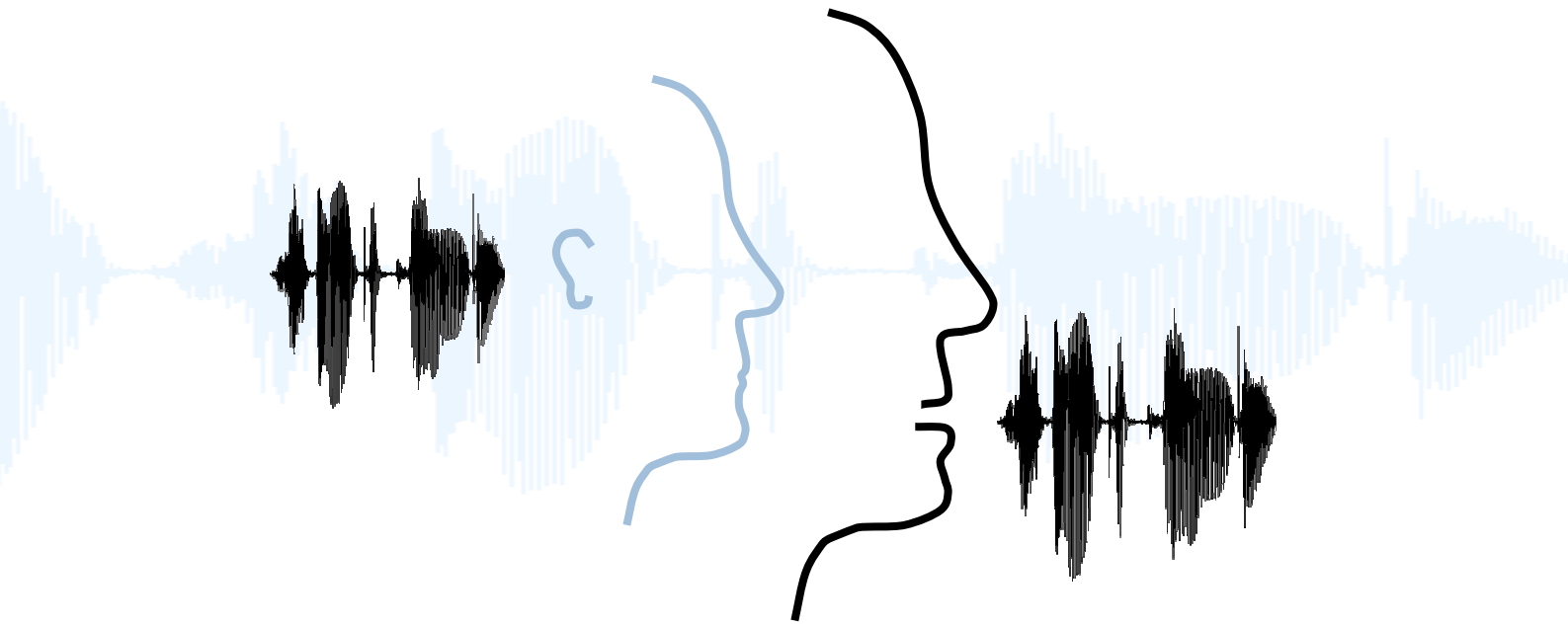
Bestemmelse af taleforståelighed i støj

Determination of Speech Intelligibility in Noise

Af Ellen Raben Pedersen
Studienr. 167595



Synopsis
Denne rapport behandler forskellige parametre, der indgår i opbygningen af en test til bestemmelse af taleforståeligheden i støj. Testen, som er designet til afvikling af materialet Dantale II, består af to forskellige versioner – den operatør- og den patientbaserede version. Overvejelser omkring parametrene er foretaget på baggrund af litteraturen, computersimuleringer og fire pilottests. Desuden er der foretaget en spørgeskema-undersøgelse og genereret et nyt taleformet støjsignal.

Kandidatprojekt – civilingeniørstudiet
Syddansk Universitet · Odense
Det Tekniske Fakultet,
Institut for Sensorer, Signaler og Elektroteknik

Intern vejleder:
Peter Møller Juhl

Eksterne vejledere:
Carsten Daugaard & **Søren L. Jørgensen**
DELTA; Dansk, Elektronik, Lys og Akustik

August 2007

Resumé

Det at forstå tale i støjfyldte omgivelser opfattes af særligt hørehæmmede som en meget vanskelig opgave. For i den kliniske praksis at bestemme et objektivt mål for denne evne er der i tidens løb blevet udviklet forskellige materialer til afvikling af tale i støj-tests. I 2001 blev der lanceret et nyt materiale kaldet Dantale II, som består af et større antal testsætninger og et taleformet støjsignal. Testsætningerne, der er konstrueret med en fast grammatisk struktur, indeholder hver fem forskellige ord og forekommer ikke som værende meningsfyldte. Ud fra dette materiale har DELTA udviklet et computerbaseret testprogram, som består af to forskellige versioner – den operatør- og den patientbaserede version. En test afviklet med den operatørbaserede version foregår ved, at FP'en (forsøgsperson) efter hver afspillet sætning skal gentage så meget af det hørte som muligt. I en test afviklet med den patientbaserede version skal FP'en i stedet via en skærm for hvert af de fem ord vælge sit svar blandt tre mulige. For begge versioner gælder, at præsentationsniveauet for den enkelte sætning løbende fastsættes på baggrund af FP'ens respons ved at optælle antallet af de ord, som vedkommende svarede korrekt.

I et sådant testprogram indgår mange parametre – valget af en række af disse behandles i denne rapport. Eksempelvis bør det for den patientbaserede version overvejes, hvor mange svarmuligheder FP'en skal præsenteres for og om vedkommende evt. skal have mulighed for at svare *ved ikke*. For begge versioner skal der tillige tages stilling til, hvordan testresultatet skal angives og herunder ved hvilke sandsynligheder for korrekt svar sætningerne skal præsenteres. I forbindelse hermed bør det bestemmes, om ændringen af præsentationsniveauet fra sætning til sætning skal foregå ved at ændre niveauet af enten tale- eller støjsignalet og om støjen skal være til stede mellem sætningspræsentationerne. Samtidigt bør det overvejes, hvor meget eller lidt præsentationsniveauet skal ændres ud fra FP'ens respons, hvor lang testen skal være samt hvordan det endelige testresultat skal beregnes. Derudover bør indlæringseffektens omfang og valgt af støjsignal undersøges.

For at få et indtryk af måden, hvorpå taletests udføres i dag og af eventuelle ønsker til en ny test, er der udsendt et spørgeskema til landets høreklinikker. Desuden er der foretaget computersimuleringer af forskellige afviklingsprocedurer og genereret et nyt taleformet støjsignal. Endelig er der udført fire pilottests, som samlet belyser mange af de parametre, der indgår i opbygningen af en taletest. Pilottestene blev udført med et nyt testprogram svarende til det udvirket af DELTA. Den væsentligste forskel de to programmer imellem er, at i det nye har FP'en i den patientbaserede version frem for tre hele ti svarmuligheder for hvert ord. Samtidigt er der designet en variant af den patientbaserede version, hvor FP'en foruden de ti svarmuligheder også har muligheden for at svare *ved ikke*.

Abstract

It can especially for hearing impaired be very difficult to understand speech in a noisy environment. Different speech materials has been developed over the years in order to fulfil the clinical practises need to have an objective measure of this ability. In 2001 a new speech material called Dantale II was launched. This consists of a large number of test sentences and a speech shaped noise signal. The test sentences, which are constructed with a fixed grammatical structure, each contains five different words and appears not to have any meaningful contents. DELTA has developed a computerized test program based on this material, their program consists of two different versions – the operator and the patient based version. In a test carried out with the operator based version the test person's objective is to verbally repeat as much as possible from the sentence just played. In a test carried out with the patient based version the test person has to select his or her response to each of the five words among three alternatives shown on a screen. For both versions the presentation level of each sentence is determined based on the number of words that the test person answers correct.

Such a test program is influenced by many parameters – the choice of some of these is investigated in this thesis. For instance it should be considered how many alternatives for each word the test person should be presented within the patient based version, and if he or she should have the option of answering *don't know*. For both versions it should also be considered what final test result is wanted from the test and as a consequence of this at what levels of probability for a correct answer the presentation should be placed. In connection with this it should be decided if the change in difficulty from sentence to sentence is to be either a change in the level of the sentences or a change in the level of the noise and if the noise should be present between the sentence presentations. It should also be considered how much the presentation level should be changed based on the test person's response, how long the test should be and how the final test result should be calculated. In addition to this the extent of the learning effect and the choice of noise signal should be investigated.

In order to get an impression of the way in which speech tests are carried out today and the desires there might be in a new test, a questionnaire was send to hearing clinics all over the country. In addition there has been carried out computer simulations of various test procedures and a new speech shaped noise signal has been generated. Finally four pilot tests have been conducted that evaluates many of the parameters involved in the design of a speech test. The pilot tests was conducted with a new test program similar to the one developed by DELTA. The essential difference between the two programs is that in the new program the test person in the patient based version can pick between ten different answers for each of the five words in a sentence. Besides that a variant of the patient based version has been designed in which the test person in addition to being presented with ten different alternatives for each word in a sentence also has the option of answering *don't know*.

Forord

Denne rapport er et resultat af mit kandidatprojekt, som er udarbejdet i forbindelse med uddannelsen til civilingeniør i *Fysik og Teknologi* ved Syddansk Universitet Odense. Projektet er blevet til i samarbejde med DELTA på baggrund af et eksisterende testprogram til bestemmelse af taleforståeligheden i støj. Som fundament for at kunne gennemføre projektet har jeg inden dets start udført et indledende litteraturstudie med titlen *Psykoakustik* – se evt. Pedersen (2006)*. Dette studie gav mig en grundlæggende indføring i den psykoakustiske terminologi. Det udrustede mig så at sige med en psykoakustisk værktøjskasse indeholdende forskellige nyttige redskaber. Denne værktøjskasse er nu blevet udvidet ikke blot med viden om taletests generelt – herunder i særdeleshed med talematerialet Dantale II og udførelsen af psykoakustiske tests i både fritfelt og med hovedtelefoner, men også med et kendskab til programmeringssproget LabView og statistikprogrammet SPSS (Statistical Product and Service Solution). Projektet, som udover at have været meget spændende og ikke mindst udfordrende, har desuden givet mig nye erfaringer med programmeringssproget MatLab.

I tilknytning til projektet har Peter Møller Juhl fungeret som intern vejleder. Til trods for at psykoakustik ikke er hans hovedområde, har han været til mig til stor hjælp og støtte, hvorfor jeg gerne vil sige ham mange tak. Samtidigt vil jeg også gerne takke DELTA for via dette projekt at have givet mig mulighed for at snuse lidt til verdenen uden for skolebænken – herunder for bl.a. at have stille ekspertise og ikke mindst en bærbar computer til rådighed. I forbindelse hermed vil jeg gerne rette en særlig tak til Carsten Daugaard og Søren L. Jørgensen, som begge har fungeret som eksterne vejledere og været mig meget behjælpelige. Derudover vil også gerne takke alle de personer, der har været involveret i mine pilottests – herunder til de, der har indvilliget i at være FP'er – dvs. såvel venner, studiekammerater som patienter – og til personalet på høreklinikken på både Odense og Vejle Sygehus. Samtidigt vil jeg også gerne benytte lejligheden til at sige tak til personalet på Odense Tekniske Bibliotek, som altid har været utroligt hjælpsomme og givet mig et smil med på vejen. Slutteligt vil jeg også gerne rette en stor tak til min kæreste, Simon Nielsen og min onkel, Knud Henningsen for hhv. at have oversat resuméet til engelsk og at have læst korrektur på rapporten.

Odense, august 2007

Ellen Raben Pedersen

Indholdsfortegnelse

1 Indledning 6
1.1 Baggrund 6
1.2 Problemformulering 8
1.3 Rapportens emner og opbygning 8
1.4 Ordforklaring 10

2 Dansk taleaudiometri gennem tiden 11
2.1 Materialer baseret på enkeltord 11
2.1.1 HC-listerne 12
2.1.2 GS-listerne 12
2.1.3 CL-listerne 13
2.1.4 Dantale I 13
2.2 Sætningsbaserede materialer 15
2.2.1 Helen-materialet 15
2.2.2 HINT 15
2.3 Sammenfatning 16

3 Talematerialet Dantale II 17
3.1 Testsætningerne 17
3.1.1 Sætningsstrukturen 18
3.1.2 Udvælgelse af basislistens ord 19
3.1.3 Indspilning af basislisten 20
3.1.4 Testsætningernes generering 21
3.1.5 Materialets homogenitet 21
3.1.6 Reduktion af sætningslængden 23
3.2 Dantale II-støjen 24
3.2.1 Støjens generering 24
3.2.2 Frekvens- og modulationsanalyse 25
3.3 Sammenfatning 27

4 Responsmetoder 28
4.1 Den operatørbaserede version 28
4.1.1 Forventet respons 29
4.1.2 Taletærsklen 30
4.1.3 Placering af præsentationer 32
4.2 Den patientbaserede version 34
4.2.1 Forventet respons 34
4.2.2 Placering af præsentationer 40
4.2.3 Alternativ afvikling 43
4.3 Sammenfatning 44

5 Testafvikling 45
5.1 Præsentationsformen 45
5.1.1 Variation af tale eller støj 45
5.1.2 Kontinuert eller afbrudt støj 46

5.2 Afviklingsniveauet 47
5.2.1 Niveauet af støjsignalet 47
5.2.2 Præsentationsniveauet af den første sætning 47
5.3 Afviklingsprocedurer 48
5.3.1 Hansen og Ludvigsens procedure 48
5.3.2 Brand og Kollmeiers procedure 49
5.3.3 MML-proceduren 51
5.3.4 Computersimuleringer 52
5.4 Indlæringseffekten 60
5.4.1 Indlæringen inden for en test 60
5.4.2 Indlæringen mellem to tests 61
5.5 Sammenfatning 62

6 Alternativer til Dantale II-støjen 64
6.1 ICRA-støj 64
6.1.1 Karakteristika 65
6.1.2 Generering 68
6.1.3 Anvendelse i forbindelse med taletests 70
6.1.4 Modificering af *icra5* 71
6.2 Forslag til nyt støjsignal 74
6.2.1 Generering 74
6.2.2 Det nye støjsignal vs. Dantale II-støjen 75
6.3 Sammenfatning 76

7 Afslutning 77
7.1 Konklusion 77
7.2 Diskussion 79
7.3 Perspektivering 79

Forkortelsesliste 80

Symbolliste 81

Litteraturliste 82

Stikordsregister 86

Appendiks De danske Hagerman sætninger 87

Bilagsfortegnelse 89
Bilag A Spørgeskema-undersøgelse 90
Bilag B Testprogram 98
Bilag C Pilottest I-III 104
Bilag D Pilottest IV 164

Indhold af cd-rom 180

1 Indledning

Idet følgende fremgår en beskrivelse af baggrunden for projektet og den tilhørende problemformulering. Desuden beskrives hvilke emner rapporten behandler og hvordan denne er opbygget. Derefter følger en ordforklaring, som definerer begreberne; reliabilitet, sensitivitet og validitet.

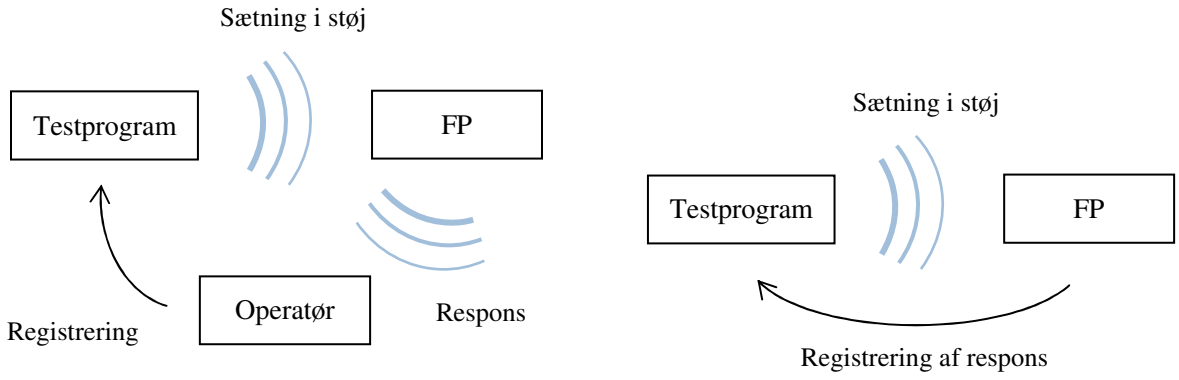
1.1 Baggrund

Baggrunden for dette projekt er et testprogram kaldet HearVal, hvis udvikling DELTA påbegyndte i 2004. Programmet, som er opbygget omkring talematerialet Dantale II, har til formål at teste og derved give et objektivt mål for evnen til at forstå tale i støj. Det primære sigte er, at testen skal kunne indgå i den kliniske praksis i forbindelse med bl.a. diagnosticering, udvælgelse/tilpasning af HA'er (høreapparat) og evaluering af HA-udbyttet. Programmet, som er konstrueret i LabVIEW, består af to forskellige versioner – den operatør- og den patientbaserede version. I de to versioner er responsmetoden forskellig – dvs. måden, hvorpå FP'en afgiver sit svar. Fælles for versionerne er, at testen forløber ved, at FP'en parallelt med en kontrolleret baggrundsstøj præsenteres for et givet antal sætninger, som alle består af fem ord. Testresultatet, der for begge versioner angives som en værdi af SRT_{50} (Speech Reception Threshold), svarer til det præsentationsniveau, hvorved den pågældende FP forventes at forstå 50% af de afspillede ord.

I den operatørbaserede version skal FP'en efter hver afspillet sætning mundtligt gentage så meget af det hørte som muligt, hvorefter en operatør via testprogrammet registrerer de korrekte ord. Da FP'en ikke er bundet til at vælge sit respons ud fra et begrænset antal svarmuligheder, kaldes måden, hvorpå FP'en angiver sit svar, en åben responsmetode. I den patientbaserede version, som forløber på knap så traditionel vis, placeres FP'en foran en trykfølsom skærm og skal efter hver afspillet sætning selv registrere sit svar via skærmen. Det foregår ved, at vedkommende for hvert af sætningens fem ord har tre forskellige svarmuligheder. Da det ikke er muligt at svare *ved ikke*, kan denne version af testprogrammet for hver afspillet sætning betragtes som fem på hinanden følgende 3AFC-tests (Alternative Forced Choice). Med andre ord er der her tale om en lukket responsmetode, idet FP'en vælger sit respons blandt et begrænset antal af svarmuligheder. Den i de to versioner forskellige responsmetode er illustreret i figur 1-1.

For begge versioner gælder, at den første sætning afspilles ved et niveau på 0 dB SNR (Signal to Noise Ratio), hvor såvel tale- som støjsignalet er 65 dB(C) SPL (Sound Pressure Level). Præsentationsniveauet af de efterfølgende sætninger fastsættes løbende i henhold til en procedure, der i denne rapport kaldes Brand & Kollmeiers procedure (Brand & Kollmeier, 2002). Dette gøres ved ud fra det antal ord, som FP'en i den forrige sætning gengav korrekt, at ændre niveauet af talesignalet, mens støjniveauet holdes konstant. Støjsignalet afspilles kontinuerligt således, at dette også er til stede mellem sætningspræsentationerne. I begge versioner forsøges sætningerne via afviklingsproceduren afspillet ved et niveau, hvor den enkelte FP forventes at kunne forstå i gennemsnit halvdelen af de afspillede ord korrekt – dvs. 2,5 ord. I den operatørbaserede version præsenteres sætningerne derved ved et niveau, hvor FP'en forventes at kunne gengive 50% af de afspillede ord korrekt, mens sætningerne i den patientbaserede version præsenteres ved et niveau, hvor FP'en forventes at kunne gengive hele 66,7% af de afspillede ord korrekt. Grunden

til sidstnævnte er, at den patientbaserede version som skrevet afvikles som 3AFC-tests, hvorfor der her optræder en sandsynlighed på 1/3 for, at FP'en – på trods af, at vedkommende ikke har hørt et givet ord – gætter ordet korrekt.



a) Den operatørbaserede version b) Den patientbaserede version

Figur 1-1: Illustration af responsmetoden i a) den operatør- og b) den patientbaserede version. I den operatørbaserede version skal FP'en efter hver afspillet sætning gentage, så meget af det hørte som muligt. En operatør registrerer de korrekte ord i FP'ens respons via testprogrammet, som derefter fastsætter præsentationsniveauet af den efterfølgende sætning. I den patientbaserede version skal FP'en i stedet selv registrere sit respons, hvorefter præsentationsniveauet af den efterfølgende sætning fastsættes.

I hver af to versioner gælder, at hvis FP'en gengiver flere ord korrekt end den angivne procentsats sænkes præsentationsniveauet af den følgende sætning, mens præsentationsniveauet hæves, hvis FP'en gengiver færre ord korrekt end den angivne procentsats. Ændringen af præsentationsniveauet – den såkaldte trinstørrelse – kan altså både være negativ og positiv. Et fortegnsskift af trinstørrelsen kaldes en reversering og er medbestemmende for antallet af sætninger, som FP'en præsenteres for. Det skyldes, at testen stoppes, når enten ti reverseringer eller en præsentation af 30 sætninger er opnået – alt afhængig af, hvilken betingelse, der først opfyldes. Når testen er færdigafviklet bestemmes testresultatet – dvs. SRT_{50} – som et gennemsnit af præsentationsniveauerne ved de sidste fire reverseringer.

Tidligere evaluering af testen

Med testen er der af DELTA blevet foretaget tre mindre undersøgelser. I den første undersøgelse blev seks normalhørende FP'er hver præsenteret for to tests med hver af de to versioner. Resultatet heraf viste, at SRT_{50} opnået med den operatørbaserede version i gennemsnit var 2,3 dB højere og derved dårligere end SRT_{50} opnået med den patientbaserede version – denne forskel var statistisk signifikant. Samtidigt blev der for SRT_{50} med den operatør- og den patientbaserede version fundet en re-test variation[1] på hhv. 0,7 og 0,6 dB. I den anden undersøgelse indgik 16

[1] Re-test variationen bestemmes fast gennem denne rapport som gennemsnittet af den numeriske forskel mellem to på hinanden følgende målinger – gennemsnittet beregnes på baggrund af værdier opnået med flere forskellige FP'er.

hørehæmmede FP'er, som hver blev testet først uden og derefter med HA – begge tests blev foretaget med den patientbaserede version. Resultatet heraf viste, at FP'erne i gennemsnit opnåede en værdi af SRT_{50}, der var 2,8 dB bedre med end uden HA – heriblandt var der dog flere, som ikke opnåede en forbedring i SRT_{50} ved at anvende HA. Den tredje og sidste undersøgelse blev foretaget i samarbejde med høreklinikken på Århus Sygehus – her blev der testet et større antal hørehæmmede FP'er udelukkende med den patientbaserede version. De hermed opnåede værdier af SRT_{50} indikerede, at testen ikke syntes at kunne differentiere fornuftigt mellem FP'er med forskellig evne til at forstå tale i støj.

