelijkbaar. Deze waarden zijn voor alle testen ongeveer gelijk, maar zijn het minst gunstig voor de condities waarbij High-pass gefilterde ruis gebruikt is.

Deze condities met High-pass gefilterde ruis vertonen een lage correlatie met het toonaudiogram, welke voor alle frequenties ongeveer hetzelfde is. De beste correlatie tussen de testuitslag en de gemeten gehoordrempels wordt gevonden voor de Oorcheck met Low-pass gefilterde ruis, juist voor de frequenties die relevant zijn voor NIHL.

Deze testen met Low-pass gefilterde ruis maken een beter onderscheid tussen normaalhorenden en de groepen met een verschillende mate van gehoorverlies en lijken daarom gevoeliger voor NIHL en dus beter bruikbaar bij het screenen op lawaaislechthorendheid.

5.5 NORMERING VAN DE EXPERIMENTELE OORCHECKS

De bovenstaande analyses laten zien dat Oorcheck met een Low-pass gefilterde maskeerruis, al dan niet gecombineerd met een fluctuerende ruis in de hoge frequenties, beter onderscheid kan maken tussen normaalhorenden en personen met een verschillende mate van slechthorendheid dan de oorspronkelijke versie van de Oorcheck.

Deze testen dienen echter eerst nog genormeerd te worden, voordat deze als alternatief voor de huidige Oorcheck gebruikt kunnen worden. Daarom wordt er ook voor deze testen een ROC analyse uitgevoerd, op basis waarvan de afkappunten voor de nieuwe Oorcheck-varianten gekozen kunnen worden om het juiste onderscheid te maken tussen normaal- en slechthorenden.

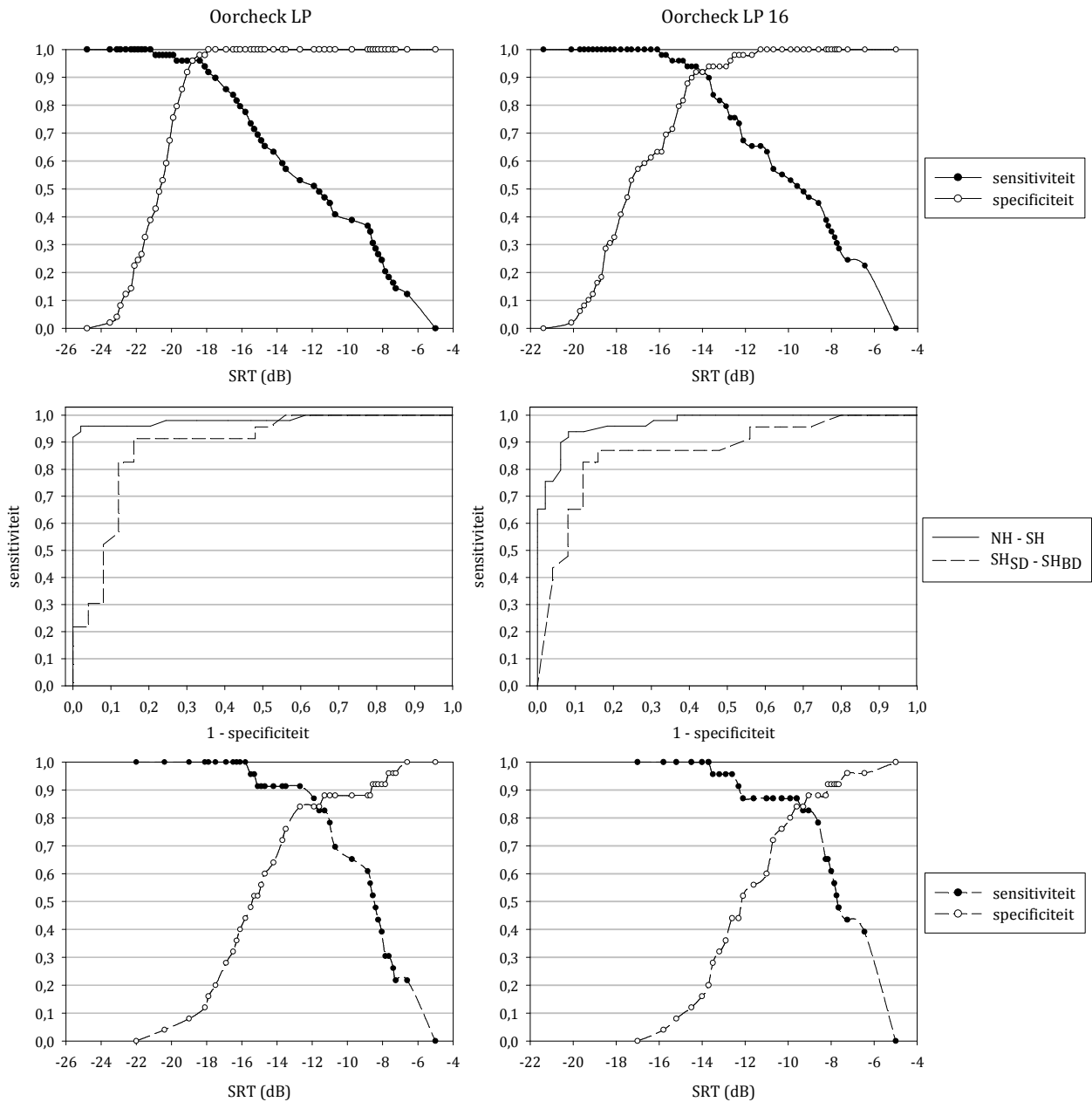
De bestaande versies van interactieve hoortesten kennen meerdere uitkomstcategorieën. Om datzelfde voor deze testen te bewerkstelligen, wordt er bij deze analyse niet alleen gezocht naar het afkappunt om onderscheid te maken tussen normaal- en slechthorenden, maar ook naar een geschikt afkappunt om onderscheid te maken naar de mate van slechthorendheid. Hiertoe wordt dezelfde ROC-procedure nogmaals herhaald, maar dan voor beide groepen slechthorenden om naar een afkappunt te zoeken dat onderscheid te maakt tussen deelnemers met een smalle dip en deelnemers met een brede dip in het audiogram. Dit resulteert uiteindelijk in het vinden van twee afkappunten, en dus in drie uitkomstcategorieën: je gehoor is goed, je gehoor is onvoldoende en je gehoor is slecht.

De resultaten van deze ROC analyse zijn weergegeven in figuur 11, waarbij zowel de verschillende waarden voor sensitiviteit en specificiteit voor de verschillende SRT uitkomsten zijn gegeven. Deze figuren bevatten zowel de resultaten van de analyse om onderscheid te maken tussen normaal- en slechthorenden (zie bovenste panels) als de resultaten om onderscheid te maken tussen slechthorenden met een smalle en een brede dip (zie onderste panels). In de middelste panels worden de ROC curven van deze analyse weergegeven voor het verkrijgen van beide afkappunten voor de twee verschillende testen.

Voor beide testen kan er een afkappunt gekozen met zowel een hoge sensitiviteit als een hoge specificiteit tot gevolg. In onderstaande tabel 17 staan de afkappunten voor beide alternatieve Oorcheck-varianten gegeven, evenals de bijbehorende waarden voor sensitiviteit en specificiteit. Ter vergelijking bevat deze tabel ook de waarden voor de vernieuwde Oorcheck met breedbandige stationaire ruis.

	afkappunt 1	afkappunt 2	sensitiviteit	specificiteit
OC nieuw	-8.7	-6.0	87%	67%
OC LP	-18.4	-12.7	95%	98%
OC LP 16	-14.5	-9.9	92%	90%

Tabel 17: Berekende afkappunten en bijbehorende sensitiviteit en specificiteit van de drie internettesten.



Figuur 11: Boven: Sensitiviteit en specificiteit tegen elkaar uitgezet, wanneer onderscheid gemaakt wordt tussen NH en SH. Midden: ROC curve, de doorgetrokken curve betreft de grenswaarde tussen NH en HI, en de onderbroken curve de grenswaarde tussen beide groepen slechthorenden. Onder: sensitiviteit en specificiteit tegen elkaar uitgezet, wanneer onderscheid gemaakt wordt voor SH met smalle dip en SH met brede dip. Het linkerpaneel geeft de gegevens weer voor OC LP, het rechterpaneel voor OC LP 16.

6 DISCUSSIE

BESTAANDE TESTEN

De hypothese dat Oorcheck in zijn originele vorm ongevoelig zou zijn voor beginnende lawaaislechthorendheid wordt in experiment 1 van dit onderzoek bevestigd. Ongeveer de helft van de gemeten slechthorende deelnemers wordt namelijk door deze spraak-in-ruistest als normaalhorend geclassificeerd. Dit betrof 76% van de slechthorenden met een smalle dip in het audiogram, en 20% van de deelnemers met een brede dip. Wanneer de Oorcheck in de thuissituatie gedaan wordt, is de sensitiviteit van de test iets hoger, maar de specificiteit is daarentegen lager, waardoor nog steeds drie op de tien personen onjuist geclassificeerd worden.