1.2 Problemformulering

På baggrund af de i forrige afsnit beskrevne undersøgelser blev det af DELTA draget tvivl om, hvorvidt de forskellige parametre, der indgår i testprogrammet, er valgt fornuftigt. Opgaven i dette projekt er derfor at klarlægge, om og evt. hvordan valget af nogle disse influerer på testresultater opnået med såvel den operatør- som med den patientbaserede version – projektoplægget formuleret af DELTA findes på den til rapporten vedlagte cd-rom. I tilknytning hertil bør det nævnes, at projektet er afgrænset til et omfang svarende til halvanden semesters arbejde – dvs. 45 ECTS-points.

1.3 Rapportens emner og opbygning

Ud fra den meget åbne problemformulering er der udarbejdet en rapport, der behandler, hvad forfatteren har fundet naturligt at undersøge i forbindelse med opbygningen af et testprogram som det udviklet af DELTA. Hovedrapporten, der er inddelt i otte kapitler, hvori der løbende er henvisninger til resultater opnået med fire pilottests, behandler følgende;

Kapitel 2 giver en historisk gennemgang af forskellige danske talematerialer – herunder beskrives materialet Dantale I, der i dag er det hyppigst anvendte.

Kapitel 3 beskriver Dantale II materialet – herunder både de egentlige testsætninger og det tilhørende taleformerede støjsignal.

Kapitel 4 behandler betydningen af valget af responsmetode – herunder hvad der grundlæggende adskiller den operatør- og den patientbaserede version.

Kapitel 5 omhandler forskellige parametre i relation til testafviklingen – herunder fremgår computersimuleringer af forskellige afviklingsprocedurer.

Kapitel 6 præsenterer forskellige alternativer til Dantale II-støjen – herunder et forslag til et nyt taleformet støjsignal, som er genereret i forbindelse med dette projekt.

Kapitel 7 afslutter hovedrapporten med en konklusion, diskussion og perspektivering baseret på de emner, der er behandlet i de øvrige kapitler.

I tilknytning til hovedrapporten fremgår af Bilag A en spørgeskema-undersøgelse omhandlende taletests i den kliniske praksis i dag og hvorvidt der er ønske om en ny test – evt. baseret på patientselvbetjening. I forbindelse med udførelsen af de fire pilottests er der i MatLab blevet konstrueret et nyt testprogram, der ligesom DELTA's program består af to forskellige versioner – den operatør- og den patientbaserede version. Den væsentligste forskel de to programmer imellem er, at der i den patientbaserede version af det nye program er valgt at inddrage hele ti svarmuligheder for hvert af sætningernes fem ord. Denne version kan derved i stedet for fem på hinanden følgende 3AFC-tests betragtes som fem på hinanden følgende 10AFC-tests. Samtidigt blev der designet en variant af den patientbaserede version, hvor det for hvert af de afspillede ord foruden de ti svarmuligheder også er muligt at svare *ved ikke*. I Bilag B findes en beskrivelse af det nye testprogram.

Pilottestene, som benævnes Pilottest I-IV, belyser mange af de parametre, der indgår i opbygningen af en taletest. En beskrivelse af Pilottest I-III og Pilottest IV fremgår af hhv. Bilag C og Bilag D. Til sammenligning kan nævnes, at der i Pilottest I-III indgik 24 normalhørende FP'er, som i de tre tests blev inddelt i to eller tre grupper, mens der i Pilottest IV indgik 35 hørehæmmede FP'er. I forbindelse med den til pilottestene hørende databehandling er der udført en række statistiske tests og analyser vha. statistikprogrammet SPSS. Der er her fast valgt et signifikansniveau på 0,05, hvorved der forekommer en sandsynlig på maks. 5% for at drage en fejlkonklusion. Bemærk; i databehandlingen er ordet signifikantniveauet valgt forkortet til sig. Teorien bag de udførte tests og analyser beskrives ikke. I stedet henvises til hjælpe-funktionen i SPSS og bogen *Complete Business Statistics* af Aczel (1996) – herunder i særdeleshed til kap. 1, 4, 7, 9, 10 og 14. I Pilottest I og III er re-test variationen bl.a. blevet undersøgt. Grundet størrelsen af denne er der for værdier generelt opnået med testene kun medtaget en decimal. Det bør bemærkes, at der i figurer hidrørende fra de forskellige versioner af testprogrammet både er anvendt en bestemt farve og et specifikt symbol. Rapporten er i øvrigt udført således, at den i sin helhed kan udskrives og kopiere i sort/hvid uden, at der mistes information.

Til rapporten er der vedlagt en cd-rom, hvorpå bl.a. al den kildekode, der i MatLab er lavet i forbindelse med projektet, er placeret. En oversigt over, hvad der i øvrigt er at finde på denne, fremgår af rapportens sidste side. I umiddelbar forlængelse af selve hovedrapporten er der placeret en litteraturliste. Referencer til denne er angivet på gængs vis med forfatternavn(e) og årstal for udgivelsen. Hvis der er flere end to forfattere, er forkortelsen *et al*. (lat.: et alli) anvendt. Ved de referencer, hvor der er angivet en asterisk (*), findes der en kopi af den pågældende kilde på den vedlagte cd-rom. Efter litteraturlisten følger en forkortelses- og en symbolliste samt et stikordsregister. Desuden er der vedlagt et appendiks, hvoraf talematerialets testsætninger fremgår.

Det bør i øvrigt bemærkes, at det i denne rapport er valgt at anvende betegnelsen SRT_{th} for taletærsklen i støj, hvor subscriptet th henviser til, ved hvilken sandsynlighed for forståelighed tærsklen opgives. I nogen litteratur anvendes betegnelsen SRT_N for netop at indikere, at der er tale om, at tærsklen opgives som et niveau i dB SNR – se evt. Ardenkjær-Madsen & Josvassen (2001) og Smits & Houtgast (2006). Dette er dog ikke valgt her, idet tærsklen ikke udelukkende behandles som hørende til én bestemt sandsynlighed for forståelighed. Til gengæld behandles kun taletærsklen i støj, hvorfor det altså her er implicit, at tærsklen opgives som et niveau i dB SNR.

1.4 Ordforklaring

For at vurdere de enkelte parametre, der indgår i testprogrammet, er i rapporten løbende anvendt begreberne reliabilitet, sensitivitet og validitet. Alle tre begreber refererer med et vist overlap til testens troværdighed. Da de derved let kan forveksles, defineres de i det følgende – de angivne definitioner er i overensstemmelse med Dillon og Ching (1995);

Reliabilitet

Begrebet reliabilitet henviser til, om testen er reproducerbar og derved til størrelsen af re-test variationen. Begrebet siger med andre ord noget om graden af overensstemmelse mellem test-resultater opnået ved gentagne målinger – dvs. ved test re-testning – under identiske omstændigheder. Det er ønskeligt, at målinger med samme FP og under identiske omstændigheder giver ensartede resultater selv ved forskellige operatører – derved opnås en lille intra-individuel spredning.

Sensitivitet

Graden af sensitivt fortæller derimod, hvorvidt testen kan vise de reelle forskelle, der måtte være mellem personer med forskellige taleopfattelsesevner. Testes personer med meget forskellige taleopfattelser under identiske omstændigheder, ønskes divergerende testresultater således, at der opnås en stor inter-individuel spredning. Begrebet sensitivitet refererer også til, at testes den selv samme person under forskellige omstændigheder – eksempelvis med og uden HA – skal testresultaterne ligeledes kunne give et reelt billede af, hvorvidt vedkommende er hjulpet med apparatet eller ej. Det bør bemærkes, at en test kun kan være sensitiv, hvis den er reliabel. Til gengæld kan en test være reliabel uden at være sensitiv. Det gælder, hvis testen er så *grov* i sin bestemmelse af testresultatet, at der ikke kan skelnes mellem selv FP'er, hvis taleopfattelser er vidt forskellige.

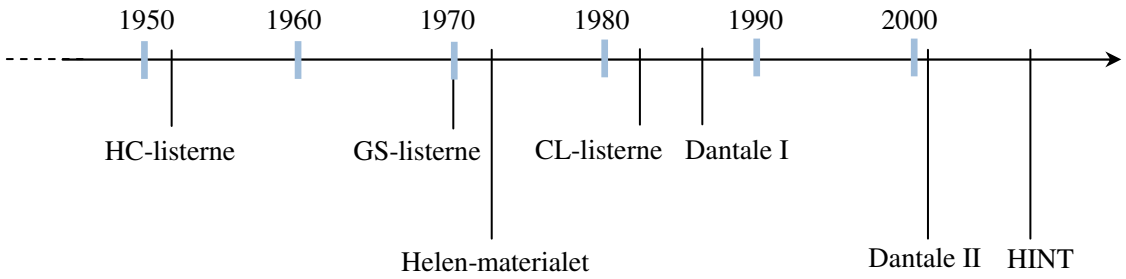
Validitet

Begrebet validitet refererer til, om testen faktisk afspejler det, der ønskes information om. Det vil med andre ord sige, hvorvidt testen simulerer den ønskede lyttesituation. Som eksempel er det ikke hensigtsmæssigt, hvis testen i sin administration er så svær, at FP'ens fokus flyttes fra selve lytteopgaven. Derved kan det risikeres, at det reelt ikke er vedkommendes taleopfattelse, der testes.

2 Dansk taleaudiometri gennem tiden

Hørelsen er en af menneskets allervigtigste sanser – den har principielt tre vigtige funktioner. For det første bruger vi den til at få oplysning om forskellige lydkilders art og funktion. Derudover bruger vi den til at retningsbestemme således, at vi kan placere os selv og de pågældende lydkilder i forhold til hinanden. Sidst med ikke mindst bruger vi hørelsen for at kunne kommunikere med vore medmennesker. I en vurdering af hørelsen er det derfor af stor vigtighed at inddrage evnen til at høre og forstå det talte sprog. For at kunne udføre en egentlig undersøgelse af taleopfattelsesevnen er der derfor i tidens løb udviklet flere forskellige talematerialer – disse har enten været baseret på enkeltord eller hele sætninger.

Selvom denne rapport omhandler konstruktion af et testprogram byggende på materialet Dantale II er det interessant at have et vist indblik i, hvordan udviklingen af taletests har forløbet. I de følgende afsnit fremgår derfor en historisk gennemgang af forskellige danske talematerialer, der alle i større eller mindre omfang har været anvendt i den kliniske praksis. Derudover beskrives HINT-testen, der først netop er blevet udviklet, mens en dybdegående gennemgang af Dantale II materialet er henlagt til kapitel 3. Af figur 2-1 ses en oversigt over, hvornår de enkelte danske materialer er blevet udviklet. Ud over de danske materialer findes der også en lang række udenlandske – for en beskrivelse af nogle af disse henvises til artikelsamlingen af Martin (1997), herunder i særdeleshed til artiklen af Knight.



Figur 2-1: Tidsaksen illustrerer, hvornår forskelligt dansk talemateriale er blevet udviklet.

2.1 Materialer baseret på enkeltord

Inden udviklingen af det første danske talemateriale, som var opbygget af enkeltord, var det almindelig praksis at vurdere en persons høreevne mere eller mindre uformelt. Det foregik typisk ved enten at bestemme fra hvilken afstand, det for vedkommende var muligt at forstå tale med normal styrke eller ved at bestemme ved hvor svag en hvisken vedkommende kunne forstå det talte (Ludvigsen, 2005). Forgangsmanden inden for dansk taleaudiometri var den nu afdøde Christian Røjskjær, der i en længere periode var overlæge ved Statens Hørecentral på Odense Sygehus – i dag kaldet Høreklinikken. Via sin doktorafhandling *Monaural Speech Audiometry*

(Røjskjær, 1952) konstruerede han et større antal ordlister, som indbefattede spondæer[2], enstavelsesord med forskellige konsonantkombinationer og talord opdelt i grupper af tre. Selvom dette materiale ikke direkte har været anvendt klinisk, lagde det grundstenen for udviklingen af det første danske materiale kaldet HC-listerne – dette materiale beskrives kort i det følgende. Derudover beskrives GS- og CL-listerne samt Dantale I materialet – i sidstnævnte indgår der talordslister svarende til de konstrueret af Røjskjær. Dantale I materialet består i øvrigt ikke udelukkende af enkeltord, idet materialet også indeholder løbende tale – på trods heraf beskrives materialet i nærværende afsnit.

2.1.1 HC-listerne

Stort set samtidigt med oprettelsen af de første danske høreklinikker – i starten af 1950'erne – blev der i umiddelbar forlængelse af Røjskjærs omfattende arbejde af to sprogforskere fra Københavns Universitet konstrueret syv kontinuerligt indspillede ordlister af 25 en- og tostavelsesord (Parving *et al*., 1976). Listerne, som kaldes Hørecentral-listerne eller blot HC-listerne, er fonetisk balancerede[3] – dvs. ordene er valgt således, at hyppigheden af ordenes enkelte sproglyde er repræsentative for det danske sprog. Listerne blev indspillet med en mandlig taler og blev anvendt ved, at FP'en blev præsenteret for én hel liste ved et for vedkommende tilstrækkeligt niveau. Ordene blev præsenteret gennem hovedtelefoner uden en tilhørende baggrundsstøj og efter hvert afspillet ord skulle FP'en på bedste vis gengive det hørte. Testresultatet blev derefter opgjort som en DL-værdi (Discrimation Loss) – svarende til procentsatsen af andelen af ikke korrekte gentagede ord. Som en bibemærkning kan nævnes, at en værdi af DL også kaldes skelnetabet eller misforståelsesprocenten.

Selvom HC-listerne var meget udbredte, var der ikke en udelt tilfredshed med disse. Mange mente, at ordenes sværhedsgrad ikke var ens listerne imellem, hvorved der var en vis skepsis over for testresultaternes troværdighed – og derved over for testens reliabilitet. Samtidigt bredte der sig et større og større ønske om en taletest, der også indbefattede en baggrundsstøj. Grunden for dette ønske var, at særligt det at forstå tale i støjende omgivelser, opfattes af mange høre-hæmmede som en meget vanskelig opgave. På trods af disse modstande over for materialet blev det anvendt igennem næsten to årtier.

2.1.2 GS-listerne

I 1970 blev der på Gentofte Amtssygehus udviklet et nyt materiale og derved en erstatning for de ikke så velansete HC-lister (Danielsen *et al*., 1971 og Parving *et al*., 1976). Dette materiale, der er tiltænkt voksne, kaldes Gentofte Standard-listerne eller blot GS-listerne og består af ti lister af hver 25 enstavelsesord – hovedsageligt af typen CVC (Consonant-Vowel-Consonant)[4]. Materialet, som ikke er fonetisk balanceret, er indspillet således, at det enten kan anvendes rent auditivt

[2] En spondæ er et tostavelsesord med tryk på begge stavelser. Eksempler herpå fra Røjskjærs ordlister er ordene *isbjerg*, *øjenbryn* og *ølkrus* – se evt. Røjskjær (1952 – Kap. 3).

[3] Ordet *fonetisk* kommer af ordet *fonem*. Et fonem er det mindste sproglige udtrykselement, som har en betydningsadskillende funktion i det talte.

[4] Ord af typen CVC er eksempelvis ordene *tyv*, *fik* og *lim*.

eller audiovisuelt. Ved at anvende en audiovisuel præsentation bliver testsituationen mere virkelighedsnær og testen derved mere valid, idet FP'en får mulighed for at kombinere høreopfattelsen med mundaflæsning. Listerne er i øvrigt indtalt af en kvinde og afvikles typisk ved anvendelse af fonemscore. Dvs. i stedet for at angive resultatet som antallet af korrekt gengivne ord, angives det ud fra antallet af korrekt gengivne fonemer. I forbindelse med udviklingen af materialet blev der også konstrueret en længe ønsket baggrundsstøj – den såkaldte Gunnar Nu-støj. Støjen blev dannet ud fra en optagelse af den nu afdøde sportskommentator Gunnar Nu Hansen, som blev overspillet med sig selv fire tidsforskudte gange.

2.1.3 CL-listerne

I starten af 1980'erne blev der også på Høreklinikken på Bispebjerg Hospital udviklet et talemateriale til erstatning for HC-listerne. Materialet, som kaldes CL-listerne – opkaldt efter Carl Ludvigsen, der i dag er leder af audiologisk forskningsafdeling på Widex – består af otte forskellige lister. Listerne er ligesom GS-listerne ikke fonetisk balancerede. De er ligeledes indtalt af en kvinde og indeholder også hver 25 enstavelsesord – dog udelukkende navneord. Som for GS-listerne er hensigten, at materialet skal anvendes ved brug af fonemscoring og at der præsenteres en hel liste ved MCL (Most Comfortable Level) med en tilhørende baggrundsstøj. For begge materialer gælder tillige, at der pr. liste i gennemsnit er 75 fonemer.

Med henblik på at få belyst, om GS- eller CL-listerne var mest egnet til kliniske målinger blev der af Nielsen & Parving (1986) foretaget en sammenlignende undersøgelse af de to materialer. Undersøgelsen viste imidlertid kun mindre forskelle de to materialer imellem, som alle kunne tilskrives, at forsøgslederne ikke var vant til anvende GS-listerne. I undersøgelsen blev der både anvendt fornem- og ordscoring. Selvom fonemscoring giver flere scoringselementer pr. ord, hvorved der principielt kan opnås en øget reliabilitet, viste dette sig ikke at være tilfældet. Grunden herfor blev tillagt, at tolkningen og deraf optællingen af antallet af korrekt gentagne fonemer varierende både mellem forsøgelederne og inden for den selv samme forsøgsleder. Selvom fonem- frem for ordscoring statistisk øger præcisionen, er dette altså ikke nødvendigvis ensbetydende med, at testen har en bedre reliabilitet i praksis. Denne problematik kan måske omgås, hvis forsøgslederne gennemgår en grundig optræning inden en egentlig test således, at termerne *fonem* og *lyd* ikke anvendes i flæng. I øvrigt kan nævnes, at CL-listerne i modsætning til GS-lsiterne aldrig har opnået en stor udbredelse.

2.1.4 Dantale I

Selvom der var gjort flere forsøg på at lave et talemateriale, der var alment accepteret, var dette ikke lykkedes. Derfor blev der udviklet endnu et materiale – det såkaldte Dantale I materiale. Materialet, som blev udgivet på cd i 1986, består bl.a. af otte ordlister beregnet til voksne (Elberling *et al*., 1989). Disse indeholder hver 25 ord fordelt på navne-, tillægs- og udsagnsord. Derudover er der i materialet indbefattet tre lister med talord, fire lister med ord til børn, en enkelt liste med ord tiltænkt småbørn og tre lydspor med løbende tale omhandlede Samsø. Som tidligere skrevet svarer talordslisterne til dem konstrueret af Røjskjær. Cd'en er konstrueret

således, at der kun er tale i venstre kanal, mens der i højre er et støjsignal, som er en simulering af den føromtalte Gunner Nu-støj – se evt. afsnit 2.1.2. På cd'en er der samtidigt indlagt forskellige kalibreringssignaler.

Jf. spørgeskema-undersøgelsen, der er foretaget i forbindelse med dette projekt, er det hyppigst anvendte materiale i den kliniske praksis i dag netop ord- og talordslisterne fra Dantale I – se evt. Bilag A. På trods for, at der af Elberling *et al*. (1988) og Keidser (1991) fremgår en omfattende dokumentation af materialet, findes der ingen egentlig vejledning i, hvordan materialet bør afvikles. For en beskrivelse af til hvilke formål en test baseret på dette materiale anvendes og hvordan en test typisk afvikles, henvises til det angivne bilag.

Af Keidser (1994b) er der blevet udført en undersøgelse, hvori materialet blev afviklet som en 4AFC-test. Af ordene i de otte ordlister til voksne blev der dannet 25 grupper af fire ord – de fire ord udgjorde tilnærmelsesvis minimalpar[5]. Et eksempel på en sådan ordgruppe er ordene *hest*, *fest*, *gæst* og *rist*. I undersøgelsen skulle FP'en afgive sit svar ved at trykke på en trykfølsom skærm, der viste de fire svarmuligheder – dvs. i stil med den patientbaserede version af det testprogram, hvis parametre undersøges i dette projekt. Var det præsenterede ord eksempelvis *mand*, kunne svarmulighederne være *sand*, *mand*, *spand*, *tand*, *and*, *land* og *vand*. I undersøgelsen blev det observeret, at hvis en eller flere af svarmulighederne indeholdt en let opfattelig lyd, som eksempelvis lyden *s* kunne denne lyd detekteres før, det afspillede ord kunne identificeres. Selv samme ord kan altså få varierende sværhedsgrad, hvis dette gennem en test ikke præsenteres med de samme svarmuligheder. Med Dantale I materialet er der i øvrigt også udført andre dokumenterede undersøgelser – se evt. Olsen, (1996), Keidser (1993) og (1994a).

For at kunne præsentere Dantale I materialet audiovisuelt blev det i 1998 udgivet på vhs og i 1999 på dvd. Grundet en mistanke om at der forekom visse niveauforskelle mellem cd-udgaven og dvd-udgaven, er der af DELTA (2003)* blevet udført en undersøgelse heraf. Undersøgelsen bekræftede mistanken, hvorfor resultater fra de to medier må sammenlignes med forsigtighed – dette gælder dog kun, hvis dvd'ens surround-signal anvendes frem for stereo-signalet. De audiovisuelle indspilninger anvendes jf. spørgeskema-undersøgelsen bl.a. til at vurdere graden af et kommunikativt handicap grundet en evt. arbejdsbetinget høreskade – herunder ved indberetninger til arbejdsskadestyrelsen. I forbindelse med en sådan vurdering testes FP'en både rent auditivt og audiovisuelt – baggrunden er, at der jf. Salomon *et al*. (1985) bør skelnes mellem et egentligt høre- og et kommunikationshandicap. Imidlertid kan der herske tvivl om, hvorvidt et resultat af en sådan test er sammenligneligt med FP'ens egen oplevelse af sit handicap (Ludvigsen, 2005).

[5] Et minimalpar er ord, som kun adskiller sig fra hinanden ved et enkelt fonem.

2.2 Sætningsbaserede materialer

Ud over talematerialer baseret på enkeltord, findes der også materialer indeholdende hele sætninger. Det første danske materiale baseret udelukkede på sætninger kaldes Helen-materialet og blev konstrueret i starten af 1970'erne – stort set samtidigt med udviklingen af GS-listerne. I det følgende beskrives både Helen-materialet og HINT-testen – som skrevet behandles Dantale II materialet i kapitel 3. Generelt kan en testsituation opbygget omkring et sætningsbaseret materiale frem for et materiale baseret på enkeltord anses som mere virkelighedsnær – med andre ord opnås en mere valid test. Det skyldes, at der i en sætningsbaseret test også indgår FP'ens evne til at bruge kontekstuelle informationer i taleopfattelsen. Hvorvidt dette er tilfældet for netop Dantale II materialet kan dog diskuteres, idet sætningerne her kan betragtes som enkeltord i sekvens – se evt. afsnit 3.1.1.

2.2.1 Helen-materialet

Formålet med udviklingen af Helen-materialet var, at det skulle kunne anvendes i forbindelse med test af svært hørehæmmede personer, hvorfor dette også er indspillet så en audiovisuel præsentation er mulig (Ludvigsen, 1974). En omfattende beskrivelse af selve materialet og dets opbygning fremgår i to monografier – Ewertsen (1973) og Ludvigsen (1981). Materialet består af otte lister af hver 25 forholdsvis simple spørgende sætninger og en tilhørende baggrundsstøj. Som eksempel lyder to af disse sætninger; *Hvilken dag kommer før fredag?* og *Hvor mange ben har en hest?* Responset fra FP'en kræver blot et enkelt ord, der afslører hvorvidt FP'en har forstået og derved hørt den pågældende sætning.

Materialet er blevet brugt i en kortere periode på Bispebjerg Hospital. Herefter blev det ikke brugt længe, indtil det på ny blev taget i anvendelse i midten af 1980'erne på Århus Hospital i forbindelse med behandling af voksne med cochleaimplantat. I den forbindelse blev materialet genindspillet, idet det format, den visuelle del oprindeligt var optaget i, ikke længere blev brugt (Ardenkjær-Madsen & Josvassen, 2001). Ifølge spørgeskema-undersøgelsen, som er beskrevet er i Bilag A, anvendes materialet stadig i et begrænset omgang den dag i dag. Der en dog en oplagt ulempe ved at udføre en høretest vha. spørgende sætninger. FP'en kan risikere at få den opfattelse, at det er vedkommendes intelligens eller almenviden frem for hørelsen, der testes. For at dette ikke skal kunne påvirke testens validitet og derved reliabilitet, må denne problematik vurderes.

2.2.2 HINT

I 2004 påbegyndte Danmarks Tekniske Universitet (DTU) under *Center for Anvendt Høreforskning* arbejdet med at konstruere et materiale til den såkaldte danske HINT-test (Hearing In Noise Test). Tilsvarende materialer findes på bl.a. tysk, hollandsk og engelsk – for en beskrivelse af hver af disse henvises hhv. til Kollmeier & Wesselkamp (1997), Plomp & Mimpen (1979) og Nilsson *et al*. (1994). Det danske materiale består af tre lister tiltænkt træning og 25 egentlige testlister med hver 13 sætninger, der er forsøgt udformet som hverdagssætninger (Nielsen, 2005).

Hver sætning indeholder fem ord, som samlet består af otte-ni stavelser, hvoraf det enkelte ord højst indeholder fire. Hver sætning er opbygget således, at der indgår et udsagnsord, som dog ikke er på navneform, mens der i modsætning til Dantale II materialet ikke indgår navneord. Sætningerne er grammatisk korrekte og er ikke udformet som hverken et ordsprog eller et spørgsmål – indholdet er så vidt muligt meningsfyldt og semantisk neutralt. De er indspillet direkte med en mandlig indtaler, hvilket er i modsætning til Dantale II materialet, som er indspillet for at blive segmenteret og derpå sat sammen på ny – se evt. afsnit 3.1.4. Som en afsluttende bemærkning kan nævnes, at der i skrivende stund er ved at blive udarbejdet en tidsskiftsartikel om netop den danske HINT-test (Jens Bo Nielsen – DTU, pers. medd. maj 2007).

2.3 Sammenfatning

I tidens løb er der blevet udviklet forskellige danske talematerialer – nogle af dem er her blevet beskrevet. De materialer der typisk har været anvendt i den kliniske praksis har været baseret på enkeltord. Dette gælder også den dag i dag, idet det hyppigst anvendte materiale er ord- og talordslisterne fra Dantale I. Desuden synes der ikke at være tradition for at lade testafviklingen være baseret på patientselvbetjening, som det gør sig gældende for den patientbaserede version af det testprogram, som undersøges i forbindelse med dette projekt. På trods heraf er der med ordlisterne til voksne i Dantale I blevet udført en ganske interessant undersøgelse, hvori materialet blev afviklet som en 4AFC-test. Af undersøgelsen blev det fundet, at selv samme ord kan få varierende sværhedsgrad, hvis dette gennem en test ikke præsenteres med de samme svarmuligheder.

3 Talematerialet Dantale II

Dantale II er et dansk sætningsbaseret talemateriale, som blev lanceret på cd i efteråret 2001. Benævnelsen Dantale II anvendes undertiden som en fællesbetegnelse for selve materialet og en tilhørende afviklingsprocedure – i denne rapport anvendes benævnelsen imidlertid kun om selve materialet. Materialet består af i alt 16 lister med hver ti testsætninger – samtlige lister er vedlagt i Appendiks. Foruden testsætningerne, der af og til kaldes de danske Hagerman sætninger, indbefatter materialet et taleformet støjsignal – dette samt en indspillet version af listerne findes på vedlagte cd-rom. Da det ikke har været muligt at finde en egentlig betegnelse for støjsignalet, er det i denne rapport valgt at referere til det pågældende som Dantale II-støjen.

Dantale II materialet er opbygget efter hhv. et tilsvarende svensk og tysk materiale. Det svenske materiale blev udviklet i staren af 1980'erne af Björn Hagerman – heraf benævnelsen de danske Hagerman sætninger. Det tyske materiale, som indgår i Oldenburger Satztest, blev derimod først udviklet i slutningen af 1990'erne af Kirsten Wagener. For en beskrivelse af disse materialer henvises til hhv. Hagerman (1984) og Wagener *et al*. (1999a-c). Materialerne, der løbende gennem rapporten omtales som det svenske og tyske materiale, adskiller sig ved måden, hvorpå disse er indspillet – mere herom i afsnit 3.1.3. Udover materialer på svensk, tysk og dansk er der i skrivende stund i forbindelse med et internationalt projekt kaldet HearCom[6] ved at blive konstrueret tilsvarende materialer på engelsk, hollandsk og fransk. Samtidigt er det også under overvejelse at konstruere et materiale på polsk (Johannes Lyzenga – HearCom, pers. medd. feb. 2007). I de følgende afsnit beskrives Dantale II materialet – såvel testsætningerne som støjsignalet. Beskrivelserne er i store træk baseret på Ardenkjær-Madsen & Josvassen (2001) og Wagener *et al*. (2003).

3.1 Testsætningerne

Talematerialet Dantale II er udarbejdet af to specialestuderende i faget audiologopædi ved Københavns Universitet. Selve indspilningen og den efterfølgende generering af de egentlige testsætninger blev imidlertid foretaget med hjælp fra Kirsten Wagener. I de følgende afsnit beskrives ganske kort testsætningernes struktur og deres tilblivelse – herunder baggrunden for ordvalget i den såkaldte basisliste samt hvordan indspilningen af denne og genereringen de endelig sætninger er foretaget. Derudover beskrives målgruppen og hvordan denne kan udvides ved at reducere sætningslængden. Det bør nævnes, at den patientbaserede version af det i dette projekt undersøgte testprogram imidlertid kan sætte en begrænsning i målgruppen.

[6] Projektet HearCom (Hearing in the Communication Society) startede i 2004 og har til formål at harmonisere en række audiologiske tests på tværs af Europa – herunder taletests. Projektet er en videreudvikling og udvidelse af det nu afsluttede projekt NATASHA (Network and Tools for the Assessment of Speech/Language and Hearing Ability).

3.1.1 Sætningsstrukturen

Samtlige testsætninger er opbygget af fem nøje udvalgte ord – udelukkende af en- og tostavelsesord. De er samtidigt konstrueret med samme grammatiske struktur således, at de fem ord tilhører bestemte ordklasser, der i hver sætning angives i følgende rækkefølge; egennavn, udsagnsord, talord, tillægsord og til sidst et navneord. Eksempelvis lyder den første sætning i liste 1; *Ingrid finder syv røde huse*. Alle sætningerne er syntaktisk mulige men semantisk usandsynlige – dvs. sætningerne forekommer ikke som værende meningsfyldte og de enkelte ord kan derfor ikke forudsiges ud fra konteksten. Da mængden af redundans er mindsket i forhold til naturlig tale, kan ordene i disse nonsens-sætninger i en vis grad betragtes som enkeltord i sekvens. I forbindelse hermed bør der i en egentlig test derfor scores på antallet af korrekt gengivne ord – dvs. der bør anvendes ordscoring. Alternativt kunne der anvendes fonem- eller sætningsscore, hvilket dog ikke er materialets hensigt. Da sætningerne kan betragtes som enkeltord i sekvens, vil en test opbygget omkring disse være hurtigere at afvikle end en decideret enkeltordstest, som eksempelvis en test baseret på ordlisterne fra materialet Dantale I. Dette gælder, hvis der med de to tests ønskes det samme antal ord præsenteret.

For at beskrive forudsigeligheden af testsætningernes enkelte ord ud fra konteksten – dvs. graden af sætningernes redundans – anvendes begrebet *j*-faktoren. Denne faktor, som blev indført af Boothroyd & Nittrouer (1988), angiver antallet af uafhængige elementer i en sætning og kan beregnes ud fra udtrykket i formel (3-1). I udtrykket er $p_{\text{sætning}}$ og p_{ord} sandsynligheden for hhv. korrekt gengivelse af en hel sætning og korrekt gengivelse af et enkelt ord.

$$j = \frac{\log(p_{\text{sætning}})}{\log(p_{\text{ord}})} \tag{3-1}$$

Jo højere værdi *j*-faktoren har, desto mere uforudsigelige er sætningernes enkelte ord. Eksempelvis vil ordsproget; *Alle veje fører til Rom* have en højere forudsigelighed og derved en lavere *j*-faktor end testsætningen; *Ingrid finder syv røde huse*. Da et sætningsmateriale med en høj *j*-faktor indeholder flere uafhængige scoringselementer end et materiale med en lav *j*-faktor, formodes materialet med den høje *j*-faktor at resultere i den mest reliable test.

I litteraturen findes der ingen opgørelse over, hvad *j*-faktoren er i forbindelse de danske testsætninger, hvorfor denne er blevet bestemt i forbindelse med Pilottest II. Af testen blev det bestemt, at for såvel den operatør- som for den patientbaserede version kan sætningerne betragtes som bestående af fire uafhængige elementer – se evt. figur C-24, s. 147. Dette er i overensstemmelse med det, der i litteraturen er opgivet for det svenske og tyske materiale (Wagener *et al*., 1999b). Til sammenligning kan nævnes, at for et tysk materiale kaldet Göttingen, hvis opbygning svarer til den danske HINT-test, er *j*-faktoren blevet bestemt til omkring to (Kollmeier & Wesselkamp, 1997). Sætningerne i dette materiale, som også består af hver fem ord, kan altså betragtes som bestående af to uafhængige elementer.

3.1.2 Udvælgelse af basislistens ord

Som de svenske og tyske testsætninger er de danske dannet ud fra en såkaldt basisliste bestående af en pulje af 50 forskellige ord – ligeligt fordelt på de fem ordklasser. Som skrevet er testsætningerne grupperet i 16 lister af hver ti sætninger, hvor hver sætning består af fem ord. I hver liste indgår altså 50 ord svarende til, at hvert af basislistens ord netop er repræsenteret én gang. Basislisten fremgår af tabel 3-1.

Tabel 3-1: Basislisten hvorfra samtlige testsætninger er dannet.
(Gengivelse af Wagener *et al*., 2003 – Tabel 2).

Egennavn	Udsagnsord	Talord	Tillægsord	Navneord
Anders	ejer	ti	gamle	jakker
Birgit	havde	fem	røde	kasser
Ingrid	ser	syv	pæne	ringe
Ulla	købte	tre	nye	blomster
Niels	vandt	seks	fine	skabe
Kirsten	får	tolv	flotte	masker
Henning	solgte	otte	smukke	biler
Per	låner	fjorten	store	huse
Linda	valgte	ni	hvide	gaver
Michael	finder	tyve	sjove	planter

For at undgå en evt. lydmæssig ubalance i forhold til de relative lydforekomster i det danske sprog er basislistens ord hverken en oversættelse af ordene i det svenske eller tyske materiale. I stedet er listens ord udvalgt på baggrund af en analyse fra 1992 af de 5.000 hyppigste skrevne danske ord. Ud fra denne analyse er ordene forsøgt valgt således, at disse hverken er følelsesladede, stødende eller på anden måde geografisk, socialt eller fagligt betingede. Derved undgås, at sådanne ord måske vil kunne påvirke FP'ens respons og derved gøre denne inkonsistent i forhold til præsentationsniveauet.

Ordene er i øvrigt valgt således, at halvdelen af sætningerne er i nutid, mens den anden halvdel er i datid. Derved undgås samme tempusform, hvorved endelserne i udsagnsordene ellers ville have været ens. Desuden er egennavnene ligeligt fordelt på drenge- og pigenavne. For at sikre at de valgte ords sproglyde er repræsentative for det danske sprog, er der i Ardenkjær-Madsen & Josvassen (2001 – Bilag 3) udført en sammenligning mellem den relative forekomst af talematerialets enkelte sproglyde og den relative forekomst af sproglydene i de 5.000 hyppigste skrevne danske ord. Denne sammenligning blev udført ved at foretage en Pearsons korrelationsanalyse, hvoraf korrelationsfaktoren blev bestemt til 0,97. Værdien af denne faktor fremgår også af Wagener *et al*. (2003).

Ud fra Ardenkjær-Madsen & Josvassens data er det i forbindelse med udarbejdelsen af dette projekt vha. to Kolmogorov-Smirnov tests blevet konstateret, at hverken værdierne af den relative forekomst af talematerialets enkelte sproglyde eller den relative forekomst af sprog-

lydene i de 5.000 hyppigste skrevne danske ord kan antages at være normalfordelt. Derfor bør der i stedet for at bestemme Pearsons korrelationsfaktor udføres en Spearmans korrelationsanalyse. Denne analyse er et alternativ til Pearsons korrelation og er i stedet for gennemsnit og spredninger baseret på såkaldte rangeringer. Ved at udføre en sådan analyse er Spearmans korrelationsfaktor blevet bestemt til 0,91, hvilket er en værdi lidt lavere end den bestemt af Ardenkjær-Madsen & Josvassen. På trods heraf er sammenhængen mellem de to grupper af sproglyde fundet statistisk signifikant, hvorfor materialet kan betragtes som værende fonetisk balanceret.

I forbindelse med udvælgelsen af ordene blev der af Ardenkjær-Madsen & Josvassen udført en lyttetest, hvori ti normalthørende FP'er blev præsenteret for ti tilfældige lister – dvs. i alt 100 sætninger. Formålet med testen var, at FP'erne skulle vurdere, hvorvidt sætningernes meningsindhold var så ualmindeligt, at opmærksomheden blev flyttet fra selve lytteopgaven. På baggrund af testen blev det besluttet at udskifte navneordet *marker*, idet det ved kombination af de øvrige ord – særligt udsagnsordene – resulterede i et tilsyneladende ualmindeligt meningsindhold af de pågældende sætninger. At dette var tilfældet begrundes med, at det umiddelbart virker underligt, at en person *låner*, *finder* etc. et givet antal *marker*. I stedet for ordet *marker* blev ordet *biler* indført. Denne ordændring har imidlertid ikke haft indflydelse på den fonetiske balancering. I Pilottest II blev ordet *biler* dog fundet sværere end de øvrige navneord – se evt. figur C-25, s. 150.

3.1.3 Indspilning af basislisten

Det svenske materiale er i modsætning til det tyske indspillet ved kun at medtage basislistens ti sætninger. Disse er indspillet med en kort pause mellem ordene, hvorved der ikke er taget hensyn til koartikulationen – dvs. til nabolyde på tværs af de forskellige ordgrænser. Med andre ord er der altså ikke taget højde for den gensidige påvirkning af udtalen, der naturligt kan optræde mellem tilstødende ord. Efter indspilningen blev ordene klippet fra hinanden, hvorefter disse på passende vis – i overensstemmelse med den fastsatte syntaks – blev kombinerede til nye sætninger (Hagerman, 1984).

Det tyske materiale er imidlertid indspillet således, at der er taget højde for koartikulationen (Wagener, 1999c). Den tyske indspilningsteknik er dog vanskeligere og dermed mere tidkrævende end den svenske. Teknikken har til gengæld resulteret i, at det tyske materiale har en mere naturlig intonation – dvs. tonegang – og derved i højere grad afspejler en virkelig kommunikationssituation end hvad, der er tilfældet for det svenske materiale – dvs. den endelig test er mere valid. Grundet sidstnævnte er den tyske teknik derfor blevet anvendt i forbindelse med indspilningen af det danske materiale. Denne blev udført i år 2000 på Carl von Ossietsky Universität i Oldenburg, som er beliggende i Nordtyskland. Oprindeligt var hensigten, at det danske materiale skulle indspilles audiovisuelt, men på nuværende tidspunkt har der dog kun fundet en ren auditiv indspilning sted.

Indspilningen forgik ved, at der ud fra basislisten blev dannet i alt 100 forskellige sætninger. Disse blev konstrueret således, at de enkelte ord kun forekom én gang i kombination med de ti mulige foregående og/eller efterfølgende ord. Den første sætning, der blev indspillet, lød således; *Anders ejer ti gamle jakker*, den næste; *Anders havde ti røde jakker*, den næste igen; *Anders ser*

ti pæne jakker og så fremdeles. Ved at indspille alle 100 sætninger er samtlige koartikulationer af nabolyde på tværs af de forskellige ordgrænser medtaget. Efter indspilningen blev sætningerne segmenteret, hvorefter de endelige testsætninger blevet genereret – mere herom i følgende afsnit. Det kan i øvrigt nævnes, at det danske materiale blev indspillet med kvindelig indtaler. Dette gør sig også gældende for det svenske materiale, mens det tyske er indspillet med en mandlige indtaler.

3.1.4 Testsætningernes generering

Ud fra indspilningerne af basislisten blev testsætningerne genereret ved at kombinere et af hvert af de ti alternative ord fra hver ordklasse. I forbindelse hermed blev hver af de 100 indspillede sætninger opdelt i fire segmenter. Opdelingen foregik ved at adskille ordene tæt på begyndelsen til det efterfølgende ord, hvorved koartikulationen mellem det pågældende ord og det efterfølgende ord blev medtaget. Sætningernes to sidste ord blev ikke adskilt. Et eksempel på denne segmentering fremgår af følgende illustration.

Anders ~ ✂ ejer ~ ✂ ti ~ ✂ gamle ~ jakker

Figur 3-1: Opdelingen af den første sætning i basislisten. Koartikulationen mellem og adskillelsen af de enkelte segmenter er angivet med hhv. en bølgestreg og en saks.

(Forsimplet gengivelse af Wagener *et al*., 2003 – Figur 2).

Ved at kombinere segmenterne således, at der blev taget højde for den korrekte koartikulation, blev de egentlige testsætninger konstrueret. Som eksempel fremgår det af figur 3-2, at den første sætning i liste 1 er konstrueret ved at vælge den indspilning af ordet *Ingrid*, som er indspillet sammen med ordet *finder*, mens ordet *finder* skal være indspillet sammen med ordet *syv* og så fremdeles. Det kan i øvrigt nævnes, at hver sætning tager fra 2 til 4 sek. at afspille.

Ingrid ~ (finder) + finder ~ (syv) + syv ~ (røde) + røde ~ huse

Figur 3-2: Konstruktion af den første sætning i liste 1, hvor der er taget højde for koartikulationen. Som i figur 3-1 er koartikulationen angivet med en bølgestreg, mens sammenkædningen af de enkelte segmenter er angivet med et additionstegn.

(Forsimplet gengivelse af Wagener *et al*., 2003 – Figur 2).

3.1.5 Materialets homogenitet

Selvom der ud fra indspilningerne er potentiale for at danne i alt 10^5 forskellige sætninger, blev der kun konstrueret 25 testlister af hver ti sætninger. I hver liste optræder som tidligere skrevet hvert af basislistens 50 ord netop én gang. På baggrund af en lyttetest med 16 normalthørende FP'er, blev eventuelle forskelle mellem ordenes sværhedsgrad forsøgt udlignet (Wagener *et al*., 2003). Dette blev gjort med en operatørbaseret afvikling ved at bestemme den såkaldte S-kurve

for hvert enkelt ord. S-kurven, som beskriver sandsynligheden for korrekt svar som funktion af præsentationsniveauet, behandles i afsnit 4.1.1. Ud fra de fundne S-kurver blev niveauet af nogle af ordene justeret. For at bevare en naturlig intonation blev niveauet af hvert ord maks. justeret med ±4 dB. Derefter blev de 16 mest homogene lister udvalgt – dvs. i alt 160 sætninger.

Denne optimeringsproces resulterede i, at hældningen af materialets samlede S-kurve ved kurvens midtpunkt – kaldet s_{50} – blev større end før optimeringsprocessen, hvorved der kan opnås en test med en større grad af reliabilitet grundet en mindre re-test variation – se evt. kapitel 4. At værdien af s_{50} blev større kan forklares med følgende formel, som stammer fra Wagener *et al*. (2003). I formlen angiver $\sigma_{SRT_{50,\mathrm{ord}}}$ og $s_{50,\mathrm{ord}}$ hhv. spredningen af ordenes SRT_{50} og den gennemsnitlige hældning af ordenes S-kurver ved SRT_{50}. Hvert ords værdi af SRT_{50} svarer til det præsentationsniveau, hvorved FP'erne forventes at kunne forstå det pågældende ord 50% af de gange, det præsenteres. Som det ses af formlen, kræver en stejl hældning af materialets samlede S-kurve en lille spredning af ordenes SRT_{50} samt en stejl gennemsnitlig hældning af ordenes S-kurver. Til formlen kan nævnes, at denne er udledt ved at antage, at materialets samlede S-kurve kan beskrives som en foldning mellem den for ordene gennemsnitlige S-kurve og fordelingen af de tilsvarende værdier af SRT_{50} – for en udledning henvises til Wagener *et al*. (1999a).

$$s_{50} \approx \frac{s_{50,\mathrm{ord}}}{\sqrt{1+\frac{16\cdot {s_{50,\mathrm{ord}}}^2\cdot {\sigma_{SRT_{50,\mathrm{ord}}}}^2}{\left(\ln\left(2\cdot e^{1/2}-1+2\cdot e^{1/4}\right)\right)^2}}} \tag{3-2}$$

På trods af optimeringsprocessen kan der stilles spørgsmålstegn ved, hvorvidt materialet i en egentlig testafvikling kan antages at være homogent og om graden af homogenitet evt. er afhængig af responsmetoden. I tilknytning til sidstnævnte virker det ikke urimeligt at tro, at nogle specifikke ord måske vil være relativt sværere i den operatør- end i den patientbaserede version. Da en manglende homogenitet vil kunne påvirke en tests samlede reliabilitet, er homogeniteten derfor blevet undersøgt i forbindelse med Pilottest II. Af testen blev der imidlertid ikke fundet nogen forskel mellem homogeniteten i den operatør- og den patientbaserede version – se evt. figur C-26, s. 152. Som følge heraf blev der ikke fundet nogen statistisk signifikant forskel mellem hældningen af S-kurven versionerne imellem – se evt. figur C-22, s. 145. På tværs af versionerne blev det dog fundet, at nogle ord generelt synes at være sværere eller lettere end den gennemsnitlige sværhedsgrad – se evt. figur C-25, s. 150. Dette gjaldt som tidligere skrevet bl.a. for *ordet* biler, som synes at være sværere end de resterende navneord. Desuden synes navnet *Anders* at være lettere, mens navnet *Ingrid* synes at være sværere end de resterende egennavne.

I forbindelse med Pilottest I blev det undersøgt, om det er rimeligt at antage, at de 16 lister er lige svære. For både den operatør- og den patientbaserede version blev der ikke fundet nogen signifikant forskel mellem listernes sværhedsgrad. Hertil bør det dog bemærkes, at den beregnede spredning af værdierne af SRT_{50} for hver liste måske er større end den faktiske – se forklaring i forbindelse med figur C-10, s. 125. At der ikke er forskel mellem listerne stemmer i øvrigt fint overens med en undersøgelse af Wagener *et al*. (2003) foretaget med en operatørbaseret afvikling, hvoraf det fremgår, at spredningen af listernes SRT_{50} er mindre end spredningen af SRT_{50} for 60 normalthørende FP'er.