Hierbij moet worden opgemerkt dat Oorcheck in dit onderzoek monuraal en onder gecontroleerde omstandigheden is gemeten, op een vooraf vastgesteld ruisniveau. Bij het online afnemen van de Oorcheck in de praktijk kunnen respondenten het afspeelniveau zelf instellen, waardoor de testen in dit experiment in sommige gevallen op een ander, wellicht lager, afspeelniveau afgenomen zijn dan wat in de thuissituatie het geval zal zijn. Er is echter bij het instellen van het volume wel voor gezorgd dat het afspeelniveau voldoende boven de spraakverstaanbaarheidsdrempel in stilte lag. Verwacht mag worden dat de respondenten daar bij het zelf instellen in de thuissituatie ook op zullen letten.

De vergelijkende studie van de bestaande testen laat zien dat er niet veel verschillen tussen deze testen zijn voor wat betreft de resultaten voor normaal- en slechthorenden en voor wat betreft de betrouwbaarheid. Twee van de bestaande testen, BOC en OC, laten een significant leereffect zien, wat echter acceptabel lijkt gezien de meeton nauwkeurigheid.

Wel blijkt dat de drie verschillende testen allemaal onvoldoende gevoelig zijn om beginnende lawaaislechthorendheid te ontdekken. De BOC lijkt het beste onderscheid tussen de drie groepen te maken, hoewel alle normaalhorenden buiten de categorie 'goed' vallen. Dit zou verklaard kunnen worden door andere classificatie van de uitslagcategorieën en het andere spraakmateriaal, waarbij gebruik gemaakt wordt van hoogfrequente medeklinkers en gepaarde, in plaats van unieke, klinkers.

VERBETERINGEN VAN DE BESTAANDE OORCHECK

De lage sensitiviteit van Oorcheck zou op basis van de uitkomsten van de ROC analyse verhoogd kunnen worden door het eerste afkappunt in de test iets lager te kiezen. Wanneer we echter naar de resultaten van Oorcheck in deze populatie kijken, zien we dat de SRT van de deelnemers met een smalle dip niet significant afwijkt van de SRT van normaalhorenden. Het onderscheidend vermogen van deze spraak-in-ruistest blijft, voor NIHL, dus beperkt.

Door het spraakmateriaal van de Oorcheck te homogeniseren zien we dat het vermogen van de test om onderscheid te maken tussen normaalhorenden en personen met een verschillende mate van hoge-tonen gehoorverlies beter wordt dan dat van de oorspronkelijke Oorcheck. Daarnaast heeft deze herkalibreerde versie van Oorcheck een sterkere correlatie met het toonaudiogram en met de SRT met Plomp zinnen, de gouden standaard.

Echter, voor deze nieuwe Oorcheck is er sprake van een redelijk groot leereffect van 1.6 dB, wat met de huidige analyses niet verklaard kan worden. Wanneer dezelfde test gedaan wordt, maar dan met een lagere intensiteit (OC -15) is dit effect kleiner, 0.7 dB. Ook voor alle andere experimentele versies, met hetzelfde herkalibreerde spraakmateriaal, wordt slechts een klein leereffect (<1 dB) gevonden.

Ook de verwachte verbetering van de testnauwkeurigheid wordt niet gevonden voor de nieuwe Oorcheck, deze blijft vergelijkbaar aan de oorspronkelijke test. Wellicht dat de combinatie van het gehomogeniseerde spraakmateriaal met de geoptimaliseerde ruis toch nog verschillen in verstaanbaarheid tussen de woorden met zich meebrengt. Dit zou nagegaan kunnen worden door op basis van de resultaten van de nieuwe Oorcheck opnieuw woordspecifieke discriminatiecurven te bepalen. Op basis van eventuele veranderingen in deze functies kan het proces van homogeniseren herhaald worden, om de testresultaten nauwkeuriger te maken.

Bovendien zou de testnauwkeurigheid van Oorcheck door aanpassingen aan de testprocedure verbeterd kunnen worden, bijvoorbeeld door het aantal aangeboden stimuli te verhogen, door de startwaarde aan te passen, door een langere aanloop naar de drempel te kiezen (deze bestaat nu uit 7 van de 27 aanbiedingen), of door de meting pas te laten starten na de eerste incorrecte respons zoals bij BOC. De invloed van een aantal van deze aanpassingen op de eigenschappen van de Oorcheck zou met aparte analyses op de verzamelde data nagegaan kunnen worden.

EXPERIMENTELE RUISCONDITIE

Naast het aanpassen van het spraakmateriaal wordt er in dit onderzoek vooral gekeken naar het effect van ruismodificaties op de uitslag van de herkalibreerde Oorcheck. De vraag die hierbij centraal staat is of de gevoeligheid van de test voor beginnende lawaaislechthorendheid verbeterd wordt door gebruik te maken van een gemodificeerde maskeerruis. De verschillende analyses laten zien dat dit voor een aantal experimentele ruiscondities inderdaad het geval is; met name de testen met de Low-pass gefilterde ruis, de fluctuerende ruis en de ruis waarin deze twee eigenschappen gecombineerd worden leveren een groter onderscheid op tussen de normaal- en slechthorende groepen. Bovendien laten de SRTs gemeten met de testen met deze ruiscondities een hogere correlatie zien met de gehoordrempels in de hoge frequenties van het toonaudiogram. Deze correlatie is het hoogst voor de Low-pass gefilterde ruis, en dan voornamelijk in de 2-6 kHz regio. Voor lawaaislechthorendheid is dit bij uitstek het gebied waarin wij geïnteresseerd zijn.

Voor de testen met de 16 Hz fluctuerende ruis geldt dat er sprake is van een grotere spreiding in de testuitslagen, voornamelijk in de groep slechthorenden met een smalle dip. Een aantal personen profiteert in bijna dezelfde mate als normaalhorenden van de fluctuaties in de ruis, terwijl dit voordeel voor anderen veel kleiner is. Deze toename in spreiding geldt voor de Low-pass gefilterde continue ruizen niet.

Deze ruiscondities zijn dus beter in staat te discrimineren tussen normaal- en slechthorenden en zijn daarom gevoeliger wat betreft het te screenen op lawaaislechthorendheid. Echter, hierbij dient wel opgemerkt te worden dat de gepresenteerde resultaten behaald zijn onder gecontroleerde omstandigheden; de testen werden allemaal afgenomen met dezelfde hoofdtelefoon, hetzelfde afspeelniveau en in een geluidsarme ruimte. Aspecten die allemaal minder onder controle zijn wanneer de experimentele Oorcheck online aangeboden wordt.

De onderzoekspopulatie heeft de originele Oorcheck voorafgaand aan het experiment ook thuis uitgevoerd. Wanneer we de gemiddelde SRT van deze test in een thuissituatie vergelijken met de gemiddelde SRT van de test gedaan onder gecontroleerde omstandigheden wordt er een significant verschil gevonden. Dit suggereert dat de resultaten verkregen in de experimentele testomgeving niet één op één te extrapoleren zijn naar de praktijk.

Bovendien geldt voor de huidige Oorcheck dat de test in de praktijk met beide oren afgenomen wordt. Ondanks het verwachte voordeel van binauraal testen van 1.4 dB (Smits 2006) wordt er in dit onderzoek een nadelig effect van de thuissituatie gevonden, waarschijnlijk veroorzaakt door de minder gecontroleerde meetopstelling in de thuissituatie, met verschillen in apparatuur en afspeelniveau. De SRTs gemeten in het lab zijn gemiddeld 1.2 dB beter dan de SRTs gemeten in de thuissituatie met hoofdtelefoon. Wanneer de test in de thuissituatie met luidsprekers uitgevoerd wordt is dit verschil zelfs groter, 2.0 dB, waarschijnlijk door de grotere invloed van achtergrondlawaai.

Gebaseerd op bovenstaande resultaten verwachten we dat de testen met gehomogeniseerd spraakmateriaal en gemodificeerde ruis in de praktijk zullen resulteren in een hogere SRT dan gemeten in de gecontroleerde testomgeving. Hoe groot dit verschil zal bedragen is op basis van deze testresultaten niet bekend. Of dit verschil groter is wanneer de online test met luidsprekers uitgevoerd wordt is evenmin bekend.

Omdat de verschillen tussen de experimentele en praktijksituatie kleiner zijn bij het gebruik van hoofdtelefoons zou aanbevolen moeten worden om de test uitsluitend per hoofdtelefoon uit te voeren. Dit ook in verband met nadelige effecten van eventueel achtergrondgeluid. Overwogen kan worden om de testen per oor uit te voeren, echter zal de online Oorcheck door het wegvallen van het voordeel van binauraal luisteren waarschijnlijk resulteren in nog hogere SRT waarden dan in dit onderzoek gemeten.

DE INVLOED VAN HET PRESENTATIENIVEAU

De oorspronkelijke versie van Oorcheck maakt gebruik van stationaire breedbandige ruis onafhankelijk van het afspeelniveau. Het wegvallen van delen van de ruis, ofwel door filtering ofwel door modulaties maakt de test gevoeliger voor achtergrondlawaai. Dit is voornamelijk het geval wanneer de respondent de test met luidsprekers uitvoert. Door het toevoegen van een ruisvloer wordt getracht het effect van achtergrondgeluid te reduceren.