3.1.6 Reduktion af sætningslængden

Ifølge Ardenkjær-Madsen & Josvassen (2001) er talematerialet Dantale II designet til at teste voksne HA-kandidater og -brugere. Pga. den store mængde af uforudsigelige testsætninger, er disse så godt som umulige at huske udenad, hvorfor materialet kan anvendes gentagne gange med samme FP. Dette er en klar fordel i forbindelse med bl.a. evaluering af HA'er og HA-tilpasninger. I tilknytning hertil kan dog tænkes, at hvis en FP først testes med den patient- og umiddelbart derefter med den operatørbaserede version vil testresultatet af sidstnævnte kunne blive påvirket af, at FP'en via den patientbaserede version har set hvilke ord, der indgår i materialet. Da testsætningerne er indspillet med en forholdsvis langsom talehastighed, er materialet jf. Wagener *et al*. (2003) desuden velegnet til vurdering af cochleaimplanteredes taleopfattelse.

I modsætning til Dantale I er der i Dantale II materialet ikke medtaget specifikke lister til afvikling af tests med børn og evt. udviklingshæmmede. Selvom målgruppen i første omgang altså ikke er tiltænkt at indbefatte disse persongrupper, kan materialet muligvis anvendes alligevel – en test med sætninger af fem ord kan dog være for uoverskuelig og derved for svær. I forbindelse med det svenske materiale nævnes, at sætningslængden med fordel kan reduceres – enten kan tillægsordene eller både egennavnene og udsagnsordene fjernes (Hagerman, 1984). Pga. den i afsnit 3.1.3 beskrevne indspilningsteknik, der tager højde for koartikulationen, er det imidlertid ikke muligt kun at fjerne tillægsordene i de danske testsætninger. Med det tyske materiale, der er indspillet på samme vis som det danske, er der blevet foretaget en undersøgelse med en reduceret sætningslængde, hvor både egennavnene og udsagnsordene var fjernet (Wagener, 2005). Af undersøgelsen blev det konkluderet, at aldersgrænsen for test med den fulde sætningslængde er 8-10 år og at yngre børn bør testes med den reducerede sætningslængde – dog nævnes ingen absolut nedre aldersgrænse.

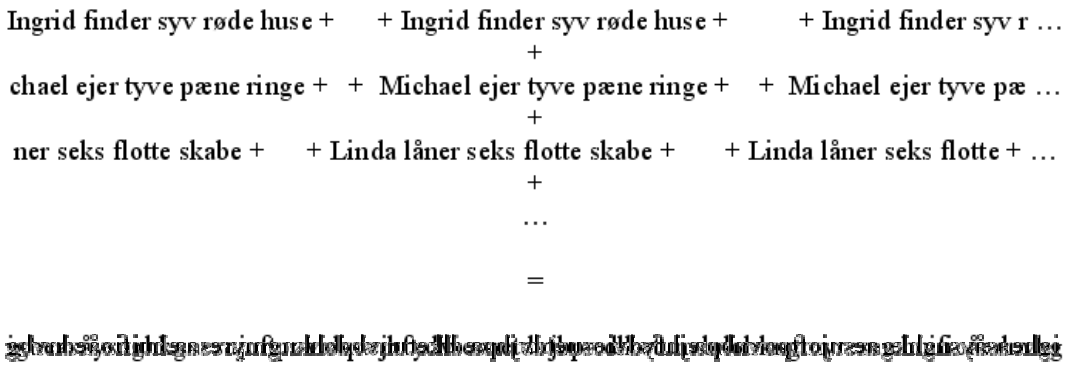
Ved en reduktion af sætningslængden bør den fonetiske balancering kontrolleres. En sådan kontrol har det imidlertid ikke været mulig at finde i litteraturen – hverken for det svenske eller det tyske materiale. Ud fra Ardenkjær-Madsen & Josvassen (2001 – Bilag 2) er det dog muligt, at optælle den samlede relative forekomst af sproglydene i det danske materiale, når egennavnene og udsagnsordene fjernes – en sådan optælling er derfor blevet udført i forbindelse med dette projekt. Efterfølgende er der tilsvarende afsnit 3.1.2 blevet foretaget en sammenligning mellem værdierne af de relative forekomster af disse sproglyde og den relative forekomst af sproglydene i de 5.000 hyppigste skrevne danske ord. Sammenligningen, som blev foretaget ved at udføre en Spearmans korrelationsanalyse, resulterede i korrelationsfaktoren på 0,69. Denne værdi er lidt lavere end værdien på 0,91, som blev bestemt i afsnit 3.1.2 gældende for det samlede materiale. Ved at fjerne de to ordklasser bliver materialets sproglyde altså mindre sammenlignelige med sproglydene i de 5.000 hyppigste skrevne danske ord, hvilket evt. kan have betydning for testens validitet.

3.2 Dantale II-støjen

Ofte er der i forbindelse talematerialer genereret et tilhørende støjsignal, hvorved støjsignaler på tværs af forskellige tests ikke er ens – dette gør sig også gældende for Dantale II materialet. Støjsignaler dannes typisk ved at foretage gentagne tidsforskudte overspilninger af materialets testsætninger. Derved opnås et signal med samme langtidsspektrum som det for summen af alle sætningerne. Alternativt kan støjsignalet dannes ved at generere et signal, hvis frekvensspektrum matcher sætningernes – se evt. Nilsson *et al*. (1994). Et taleformeret støjsignal er ideelt rent frekvensmæssigt til at maskere testsætningerne, hvilket har betydning for testens reliabilitet – herunder ordenes indbyrdes sværhedsgrad. Imidlertid er den maskerende effekt også afhængig af niveauet af tale- og støjsignalet samt af modulationen over tid. I det følgende afsnit fremgår, hvordan Dantale II-støjen er genereret samt en analyse af støjens frekvensindhold og modulationsgrad.

3.2.1 Støjens generering

Dantale II-støjen er genereret ud fra testsætningerne. Det foregik ved, at der for hver af de i alt 160 sætninger blev dannet en sekvens af den pågældende sætning på 2,5 min. Mellem hver gentaget sætning i sekvensen blev der indlagt et stilleinterval, som tilfældigt blev valgt mellem 5 ms og 2 s. Samtidigt blev startpunktet af den første sætning i hver sekvens valgt tilfældigt. Derefter blev sekvenserne afspillet oveni hinanden 30 gange, hvorved det egentlige støjsignal fremkom. Dette er forsøgt illustreret i figur 3-3. Grundet testsætningernes komplekse sammenkopiering minder støjsignalet ved lytning imidlertid ikke meget om virkelige tale, men mere om hvidstøj – som det vil fremgå af følgende afsnit har støjen dog et frekvensspektrum meget nær det for sætningerne. I stedet for at anvende Dantale II-støjen kan andre støjsignaler imidlertid også tænkes anvendt – forskellige forslag hertil er henlagt til kapitel 6.



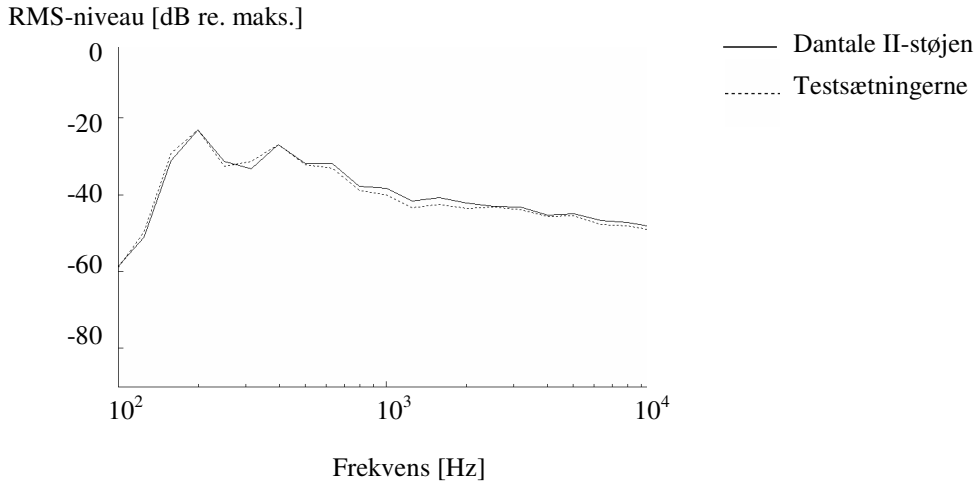
Figur 3-3: Illustration af hvordan Dantale II-støjen er genereret. Ud fra hver sætning blev der dannet en sekvens af den pågældende sætning på 2,5 min. Mellem hver gentaget sætning blev der indlagt et stilleinterval og startpunktet af den første sætning i hver sekvens blev valgt tilfældigt. Ved at afspille sekvenserne oveni hinanden fremkom det egentlige støjsignal. Sammenkædningen af sætninger og stilleintervaller samt sammenlægningen af de heraf kommende sekvenser er angivet med et additionstegn.

(Modificeret gengivelse af Wagener *et al*., 2003 – Figur 3).

3.2.2 Frekvens- og modulationsanalyse

For at beskrive Dantale II-støjen er dets frekvensspektrum beregnet med 1/3 oktavbånds IIR-filtre (Infinite Impulse Response)[7] med en filterorden på tre – såvel centerfrekvenser som båndbredder er beregnet som angivet i IEC 61260, § 3. Af figur 3-4 fremgår både støjens og sætningernes samlede frekvensspektrum – dvs. de såkaldte langtidsspektre. Frekvensen er angivet fra 100 Hz til 10 kHz, mens RMS-niveauet (Root Mean Square) er opgivet relativt til det maksimale niveau i computeren. I øvrigt bør nævnes, at de i figuren angivne frekvensspektre også er beregnet med 1/3 oktavfiltre med en båndbredde på en 1/1 oktav. Dette resulterede imidlertid i spektre, som var mere udglattede og derved ikke indeholdt så mange detaljer som de i figuren – bl.a. var dykket ved ca. 300 Hz ikke synligt. Disse spektre er derfor ikke valgt gengivet.

Som det ses af figuren, er der en lille forskel mellem niveauet af de to spektre. Denne forskel kan tillægges, at støjen er genereret ud fra sætningerne inden den i afsnit 3.1.5 omtalte optimeringsproces blev udført. Det bør i den forbindelse bemærkes, at optimeringsprocessen ikke kan udføres førend støjen er genereret. Imidlertid kunne støjen korrigeres således, at niveauet ved samtlige frekvenser er identisk med det for sætningerne. Forskellen mellem de to spektre er dog lille, hvorfor en sådan korrektion ikke vurderes til at have megen betydning for resultatet af en egentlig testafvikling.



Figur 3-4: Langtidsspekteret for Dantale II-støjen og testsætningerne – disse er hhv. angivet med en fuldtoptrukken og prikket linje. Begge spektre er beregnet med 1/3 oktavbånd og RMS-niveauet er opgivet relativt til det maksimale niveau i computeren.

[7] Et IIR-filter har en uendelig impulsrespons og er kendetegnet ved at have poler, hvilket kan gøre det ustabilt – til gengæld kan der opnås en stor flankestejlhed i forhold til den valgte filterorden. For en uddybende beskrivelse af IIR-filtre henvises til Hüche (1998 – Kap.7).

De i figuren angivne spektre er i overensstemmelse med tilsvarende spektre bestemt af Wagener *et al*. (2003). Af Wagener *et al*. fremgår desuden, at spektrene har samme form som LTASS (Long-Term Average Speech Spectrum) svarende til middel langtidsspektret for gennemsnitlig tale på tværs af 12 forskellige sprog herunder dansk. For en beskrivelse af hvordan LTASS er bestemt henvises til Byrne *et al*. (1994). Da langtidsspektret stort ses er ens for forskellige sprog, kan det overvejes at anvende det samme støjsignal på tværs af disse. Et sådant universelt støjsignal kunne være en af de ni såkaldte ICRA-støje (International Collegium of Rehabilitative Audiology) – disse behandles i afsnit 6.1. De forskellige ICRA-støje adskiller sig fra hinanden ved bl.a. at have en forskellig modulationsgrad. Modulationsgraden beskriver, hvor meget amplituden af indhylningskurven varierer i forhold til den gennemsnitlige amplitude.

Af den opgivne litteratur har det desværre ikke været muligt at finde dokumentation for Dantale II-støjens modulationsgrad, hvorfor den her er forsøgt bestemt. Måden hvorpå denne er bestemt forklares i det følgende og svarer i store træk til den angivet i Dreschler *et al*. (2001). Først blev indhylningskurven af signalet bestemt – denne blev beregnet ud fra det analytiske signal[8]. Derefter blev signalet lavpasfiltreret med et 2048-ordens FIR-filter (Finite Impulse Response)[9] med en knæk-frekvens på 50 Hz. Signalet blev lavpasfiltreret, idet det netop er den lavfrekvente modulation, som har betydning for, om signalets intensitet opfattes som værende varierende. I forbindelse hermed bør det nævnes, at modulationer i almen tale befinder sig området 3-4 Hz (Wagener *et al*., 2006). Ordenen på lavpasfiltret blev valgt som et kompromis mellem at undgå aliasing[10] og en rimelig beregningstid, mens knæk-frekvensen blev valgt på 50 Hz som et kompromis mellem filterordnen og en minimal dæmpning ved de frekvenser, der ønskes analyseret.

Efter denne filtrering blev signalet downsamplet med en faktor 210. Samplefrekvensen var derved 210 Hz og ikke 44,1 kHz, som er standard for bl.a. lyd lagret på cd – herunder Dantale II-cd'en. Denne downsampling var nødvendig for at kunne analysere i hvor høj grad, de enkelte frekvenser bidrager til modulationen. Analysen af indhylningskurven er udført med 1/3 oktavbånd på samme måde som frekvensanalysen af støjsignalet – frekvenserne er dog her opgivet fra 0,5 til 20 Hz. Niveauet er samtidigt valgt relativt til indhylningskurvens gennemsnitlige niveau dvs. svarende til netop modulationsindekset for de behandlede frekvensbånd.

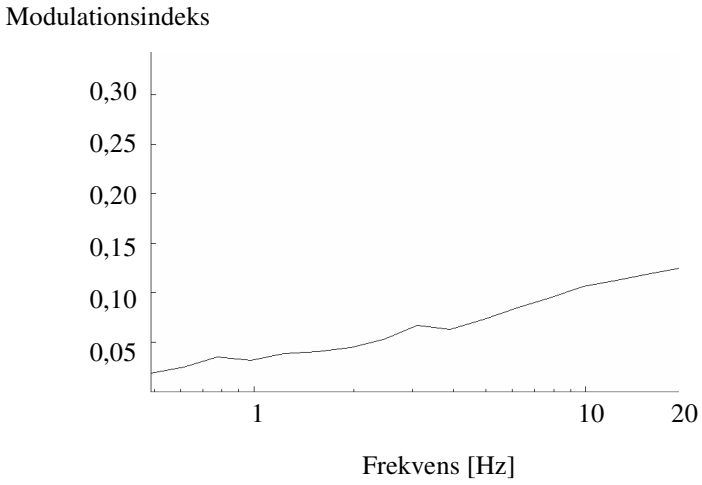
Af figur 3-5 fremgår modulationsindekset for Dantale II-støjen. I afsnit 6.1.1 og 6.2.2 findes modulation af hhv. de ni ICRA-støje og for et nyt støjsignal. Det nye støjsignal, der ligesom Dantale II-støjen er konstrueret ud fra testsætningerne, er genereret i forbindelse med dette projekt. Sammenlignes indekserne ses, at Dantale II-støjen kun er forholdsvist svagt moduleret. For at kunne opgive kun én talværdi for modulationen er der beregnet et samlet modulationsindeks for den lavfrekvente modulation – dvs. op til 50 Hz. Dette samlede indeks er for Dantale II-støjen blevet bestemt til 0,36.

---

[8] Det analytiske signal indeholder information om det oprindelige signals øjeblikkelige egenskaber – herunder amplituden og fasen. Det er bestemt ved at udføre en Hilbert-transformation på det oprindelige signal. For en beskrivelse af Hilbert-transformationen henvises til Goodman (2000 – Afsnit 3.8.2).

[9] Et FIR-filter har en endelig impulsrespons og er kendetegnet ved udelukkende at have nulpunkter, hvorved det altid vil være stabilt – til gengæld kræves en høj filterorden for at opnå en stor flankestejlhed. For en uddybende beskrivelse af FIR-filtre henvises til Hüche (1998 – Kap. 8).

[10] Aliasing er uønskede frekvenskomposanter. Disse opstår, når signalet, der ønskes samplet, indeholder frekvenskomposanter, som er højere end den halve samplingsfrekvens – dvs. den såkaldte Nyquist-frekvens.



Figur 3-5: Modulationsindekset for Dantale II-støjen med en opløsning på 1/3 oktav i frekvensområdet fra 0,5 til 20 Hz.

3.3 Sammenfatning

Talematerialet Dantale II består af testsætninger og et taleformet støjsignal. Testsætningerne er grupperet i 16 lister af hver ti sætninger. Alle sætningerne er konstrueret med samme grammatiske struktur og indeholder hver fem forskellige ord hørende til bestemte ordklasser. Sætningerne forekommer ikke som værende meningsfyldte og de enkelte ord kan i høj grad antages som værende uafhængige af hinanden. Ordene er nøje udvalgt og hver liste indeholder præcis de samme ord, som i de enkelte sætninger er kombineret på forskellig vis. Materialet er indspillet med en kvindelig indtaler, hvor der er taget højde for koartikulation på tværs af ordgrænserne. Materialet er designet til at teste voksne HA-kandidater og -brugere. Ved at reducere sætningslængden således, at hver sætning kun består af tre ord, kan materialet imidlertid også anvendes til at teste bl.a. børn. Støjsignalet – her betegnet Dantale II-støjen – er genereret ud fra testsætningerne, hvorfor støjen og sætningerne har samme frekvensindhold. Støjen er i øvrigt kun forholdsvist svagt moduleret og minder ved lytning ikke meget om egentlig tale.

4 Responsmetoder

DELTA's testprogram til afvikling af Dantale II materialet består som skrevet i afsnit 1.1 af to forskellige versioner – den operatør- og den patientbaserede version. Det samme er tilfældet for det nye testprogram, der er konstrueret i forbindelse med dette projekt. I den operatørbaserede version skal FP'en gentage så meget af det hørte som muligt, hvorefter en operatør via testprogrammet noterer de korrekt gengivne ord. I forbindelse hermed bør det bemærkes, at registreringen af FP'ens respons bygger på operatørens vurdering af hvilke ord, FP'en gengav korrekt. En sådan vurdering kan imidlertid variere både mellem forskellige operatører og inden for den selv samme operatør. Da FP'en i den operatørbaserede version ikke er bundet til at vælge sit respons ud fra et begrænset antal svarmuligheder, kaldes måden, hvorpå FP'en afgiver sit respons en åben responsmetode.

I den patientbaserede version skal FP'en i stedet selv registrere sit respons vha. en trykfølsom skærm, hvilket foregår ved at vedkommende har et fast antal svarmuligheder for hvert af den afpillede sætnings fem ord. I det nye testprogram er der konstrueret to varianter af den patientbaserede version. I den ene har FP'en ti svarmuligheder for hvert ord, mens FP'en i den anden foruden de ti svarmuligheder også har muligheden for at svare *ved ikke*. Da førstnævnte svarer til fem på hinanden følgende 10AFC-tests er responsmetoden lukket, mens den er semi-lukket i den variant, hvori svarmuligheden *ved ikke* er tilgængelig. De ti svarmuligheder svarer til de ti ord, der i basislisten er for hver af de fem ordklasser – se evt. tabel 3-1, s. 19 eller figur B-5, s. 100. Mht. den patientbaserede version bør det bemærkes, at denne stiller krav til FP'ens læse- og til dels motoriske færdigheder[11].

I det følgende behandles, hvordan responsmetoden i både den operatør- og den patientbaserede version influerer på testafviklingen. Det bør bemærkes, at der i dette kapitel ikke tages højde for *j*-faktoren, hvorved de fem ord i hver sætning betragtes som fem uafhængige elementer.

4.1 Den operatørbaserede version

For den operatørbaserede version fremgår af det følgende, hvordan FP'ens forventede respons som funktion af præsentationsniveauet kan beskrives vha. en såkaldt S-kurve. I forbindelse hermed vises, at taletærsklen bestemt som SRT_{50} giver den teoretisk mindste spredning af SRT_{th}, hvor th er den procentsats forståelighed, som tærsklen opgives ved. Værdien af SRT_{50} bestemmes i den operatørbaserede version ved at lade en afviklingsprocedure afspille sætningerne ved det niveau, hvor FP'en forventes at kunne gengive 50% af de afspillede ord korrekt. Ønskes foruden SRT_{50} også hældningen af S-kurven bestemt, skal sætningerne imidlertid afspilles ved to forskellige niveauer. Af afsnit 4.1.3 fremgår en beregning af ved hvilke sandsynligheder for korrekt svar disse niveauer bør placeres. Denne beregning er baseret på artiklen af Brand & Kollmeier (2002).

[11] Hertil kan nævnes, at ifølge en undersøgelse fra 1995 anslås, at ca. 3% af den voksne danske befolkning har store læsevanskeligheder – herunder er ordblinde, mens yderligere ca. 9% skønnes at have moderate læsevanskeligheder (Elbro *et al*., 1995).

4.1.1 Forventet respons

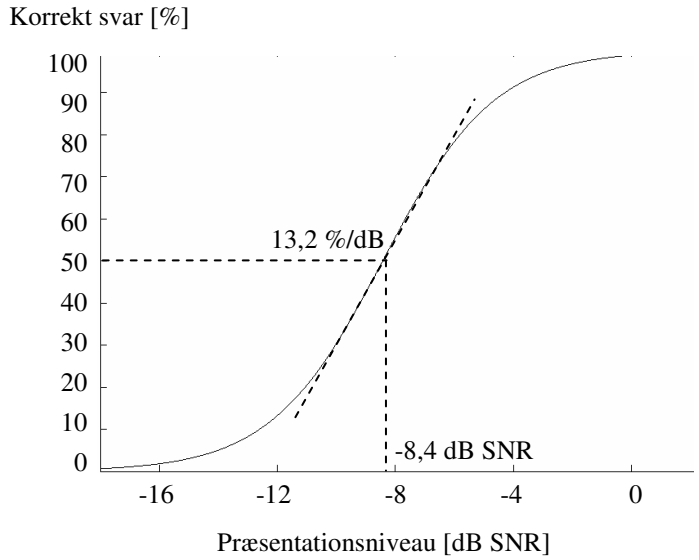
Som eksempelvis høretærsklen for en ren tone er tærsklen for korrekt opfattelse af tale ikke en skarp grænse, der adskiller hvorvidt et givet ord kan forstås eller ej. For i den operatørbaserede version at beskrive sandsynligheden for et korrekt svar som funktion af præsentationsniveauet kan formel (4-1) anvendes – se evt. Wagener *et al*. (2003) eller Brand & Kollmeier (2002). Som tidligere skrevet refererer værdien af SRT_{50} til det niveau, hvorved der forekommer en sandsynlighed på 50% for at forstå de afspillede ord. For den operatørbaserede version antages, at dette niveau er ensbetydende med en sandsynlighed på 50% for afgive et korrekt svar. Værdien af s_{50} er hældningen af kurven ved SRT_{50} og er afhængig af ordenes indbyrdes sværhedsgrad – se evt. formel (3-2), som er angivet i afsnit 3.1.5. Formel (4-1) er en såkaldt logistisk funktion, hvis kurve har form som kurven for en kumulativ normalfordeling. Da kurvens form minder om et *S* kaldes denne derfor ofte en S-kurve.

$$p_{OPE}(SNR) = \frac{1}{1+\exp\left(4 \cdot s_{50} \cdot \left(SRT_{50} - SNR\right)\right)} \tag{4-1}$$

Referenceværdien af SRT_{50} og s_{50} for Dantale II materialet, der af Wagener *et al*. (2003) er bestemt på baggrund af en undersøgelse med 60 normalthørende FP'er, er hhv. -8,4 dB SNR (spredning: 1,0 dB) og 13,2 %/dB (spredning: 1,9 %/dB). I Pilottest I-III er der bestemt værdier af SRT_{50} svarende til den fundet af Wagener *et al*.– se evt. figur C-5, s. 119; figur C-21, s. 144 og figur C-31, s. 162. Det samme gør sig i Pilottest II og III gældende for værdien af s_{50} – se evt. figur C-22, s. 145 og figur C-32, s. 162. Det bør i øvrigt bemærkes, at referenceværdierne gælder, når testen afvikles med Dantale II-støjen – udskiftes støjsignalet kan værdierne ændres. Derudover kan nævnes, at der for det svenske og tyske materiale er bestemt referenceværdier, hvis størrelsesordner modsvarer de for det danske – se evt. Wagener (2003 – Tabel 3.1)*. Af figur 4-1 fremgår S-kurven bestemt ved i formel (4-1) at indsætte de to omtalte referenceværdier. Heraf ses, at kurven løber fra en sandsynlighed på 0 til 100% for korrekt svar og at den er symmetrisk – dvs. hvis kurven roteres 180^{o} i papirets plan omkring dens midtpunkt vil den dække sig selv.

Selvom formel (4-1) ofte i litteraturen antages at kunne beskrive en FP's forventede respons, har det ikke været mulig at finde oplysning om, hvorvidt denne antagelse er rimelig – herunder om kurven virkelig er symmetrisk. Umiddelbart synes det ikke urimeligt at tro, at den virkelige kurve ved niveauer, der for FP'en synes meget lave, måske er anderledes i forhold til den angivet i figuren. I forbindelse med Pilottest II blev det derfor undersøgt, om en virkelig FP's respons kan beskrives ud fra formel (4-1). Dette vistes sig at være tilfældet, idet der af testen blev fundet en statistisk signifikant sammenhæng mellem hver FP's reelle respons og den bestemt ud fra formlen – se evt. figur C-13, s. 136 og figur C-16, s. 139. I Pilottest II indgik kun normalthørende, hvorfor resultatet af testen ikke fortæller noget om, hvorvidt formlen også er rimelig til at beskrive responset fra en hørehæmmet FP. I forbindelse hermed findes der i litteraturen anmærkninger om, at ikke alle hørehæmmede opnår en sandsynlighed på 100% for korrekt svar og at hældningen af S-kurven tilsvarende er mindre stejl end den for normalthørende FP'er – se evt. Hagerman & Kinnefors (1995) og Evans (1997 – Figur 7.1). Sidstnævnte blev

også fundet i Pilottest IV – se evt. figur D-9, s. 175. Det at S-kurven måske ikke når 100% for hørehæmmede, er der i dette projekt ikke taget højde for – herunder heller ikke i forbindelse med Pilottest IV.



Figur 4-1: Forløbet af den teoretiske S-kurve for den operatørbaserede version. Kurven, som angiver den forventede sammenhæng mellem præsentationsniveauet og responset for en normalthørende FP, er bestemt ved i formel (4-1) at sætte SRT_{50} og s_{50} til hhv. -8,4 dB SNR og 13,2 %/dB.

4.1.2 Taletærsklen

Tærsklen for korrekt opfattelse af tale opgives ofte som SRT_{50}. Imidlertid kan det vælges at opgive et hvilket som helst punkt på S-kurven som værende taletærsklen. For at afgøre ved hvilken sandsynlighed for korrekt svar tærsklen bør opgives, betragtes præsentationen af hvert enkelt ord som et Bernoulli-forsøg. Dette gælder, idet der kun er to udfald – enten gengiver FP'en ordet korrekt eller forkert. FP'ens respons kan derved beskrives som værende binomialfordelt, hvorved den teoretiske spredning af SRT_{th} kan udtrykkes ved formel (4-2). I formlen svarer p_{OPE} til formel (4-1), mens n er antallet af ord, som FP'en præsenteres for. Da spredningen af SRT_{th} udtrykker testens reliabilitet, skal denne være så lille som muligt.

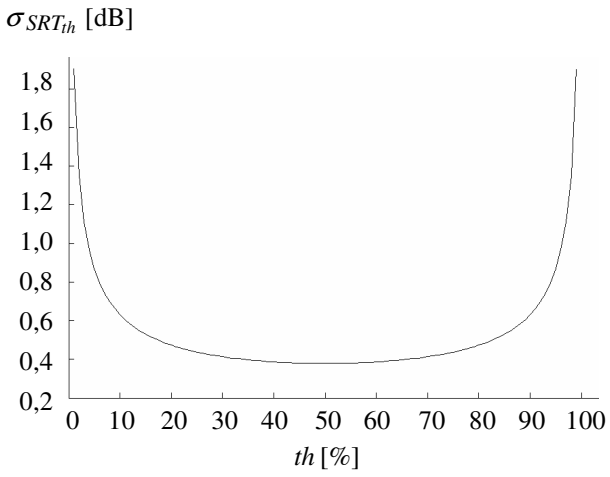
$$\sigma_{SRT_{th}} = \sqrt{\frac{th \cdot (1-th)}{n}} \cdot \left(\left. \frac{\partial p_{OPE}}{\partial SNR} \right|_{SNR=SRT_{th}} \right)^{-1} \tag{4-2}$$

Ved at aflede formlen mht. th og derpå at sætte det fundne udtryk lig nul fås, at den mindste spredning opnås, når th er lig 50%. Den teoretisk mindste spredning $\sigma_{SRT_{50}}$ kan derved udtrykkes ved følgende formel, hvoraf det ses, at spredningen falder, når enten s_{50} eller n stiger. Da denne spredning er relateret til størrelsen af re-test variationen af SRT_{50} betyder dette, at der for hørehæmmede kan forventes en større re-test variation end den for normalthørende.

$$\sigma_{SRT_{50}} = \frac{0{,}5}{s_{50} \cdot \sqrt{n}} \tag{4-3}$$

Da det er ved en sandsynlighed på 50% for korrekt svar, at spredningen af SRT_{th} er mindst, bør taletærsklen opgives som SRT_{50}. Dette gælder, når der tages udgangspunkt i formel (4-1). Som skrevet i forrige afsnit er der imidlertid flere, heriblandt Hagerman & Kinnefors (1995), som har observeret, at denne funktion ikke synes at kunne beskrive alle FP'er. Da ikke alle hørehæmmede vil opnå en sandsynlighed på 100% for korrekt svar, angiver Hagerman & Kinnefors i stedet taletærsklen som SRT_{40}.

Af figur 4-2 fremgår spredningen af SRT_{th} som funktion af th bestemt ud fra formel (4-2) – i formlen er s_{50} sat lig 13,2 %/dB, mens n er sat lig 100 svarende til 20 sætninger af hver fem ord. Af figuren ses, at den mindste værdi af spredningen af SRT_{th} som beregnet opnås, når th er lig 50%, mens spredningen er maks. når th er 0 eller 100%. Samtidigt ses, at spredningen bliver en anelse større ved at bestemme SRT_{80} frem for SRT_{50}, hvilket er i overensstemmelse med den i Pilottest I fundne re-test variation – se evt. figur C-9, s. 124. I pilottesten blev forskellen mellem re-test variationen for hhv. SRT_{50} og SRT_{80} imidlertid ikke fundet statistisk signifikant. Ved at opgive testresultatet som SRT_{80} frem for SRT_{50} kan der til gengæld forventes en mere sensitiv test. Det skyldes, at testes en FP med en god evne til at forstå tale i støj formodes vedkommende at opnå en forholdsvis lav værdi af SRT_{50} og en høj værdi af s_{50}, mens en FP med en dårlig evne formodes at opnå en forholdsvis høj værdi af SRT_{50} og en lav værdi af s_{50}. Ved at opgive SRT_{80} frem for SRT_{50} formodes testen derved bedre at kunne skelne mellem de to FP'er. At dette netop er tilfældet blev fundet i forbindelse med Pilottest IV – se evt. figur D-10, s. 176.



Figur 4-2: Spredningen af SRT_{th} som funktion af th bestemt ud fra formel (4-2), hvor s_{50} og n hhv. er sat lig 13,2 %/dB og 100.

4.1.3 Placering af præsentationer

Hvis alene SRT_{50} ønskes bestemt skal sætningerne afspilles ved et præsentationsniveau svarende til en sandsynlighed på 50% for korrekt svar. Ønskes testresultatet imidlertid opgivet som både en værdi af SRT_{50} og s_{50} skal sætningerne afspilles ved to forskellige niveauer, som i det følgende benævnes SNR_1 og SNR_2. De to tilhørende sandsynligheder for korrekt svar, som sætningerne præsenteres ved, benævnes tar_1 og tar_2, mens antallet af ord, som FP'en præsenteres for, benævnes n_1 og n_2. Ved at afspille sætningerne ved to forskellige sandsynligheder for korrekt svar kan der, når testen er afviklet, laves et maksimum-likelihood-estimat for S-kurvens forløb – dette vil blive forklaret i afsnit 5.3.2 og 5.3.3. Alternativt kan der beregnes et gennemsnit af FP'ens respons ved de to niveauer og derpå løses to ligninger med to ubekendte vha. formel (4-1), hvor de ubekendte er SRT_{50} og s_{50}. For at afgøre ved hvilke sandsynligheder for korrekt svar, sætningerne bør præsenteres for at opnå en lille teoretisk spredning af såvel SRT_{50} som s_{50}, betragtes i det følgende først spredningen af s_{50}. Denne kan udtrykkes ved følgende formel, som er baseret på den gaussiske fejllov (eng: the Gaussian error law):

$$\sigma_{s_{50}} = \sqrt{\frac{1}{n_1}\cdot\left(\frac{\partial s_{50}}{\partial tar_1}\right)^2 \cdot tar_1 \cdot (1-tar_1) + \frac{1}{n_2}\cdot\left(\frac{\partial s_{50}}{\partial tar_2}\right)^2 \cdot tar_2 \cdot (1-tar_2)} \tag{4-4}$$

I udtrykket er s_{50} bestemt ud fra følgende to formler:

$$tar_1 = \frac{1}{1+\exp(4\cdot s_{50}\cdot(SRT_{50}-SNR_{tar1}))} \quad \Leftrightarrow$$

$$SRT_{50} = \frac{1}{4\cdot s_{50}}\cdot\ln\left(\frac{1}{tar_1}-1\right)+SNR_{tar1} \tag{4-5}$$

$$tar_2 = \frac{1}{1+\exp(4\cdot s_{50}\cdot(SRT_{50}-SNR_{tar2}))} \quad \Leftrightarrow$$

$$SRT_{50} = \frac{1}{4\cdot s_{50}}\cdot\ln\left(\frac{1}{tar_2}-1\right)+SNR_{tar2} \tag{4-6}$$

Ved at sætte disse to formler lig hinanden fås efter lidt regning:

$$s_{50} = \frac{1}{4\cdot(SNR_{tar1}-SNR_{tar2})}\cdot\ln\left(\frac{tar_1\cdot(1-tar_2)}{tar_2\cdot(1-tar_1)}\right) \tag{4-7}$$

Indsættes formel (4-7) i formel (4-4) kan det bestemmes, at spredningen af s_{50} er mindst, når tar_1 = 8,3% og tar_2 = 91,7% – eller alternativt, når tar_1 = 91,7% og tar_2 = 8,3%. Indsættes de fundne sandsynligheder i formel (4-4) fås udtrykket i formel (4-8), som derved angiver den teoretiske mindste spredning af s_{50}. I forbindelse hermed er både n_1 og n_2 sat lig n, hvorved n angiver antallet af ord, som FP'en præsenteres for ved hver af de to sandsynligheder. Af udtryk-

ket ses, at spredningen falder, når enten s_{50} falder eller n stiger. Med andre ord forventes en mindre re-test variationen af s_{50} for hørehæmmede end den for normalthørende. Det bør i øvrigt bemærkes, at jf. formel (4-3) falder spredningen af SRT_{50}, når s_{50} stiger.

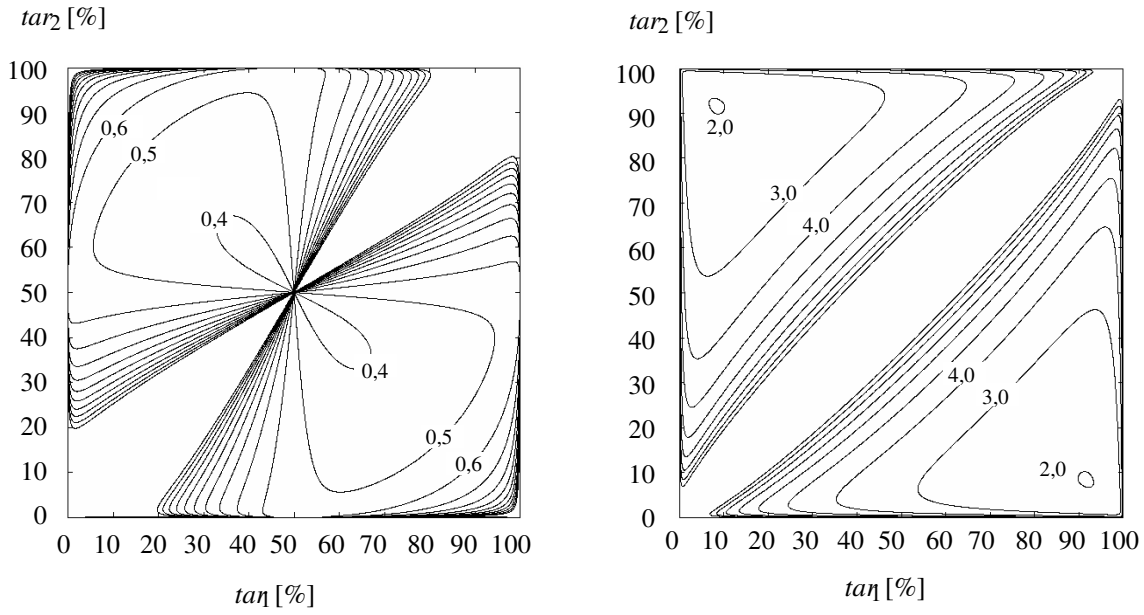
$$\sigma_{s_{50}} = \frac{1{,}07 \cdot s_{50}}{\sqrt{n}} \tag{4-8}$$

Til sammenligning opnås den mindste spredning af SRT_{50}, når både tar_1 og tar_2 er lig 50%. Sandsynlighederne for korrekt svar, hvorved præsentationsniveauerne bør præsenteres, afhænger altså af, om det er SRT_{50} eller s_{50}, der ønskes bestemt med den teoretisk mindste spredning. Hvis både SRT_{50} og s_{50} ønskes bestemt, bør der findes et kompromis mellem de pågældende spredninger. Et sådant kan findes ved at beregne den kvadratiske middelværdi af spredningen af SRT_{50} og s_{50} relativt til de respektive minimumsværdier. Et udtryk herfor fremgår af formel (4-9), hvoraf de to spredninger vægtes ens. Ønskes de to spredninger imidlertid vægtet forskelligt, kan der foran hver af de to parenteser indsættes en såkaldt vægtfaktor. Af udtrykket fås, at σ_{komp} er mindst, når $tar_1 = 19{,}0\%$ og $tar_2 = 81{,}1\%$ – eller alternativt, når $tar_1 = 81{,}1\%$ og $tar_2 = 19{,}0\%$.

$$\sigma_{komp} = \frac{1}{\sqrt{2}} \cdot \sqrt{\left(\frac{\sigma_{SRT_{50}}}{\min(\sigma_{SRT_{50}})}\right)^2 + \left(\frac{\sigma_{s_{50}}}{\min(\sigma_{s_{50}})}\right)^2} \tag{4-9}$$

Af figur 4-3 fremgår konturlinjer for spredningen af SRT_{50} og s_{50} som funktion af tar_1 og tar_2. Værdierne af spredningen af SRT_{50} er bestemt ved i formel (4-4) at erstatte s_{50} med SRT_{50}. I den heraf fremkommende formel blev SRT_{50} bestemt ud fra formel (4-5) og (4-6) ved i hvert af de to udtryk at isolere s_{50} og derpå sætte de fundne udtryk lig hinanden. Værdierne af spredningen af s_{50} er derimod fundet direkte ud fra formel (4-4). For værdierne af de to spredninger gælder, at SRT_{50} og s_{50} hhv. er sat lig. -8,4 dB SNR og 13,2 %/dB, mens både n_1 og n_2 er sat lig 50, hvor sidstnævnte svarer til ti sætninger af hver fem ord ved hvert af de to præsentationsniveauer.

Konturlinjerne viser dels, ved hvilke sandsynligheder for korrekt svar den mindste spredning af SRT_{50} og s_{50} teoretisk kan opnås og dels, hvor meget de pågældende spredninger varierer, når procentsatserne afviger fra de optimale værdier. I forbindelse hermed fås den teoretisk mindste spredning af SRT_{50} og s_{50} til hhv. 0,38 dB og 1,99 %/dB. Som en ekstra bemærkning kan nævnes, at ønskes en teoretisk spredning af SRT_{50} på 0,38 dB ved at præsentere sætningerne ved to sandsynligheder på hhv. 20 og 80% for korrekt svar, skal n_1 og n_2 være lig 78 svarende til knap 16 sætninger af hver fem ord. Det bør i øvrigt nævnes, at de to delfigurer i store træk svarer til figurer angivet i Brand & Kollmeier (2002 – Figur 1 og 2) – den væsentligste forskel er dog, at abscissen og ordinaten her angiver sandsynligheder og ikke præsentationsniveauer.



a) Spredningen af SRT_{50}

b) Spredningen af s_{50}

Figur 4-3: Illustration af hvordan den teoretiske spredning af SRT_{50} og s_{50} varierer ved forskellige sandsynligheder for korrekt svar i en operatørbaseret afvikling. I delfigur a) fremgår et konturplot af spredningen af SRT_{50} i dB bestemt ved i formel (4-4) at erstattet s_{50} med SRT_{50}. I delfigur b) fremgår et konturplot af spredningen af s_{50} i %/dB bestemt direkte ud fra formel (4-4). For begge delfigurer gælder, at SRT_{50} og s_{50} er sat lig hhv. -8,4 dB SNR og 13,2 %/dB, mens både n_1 og n_2 er sat lig 50.

4.2 Den patientbaserede version

Som for den operatørbaserede version fremgår i det følgende, hvordan FP'ens forventede respons i den patientbaserede version kan beskrives vha. en S-kurve. Selvom den patientbaserede version i det nye testprogram er opbygget omkring en 10AFC-test, vil beskrivelsen gælde generelt for en *x*AFC-test, hvorved betydningen af antallet af svarmuligheder tillige behandles. Samtidigt diskuteres svarmuligheden *ved ikke*. Derudover beregnes ved hvilke sandsynligheder for korrekt svar sætningerne i en *x*AFC-test bør afspilles for at opnå en teoretisk lille spredning af såvel SRT_{50} som s_{50}. I forbindelse hermed bør det bemærkes, at artiklen af Brand & Kollmeier (2002), hvorpå beregningen er baseret, kun omhandler en operatørbaseret afvikling. Afslutningsvist berøres, hvordan den patientbaserede version alternativt kan afvikles således, at responsmetoden bliver åben. Hvis responsmetoden gøres åben, som i den operatørbaserede version, bør FP'ens respons beskrives vha. formel (4-1), der er angivet i afsnit 4.1.1.

4.2.1 Forventet respons

I den patientbaserede version forekommer der en vis sandsynlighed for, at FP'ens respons indeholder flere korrekte ord end hvad, vedkommende rent faktisk har hørt. Med andre ord forekommer der altså ikke den samme sandsynlighed for at kunne forstå de afspillede ord og for at kunne

afgive et korrekt svar. Dette skyldes, at denne version er opbygget omkring en 10AFC-test, hvorfor FP'en via gæt kan svare flere ord korrekt end hvad, vedkommende har hørt. I det følgende betragtes først den variant, hvori det ikke er muligt at svare *ved ikke*. FP'en er her for hvert de afspillede ord tvunget til at vælge netop ét blandt de ti svarmuligheder. Med andre ord er FP'en, hvad enten vedkommende slet ikke har hørt ordet eller ikke er helt sikker på, hvilket ord der blev afspillet, tvunget til at gætte. Der er her tale om to forskellige slags gæt; *rene gæt* og *motiverede gæt*. I det tilfælde hvor FP'en slet ikke har hørt det afspillede ord, kaldes gættet et *rent gæt*. Hvis FP'en derimod har hørt en del af ordet, øges vedkommendes sandsynlighed for at gætte det korrekte ord. Det skyldes den relative store fonologiske afstand mellem de ti svarmuligheder, hvorfor et sådant gæt kaldes et *motiveret gæt*.

For at beskrive hvordan FP'ens respons afhænger af præsentationsniveauet i varianten uden svarmuligheden *ved ikke*, udledes i det følgende en funktion baseret på formel (4-1), som er gældende for den operatørbaserede version. I forbindelse hermed er det valgt at lade variablen x angive antallet af svarmuligheder således, at udtrykket ikke kun er begrænset til en 10AFC-test. I det følgende antages, at FP'en svarer korrekt på de ord, som vedkommende har hørt og at FP'en enten har hørt eller ikke hørt det enkelte ord. Derved tages der kun højde for vedkommendes *rene gæt,* hvilket betyder, at hvis FP'en ikke har hørt et givet ord, vil vedkommendes gæt til det pågældende ord være fuldstændigt tilfældigt.