Toch blijven de gemoduleerde ruizen hiervoor het meest gevoelig. Ook geldt bij deze ruizen een relatief sterke afhankelijkheid van het afspeelniveau: de SRT wordt lager bij toenemend afspeelniveau, omdat men dieper in de gaten kan luisteren. De oncontroleerbaarheid van deze effecten bij het gebruik als online test maken dat er een sterke voorkeur is om de Low-pass gefilterde continue ruis uit te roepen tot de meest geschikte conditie (binnen de set van onderzochte ruisvarianten) om Oorcheck gevoeliger te maken als een online test voor lawaaislechthorendheid.

De zeer hoge sensitiviteit en specificiteit, en de hoge correlatie met de hoogfrequente drempels in het toonaudiogram in deze populatie bevestigen dit. Het onderzoek geeft voldoende aanwijzingen om de afkappunten verstandig te kunnen kiezen. Na implementatie lijkt een gedetailleerde retrospectieve analyse van de eerste 50.000 online uitgevoerde tests aan te bevelen.

7 CONCLUSIES

Hoewel de veelgebruikte online hoortest Oorcheck een belangrijke rol speelt bij het waarschuwen voor en ontdekken van gehoorverlies door lawaai, is de test in zijn huidige hoedanigheid niet gevoelig genoeg om beginnende lawaaislechthorendheid te ontdekken.

Het homogeniseren van het gebruikte spraakmateriaal levert een test met vergelijkbare betrouwbaarheid, maar met een kleine toename in het discriminerend vermogen tussen normaal- en slechthorenden en in de correlaties met het toonaudiogram en de Plomptest.

De grootste winst wordt echter behaald wanneer er gebruik gemaakt wordt van slim gekozen ruismateriaal. Wanneer de maskeerruis Low-pass gefilterd wordt ontstaan er grote verschillen in de gemeten SRT voor normaalhorenden en personen met een verschillende mate van lawaaislechthorendheid.

Deze test heeft een vergelijkbaar leereffect en testbetrouwbaarheid als de oorspronkelijke Oorcheck, maar laat een sterke toename zien in sensitiviteit van 51% naar 95%, en gelijktijdig een verdere toename van de specificiteit van 90% naar 98%.

De Oorcheck met Low-pass gefilterde continue maskeerruis is daarom bij uitstek het alternatief voor de huidige Oorcheck, met de hoogste gevoeligheid voor beginnende lawaaislechthorendheid binnen de in dit onderzoek geteste Oorcheck-varianten.

8 DANKWOORD

De auteurs willen allereerst de collega's van het Leids Universitair Medisch Centrum (o.l.v. dr.ir. J.A.P.M. de Laat) en van het Universitair Medisch Centrum Nijmegen (o.l.v. prof.dr. A.F.M. Snik) danken voor hun bijdrage aan deze multi-center studie.

De auteurs zijn verder veel dank verschuldigd aan mw. C. Jacobs, die tijdens haar afstudeerstage veel voorbereidend werk heeft verricht, en aan dr. K.S. Rhebergen voor zijn hulp bij de opzet van dit project en bij het maken van het ruismateriaal en de SII berekeningen.

Verder is nauw samengewerkt met ZICHT, die de implementatie van de experimentele tests op het internet heeft verzorgd.

Dit project is mogelijk geworden dankzij financiële ondersteuning van de Stichting Hoortoestelbranche en van het Revalidatiefonds aan de Nationale Hoorstichting.

9 REFERENTIES

- Bosman A.J., Smoorenburg G.F.(1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Int J Audiol*; 34(5):260-284.
- International Organization for Standardization (1983) ISO 6189:1983, Acoustics -- Pure tone air conduction threshold audiometry for hearing conservation purposes. Geneva ISO.
- Jacobs C. (2008). Internettesten voor Noise-Induced Hearing Loss. Scriptie van afstudeeronderzoek binnen het AMC voor de studie logopedische en audiologische wetenschappen aan KU Leuven.
- Jongmans P., Martens M.K., Dreschler W.A. (2009). Analyse van de hoortesten van de Nationale Hoorstichting 2008.
- Jongmans P., Martens M.K., Dreschler W.A. (2008). Analyse van de hoortesten van de Nationale Hoorstichting 2007.
- Martens M.K., Perenboom R.J.M., Ploeg C.P.B. van der, Dreschler W.A. (2005). Oorcheck: analyse van gehoortesten voor jongeren over de periode april 2004-oktober 2004. TNO-rapport.
- Phillips D.P., Rappaport J.M., Gulliver J.M. (1994) Impaired word recognition in noise by patients with noise-induced cochlear hearing loss: contribution of temporal resolution defect. *Am J Otol*; 15(5):679-686.
- Plomp R., Mimpen. (1979) Improving the reliability of testing the speech reception threshold for sentences. *Audiol*: 18(1):43-52.
- Smits C., Kapteyn T.S., Houtgast T. (2004). Development and validation of an automatic speech-in-noise screening test by telephone. *Int J Audiol*; 43(1):15-28.
- Smits C., Merkus P., Houtgast T. (2006). How we do it: The Dutch functional hearing-screening tests by telephone and internet. *Clin Otolaryngol*; 31(5):436-40.
- Smits C., Houtgast T. (2007). Recognition of digits in different types of noise by normal-hearing and hearing-impaired listeners. *Int J Audiol*; 46(3):134-144.
- Smoorenburg G.F. (1992). Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *J Acoust Soc Am*; 91(1):421-437.
- Versfeld N.J., Daalder L., Festen J.M., Houtgast T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*; 107(3):1671-84.
- Wagenaar W.A. (1969). Note on the construction of digram-balanced Latin squares. *Psychol Bull*; 72(6):384-386.

BIJLAGEN

- BIJLAGE I: GELUIDSNIVEAU IN MEETRUIMTE LUMC
- BIJLAGE II: BEREKENEN VAN WOORDSPECIFIEKE DISCRIMINATIECURVEN
- BIJLAGE III: BEREKENINGEN MET BEHULP VAN HET SII MODEL
- BIJLAGE IV: RESULTATEN PILOTMETINGEN

BIJLAGE I: GELUIDSNIVEAU IN MEETRUIMTE LUMC

Omdat het in het LUMC niet mogelijk was de internettesten in een geluidsdichte cabine uit te voeren, vond dit deel van het experiment plaats in een andere meetruimte. Deze was relatief stil, zonder afleiding, maar niet geluidsdicht. Om te controleren of het achtergrondgeluid niet hinderlijk is geweest tijdens de uitvoer van de metingen, wordt het geluidsniveau in deze kamer gemeten (L_{FZ}), en vergeleken met het maximaal toelaatbare niveau van achtergrond ruis (L_{max}) volgens ISO 6189 (1983).

Per proefpersoon is het geluidsniveau in de spreekkamer gemeten op drie tijdstippen tijdens het uitvoeren van de internettesten. De metingen duurden 30 seconden en vonden plaats ter hoogte van de proefpersoon en met de testleider in de spreekkamer.

Het geluidsniveau in de spreekkamer is gemeten in octaafbanden van 16 Hz t/m 16 kHz. De verkregen waarden voor het gemeten geluidsniveau zijn energetisch gemiddeld, volgens onderstaande formule, en weergegeven in tabel I-1:

$$LFZ_{gem} = 10 \log \left(\frac{\sum^n 10^{\frac{LFZ}{10}}}{n} \right) = 10 \log \left(\frac{10^{\frac{LFZ1}{10}} + 10^{\frac{LFZ2}{10}} + K + 10^{\frac{LFZn}{10}}}{n} \right)$$

frequentie	geluidsniveau
16	60,6
31,5	53,8
63	51,6
125	53,6
250	50,3
500	47,0
1000	42,3
2000	41,5
4000	40,0
8000	34,7
16000	31,2

Tabel I-1: De geluidsniveaus gemeten in de spreekkamer, per octaafband.

De gevonden waarden worden vervolgens vergeleken met de grenswaarden voor het niveau van achtergrondgeluid, gegeven in ISO 6198 (1983). Deze ISO norm is opgesteld voor het meten van gehoordrempels vanaf 500 Hz in het kader van gehoorscreening. Het gaat hierbij om het afnemen van de (ongemaskeerde) luchtgeleidingsdrempels in het toonaudiogram.

In de norm worden voor geluidsniveaus gemeten in octaafbanden tabellen met grenswaarden voor maximaal achtergrondgeluid gegeven, onder andere een tabel met geluidsniveaus waarboven geen toonaudiometrie tot minimaal 0 dB HL mag worden uitgevoerd (ISO tabel 3). Bij deze tabel wordt aangegeven dat indien men een ander minimum zou willen meten, het verschil van dit gewenste minimum met 0 dB HL bij de grenswaarden opgeteld kan worden.