FP'ens gæt efter en sætningspræsentation betragtes som $n_{gæt}$ Bernoulli-forsøg, hvor $n_{gæt}$ angiver et helt antal ord, som vedkommende ikke har hørt. Som i afsnit 4.1.2 er der her tale om Bernoulli-forsøg, idet der kun er to udfald – enten gætter FP'en ordet korrekt eller forkert. Da hvert af de $n_{gæt}$ uafhængige forsøg har samme sandsynlighed på $1/x$ for, at FP'en afgiver et korrekt svar, vil antallet af de ord, som FP'en gætter korrekt, være binomialfordelt. I binomialfordelingens tæthedsfunktion, som fremgår af følgende formel, angiver den stokastiske variabel $c_{gæt}$, antallet af de ord, som FP'en gætter korrekt ud af de $n_{gæt}$ ord.

$$P_{n_{gæt},x\text{AFC}}(X = c_{gæt}) = \binom{n_{gæt}}{c_{gæt}} \cdot \left(\frac{1}{x}\right)^{c_{gæt}} \cdot \left(1-\frac{1}{x}\right)^{n_{gæt}-c_{gæt}} \tag{4-10}$$

Udtrykkets første koefficient ($n_{gæt}$ over $c_{gæt}$) er den såkaldte binomialkoefficient, der angiver antallet af måder, hvorpå FP'en kan gætte $c_{gæt}$ ord korrekt blandt $n_{gæt}$ mulige. Denne værdi ganges med sandsynligheden for, at hvert af de mulige udfald optræder.

$$\binom{n_{gæt}}{c_{gæt}} = \frac{n_{gæt}!}{c_{gæt}!\left(n_{gæt}-c_{gæt}\right)!} \tag{4-11}$$

Tilsvarende kan antallet af mulige udfald for at gætte $c_{gæt}$ ord korrekt udtrykkes som følgende:

$$N_{n_{gæt},x\text{AFC}}(X = c_{gæt}) = P_{n_{gæt},x\text{AFC}}(X = c_{gæt}) \cdot x^{n_{gæt}} = \binom{n_{gæt}}{c_{gæt}} \cdot (x-1)^{n_{gæt}-c_{gæt}} \tag{4-12}$$

I udtrykket angiver $x^{n_{gæt}}$ det samlede antal af mulige udfald, når der er x svarmuligheder og $n_{gæt}$ ord, der ikke er hørt. Som skrevet er den patientbaserede version valgt designet således, at den svarer til fem på hinanden følgende 10AFC-tests pr. afspillet sætning, hvorfor x i det følgende sættes lig ti. Ud fra formel (4-10) er sandsynligheden beregnet for, at FP'en svarer et bestemt antal ord korrekt, når vedkommende har hørt nul, et, to, tre, fire eller fem ord – se tabel 4-1. Bemærk, antallet af korrekt svarende ord er summen af de ord, som FP'en har hørt og $c_{gæt}$. Som tidligere skrevet antages, at hvis FP'en har hørt ordet, gengiver vedkommende det korrekt. Af tabellen fremgår, at hvis FP'en ikke har hørt et eneste ord er sandsynligheden for at vedkommende gætter fire eller alle ord korrekt forsvindende. Samtidigt fremgår, at sandsynligheden for at svare alle ord korrekt, som ventet er proportionel med antallet af ord, som FP'en har hørt.

Tabel 4-1: Sandsynligheden for at svare et bestemt antal ord korrekt, når FP'en har hørt hhv. nul ord ($n_{gæt}$ = 5), et ord ($n_{gæt}$ = 4), to ord ($n_{gæt}$ = 3), tre ord ($n_{gæt}$ = 2), fire ord ($n_{gæt}$ = 1) eller fem ord ($n_{gæt}$ = 0). De grå felter refererer til det antal ord, som FP'en har hørt. Hvis FP'en eksempelvis har hørt fire af de afspillede ord, svarende til $n_{gæt}$ = 1, er sandsynligheden for, at vedkommendes respons indeholder tre korrekte ord eller færre, lig med nul. Hertil antages, at hvis FP'en har hørt ordene, gengiver vedkommende dem korrekt, hvorfor mindst fire korrekte ord.

Antal af korrekt svarede ord	0	1	2	3	4	5
$P_{5,\ 10AFC}$ [%]	59,0	32,8	7,3	0,8	0	0
$P_{4,\ 10AFC}$ [%]	0	65,6	29,2	4,9	0,4	0
$P_{3,\ 10AFC}$ [%]	0	0	72,9	24,3	2,7	0,1
$P_{2,\ 10AFC}$ [%]	0	0	0	81,0	18,0	1,0
$P_{1,\ 10AFC}$ [%]	0	0	0	0	90,0	10,0
$P_{0,\ 10AFC}$ [%]	0	0	0	0	0	100,0

Antallet af ord, som FP'en i gennemsnit vil svare korrekt, når vedkommende har hørt nul, et, to, tre, fire eller fem ord, kan beskrives ud fra en modificeret gennemsnitsværdi, som er angivet i formel (4-13) – dvs. FP'ens mest sandsynlige respons. Grunden til at værdien er modificeret er, at den tager højde for de ord, som FP'en har hørt. I udtrykket henfører tallet 5 derfor til, at sætningernes længde fast er på fem ord.

$$\mu_{x\mathrm{AFC}} = 5 - n_{gæt} + n_{gæt} \cdot \frac{1}{x} = 5 + n_{gæt} \cdot \left(\frac{1}{x} - 1 \right) \tag{4-13}$$

Med udgangspunkt i formel (4-1), som beskriver S-kurven for den operatørbaserede version, kan formel (4-13) omskrives til at beskrive S-kurven for en xAFC-test. Som tidligere skrevet angiver S-kurven sandsynligheden for et korrekt svar som funktion af præsentationsniveauet – dvs. her den samlede sandsynlighed for at høre ordet og gætte ordet korrekt, hvis dette ikke er hørt. Omskrivningen svarer til at normere formel (4-13) og derpå definere værdien af $n_{gæt}$ således, at den beskrives ud fra p_{OPE}.

$$p_{x\mathrm{AFC}} = 1 + (1 - p_{\mathrm{OPE}}) \cdot \left(\frac{1}{x} - 1 \right) = p_{\mathrm{OPE}} \cdot \left(1 - \frac{1}{x} \right) + \frac{1}{x} \tag{4-14}$$

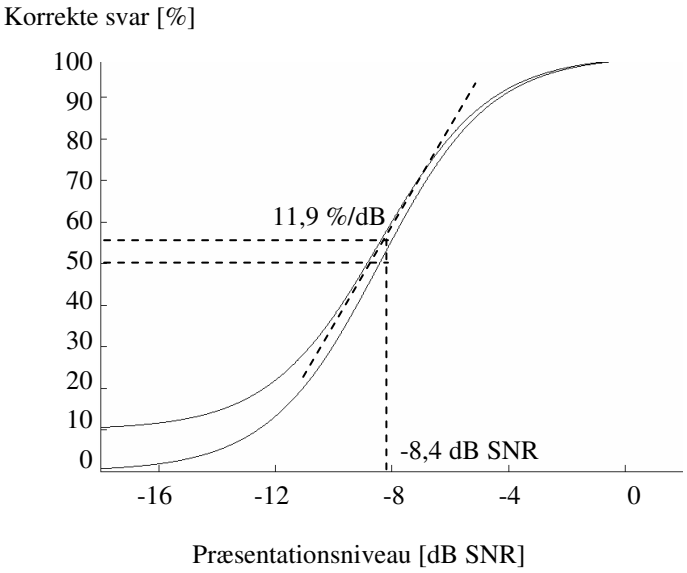
Ved at indsætte formel (4-1) vil udtrykket for $p_{x\mathrm{AFC}}$ indeholde hældningen af S-kurven for den operatørbaserede version. Dette er imidlertid ikke ønskeligt. Formel (4-14) – med formel (4-1) indsat – afledes derfor mht. SNR. Ved derpå at sætte SNR lig SRT_{50} fås at hældningen af S-kurven for en *x*AFC-test kan beskrives ved at gange hældningen af S-kurven for den operatørbaserede version med (1–1/*x*). S-kurven for den patientbaserede version kan derved beskrives med følgende logistiske funktion, hvori s_{50} er hældningen af den pågældende kurve:

$$p_{x\mathrm{AFC}}(SNR) = \frac{1}{1 + \exp\left(4 \cdot \left(1 - \frac{1}{x} \right)^{-1} \cdot s_{50} \cdot \left(SRT_{50} - SNR \right) \right)} \cdot \left(1 - \frac{1}{x} \right) + \frac{1}{x} \tag{4-15}$$

Af formlen ses, at funktionens afhængige variabel befinder sig i intervallet fra 1/*x* til 1. Det betyder, at jo færre svarmuligheder FP'en har at vælge imellem, desto mindre bliver dynamikområdet for den afhængige variabel. Samtidigt ses, at jo færre svarmuligheder, desto mindre bliver hældningen af kurven, hvorved FP'ens mest sandsynlige respons bliver mindre følsom over for en ændring i SNR. Med andre ord komprimeres kurven for den operatørbaserede version som funktion af, at antallet af svarmuligheder reduceres. For en 10AFC-test kan udtrykket i formel (4-15) skrives:

$$p_{10\mathrm{AFC}}(SNR) = \frac{1}{1 + \exp\left(\frac{40}{9} \cdot s_{50} \cdot \left(SRT_{50} - SNR \right) \right)} \cdot \frac{9}{10} + \frac{1}{10} \tag{4-16}$$

Af figur 4-4 fremgår en normalthørende FP's forventede respons som funktion af præsentationsniveauet bestemt ved i formel (4-16) at sætte SRT_{50} og s_{50} lig hhv. -8,4 dB SNR og 11,9 %/dB – sidstnævnte er fremkommet ved at gange referenceværdien af s_{50} for den operatørbaserede version med 9/10. Til sammenligning fremgår af figuren også den tilsvarende S-kurve for den operatørbaserede version, som svarer til den angivet i figur 4-1. Området mellem de to kurver repræsenterer de *rene gæt* i den patientbaserede version uden svarmuligheden *ved ikke*. Området mellem de to kurver bliver større jo færre svarmuligheder der medtages, hvorved omfanget af *rene gæt* øges. Ved at reducere antallet af svarmuligheder formodes omfanget af *motiverede gæt* imidlertid også at kunne øges, idet FP'en derved har færre muligheder at vælge imellem. Med andre ord, jo flere svarmuligheder desto mere går den lukkede mod den åbne responsmetode.



Figur 4-4: Forløbet af den teoretiske S-kurve for den patientbaserede version uden svarmuligheden *ved ikke* – dvs. hver afspillet sætning kan betragtes som fem på hinanden efterfølgende 10AFC-tests. Kurven, som angiver den teoretiske sammenhæng mellem præsentationsniveauet og responset for en normalthørende FP, er bestemt ved i formel (4-16) at sætte SRT_{50} og s_{50} til hhv. -8,4 dB SNR og 11,9 %/dB. Til sammenligning fremgår også S-kurven for den operatørbaserede version svarende til den, der er angivet i figur 4-1. Området mellem de to kurver repræsenterer de såkaldte *rene gæt*.

I det følgende betragtes taletærsklen i en 10AFC-test opgivet som SRT_{50} – dvs. som det niveau, hvor FP'en forventes at kunne forstå 50% af de afspillede sætninger. Hvis der ønskes sammenlignelige resultater i den operatør- og den patientbaserede version uden svarmuligheden *ved ikke*, skal tærsklen ifølge figuren i den patientbaserede version bestemmes som det niveau, hvor FP'en forventes at kunne gengive 55% af de afspillede ord. Dette gælder, idet der her optræder en sandsynlighed på 1/10 for, at FP'en – på trods af, at vedkommende ikke har hørt et givet ord – gætter på det korrekte ord. Hertil bør huskes, at der i udledningen af formel (4-16) kun er taget højde for *rene* og ikke for *motiverede gæt,* hvorfor ovenstående betragtning kun gør sig gældende, hvis omfanget af *motiverede gæt* er forsvindende.

I forbindelse med Pilottest II blev omfanget af *motiverede gæt* undersøgt. Af testen blev det fundet, at værdier af SRT_{50} bestemt med den operatørbaserede version var 1,1 dB højere og derved dårligere end tilsvarende værdier bestemt med den patientbaserede version uden svarmuligheden *ved ikke* – se evt. figur C-20, s. 143 og figur C-21, s. 144. Forskellen mellem værdierne af SRT_{50} bestemt med hver af de to versioner blev fundet statistisk signifikant. Da der var taget højde for *rene gæt*, kan denne forskel antages at hidrøre fra *motiverede gæt*, hvorfor testresultaterne de to versioner imellem ikke er sammenlignelige. På trods heraf blev det fundet, at formel (4-16) synes meget rimelig til at beskrive en FP's forventede respons, idet der blev fundet en statistisk signifikant sammenhæng mellem hver FP's reelle respons og den bestemt ud fra formlen – se evt. figur C-15, s. 138 og figur C-16, s. 139. Til sammenligning kan nævnes, at jf. afsnit 1.1 blev der også i DELTA's undersøgelse fundet en statistisk signifikant forskel mellem SRT_{50} bestemt med hver af de to versioner, hvoraf der i den patientbaserede version var indeholdt tre svarmuligheder for hvert afspillet ord. Forskellen, som var på 2,3 dB, antyder sammen med de 1,1 dB netop, at jo færre svarmuligheder desto større omfang af *motiverede gæt*.

I Pilottest II indgik også den patientbaserede version med muligheden for at svare *ved ikke*. Denne svarmulighed blev tilføjet for at undgå eventuelle frustrationer hos FP'en over at skulle afgive et ord som svar til noget, der ikke er hørt. Forskellige FP'er kan imidlertid gøre forskelligt brug af denne svarmulighed. Den *forsigtige* FP vil først afgive et svar, når vedkommende er helt sikker på at have hørt det pågældende ord, mens *gambleren* ved blot den allermindste antydning af et givet ord vil afgive dette som svar. For at tage højde herfor blev hvert *ved ikke*-svar konverteret til en 1/10 korrekt svar og taletærsklen blev derefter opgivet, som det niveau, hvor FP'en i en 10AFC-test forventes at kunne gengive 55% af de afspillede ord – dvs. svarende til 50% forståelighed. De heraf fundne værdier af SRT_{50} blev ikke fundet statistisk signifikant forskellige fra de bestemt med den operatørbaserede version – se evt. igen figur C-20, s. 143 og figur C-21, s. 144. Dette antyder, at idet FP'erne ikke er tvunget til at angive et ord som svar, reduceres omfanget af *motiverede gæt*. Ved at betragte *ved ikke*-svar som 1/10 korrekt svar blev det som for den patientbaserede version uden svarmuligheden *ved ikke* fundet at formel (4-16) er rimelig til at beskrive en FP's forventede respons – se evt. figur C-14, s. 137 og figur C-16, s. 139.

Idet de FP'er, der indgik i pilottesten var forholdsvis mådeholdne overfor blot at gætte sig frem, kan deres respons imidlertid også beskrives med S-kurven for den operatørbaserede version – se evt. figur C-16, s. 139 og figur C-18, s. 141. I forbindelse hermed betragtes *ved ikke*-svar som forkerte svar. Et forkert svar er, når FP'en afgiver et andet end det afspillede ord som svar. Hvis FP'en eksempelvis afgiver svaret *Henning*, men det afspillede ord i virkeligheden var *Per*, er det et forkert svar. Til sammenligning svarer et forkert svar i den operatørbaserede version til, at FP'en enten gengiver et andet end det afspillede ord som svar eller at vedkommende slet ikke gengiver noget.

Som et mål for hvor ofte den enkelte FP gør brug af muligheden for at svare *ved ikke*, kan formel (4-17) anvendes – denne giver et tal for, hvor chancetagen (*Ch*) FP'en kan antages at være. I formlen angiver *forkert ord* og *ved ikke* antallet af de gange, hvor FP'en har afgivet hhv. et forkert ord som respons og svaret *ved ikke*. Hvis FP'en er meget *forsigtig* og altid vil være 100% sikker inden vedkommende svarer et specifik ord, vil antallet af *forkert ord* principielt være nul, hvorved *Ch* ligeledes er nul. Hvis FP'en derimod er en *gambler* og derfor aldrig bruger svarmuligheden *ved ikke*, vil *Ch* være lig én – vedkommende vil derved afgive sit respons som i den patientbaserede version, hvor det ikke er muligt at svare *ved ikke*. Det bør bemærkes, at en FP godt kan være meget chancetagen uden at opnå en høj værdi af *Ch*. Det gælder, hvis FP'en ved sine gæt afgiver korrekte svar – eksempelvis via *motiverede gæt*.

$$Ch = \frac{forkert\ ord}{forkert\ ord + ved\ ikke} \tag{4-17}$$

Slutteligt kan det nævnes, at der i forbindelse med det tyske materiale er blevet udført en undersøgelse af Brand *et al.* (2004)*, hvor materialet blev afviklet som i den patientbaserede version med svarmuligheden *ved ikke*. I denne undersøgelse blev *ved ikke*-svar betragtet som forkerte svar (Thomas Brand – Carl von Ossietzky Universität Oldenburg, pers. medd. feb. 2007). Af undersøgelsen blev der opnået værdier af SRT_{50}, som efterfølgende blev sammenlignet med tilsvarende værdier opnået med en operatørbaseret afvikling. Da der ikke blev fundet nogen forskel værdier imellem antydes, at FP'erne også her har været mådeholdne overfor blot at gætte

sig frem. Det har i øvrigt ikke været muligt at opspore mere end den ene undersøgelse, hvor et talemateriale som det danske afvikles vha. patientselvbetjening.

4.2.2 Placering af præsentationer

I afsnit 4.1.3 blev det fundet, at den teoretisk mindste spredning af SRT_{th} kan bestemmes, når sætningerne præsenteres ved S-kurvens midtpunkt – dvs. hvor kurven er stejlest svarende til 50% forståelighed. Hvis alene SRT_{50} ønskes bestemt skal sætningerne i den operatørbaserede version afspilles ved et præsentationsniveau svarende til en sandsynlighed på 50% for korrekt svar – dvs. tar er lig 50%. Midtpunktet for en *x*AFC-test befinder sig imidlertid ikke ved en sandsynlighed på 50% for korrekt svar, hvorfor tar i stedet kan bestemmes ud fra følgende formel, hvoraf det som forventet ses, at tar stiger, når antallet af svarmuligheder falder. Med andre ord skal sætningerne i en *x*AFC-test altså afspilles ved et niveau, hvorved FP'en formodes kunne forstå 50% af de afspillede sætninger, hvilket svarer til mere end 50% korrekt gengivet. Dette blev i øvrigt også omtalt i forrige afsnit.

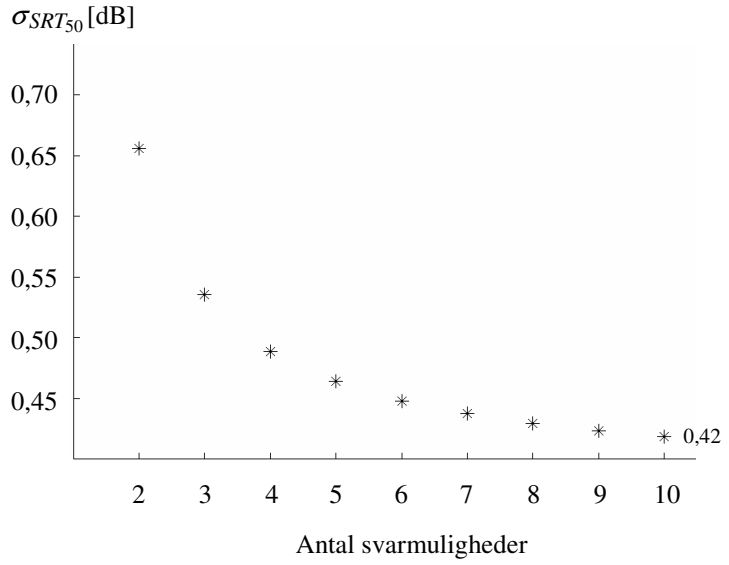
$$tar = \left(1 - \frac{1}{x}\right) \cdot \frac{1}{2} + \frac{1}{x} \tag{4-18}$$

Ved at i formel (4-2) at udskifte p_{OPE} med p_{xAFC} og derpå indsætte formel (4-15) og formel (4-18) på th's plads fås, at spredningen af SRT_{50} efter lidt regning kan beskrives ved:

$$\sigma_{SRT50} = \frac{0{,}5}{s_{50}} \cdot \sqrt{\frac{x^2 - 1}{n \cdot x^2}} \tag{4-19}$$

I figur 4-5 fremgår, hvordan spredningen af SRT_{50} afhænger af antallet af svarmuligheder. Værdierne af spredningen er bestemt ud fra formel (4-19) ved at sætte s_{50} lig 13,2 %/dB gange (1–1/*x*), mens n er sat til 100 svarende til 20 sætninger af hver fem ord. Af figuren ses, at spredningen bliver mindre, jo flere svarmuligheder, der medtages. Ved en 10AFC-test er spredningen 0,42 dB. Til sammenligning fås en værdi af spredningen på 0,38 dB, når *x* er gående imod uendeligt – denne værdi svarer til den angivet i afsnit 4.1.3. Ønskes i en 10AFC-test med den samme spredning som for den operatørbaserede version kan det ud fra formel (4-19) beregnes, at FP'en frem for 100 skal præsenteres for godt 122 ord – dvs. godt 24 sætninger. Tilsvarende gælder, at hvis den pågældende spredning ønskes i en 3AFC-test skal FP'en præsenteres for knap 200 ord – dvs. knap 40 sætninger. Præsenteres FP'en for et fast antal sætninger kan valget af antallet af svarmuligheder altså influere på testens reliabilitet. Hertil kan nævnes, at afviklingstiden, som både er tiden for afspilningen af sætningerne og FP'ens afgivelse af svar, i Pilottest I er fundet betydelig længere for den patient- end for den operatørbaserede version. Dette gjaldt, selvom der i de to versioner var medtaget det samme antal sætninger – se evt. figur C-11, s. 127. Imidlertid kan afviklingstiden i den patientbaserede version tænkes at være afhængig af antallet af svarmuligheder således, at jo færre svarmuligheder, desto kortere afviklingstid.

For både den operatør- og den patientbaserede version uden svarmuligheden *ved ikke* blev re-test variationen i Pilottest I også bestemt. For versionerne blev der ikke fundet nogen statistisk signifikant forskel re-test variationerne imellem – se evt. figur C-9, s. 124. Det bør hertil bemærkes, at forskellen mellem de to teoretiske spredninger blot er 0,04 dB – denne forskel mellem de to spredninger er kun afhængig af antallet af svarmuligheder og derved ikke at antallet af ord, som FP'en præsenteres for. De teoretiske spredninger må i øvrigt ikke sammenlignes direkte med de i pilottesten fundne re-test variationer. Til gengæld kan nævnes, at de i pilottesten fundne re-test variationer er af samme størrelsesorden som de bestemt i DELTA's undersøgelse, som er omtalt i afsnit 1.1. Bemærk; i DELTA's testprogram indgår kun tre svarmuligheder for hvert afspillet ord – til gengæld blev hver enkelt FP præsenteret for op til 30 sætninger i hver måling. Til sammenligning blev FP'erne i Pilottest I kun præsenteret for 20 sætning i hver måling.



Figur 4-5: Spredningen af SRT_{50} som funktion af x bestemt ud fra formel (4-19), hvor s_{50} er sat lig 13,2 %/dB gange (1–1/x), mens n er sat lig 100.

Ønskes testresultatet opgivet som både en værdi af SRT_{50} og s_{50}, skal der som for den operatørbaserede version anvendes to forskellige præsentationsniveauer. De to tilhørende sandsynligheder for korrekt svar, som sætningerne præsenteres ved, benævnes i det følgende tar_1 og tar_2. Fremgangsmåden for at bestemme værdierne af tar_1 og tar_2, hvis enten kun s_{50} eller både SRT_{50} og s_{50} er den samme som i afsnit 4.1.3 – dog med den forskel, at S-kurven beskrives med formel (4-15) frem for formel (4-1). Derved fås de værdier som fremgår af tabel 4-2 for forskellige antal af svarmuligheder. De angivne sandsynligheder for tar, når kun SRT_{50} ønskes bestemt, er beregnet ud fra formel (4-18). Det bør bemærkes, at det kun er for x gående imod uendeligt, at tar_1 og tar_2 er beliggende symmetrisk omkring S-kurvens midtpunkt – dvs. omkring SRT_{50}. For tabellens værdier gælder i øvrigt, at disse er i overensstemmelse med O'Regan & Humbert (1989), hvori tilsvarende værdier for x gående imod uendeligt, $x = 4$, $x = 3$ og $x = 2$ er bestemt. Fremgangsmåden for bestemmelsen af de i litteraturen fundne værdier er en

smule mere kompliceret end den her udførte, hvorfor der for en beskrivelse af denne henvises til O'Regan & Humbert (1989 – Appendiks A).