De ISO norm geeft een formule om het maximaal toelaatbare niveau van het omgevingsgeluid te berekenen bij gebruik van een hoofdtelefoon. Met deze formule wordt er gecorrigeerd voor de demping van de betreffende hoofdtelefoon, in dit onderzoek een Sennheiser HDA 200. De formule voor het maximale toelaatbare geluidsniveau is als volgt:

$$L_{\max} = k + A$$

k = gegeven constante in de ISO-tabellen

A = gemiddelde demping van de hoofdtelefoon

Omdat het in dit onderzoek niet gaat om een detectie van tonen van een dergelijk lage intensiteit, worden de waarden van k gegeven in de ISO-tabel 3 gebruikt als indicatie voor het geluidsniveau in de spreekkamer waar de internettesten worden afgenomen. Deze waarden voor k, en de bijbehorende L_{\max} zijn gegeven in tabel I-2.

frequentie (Hz)	demping (A) HDA200 (dB)	k (tabel 3, ISO-6189)	L_{\max} obv A + k (dB SPL)	gemeten L_{FZ} (dB SPL)
125	14.3	55	69.3	53.6
250	15.9	39	54.9	50.3
500	22.5	19	41.5	47.0
1000	28.6	13	41.6	42.3
2000	32.0	11	43.0	41.5
4000	45.7	13	58.7	40.0
8000	43.8	16	59.8	34.7

Tabel I-2. Demping van de gebruikte hoofdtelefoon, de in ISO-6189 gegeven waarde voor k en berekende maximaal toegestane L_{\max} en de niveaus gemeten in de spreekkamer, per octaafband. De geluidsniveaus in de grijs gearceerde octaafbanden overschrijden de L_{\max} .

Over het algemeen ligt het geluidsniveau in de spreekkamer dus onder het maximaal toegestane niveau, maar voor 500 Hz en 1000 Hz worden overschrijdingen gevonden met respectievelijk 5.5 dB en 0.7 dB.

In dit onderzoek dient er niet tot minimaal 0 dB gemeten te worden, aangezien het gaat om internettesten met een minimaal afspeelniveau van 65 dB(A). Bovendien gaat het in dit onderzoek niet om het meten van gehoordrempels, en daarom gaan we er vanuit dat het gemeten achtergrondgeluid voor het meten van de SRT met behulp van de internettesten geen rol zal spelen.

BIJLAGE II: BEREKENEN VAN WOORDSPECIFIEKE DISCRIMINATIECURVEN

Om het spraakmateriaal van de bestaande Oorcheck te optimaliseren wordt er door middel van woordspecifieke discriminatiecurven gekeken naar de benodigde aanpassing van de intensiteit, om een homogene set spraakstimuli te verkrijgen. Op deze manier wordt er inzicht verkregen in het 50% punt van alle woorden en hoe deze ten opzichte van elkaar liggen, en in de steilheid (en dus het onderscheidingsvermogen) van de verschillende testen.

Voor het maken van deze curven wordt er gebruik gemaakt van de resultaten van de originele Oorcheck versie, verkregen vanaf de implementatie van de test in 2004. Van deze resultaten worden de testen geselecteerd die door de beoogde doelgroep in de leeftijd van 12 t/m 24 jaar is uitgevoerd. Vervolgens zijn testen geselecteerd die voldoende betrouwbaar bevonden waren. Dit gebeurde op basis van de interne standaarddeviatie, deze mocht niet hoger zijn dan 2 dB.

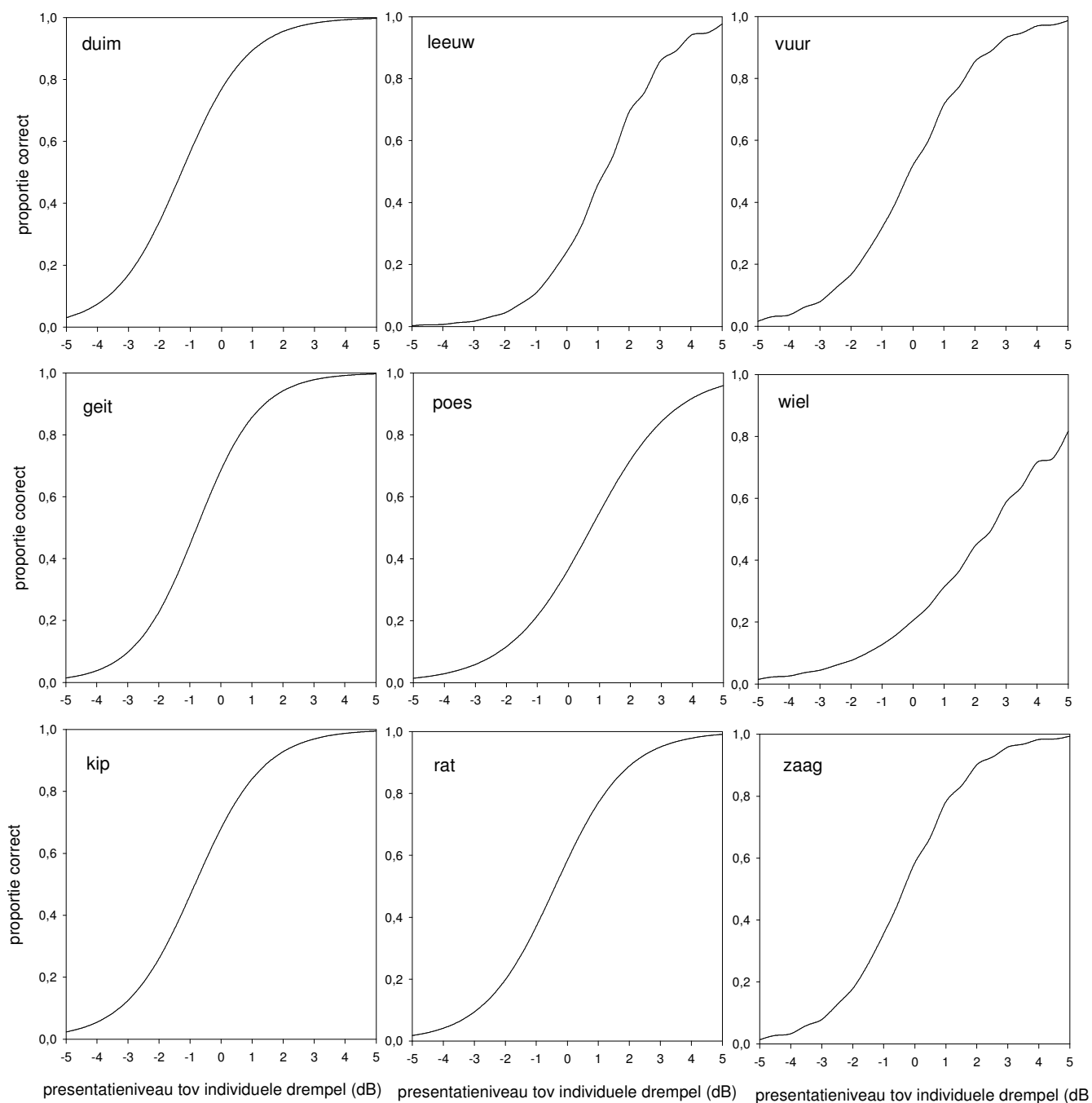
De curven worden gemaakt op basis van de aangeboden stimulus, de gegeven respons en of deze juist is. Bovendien wordt de intensiteit van de stimulus ten opzichte van de individuele drempel (het relatieve dB niveau) berekend, om rekening te houden met de individuele verschillen in spraakverstaan. Door dan het proportie correcte antwoorden als functie van het presentatieniveau ten opzichte van de individuele SRT-drempel uit te zetten, kan per woord een psychometrische curve berekend worden.

Om alle data van één woord te bundelen worden de woorden gesorteerd en wordt er per woord een file gemaakt, wat resulteert in negen files voor elk jaar. Per stimulus worden de woorden vanaf de achtste aanbieding, deze tellen mee voor de uiteindelijke SRT-drempel berekening, geselecteerd.

Dan wordt er per stimulus berekend worden hoeveel het niveau van de stimulus afwijkt van de individuele SRT-drempel. Daarnaast moet gekeken worden of een antwoord correct is of niet. Indien stimulus en respons overeenkomen krijgt de variabele 'correct' de waarde 1, zo niet krijgt de variabele de waarde 0.

Voor een juiste fit van de discriminatiefunctie, en om de steilheid en de moeilijkheid van de woorden te bepalen wordt er een logistische regressie uitgevoerd. Hiermee kan het verband tussen de snr van de stimulus, relatief ten opzichte van de individuele SRT, en de kans op correcte score worden weergegeven. Bij deze logistische regressieanalyse geldt de proportie correct als afhankelijke y-variabele en de signaal-ruis verhouding ten opzichte van de individuele drempel als onafhankelijke x-variabele.

De fit van deze functie voor elk woord is weergegeven in figuur II-1. Door middel van de logistische regressie formule kan er een constante en een regressiecoëfficiënt berekend worden, welke de relatie tussen de twee variabele beschrijven.



Figuur II-1: Woordspecifieke discriminatiecurven berekend d.m.v. logistische regressie, o.b.v. de resultaten van de testen uitgevoerd in 2004 t/m 2007

In de figuren staat op de X-as het presentatieniveau t.o.v. de individuele drempel en op de Y-as het percentage correct beantwoord. Als het presentatieniveau en de individuele drempel gelijk zijn wordt er verwacht dat in 50% van de gevallen er een juiste respons is. Als dit percentage hoger ligt dan is het woord makkelijk en als het percentage lager ligt is het woord moeilijk.

Dit gebeurt door middel van:

$$P(x) = \frac{1}{1 + e^{-(b_0 + b_1 x)}}$$

met b_0 = constante van regressievergelijking

en b_1 = regressiecoëfficiënt.

De constante en regressiecoëfficiënt zeggen wat over de relatie tussen de relatieve signaal-ruis verhouding van de stimuli en het percentage correct en over de bijbehorende curve, maar ze kunnen nog geen begrijpelijke karakterisering van de curve geven.

Door deze parameters om te rekenen kunnen echter wel de helling en de verschuiving van de psychometrische curve als volgt berekend worden:

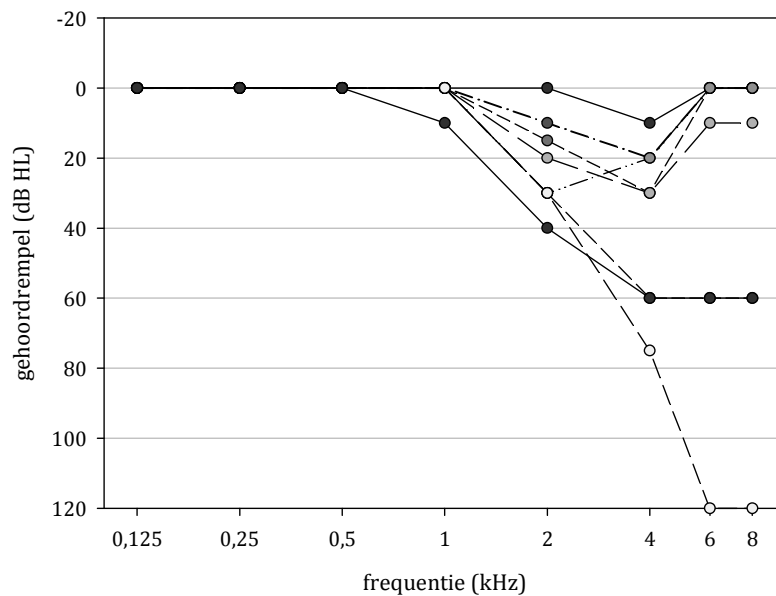
$$S = \frac{\text{regressiecoëfficiënt}}{4}$$

$$\text{verschuiving} = - \frac{\text{constante}}{\text{regressiecoëfficiënt}}$$

Door voor elke stimulus een logistische regressie uit te voeren kunnen de regressiecoëfficiënt en de constante berekend worden. Hier vloeien de karakteristieken van de psychometrische curve dan weer uit voort, en kunnen er correcties toegepast worden op de intensiteit waarop de verschillende woorden worden aangeboden.

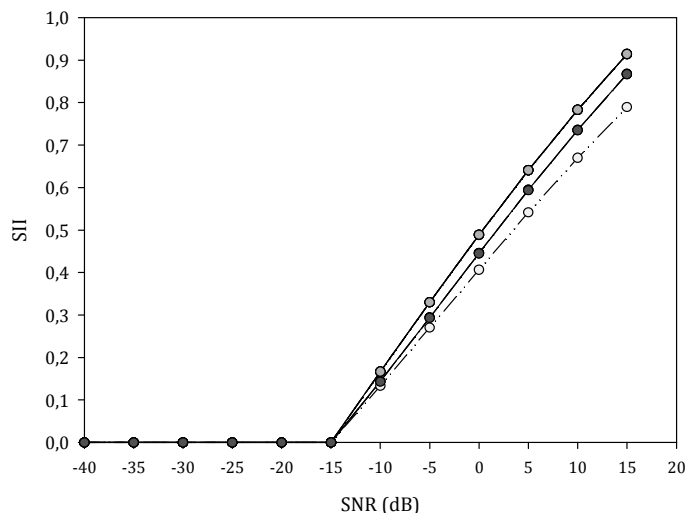
BIJLAGE III: SII VOORSPELLINGEN

De verwachte effecten van ruisfiltering op de SRT werden voorafgaand aan de pilotmetingen en het daadwerkelijke experiment geschat met behulp van het SII model. Dit model kan de spraakverstaanbaarheid voorspellen, door op basis van een gegeven spraak-in-ruisconditie te berekenen welk percentage spraakinformatie beschikbaar is voor een luisteraar met een bepaalde gehoordrempel. De audiogrammen uit figuur III-1 van personen met lawaaislechthorendheid in verschillende mate werden gebruikt voor deze SII-voorspellingen.



Figuur III-1: De audiogrammen met verschillende mate van NIHL, gebruikt voor de SII voorspellingen.

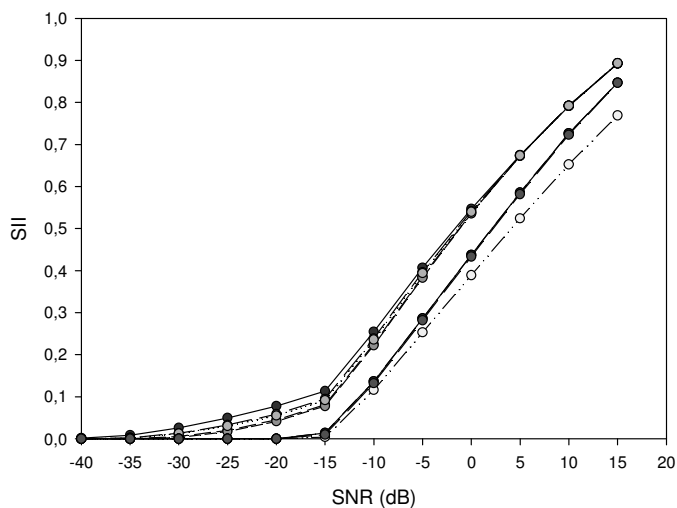
Normaalhorenden hebben ongeveer 33% van de totale spraakinformatie nodig om in stationaire ruis de helft van de zinnen foutloos te kunnen verstaan. Voor het spraakmateriaal van Oorcheck, negen CVC-woorden, kan dit percentage nog wat lager liggen. In figuur III-2 staat de SII-voorspelling weergegeven voor de huidige Oorcheck met breedbandige continue ruis, berekend voor de verschillende audiogrammen uit figuur III-1. Deze figuur laat zien dat er in een breedbandige ruis weinig verschil voorspeld wordt tussen de verschillende audiogramconfiguraties voor wat betreft de SII waarde.



Figuur III-2: Voorspellingen voor de SII als functie van SNR, voor OC met breedbandige continue ruis.

Een eerste optie voor filtering is een notch met een breedte van een octaaf rond 4 kHz. De SII-voorspelling van deze ruis (fig. III-3) verschilt slechts minimaal met die van de breedband ruis. Te zien is dat bij een SII van 0.3 de verschillende audiogrammen in twee groepen uit elkaar getrokken zijn en met een verschil van ongeveer 4 dB. Waarschijnlijk is de dip te smal omdat een lokaal gehoorverlies op 4 kHz alleen niet zo belangrijk is voor spraakverstaan. Daarom zullen we deze ruis niet gaan gebruiken.

Een tweede mogelijkheid is de ruis te filteren door middel van een laagdoorlaat filter. Er kan bij deze mogelijkheid voor meerdere afsnijfrequenties gekozen worden. In eerste instantie wordt er gekeken naar 1.4 kHz (midden van de SPIN-weging) en 1,6 kHz.

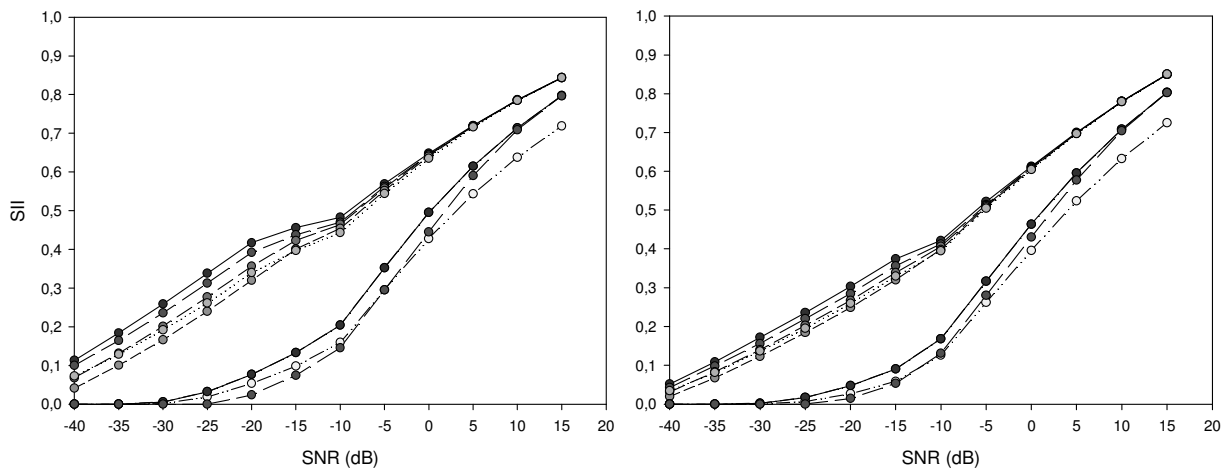


Figuur III-3: Voorspellingen volgens SII model voor een dip van 1 octaaf breed rond 4 kHz.