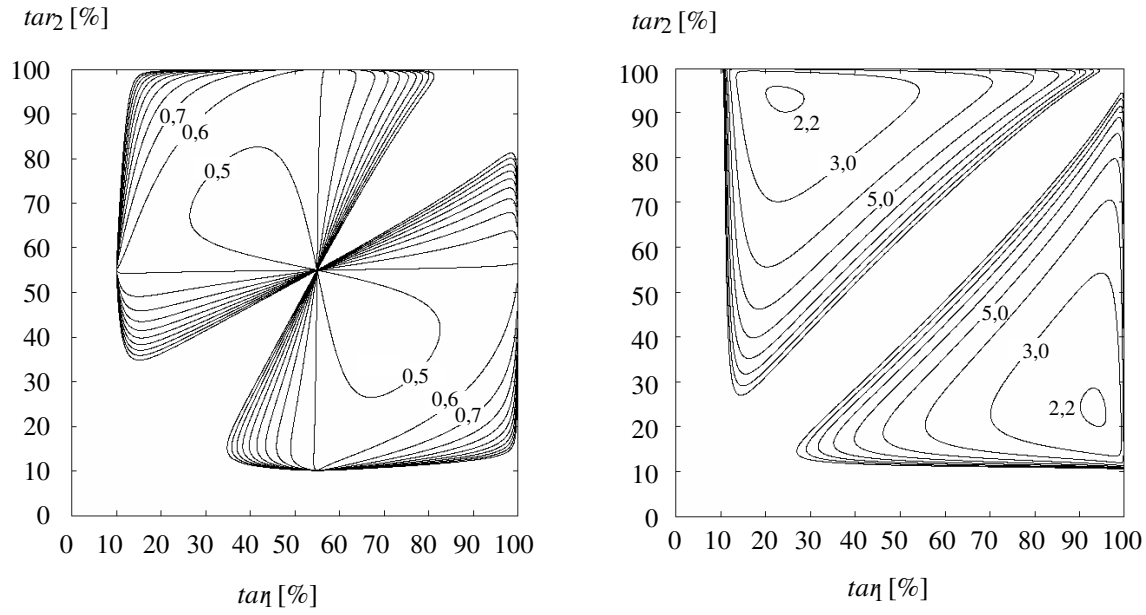
Tabel 4-2: Tabellen angiver ved hvilke sandsynligheder i % for korrekt svar, sætningerne bør præsenteres ved for at opnå en så lille teoretisk spredning som mulig, når enten SRT_{50}, s_{50} eller både SRT_{50} og s_{50} ønskes bestemt. Procentsatserne for x gående imod uendeligt er identiske med de angivet i afsnit 4.1.3.

	Kun SRT_{50}	**Kun s_{50}**		**Både SRT_{50} og s_{50}**	
x	***tar***	***tar_1***	***tar_2***	***tar_1***	***tar_2***
uendelig	50,0	8,3	91,7	19,0	81,1
10	55,0	23,7	93,5	33,9	86,1
9	55,6	25,0	93,6	35,2	86,5
8	56,3	26,6	93,8	36,6	87,0
7	57,1	28,6	94,0	38,6	87,6
6	58,3	31,2	94,3	41,0	88,3
5	60,0	34,6	94,7	44,1	89,2
4	62,5	39,6	95,2	48,7	90,5
3	66,6	47,3	96,0	55,6	92,3
2	75,0	61,6	97,3	68,2	95,2

Som eksempel kan den teoretisk mindste spredning af s_{50} for en 10AFC-test beregnes ved følgende formel, som er bestemt på tilsvarende vis som den for den operatørbaserede version – dvs. svarende til formel (4-8). Tages der højde for, at s_{50} for en 10AFC-test svarer til at gange s_{50} for den operatørbaserede version med 9/10 fås, at spredningen af s_{50} er en anelse større for en 10AFC-test end for den operatørbaserede afvikling.

$$\sigma_{s_{50}} = \frac{1{,}29 \cdot s_{50}}{\sqrt{n}} \tag{4-20}$$

Af figur 4-6 fremgår konturlinjer for spredningen af SRT_{50} og s_{50} som funktion af tar_1 og tar_2. Disse er bestemt ved at sætte SRT_{50} og s_{50} lig hhv. -8,4 dB SNR og 11,9%/dB, mens både n_1 og n_2 er sat lig 50, hvor sidstnævnte svarer til ti sætninger af hver fem ord ved hvert af de to præsentationsniveauer. Som for figur 4-3, s. 34 viser konturlinjerne dels, ved hvilke sandsynligheder for korrekt svar den mindste spredning af SRT_{50} og s_{50} teoretisk kan opnås og dels, hvor meget de pågældende spredninger varierer, når procentsatserne afviger fra de optimale værdier. For de pågældende værdier, er den teoretisk mindste spredning af SRT_{50} som skrevet 0,42 dB, mens den teoretisk mindste spredning af s_{50} tilsvarende er 2,16 %/dB.



a) Spredningen af SRT_{50}

b) Spredningen af s_{50}

Figur 4-6: Illustration af hvordan den teoretiske spredning af SRT_{50} og s_{50} varierer ved forskellige sandsynligheder for korrekt svar i en 10AFC-test. I delfigur a) og b) fremgår et konturplot af hhv. spredningen af SRT_{50} i dB og spredningen af s_{50} i %/dB. For begge delfigurer gælder, at SRT_{50} og s_{50} er sat lig hhv. -8,4 dB SNR og 11,9 %/dB, mens både n_1 og n_2 er sat lig 50.

4.2.3 Alternativ afvikling

Ligesom for den operatørbaserede version kan den patientbaserede version forløbe ved, at FP'en afgiver sit respons ved en åben responsmetode. En test, hvori FP'en afgiver sit respons ved en åben responsmetode, kan eksempelvis laves ved at lade FP'en skriftligt gengive det hørte vha. en computer. En skriftlig gengivelse kræver dog, at FP'en staver korrekt, idet gengivelsen ellers kan fejlfortolkes således, at det pågældende ord registreres som værende en fejl i taleopfattelse og ikke blot en stavefejl. Systemet kunne evt. være opbygget således, at det tager højde for gængse stavefejl – eksempelvis kunne systemet registrere ordet *Hening* (med blot et enkelt *n*) som værende en korrekt gengivelse af ordet *Henning*. Ved udvikling af et sådant system bør der tages højde for, at dansk ikke altid skrives, som det udtales. Det at nogle kan have svært ved skriftligt at gengive det hørte, vil kunne påvirke testens samlede validitet. En anden ulempe ved denne afviklingsform er, hvis FP'en får den opfattelse, at det ikke er hørelsen men derimod vedkommendes stavefærdigheder, der testes. En lignende problematik blev omtalt i afsnit 2.2.1, som beskriver Helen-materialet.

Alternativt kan FP'en som i den operatørbaserede version gentage, hvad vedkommende hørte, hvorefter en computer med indbygget talegenkender – dvs. en mikrofon og speciel designet software – registrerer, hvor mange korrekte ord, der er indeholdt i responset. En sådan løsning byggende på talegenkendelse er mulig at lave i dag – men er dog ikke helt billig. Ifølge PDC (Prolog Development Center), som er en førende dansk virksomhed inden for systemer til talegenkendelse, vil en sådan løsning næppe koste under 100.000 kr. (Carsten Christoffersen – PDC, pers. medd. marts 2007). PDC har gennem flere år udviklet skræddersyede taledialogsystemer til

forskellige større danske virksomheder. Nogle af disse systemer kræver træning af hver enkelt bruger for at kunne genkende stemmeprofil og udtale – for andre systemer er dette ikke nødvendigt.

Som skrevet i afsnit 3.1.2 består talematerialet Dantale II kun af 50 forskellige ord, hvilket er en klar fordel i denne forbindelse, idet et system derfor kun skal kunne genkende et meget begrænset antal ord. Dette vil bl.a. betyde, at der for hvert ord kan indlægges forskellige dialekter og udtalevarianter, hvorved en egentlig træning af systemet måske kan udlades. Det er af stor vigtighed at et sådant system er pålideligt. Det dur jo eksempelvis ikke, at FP'er, der taler nordjysk, opnår dårligere testresultater end FP'er, der taler rigsdansk – på trods af, at deres taleforståelsesevne er den samme. Da en løsning med talegenkendelse, som skrevet, er meget omkostelig, vil denne løsning ikke blive behandlet yderligere. Det kan dog nævnes, at et sådant system imidlertid med tiden – i takt med at udvikling af systemer baseret på talegenkendelse bliver mere udpræget – formodes, at blive billigere. Hvis der om nogle år er interesse for et system baseret på talegenkendelse, anbefales det at undersøge markedet på ny.

4.3 Sammenfatning

For både den operatør- og den patientbaserede version kan FP'ens respons beskrives vha. en S-kurve. For den variant af den patientbaserede version hvor det er muligt at svare *ved ikke* bør det vurderes, hvorvidt FP'ernes respons bedst synes at kunne beskrives med S-kurven for den operatørbaserede version eller den for en *x*AFC-test. Det gælder, at hældningen af S-kurven for en *x*AFC-test svarer til at gange hældningen af S-kurven for den operatørbaserede version med (1–1/*x*). Generelt bør taletærsklen opgives som det niveau, hvorved FP'en forventes at have en taleforståelighed på 50% – dvs. svarende til en værdi af SRT_{50}. I den operatørbaserede version betyder dette, at sætningerne skal afspilles ved det niveau, hvor FP'en kan gengive 50%, mens sætningerne i eksempelvis en 10AFC-test skal afspilles ved et niveau, hvorved FP'en kan gengive hele 55%. Ønskes en værdi af s_{50} bestemt skal sætningerne i den operatørbaserede version afspilles ved to niveauer med en sandsynlighed på hhv. 10 og 90% for korrekt svar. Ønskes i stedet både en værdi af SRT_{50} og s_{50} bør der foretages et kompromis mellem den teoretisk mindste spredning af såvel SRT_{50} som s_{50}. Derved fås, at sætningerne bør afspilles ved to niveauer med en sandsynlighed på hhv. 20 og 80% for korrekt svar. Tilsvarende beregninger kan udføres for en *x*AFC-test. Frem for at lade den patientbaserede version være opbygget omkring en *x*AFC-test, kan den alternativt afvikles ved, at FP'en skriftligt gengiver det hørte eller vha. talegenkendelse.

5 Testafvikling

Foruden fastsættelsen af ved hvilken eller hvilke sandsynligheder for korrekt svar sætningerne skal præsenteres, er der i relation til en adaptiv testafvikling flere parametre, hvis valg bør overvejes. Det være sig om præsentationsniveauet skal justeres ved enten at variere niveauet af tale- eller støjsignalet og om støjen skal være kontinuert eller afbrudt. Derudover skal det vælges ved hvilket niveau det signal, der holdes konstant, skal afspilles samt ved hvilket forhold mellem tale- og støjniveauet den første sætning skal præsenteres. Når dette er fastlagt, skal det også bestemmes, hvordan præsentationsniveauet ud fra FP'ens respons skal ændres samt hvordan det endelige testresultat skal beregnes. Endelig bør der tages højde for, at der i tilknytning til en testafvikling med et talemateriale som Dantale II er observeret en såkaldt indlæringseffekt, hvorfor FP'en bør præsenteres for et træningsforløb, inden selve testen afvikles. I de følgende afsnit beskrives forskellige betragtninger over disse parametre, der alle optræder i såvel en operatør- som i en patientbaseret afvikling.

5.1 Præsentationsformen

Inden en adaptiv testafvikling kan finde sted, skal det som skrevet afgøres, om præsentationsniveauet skal justeres ved enten at variere niveauet af tale- eller støjsignalet og om støjsignalet skal være til stede mellem sætningspræsentationerne – dvs. hvorvidt støjsignalet skal være kontinuert eller afbrudt. Af Wagener & Brand (2005) kaldes disse parametre samlet Presentation mode, mens der i denne rapport – som antydet i overskriften – er valgt en dansk oversættelse heraf. I de to følgende afsnit beskrives forskellige overvejelser omkring disse parametre, hvis valg knytter sig til testens afviklingsniveau, der behandles i afsnit 5.2.

5.1.1 Variation af tale eller støj

I forbindelse med en adaptiv testafvikling er justeringen af præsentationsniveauet i forskellige undersøgelser blevet foretaget på forskellig vis. Ved tests med det svenske materiale er det af Hagerman (1984) og Hagerman & Kinnefors (1995) eksempelvis foregået ved at variere niveauet af støjsignalet, mens taleniveauet blev holdt konstant. Ved tests med det tyske materiale er der af Wagener (1999b) imidlertid blevet anvendt den omvendte fremgangsmåde – dvs. niveauet af talesignalet blev varieret, mens støjniveauet blev holdt konstant. Som det vil fremgå af afsnit 5.3.1, lægger Hansen & Ludvigsen (2001) i forbindelse med tests med det danske materiale op til at variere niveauet af talesignalet – der gives dog ingen begrundelse herfor. For at bestemme om en evt. uoverensstemmelse mellem forskellige testresultater kan tillægges præsentationsformen, er der af Wagener & Brand (2005) blevet udført en undersøgelse heraf. Af undersøgelsen, som blev foretaget med det tyske materiale, blev det konkluderet, at hvorvidt tale- eller støjniveauet varieres, er uden betydning for både bestemmelsen af SRT_{50} og s_{50}.

Inden det afgøres, hvilket signalniveau, der skal varieres, er det dog vigtigt at have en forestilling om, i hvilket område SRT_{50} forventes at befinde sig. Det bør bemærkes, at dette område

kan være afhængigt af, med hvilket støjsignal testen afvikles. Det signal, hvis niveau holdes konstant, skal have et for FP'en behageligt hørbart niveau. Samtidigt må det signal, hvis niveau varieres, ikke kunne blive så højt, at det vækker ubehag hos FP'en. Er det støjniveauet, der varieres, må det omvendt heller ikke blive så lavt, at målerummets naturlige baggrundsstøj vil kunne bidrage til (eller måske helt overtage) maskeringen af talesignalet. Ved at variere støjniveauet kan det tilmed risikeres, at dette niveau bliver lavere end FP'ens høretærskel, hvorfor testen ikke giver et validt billede af FP'ens taleopfattelse i støj. Hvorvidt dette sker, er afhængig af både det valgte taleniveau og FP'ens taleopfattelse.

For at tage højde for ovenstående kan det være en god idé at variere niveauet af talesignalet, mens støjniveauet holdes konstant. Et konstant støjniveau blev derfor valgt i forbindelse med de fire udførte pilottests. Populært kan en forøgelse af taleniveauet beskrives ved, at taleren træder nærmere FP'en, mens taleren træder længere væk fra FP'en, når taleniveauet sænkes. Dette gælder, hvis støjen fra omgivelserne kan betragtes som værende konstant og at taleren taler direkte mod FP'en. Alternativt kan en forøgelse af taleniveauet beskrives ved, at taleren hæver stemmen, mens taleren taler mere lavmælt, når taleniveauet sænkes. I forbindelse hermed bør det dog nævnes, at talens frekvensspektrum er forskelligt ved eksempelvis råben og hvisken.

5.1.2 Kontinuert eller afbrudt støj

Mht. støjsignalet skal det afgøres, hvorvidt dette skal være kontinuert eller afbrydes mellem sætningspræsentationerne. Som for ændringen af præsentationsniveauet omtalt i forrige afsnit, er der også her i forskellige undersøgelser gjort forskellige valg – alene for det danske materiale. Af Wagener *et al*. (2003) er der eksempelvis blevet anvendt afbrudt støj, hvor støjen startede kort før afspilningen af sætningen begyndte og sluttede igen, når sætningen var færdig. Da der her altid vælges det samme udpluk af støjen, begrundes dette valg med, at støjen vil maskere de enkelte ord ens hver gang, de præsenteres. Hvorvidt dette reelt er tilfældet, kan dog diskuteres, idet ordene ikke har præcis samme længde. Fordelen er måske snarere, at der ikke er støj til stede, når FP'en afgiver sit svar – støjen vil måske unødigt kunne forstyrre FP'en. Hansen & Ludvigsen (2001) anbefaler i stedet, at der anvendes et kontinuert støjsignal, idet der derved opnås en mere realistisk lyttesituation og derved en mere valid test. Af den grund blev støjsignalet i forbindelse med de fire udførte pilottests valgt kontinuert. I øvrigt bør nævnes, at testes FP'er med HA udsættes dette derved for en konstant kendt støjpåvirkning, hvorved transiente fænomener i apparatet undgås.

Af Wagener & Brand (2005) er det blevet undersøgt, hvorvidt støjens kontinuitet har indflydelse på bestemmelsen af SRT_{50} og s_{50}. I undersøgelsen, som blev udført med det tyske materiale, indgik både normalthørende og hørehæmmede FP'er. Resultatet af undersøgelsen viste, at hvorvidt støjen er kontinuert eller afbrudt ikke har betydning for bestemmelsen af s_{50}. FP'erne opnåede imidlertid mindre – og derved bedre – værdier af SRT_{50} i kontinuert end i afbrudt støj. Denne forskel i SRT_{50} var statistisk signifikant for de normalthørende men ikke for de hørehæmmede FP'er, hvilket indikerer vigtigheden i at notere hvad, der i en given testsituation er blevet gjort. I forbindelse med undersøgelsen blev det tilmed undersøgt, om den bratte start af den afbrudte støj virkede mere maskerede på det første end på de efterfølgende ord – dvs. om forskellen mellem SRT_{50} for egennavnet og de andre ord var større i afbrudt end i kontinuert

støj. Dette viste sig imidlertid ikke at være tilfældet, hvorfor forskellen i SRT_{50} mellem de to forskellige præsentationsformer måske snarere skal tillægges, at det menneskelig øre må have inkorporeret en form for støjreduktion, der er bedre til at filtre støj fra, hvis den er til stede uafbrudt.

5.2 Afviklingsniveauet

Ved fastsættelsen af den såkaldte præsentationsform bør niveauet af det signal, der holdes på et konstant niveau også tages i betragtning. Da det i forbindelse med de udførte pilottests er valgt at variere niveauet af talesignalet, mens støjniveauet holdes konstant, vil følgende tage udgangspunkt i overvejelser omkring niveauet af støjsignalet. Samtidigt berøres også ganske kort præsentationsniveauet af den første sætning – dvs. ved hvilket forhold mellem tale- og støjniveauet denne bør afspilles. Det bør bemærkes, at valget af støjniveau evt. kan sætte begrænsninger i testens målgruppe.

5.2.1 Niveauet af støjsignalet

Vælges støjniveauet konstant, skal dette som antydet i afsnit 5.1.1 have et fornuftigt niveau således, at dette kan høres og at taleniveauet for en FP med dårlig taleopfattelse ikke bliver ubehageligt højt. Jf. Hansen & Ludvigsen (2001) vælges typisk et støjniveau på 65 dB(C) SPL, hvilket også gør sig gældende for de fire udførte pilottests. Et sådant niveau virker rimeligt, idet det af Poulsen (1993 – Kap. 2) fremgår, at det gennemsnitlige taleniveau for mænd målt i 1 m's afstand netop er 65 dB(C) SPL, mens det for kvinder er ca. 3 dB lavere – ved råben øges værdierne imidlertid med 10-15 dB. Af Wagener & Brand (2005) fremgår, at støjniveauet kan hæves op til 85 dB(C) SPL uden, at dette vil have betydning for bestemmelsen af hverken SRT_{50} eller s_{50}. I tilknytning hertil anbefaler Wagener & Brand at anvende et støjniveau på 65 dB(C) SPL ved tests af normalthørende, mens støjniveauet ved tests af hørehæmmede bør være 80 dB(C) SPL. Sidstnævnte skyldes som erfaret i forbindelse med pilottest IV, at 65 dB(C) SPL for nogle hørehæmmede uden brug af HA er for lavt til, at de kan høre signalet. Det kan i øvrigt overvejes, at indstille niveauet af støjsignalet individuelt. Dette kan evt. foregå ud fra FP'ens høretærskler for rentoner eller ved at lade vedkommende selv justere niveauet af et tale- og /eller støjsignal således, at vedkommende synes, at dette har et behageligt niveau.

5.2.2 Præsentationsniveauet af den første sætning

I forbindelse med fastsættelsen af støjniveauet bør præsentationsniveauet af den første sætning i en test også overvejes. I forbindelse med tests med det svenske materiale anbefaler Hagerman & Kinnefors (1995), at den første sætning præsenteres ved et niveau på 20 dB SNR, mens der i DS/EN ISO 8253-3, § 15.3 generelt proponeres, at denne præsenteres ved et niveau på 5 dB SNR. Som det vil fremgå af afsnit 5.3.1, skal den første sætning jf. Hansen og Ludvigsen (2001) præsenteres ved et niveau på 0 dB SNR. Sidstnævnte blev valgt i forbindelse med Pilottest I og III. Da Pilottest IV blev udført med hørehæmmede FP'er blev den første sætning her præsenteret

ved et niveau på 5 dB SNR – tale- og støjniveauet var hhv. 70 og 65 dB(C) SPL. Dette blev valgt for at sikre, at så mange af FP'erne som muligt kunne høre min. et af de afspillede ord i den første sætning. Det vurderes som vigtigt, at FP'en kan høre noget at den første sætning, idet vedkommende derved hurtigere formodes at blive fortrolig med testsituationen. I forbindelse med præsentationsniveauet kan det evt. overvejes, at lade støjen i løbet af de første sætninger gradvist stige i intensitet indtil det ønskede niveau er opnået.

5.3 Afviklingsprocedurer

I forbindelse med afvikling af Dantale II materialet er der i litteraturen omtalt to forskellige adaptive afviklingsprocedurer, som vil blive beskrevet i de følgende afsnit. Den først procedure, der beskrives, stammer fra en vejledning omkring Dantale II materialet. Denne vejledning blev udgivet i tilknytning til lanceringen af Dantale II cd'en og er forfattet af Hansen & Ludvigsen (2001), hvorfor det her er valgt at kalde den pågældende procedure Hansen og Ludvigsens procedure. Den anden procedure, der beskrives, er tidligere beskrevet i artiklen af Brand & Kollmeier (2002), hvorfor den her kaldes Brand og Kollmeiers procedure. Foruden disse to procedurer beskrives også, hvordan Dantale II materialet kan afvikles vha. en MML-procedure (Method of Maximum Likelihood). Principperne bag en sådan procedure er grundlæggende beskrevet i Poulsen (2001 – Kap. 3.3). Det bør bemærkes, at Hansen og Ludvigsens procedure ikke er blevet anvendt i de udførte pilottests. Til gengæld er Brand og Kollmeiers procedure blevet anvendt i Pilottest I, mens MML-proceduren er brugt i både Pilottest III og IV. For alle tre procedurer er der udført computersimuleringer, som ligeledes vil blive beskrevet i dette afsnit.