Door deze filtering worden de lage frequenties van het spraakmateriaal gemaskeerd en zijn deze voor iedereen even moeilijk te verstaan. De hoge frequenties worden echter niet gemaskeerd en

hier is de score dus sterk afhankelijk van het gehoorverlies in de hoge frequenties. Normaalhorenden zullen de spraak in deze ruis dus bij een lage signaal-ruis verhouding nog goed kunnen verstaan. Personen met NIHL hebben in dit gebied dat minder gemaskeerd wordt een verminderd gehoor. Ze kunnen daarom minder goed beschikken over de aanwezige spraakinformatie en hebben daarom een hogere signaal-ruis verhouding nodig om de spraak goed te kunnen verstaan. Dit leidt, afhankelijk van de gehoordrempel, tot hogere SRT waarden. In figuur III-4 (linkerpanel) is dit goed te zien, volgens het SII-model heeft een dergelijke filtering veel effect.; de voorspellingen voor de verschillende audiogrammen waaieren mooi uit elkaar en er is 25 tot 30 dB verschil bij een SII van 0.3 tussen de voorspelde SNR van het beste en van het slechtste audiogram.

Wanneer er voor een hogere afsnijfrequentie gekozen wordt, zoals 1.6 kHz, voorspelt het model dat de verschillen tussen de audiogrammen met normaal gehoor en milde NIHL kleiner worden (fig. III-4, rechterpanel). Er wordt immers meer informatie gemaskeerd, waardoor het ook voor normaalhorenden lastiger wordt spraak in ruis te verstaan.

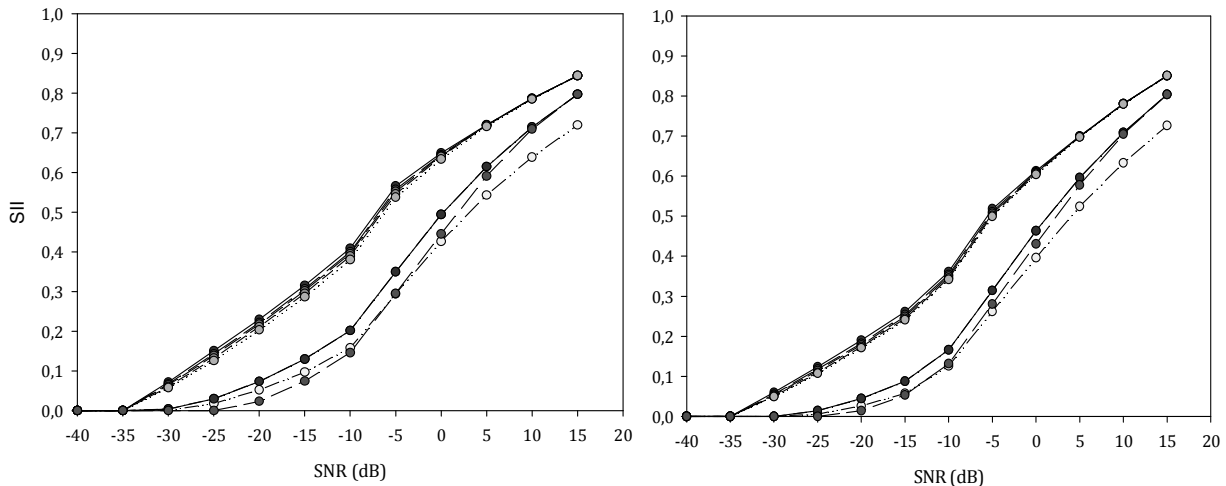


Figuur III-4: Voorspellingen volgens SII model voor een laagdoorlaat filter met afsnijfrequentie op 1.4 kHz (links) en afsnijfrequentie 1.6 kHz (rechts) en een ruisvloer van -35 dB.

Een belangrijk aandachtspunt is dat de SRT voor spraak in een breedbandige ruisonafhankelijk is van de intensiteit van de aanbieding (mits voldoende bovendrempelig), omdat het hierbij om een signaal-ruis verhouding gaat. Daarom was oorspronkelijk een spraak-in-ruistest met continue breedbandige ruis gekozen. Maar bij deze alternatieve ruizen zijn er 'gaten' in de ruizen toegevoegd zodat de uitkomst wel afhankelijk kan zijn van de intensiteit; de spraak wordt harder zonder dat de maskering op bepaalde punten toeneemt.

Omdat de internettesten in een thuissituatie uitgevoerd zullen worden, kan achtergrondgeluid tijdens het uitvoeren van de test een rol kan gaan spelen in het gedeelte van het spraakspectrum dat niet meer door ruis gemaskeerd wordt. Is de achtergrondruis hoog, kan deze de gehoordrempel benaderen. Hierdoor kan de uitslag van de test beïnvloed worden. Om deze invloed zo klein mogelijk te maken wordt door middel van een ruisvloer de achtergrondruis voor iedereen op een zelfde 'drempelniveau' gehouden.

De voorspellingen met behulp van het SII model zijn, behalve voor meerdere afsnijfrequenties, ook voor meerdere ruisvloeren gedaan. De curven in figuur III-4 zijn gemaakt op basis van Low-pass gefilterde ruizen met een ruisvloer van -35 dB. Dezelfde curven zijn weergegeven in figuur III-5, maar dan voor Low-pass gefilterde ruizen met een ruisvloer van -20 dB. Omdat de ruisvloer hier hoger ligt en dus meer spraakinformatie gemaskeerd wordt, is het voorspelde verschil tussen de verschillende maten van gehoorverlies kleiner.



Figuur III-5: Voorspellingen volgens SII model voor een laagdoorlaat filter met afsnijfrequentie op 1.4 kHz (links) en afsnijfrequentie 1.6 kHz (rechts) en een ruisvloer van -20 dB.

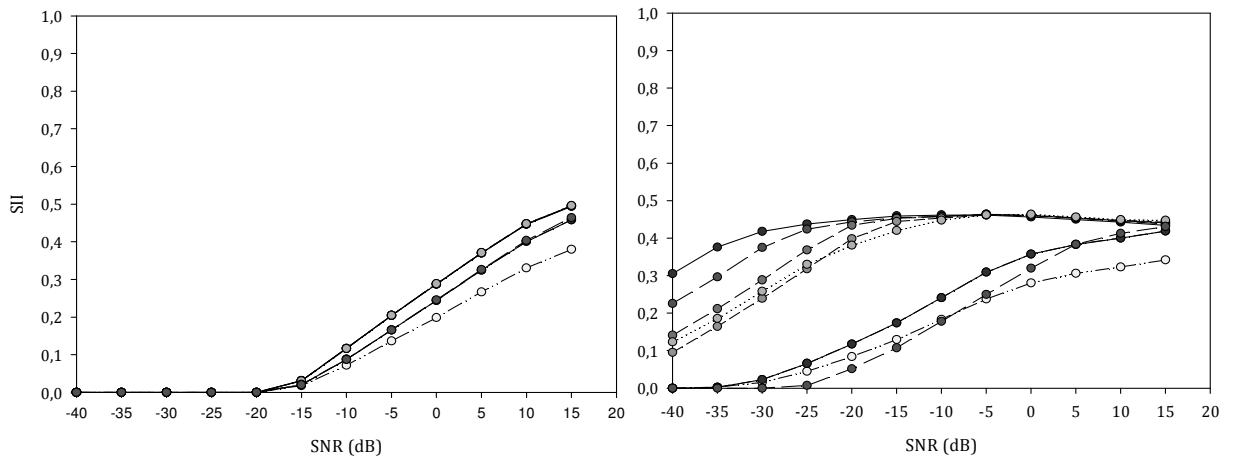
Een andere mogelijkheid om de ruizen te modificeren is door het temporele aspect van de breedbandige ruis aan te passen, door bijvoorbeeld een blokmodulatie. Omdat het SII-model alleen voorspellingen kan doen voor spraakverstaan in continue ruis kan er voor een conditie met fluctuerende geen voorspelling weergegeven worden.

Filtering van het spraaksignaal

Een andere methode is het filteren van het spraaksignaal, waardoor de nadruk op een bepaald frequentiegebied van de spraak kan worden gelegd. Deze methode werkt volgens hetzelfde principe als de filtering van het ruissignaal; de spraak wordt gefilterd zodat men gedwongen wordt meer in het hogere frequentiegebied te luisteren waardoor er een groter verschil ontstaat tussen normaal- en slechthorenden.

De meest voor de hand liggende methode is om de spraak door een hoogdoorlaat filter op 1.4 kHz te filteren, maar in breedbandige ruis heeft dit weinig effect vergeleken met gewone spraak in breedbandige ruis (fig. III.6, linkerpanel). Wanneer deze spraak wordt gecombineerd met een laagdoorlaat gefilterde ruis op 1.4 kHz, verkrijgt men volgens het SII-model ongeveer hetzelfde effect als wanneer alleen de ruis is gefilterd (zie figuur III-6, rechterpanel).

Voorspellingen van beide spraak-in-ruiscombinaties laten slechts kleine effecten zien van de spraakfiltering op het onderscheiden van normaal- en slechthorenden. Bovendien zal de gefilterde spraak kunstmatig, en dus onprettig klinken. Daarom wordt deze optie verder niet in het onderzoek meegenomen.



Figuur III-6: Voorspellingen volgens SII model voor high pass speech met afsnijfrequentie op 1.4 kHz in een breedbandige ruis (links) en in een Low-pass gefilterde ruis (rechts).

BIJLAGE IV: RESULTATEN PILOTMETINGEN

Om inzicht te krijgen in het effect van de verschillende filteringen van de maskeerruis op de uitkomst van de Oorcheck, werden er behalve voorspellingen met behulp van het SII model, eerst een aantal pilotmetingen uitgevoerd. Hierin worden ook de bestaande internettesten en de Oorcheck met herkalibreerd spraakmateriaal meegenomen, ter vergelijking met de experimentele condities. Deze pilot wordt uitgevoerd op het AMC.

Voor het eerste deel van de pilot worden acht normaalhorende personen getest, zes vrouwen en twee mannen, tussen 20-44 jaar oud. Bij deze personen worden, naast een toonaudiogram, de volgende testen afgenomen:

- de bestaande internettesten NHT, BOC en OC
- de vernieuwde Oorcheck met herkalibreerd spraakmateriaal
- vijf Oorchecks met experimentele ruiscondities
 - Fluctuerende breedbandige ruis met 16 Hz blokmodulatie en ruisvloer op -35 dB
 - Low-pass ruis met afsnijfrequentie 1.4 kHz en ruisvloer op -35
 - Low-pass ruis met afsnijfrequentie 1.4 kHz en ruisvloer op -20
 - Low-pass ruis met afsnijfrequentie 1.6 kHz en ruisvloer op -35
 - Low-pass ruis met afsnijfrequentie 1.4 kHz, High-pass fluctuerende ruis met 16 Hz blokmodulatie en ruisvloer op -35

Alle testen werden afgenomen met de opstelling die later ook in het daadwerkelijke experiment gebruikt zal worden (zie § 2.4). De testen worden allemaal twee keer afgenomen, op 65 dB(A) en in een gebalanceerde volgorde. Vijf van de onderzochte personen hebben Oorchecks met experimentele ruiscondities tijdens een tweede sessie nogmaals uitgevoerd, maar dan op een afspeelniveau van 75 dB(A). De resultaten van alle pilot metingen zullen in dit hoofdstuk besproken worden.

In tabel IV-1 zijn de gemiddelden en standaarddeviaties van de metingen weergegeven per meting. Ook het overall gemiddelde van test en retest is weergegeven, evenals de test-retest standaarddeviatie. De resultaten van de bestaande testen komen goed met elkaar overeen, alleen de NHT heeft een ander gemiddelde wat is terug te voeren op het gebruik van ander spraakmateriaal, met een beperkte bandbreedte. De geoptimaliseerde Oorcheck heeft met het opnieuw gekalibreerde spraakmateriaal een kleinere test-retest standaarddeviatie dan de originele versie. De OC_{nieuw} blijkt dus nauwkeuriger te zijn.

De pilotstudie laat zien dat de alternatieve ruizen resulteren in sterk verlaagde gemiddelde SRTs; de gemiddelden van deze testen liggen tussen -21.6 en -28.2. De verschillende ruiscondities maskeren dus 10-18 dB minder dan de originele breedbandige continue ruis. Ook is er binnen de testen met experimentele ruizen sprake van een iets grotere spreiding. Deze toename in spreiding is het grootst voor de Low-pass gefilterde ruis met de afsnijfrequentie op 1.6 kHz.

	test	gem (sd) test	gem (sd) retest	overall gem (sd)	test-retest sd
65 dB(A)	NHT	-6.3 (1.5)	-7.1 (0.9)	-6.7 (1.2)	1.1
	BOC	-10,5 (1.1)	-10.9 (1.2)	-10.7 (1.1)	1.0
	OC origineel	-10.2 (1.9)	-11.0 (1.4)	-10.6 (1.7)	1.4
	OC nieuw	-10.8 (2.4)	-11.3 (1.8)	-11.1 (2.0)	0.7
	OC 16 Hz	-28.1 (0.7)	-28.3 (0.8)	-28.2 (0.8)	0.6
	OC LP 1.4 -35	-23.7 (2.0)	-28.3 (1.0)	-26.0 (2.8)	3.2
	OC LP 1.4 -20	-22.8 (1.3)	-22.2 (3.3)	-22.5 (2.4)	1.3
	OC LP 1.6 -35	-26.9 (0.8)	-27.0 (0.6)	-27.0 (0.6)	0.8
	LP 16 -35	-23.1 (2.5)	-23.4 (1.4)	-23.2 (1.9)	1.0
75 dB(A)	OC nieuw	-11.5 (0.9)			
	OC LP 1.4 -35	-24.6 (1.8)	-27.9 (1.2)	-26.2 (2.2)	2.0
	OC LP 1.4 -20	-21.9 (2.7)	-21.4 (2.1)	-21.6 (2.2)	2.0
	OC LP 1.6 -35	-26.8 (1.0)	-26.5 (1.2)	-26.7 (1.0)	0.4
	LP 16 -35	-24.1 (1.3)	-26.7 (1.4)	-25.4 (1.8)	1.9

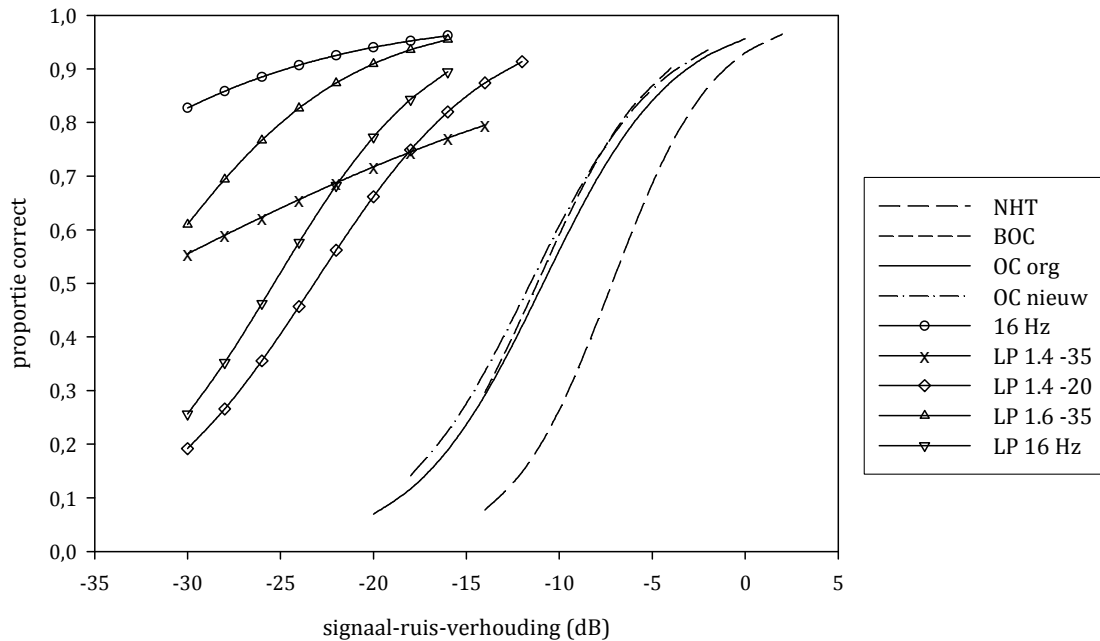
Tabel IV-1: Gemiddelde en gepoolde standaard deviaties binnen de metingen en overall gemiddelde en test-retest standaarddeviaties voor alle afgenomen testen.

Er lijkt weinig verschil te zijn tussen de test en de retest condities. Dit wordt bevestigd door een gepaarde t-test, die alleen een significant test-retest verschil aantoonst voor de Oorcheck met fluctuerende ruis. De retest van deze variant resulteert in een veel lagere gemiddelde SRT bij alle proefpersonen dan de eerste meting.

Tenslotte is er gekeken naar verschillen tussen SRTs van de testen met experimentele ruiscondities verkregen op verschillende presentatieniveaus, 65 en 75 dB(A). Dit kan inzicht verschaffen in de invloed van het presentatieniveau op de gemeten SRT-drempels. Wanneer de gemiddelden van de testen die afgenomen zijn op beide niveaus vergeleken worden komen deze erg overeen. Dit wordt bevestigd door het uitvoeren van gepaarde t-toetsen welke laten zien dat er voor geen enkele test een significant verschil is tussen de testen afgenomen op 65 dB(A) en op 75 dB(A). Het presentatieniveau van de testen lijkt dus weinig effect te hebben. Dit is echter op basis van een kleine groep normaalhorenden bepaald. De vraag is of deze bevinding ook geldt voor slechthorenden.

De lage gemiddelde signaal-ruis verhouding van de alternatieve testen is een belangrijk aandachtspunt voor het vervolg van deze pilotstudie. Het grote nadeel van deze lage snr is namelijk dat de dynamiek van de testen, die van +4 tot -30 dB loopt, niet toereikend is. Een aantal van de nieuwe testen loopt bij een SNR van -30 dB tegen de ondergrens aan, waardoor de up-down-procedure niet meer correct uitgevoerd kan worden. Dit is voornamelijk het geval bij de fluctuerende ruis en de 1.6 kHz Low-pass gefilterde ruis. Dat verklaart ook de lage standaarddeviatie die voor deze test gevonden wordt; alle proefpersonen lopen tegen de ondergrens van de test aan, waardoor de test uiteindelijk resulteert in SRTs rond dezelfde waarde. Om deze reden is deze test later niet op een ander afspeelniveau gemeten, omdat er vanwege de beperking van de test geen verschillen in resultaat verwacht worden.

Wanneer we per test een PI curve (percentage correct als functie van SNR) plotten zien we dat door het beperkte dynamisch bereik niet voor alle testen met nieuwe ruizen een 50% punt gemeten kon worden (fig IV-1).



Figuur IV-1: Performance-intensity curves per tests

Wanneer de SRT-drempels gemeten met de testen met alternatieve ruizen te laag worden, kunnen behalve de testdynamiek ook verschillen in gebruikte computersystemen en -instellingen bij thuisgebruik de resultaten van de test beïnvloeden. Bovendien gaat achtergrondlawaai dan ook een grotere rol spelen.

Daarom worden er in het vervolg van de pilot nieuwe ruizen onderzocht die wat meer maskeren en daarmee een wat minder lage SNR genereren. Hierbij kan gedacht worden aan het toevoegen van een wat hogere ruisvloer aan de hier beschreven ruisvarianten. Zoals te zien is in deze pilot, heeft de gemeten groep een hogere gemiddelde SRT voor de test met de low-pass gefilterde ruis met afsnijfrequentie 1.4 kHz en een ruisvloer van -20 dB dan voor de test met dezelfde ruis met een ruisvloer van -35 dB.

Ook een hogere afsnijfrequentie bij de low pass ruizen en een aangepaste duty cycle bij de fluctuerende ruizen (60% ruis en 40% stilte) behoren tot de mogelijkheden. Voor een low-pass gefilterde ruis met hogere afsnijfrequentie wordt in de eerdere pilotmetingen niet per se een hogere SRT-drempel gevonden, hoewel dit wel volgens SII model voorspeld wordt

De bovenstaande waarden voor de verschillende parameters worden in een nieuwe set ruizen gecombineerd tot de volgende ruiscondities die in het vervolg van de pilot gemeten worden (zie tabel IV-2):

	conditie	spectrum	afsnijfreq	ruisvloer	modulatie	duty cycle	gemiddelde
1	OC nieuw	BB	-	-	-	-	-11,3 (2,2)
2	16 Hz – DC 60	BB	-	-20	16Hz	60/40	-28,3 (1,1)
3	16 Hz- 35	BB	-	-35	16 Hz	50/50	-29,4 (0,5)
4	16 Hz – 20	BB	-	-20	16 Hz	50/50	-28,9 (0,7)
5	LP1.4 – 20	LP	1,4	-20	--	--	-23,0 (3,7)
6	LP1.6 – 35	LP	1,6	-35	--	--	-29,0 (0,6)
7	LP 16 – 35	LP	1,4	-35	16 Hz	50/50	-27,2 (1,3)
8	LP 16 – 20	LP	1,4	-20	16 Hz	50/50	-24,7 (1,7)
9	LP 16- DC 60	LP	1,4	-20	16 Hz	60/40	-24,5 (2,5)
10	LP1.8 -35	LP	1,8	-35	-	-	-29,0 (0,6)
11	LP1.8 -20	LP	1,8	-20	-	-	-23,0 (1,7)

Tabel IV-2: Nieuwe ruiscondities, met bijbehorende specificaties van de ruizen en de in de pilot gemeten gemiddelde SRT (sd).

De nieuw opgestelde Oorcheck-varianten worden getest door ze bij zes normaalhorenden af te nemen. Eerst wordt de vernieuwde Oorcheck getest, daarna volgen de tien experimentele testvarianten. Dit gebeurt volgens een gebalanceerde volgorde, gebaseerd op een Latijns vierkant van 10x10. Alle testen worden tweemaal afgenomen (test-retest) op 65 dB(A).

Proefpersonen rapporteren dat bij sommige versies de woorden nog erg goed te verstaan zijn, bij sommige versies is dat moeilijker. Dit blijkt ook uit de uitslagen (tabel IV-2). Van vier testen ligt de gemiddelde SRT tussen -20 en -25 dB en de gemiddelde SRTs van de andere testvarianten benaderen -30 dB. Dit wijst erop dat ook deze testen tegen de ondergrens van het dynamische bereik van de Oorcheck aanlopen. Dit betreft alle testen met een maskeerruis waaraan een ruisvloer van -35 dB is toegevoegd, de testen met ruizen met een ruisvloer van -20dB hebben dit probleem niet. Echter bij de breedbandige fluctuerende ruizen is dit wel het geval, ook wanneer de duty cycle is aangepast.

Omdat we de fluctuerende ruisconditie is graag in het uiteindelijke onderzoek zouden willen testen, wordt geprobeerd de lage gemiddelde SRT te verhogen door de toegevoegde ruisvloer nog 5 dB hoger te leggen, zodat de modulatie diepte 15 dB bedraagt. Ter vergelijking wordt dit ook voor een selectie van de low-pass gefilterde ruizen gedaan. Dit resulteert in een nieuwe set van vijf aangepaste experimentele ruizen met een ruisvloer van -15 dB. Ook wordt er een 'stilte'conditie opgenomen, dat is een Oorcheck waar alleen een ruisvloer van -15 dB aan toe is gevoegd.

Deze zes nieuwe testen worden afgenomen bij vier van de eerder onderzochte normaalhorende proefpersonen (tabel IV-3).

	Conditie	spectrum	afsnijfreq	ruisvloer	modulatie	duty cycle	Gemiddelde
12	16 Hz - 15	BB	-	-15	16 Hz	50/50	-28,2 (1,1)
13	16 Hz - DC 60 - 15	BB	-	-15	16 Hz	60/40	-27,8 (1,4)
14	LP1,4 - 15	LP	1,4	-15	-	-	-23,5 (1,5)
15	LP 16 - 15	LP	1,4	-15	16 Hz	50/50	-25,7 (1,3)
16	LP 16 - DC60 - 15	LP	1,4	-15	16 Hz	60/40	-22,5 (2,3)
17	OC-'stil' - 15	BB	-	-15	-	-	-29,6 (0,3)

Tabel IV-3: Nieuwe ruiscondities met ruisvloer op -15 dB, met bijbehorende specificaties van de ruizen en de in de pilot gemeten gemiddelde SRT (sd).

De gemiddelde SRTs gemeten met de aangepaste testen liggen allemaal wat hoger dan in de eerdere pilot. De beide breedbandige fluctuerende ruizen lopen soms nog tegen de grens van -30 aan, maar dat is slechts een enkele keer per persoon.

De bevindingen van de pilotmetingen laten geen grote verschillen zien tussen de low-pass ruis met afsnijfrequentie 1.4 kHz en die met afsnijfrequentie 1.6 en 1.8 kHz. Bij de filtering van de ruizen zal de afsnijfrequentie 1.4 kHz gehanteerd worden. Deze frequentie ligt in het midden van het spraakspectrum volgens de SPIN-weging en vindt volgens de SII voorspellingen de grootste verschillen tussen verschillende audiogrammen.

In de breedbandige fluctuerende ruizen wordt slechts een klein verschil gevonden in SRT tussen de ruizen met duty cycle 50/50 en met 60/40. De conventionele en meest beschreven methode is die waarbij de duty cycle van de ruis 50/50 is en daar zullen we dan ook in deze studie vanuit gaan.

Tenslotte laten de pilot-metingen zien dat, met name de fluctuerende ruizen, bij een modulatie diepte van 15 dB de meest gunstige resultaten behaald worden. Met deze modulatie diepte spelen de beperkingen van de testen geen, of slechts een geringe rol. Bovendien komen de gemiddelde SRT-drempels iets hoger uit. Daarom zal er in het experiment uitgegaan worden van een modulatie diepte van 15 dB.