5.3.1 Hansen og Ludvigsens procedure

Hansen og Ludvigsens procedure er forholdsvis simpel og kan i modsætning til de to andre procedurer, som her beskrives, administreres manuelt – dvs. uden brug af computer. Den er designet til afvikling med en åben responsmetode dvs. svarende til måden, hvorpå FP'erne i den operatørbaserede version afgiver deres svar. I forbindelse med proceduren anbefales, at der i en egentlig testafvikling anvendes i alt 30 sætninger – svarende til tre tilfældige lister. Første sætning skal præsenteres ved et niveau på 0 dB SNR. Mens støjniveauet holdes konstant, justeres niveauet af de enkelte sætninger ud fra FP'ens netop afgivne respons. Hvordan antallet af korrekte ord i FP'ens respons afgør ændringen af taleniveauet, fremgår af tabel 5-1. Skiftet mellem tabellens store og små trin foretages efter den femte sætning. Det bør bemærkes, at ønskes i stedet niveauet af støjsignalet varieret, mens taleniveauet holdes konstant, skal fortegnet på alle trinene vendes.

Trinene stifter fortegn omkring værdien for 2,5 korrekte ord, hvorfor proceduren konvergerer imod en værdi af *tar* på 50%. Hvis FP'en til den første sætning kan svare alle fem ord korrekt er præsentationen for vedkommende let, hvorfor taleniveauet sænkes 3 dB. Hvis FP'en derimod kun kan svare et enkelt ord korrekt, er præsentationen for vedkommende svær, hvorfor taleniveauet hæves 2 dB. Hvorvidt præsentationen betragtes som let eller svær skal ses i forhold til værdien af *tar* på 50%. Når alle sætninger er blevet præsenteret, beregnes testresultatet som et

gennemsnit af præsentationsniveauerne ved sætning nr. 12 til 31. Selvom sætning nr. 31 ikke præsenteres, bestemmes niveauet af denne som for de øvrige sætninger ud fra tabellens værdier. Da proceduren skal anvendes med en åben responsmetode, svarer resultatet derved til en værdi af SRT_{50}. I øvrigt kan nævnes, at hvis proceduren i stedet for 30 afvikles med eksempelvis 20 sætninger, beregnes SRT_{50} som et gennemsnit af præsentationsniveauerne ved sætning nr. 12 til 21.

Tabel 5-1: Ændring af taleniveauet bestemt ud fra antallet af korrekte ord til forrige sætning.
(Gengivelse af Hansen & Ludvigsen, 2001 – Tabel 1).

Antal korrekte ord i forrige sætning	**Ændring af taleniveauet [dB]**	
	Store trin	**Små trin**
0	3	2
1	2	1
2	1	0
3	-1	0
4	-2	-1
5	-3	-2

5.3.2 Brand og Kollmeiers procedure

I modsætning til Hansen og Ludvigsens procedure kan Brand og Kollmeiers procedure anvendes således, at der både bestemmes en værdi af SRT_{50} og s_{50}. Ved både at bestemme SRT_{50} og s_{50} kan et vilkårligt punkt på S-kurven fastlægges. Proceduren er ligesom Hansen og Ludvigsens procedure designet til en operatørbaseret afvikling, men kan også anvendes i forbindelse med en patientbaseret afvikling. I proceduren bestemmes de enkelte præsentationsniveauer ud fra følgende formel, som udtrykker ændringen af niveauet i forhold til den forrige præsentation – dvs. trinstørrelsen. Ændringen af præsentationsniveauet kan enten være i tale- eller støjniveauet.

$$\Delta SNR = -\frac{f(i)\cdot(prev - tar)}{s_{tar}} \tag{5-1}$$

I formlen angiver tar og s_{tar} hhv. den sandsynlighed for korrekt svar, hvorved sætningerne ønskes præsenteret og hældningen af S-kurven ved tar. Parameteren $prev$ er den såkaldte diskriminationsværdi og angiver andelen af de ord, som FP'en til forrige sætning gengav korrekt. Konvergensraten og derved trinstørrelsen styres af funktionen i formel (5-2), hvorved størrelsen af trinene adaptivt mindskes efter hver reversering af taleniveauet. En reversering, i opstår, når trinene skifter fortegn.

$$f(i) = a \cdot b^{-i} \tag{5-2}$$

Ifølge Brand & Kollmeier (2002) bør værdien af *a* og *b* hhv. sættes lig 1,5 og 1,41 og samtidigt bør der for funktionen i formel (5-2) indføres en minimumsværdi på 0,1. Anvendes de opgivne værdier for *a* og *b*, opnås denne minimumsværdi ved den ottende reversering. Det vil med andre ord sige, at fra den ottende reversering og frem bør de opnåede trinstørrelser holdes konstante. Ønskes proceduren anvendt i en operatørbaseret afvikling, hvor SRT_{50} ønskes bestemt, skal s_{tar} sættes lig s_{50}. Jf. afsnit 4.1.1 er s_{50} for Dantale II materialet 13,2 %/dB. Indsættes denne værdi af s_{50} fås de trinstørrelser, der angivet i tabel 5-2.

Tabel 5-2: Ændring af præsentationsniveauet bestemt ud fra andelen af korrekte ord til forrige sætning og antallet af opnåede reverseringer. Størrelsen af trinene er bestemt ved i formel (5-1) at sætte *tar* og s_{tar} lig hhv. 50% og 13,2 %/dB.

	Ændring af præsentationsniveauet [dB]								
prev	***i* = 0**	***i* = 1**	***i* = 2**	***i* = 3**	***i* = 4**	***i* = 5**	***i* = 6**	***i* = 7**	***i* = 8**
0,0	5,68	4,03	2,86	2,03	1,44	1,02	0,72	0,51	0,36
0,2	3,41	2,42	1,71	1,22	0,86	0,61	0,43	0,31	0,22
0,4	1,14	0,81	0,57	0,41	0,29	0,20	0,14	0,10	0,07
0,6	-1,14	-0,81	-0,57	-0,41	-0,29	-0,20	-0,14	-0,10	-0,07
0,8	-3,41	-2,42	-1,71	-1,22	-0,86	-0,61	-0,43	-0,31	-0,22
1,0	-5,68	-4,03	-2,86	-2,03	-1,44	-1,02	-0,72	-0,51	-0,36

Efter at responset til den sidste sætning er afgivet, bestemmes ud fra de opnåede præsentationsniveauer og tilhørende responser et maksimum-likelihood-estimat for S-kurvens forløb. Dette bestemmes ud fra følgende formel, hvori parameteren m angiver antallet af sætninger FP'en er blevet præsenteret for. Parameteren c er antallet af ord, der er svaret korrekt ved den k'te sætningspræsentation. Hvis materialet består af fem ord, svarer c med andre ord til at gange parameteren $prev$ med 5. Det bør bemærkes, at består materialet ikke af fem ord, skal 5-tallet i formlen ændres. Afhængigt af hvordan FP'ens respons forventes at kunne beskrives, sættes p lig den logistiske funktion for enten den operatørbaserede version eller for en *x*AFC-test – dvs. svarende til hhv. formel (4-1) eller (4-15).

$$p_{Like}(SNR) = \prod_{k=1}^{m} p(SNR_k)^{c_k} \cdot (1 - p(SNR_k))^{5-c_k} \tag{5-3}$$

Estimatet for S-kurvens forløb bestemmes ved at variere værdien af SRT_{50} og s_{50} indtil udtrykket $\log(p_{Like}(SNR))$ er maks. Den heraf fundne værdi af SRT_{50} og s_{50} er da den mest sandsynlige, hvorved testresultatet er fundet. Præsenteres sætningerne kun ved én værdi af *tar* bliver den teoretiske spredning af s_{50} jf. figur 4-3, s. 43 og figur 4-6, s. 43 imidlertid forholdsvis stor. Samtidigt opstår der en forskydning af bestemmelsen af s_{50}, hvilket vil fremgå i forbindelse med figur 5-3, s. 58. Ved kun at medtage en værdi af *tar* bør værdien af s_{50} derfor forkastes. Ønskes en værdi af s_{50}, skal der med proceduren som tidligere skrevet konvergeres imod to

punkter. Ifølge Brand & Kollmeier bør der i en procedure, hvormed kun SRT_{50} ønskes bestemt, indgår min. 20 sætninger. Hvis der både ønskes en værdi af SRT_{50} og s_{50}, bør der derimod indgå min. 30 sætninger. Disse antal af sætninger er af Brand & Kollmeier blevet bestemt ud fra computersimuleringer svarende til dem, der vil fremgå af afsnit 5.3.4.

Som beskrevet i afsnit 1.1 indgår denne afviklingsprocedure i DELTA's testprogram. I forbindelse hermed fastsættes præsentationsniveauet løbende ud fra formel (5-1). Selve testresultatet, der opgives som en værdi af SRT_{50}, bestemmes imidlertid ikke vha. et maksimum-likelihood-estimat. I stedet beregnes SRT_{50} som et gennemsnit af præsentationsniveauerne ved de sidste fire opnåede reverseringer. I Pilottest I blev det undersøgt, om der evt. er forskel mellem at bestemme SRT_{50} som et maksimum-likelihood-estimat eller at beregne SRT_{50} som et gennemsnit af ti præsentationsniveauer eller af de sidste fire reverseringer. For SRT_{50} bestemt med de tre forskellige beregningsmetoder blev der imidlertid ikke fundet en statistisk signifikant forskel – se evt. figur C-12, s. 128. Beregnes SRT_{50} ud fra et bestemt antal reverseringer, bør det bemærkes, at små trin giver færre reverseringer end store trin inden for det samme antal sætninger. I en sådan beregning bør der i øvrigt anvendes et lige antal reverseringer således, at testresultatet ikke trækkes skævt.

5.3.3 MML-proceduren

I en MML-procedure estimeres forløbet af den mest sandsynlige S-kurve, hver gang FP'en har afgivet et respons – et sådant estimat kaldes som i forrige afsnit et maksimum-likelihood-estimat. Proceduren kan som Brand og Kollmeiers procedure anvendes således, at der både bestemmes en værdi af SRT_{50} og s_{50}. I Pilottest III og IV blev proceduren anvendt ved først at præsentere FP'en for fire sætninger, inden det første estimat bestemmes. I Pilottest III og IV blev den første sætning præsenteret ved et niveau på hhv. 0 og 5 dB SNR, mens præsentationsniveauet af den anden til fjerde sætning blev bestemt ud fra trinene angivet i tabel 5-3 ved at ændre niveauet af talesignalet. Størrelsen af trinene er valgt således, at niveauet af den følgende præsentation formodes at befinde sig inden for S-kurvens dynamikområde – trinene svarer til afrundede værdier af trinstørrelserne i tabel 5-2 for $i = 0$.

Tabel 5-3: Ændring af taleniveauet bestemt ud fra antallet af korrekte ord til forrige sætning – trinene svarer til afrundede værdier af trinstørrelserne i tabel 5-2 for $i = 0$.

Antal korrekte ord i forrige sætning	Ændring af taleniveauet [dB]
0	6
1	3
2	1
3	-1
4	-3
5	-6

Grunden til at der først blev præsenteret fire sætninger, hvis niveau blev bestemt ud fra trinstørrelserne i tabellen, er, at der derved er *noget at regne på* inden det første estimat bestemmes. I forbindelse med de i afsnit 5.3.4 omtalte computersimuleringer blev det undersøgt om den teoretiske spredning, der dermed kan opnås, er afhængig af, om der medtages fire, seks eller otte sætninger inden bestemmelsen af det første estimat. Dette vidste sig imidlertid ikke at være tilfældet, hvorfor simuleringerne ikke er valgt gengivet.

Efter præsentationen af den fjerde sætning og hver af de følgende sætninger blev der bestemt et maksimum-likelihood-estimat for S-kurvens forløb vha. formel (5-3), som er angivet i forrige afsnit. I modsætning til ovenfor sættes m her løbende lig antallet af hidtil præsenterede sætninger – efter præsentationen af den fjerde sætning er m eksempelvis lig 4. Det bør i øvrigt bemærkes, at i Brand og Kollmeiers procedure bestemmes præsentationsniveauet af den enkelte sætning ud fra FP'ens respons til den forrige, mens niveauet her bestemmes på baggrund af samtlige af de responser, som FP'en løbende har afgivet. Med andre ord indgår der i bestemmelsen af estimatet her flere og flere værdier af c_k og SNR_k i takt med at antallet af præsenterede sætninger stiger.

Som for Brand & Kollmeiers procedure kan det vælges at sætte p lig den logistiske funktion for enten den operatørbaserede version eller for en xAFC-test – dvs. svarende til hhv. formel (4-1) eller (4-15). Tilsvarende bestemmes estimatet for S-kurvens forløb ved at variere værdien af SRT_{50} og s_{50} indtil udtrykket $\log(p_{Like}(SNR))$ er maks. Den heraf fundne værdi af SRT_{50} og s_{50} er da den mest sandsynlige, hvorfor disse indsættes i den gældende logistiske funktion. Ved på forhånd at have fastsat ved hvilken eller hvilke sandsynligheder for korrekt svar sætningerne ønskes præsenteret, kan præsentationsniveauet af den følgende sætning bestemmes. I Pilottest III og IV blev det valgt at præsentere sætningerne på skift ved niveauer med en sandsynlighed på hhv. 20 og 80% for korrekt svar – disse procentsatser blev valgt i henhold til beregningerne i afsnit 4.1.3, idet både SRT_{50} og s_{50} ønskes bestemt.

I de to pilottests er det valgt at præsentere FP'erne for et fast antal sætninger. Alternativt kunne det være valgt at stoppe testen, når værdien af SRT_{50} og/eller s_{50} havde været (næsten) uændret gennem flere præsentationer. I forbindelse med Pilottest IV er det beregnet, hvor meget værdien af såvel SRT_{50} som s_{50} varierer gennem en test og derved hvor stor en spredning af SRT_{50} og s_{50}, der kan forventes efter en præsentation af et bestemt antal sætninger – se evt. figur D-12 og D-13, s. 178.

5.3.4 Computersimuleringer

Fælles for de tre procedurer beskrevet i forrige afsnit er, at de er dobbelt adaptive, hvorved bestemmelsen af niveauændringen både er afhængig af, hvor mange ord FP'en svarende korrekt til forrige sætning og hvor langt testen er fremskredet. Da det derved er svært at vurdere, hvor stor en re-test variation der kan forventes med hver af disse producerer, er der udført computersimuleringer for at undersøge dette. Disse er kun udført for en operatørbaseret afvikling. Som for Hansen og Ludvigsens procedure blev Brand og Kollmeiers procedure og MML-proceduren simuleret med en værdi af tar på 50% – dvs. svarende til en sandsynlighed på 50% for korrekt svar. Desuden blev der for Brand og Kollmeiers procedure og MML-proceduren udført simuleringer, hvor sætningerne blev præsenteret ved to forskellige sandsynligheder for korrekt svar. I

forbindelse hermed blev sætningerne præsenteret ved sandsynligheder på hhv. 10 og 90% for korrekt svar og ved sandsynligheder på hhv. 20 og 80% for korrekt svar. Disse procentsatser, hvorved sætningerne blev præsenteret på skift, er valgt på baggrund af beregningerne i afsnit 4.1.3. I de forskellige simuleringer blev værdien af s_{tar} i Brand og Kollmeiers procedure sat forskelligt. For *tar* lig 50% blev s_{tar} sat lig 13,2 %/dB svarende til referenceværdien af s_{50}. Ud fra denne værdi og ved at differentiere formel (4-1) mht. *SNR* blev s_{tar} for *tar* lig 20 og 80% sat lig 8,4 %/dB. Tilsvarende blev s_{tar} sat lig 4,8 %/dB for *tar* lig 10 og 90%.

I hver simulering indgik 1.000 fiktive FP'er. Hver fiktiv FP blev dannet ved at tildele vedkommende en tilfældig værdi af SRT_{50} fra en normalfordeling med en gennemsnitsværdi og spredning på hhv. 0 og 5 dB. Den valgte spredning stemmer fint overens med de spredninger, der blev bestemt i Pilottest IV – se evt. s. 172. For Hansen og Ludvigsens procedure blev der foretaget tre forskellige simuleringer, hvor de fiktive FP'er fik tildelt en værdi af s_{50} på hhv. 5, 10 eller 20 %/dB. Tilsvarende blev der for Brand og Kollmeiers procedure og MML-proceduren foretaget tre simuleringer for hver af de valgte sandsynligheder for korrekt svar – se evt. figurteksten til figur 5-1 til figur 5-4.

I hver simulering blev de 1.000 fiktive FP'er præsenteret for 50 sætninger af hver fem ord. Den første sætning blev præsenteret ved et niveau på 0 dB SNR, mens niveauet af de følgende sætninger blev bestemt ud fra den pågældende afviklingsprocedure. FP'ens respons til hver enkelt sætning svarer til, at vedkommende udfører fem på hinanden følgende Bernoulli-forsøg. For et Bernoulli-forsøg gælder som tidligere skrevet i afsnit 4.1.2, at der kun er to udfald – enten gengiver FP'en ordet korrekt eller forkert. For hver FP blev der for hvert af sætningernes ord valgt et tilfældigt tal mellem 0 og 1, som antages at kunne repræsentere FP'ens indre variabel svarende til den sensoriske oplevelse, som præsentationen af det pågældende ord giver anledning til.

For at kunne tildele FP'en et respons efter hver præsenteret sætning blev hvert af de fem tal for den indre variabel sammenlignet med værdi af p_{OPE} bestemt vha. formel (4-1), som svarer til den logistiske funktion for den operatørbaserede version. Værdien af p_{OPE} blev bestemt ved i formlen at indsætte det pågældende præsentationsniveau og FP'ens tildelte værdi af SRT_{50} og s_{50}. Hvis den indre variabel var mindre end eller lig med p_{OPE}, blev FP'ens svar til det pågældende ord sat som værende korrekt. Var den indre variabel derimod større end p_{OPE}, blev FP'ens svar sat som værende forkert. Denne måde, at lave en fiktiv FP på er i overensstemmelse med litteraturen – se evt. Brand og Kollmeier (2002) og Hagerman (1984). Det bør bemærkes, at ved at anvende formel (4-1) forudsættes, at denne er rimelig til at beskrive FP'ernes forventede respons.

Ud fra FP'ens respons og den pågældende afviklingsprocedure blev præsentationsniveauet af den følgende sætning bestemt. Efter præsentationen af de første ti sætninger blev SRT_{50} og s_{50} for hver FP løbende beregnet relativt til de værdier vedkommende havde fået tildelt – disse beregnede værdier kaldes i det følgende fejlen på hhv. SRT_{50} og s_{50}. Samlet over de 1.000 FP'er blev gennemsnittet og spredningen af fejlen på SRT_{50} og s_{50} bestemt for at vurdere hhv. *forskydningen* og *usikkerheden* af bestemmelsen af de to værdier – sidstnævnte kan relateres til størrelsen af re-test variationen. Disse gennemsnit og spredninger fremgår for de tre procedurer af figurerne på de følgende sider. Af de figurer, hvoraf de fundne spredninger fremgår, er den teoretisk mindste spredning af hhv. SRT_{50} og s_{50} også indtegnet – dvs. svarende til formel (4-

3) og (4-8), som er angivet i hhv. afsnit 4.1.2 og 4.1.3. Det bør i øvrigt bemærkes, at hverken gennemsnittet eller spredningen af fejlen på s_{50} kan bestemmes for Hansen og Ludvigsens procedure – se evt. afsnit 5.3.1.

I den opgivne litteratur findes der simuleringer af Brand og Kollmeier procedure, som stort set er foretaget med samme fremgangsmåde, som de her udførte (Brand og Kollmeier, 2002). Den væsentligste forskel simuleringerne imellem er, at der i de, der fremgår af litteraturen, er taget højde for *j*-faktoren. Imidlertid er det ikke muligt at medtage *j*-faktoren i simuleringerne med Hansen og Ludvigsens procedure, idet denne procedure bygger på tabelopslag. Derfor er der i simuleringerne her fast valgt ikke at tage højde for denne faktor. Som en bibemærkning kan nævnes, at hvis der med eksempelvis MML-proceduren er et ønske om at tage højde for *j*-faktoren, skal antallet af Bernoulli-forsøg ikke sættes lig fem men *j*. Det bør i øvrigt bemærkes, at det er valgt kun at lave simuleringer med en operatørbaseret afvikling. Ved at anvende formel (4-15) frem for formel (4-1) og ændre de angivne sandsynligheder for korrekt svar kunne der være lavet tilsvarende simuleringer for en *x*AFC-test. Jf. afsnit 4.2.2 forventes spredningen af fejlen på både SRT_{50} og s_{50} imidlertid at være større end for en operatørbaseret afvikling. Endelig bør det bemærkes, at simuleringerne ikke fortæller noget om indlæringseffekten, der behandles i afsnit 5.4.

Gennemsnittet af fejlen på SRT_{50}

Af figur 5-1 fremgår gennemsnittet af fejlen på SRT_{50}. For alle tre procedurer ses, at gennemsnittet befinder sig omkring 0 dB, hvorfor der for ingen af procedurerne opstår en uønsket forskydning af bestemmelse af SRT_{50} – dette gælder uafhængigt af den valgte værdi af *tar*. For s_{50} lig 5 %/dB synes gennemsnittet dog at have større fluktuationer omkring 0 dB end hvad, der er tilfældet, når værdien af s_{50} er lig 10 eller 20 %/dB. Disse fluktuationer kan forklares ud fra spredningen af fejlen på SRT_{50}, idet denne er større for s_{50} lig 5 %/dB end for s_{50} lig 10 eller 20 %/dB.

Spredningen af fejlen på SRT_{50}

Af figur 5-2 fremgår spredningen af fejlen på SRT_{50}. Betragtes figurerne for *tar* lig 50% ses, at medtages der i en test 20-30 sætninger, kan der med Hansen og Ludvigsens procedure forventes en re-test variation, der enten er mindre eller i samme størrelsesorden som den, der kan opnås med hhv. Brand og Kollmeiers procedure og MML-proceduren. Grunden til at både Brand og Kollmeiers procedure og MML-proceduren opnår en spredning, der efter 20-30 sætninger er tættere på den teoretisk mindst mulige end hvad der er tilfældet for Hansen og Ludvigsens procedure kan forklares ud fra størrelsen af trinene de tre procedurer imellem.

For Brand og Kollmeiers procedure og MML-proceduren ses samtidigt en tendens til, at den mindste spredning som forventet jf. figur 4-3 a), s. 34 opnås, når *tar* er lig 50%, mens spredningen er størst for *tar* lig 10 og 90%. At spredningen for *tar* lig 20 og 80% er større end for *tar* lig 50% er i overensstemmelse med den i Pilottest I og III fundne re-test variation af SRT_{50} – sammenlign evt. figur C-9, s. 124 og figur C-27, s. 158. Hertil bør det huskes, at Pilottest I blev afviklet i henhold til Brand og Kollmeiers procedure indeholdende en værdi af *tar* – hver måling indbefattede 20 sætninger, mens Pilottest III blev afviklet med MML-proceduren indeholdende

to værdier af *tar*. Som det ses af figur 5-2 bliver spredningen mindre jo flere sætninger, der medtages. Det samme gjorde sig gældende i Pilottest III for re-test variationen af SRT_{50} bestemt med såvel den operatør- som den patientbaserede version. I forbindelse hermed kan nævnes, at for begge versioner blev der i pilottesten ikke fundet nogen statistisk signifikant forskel mellem at bestemme re-test variationen af SRT_{50} med 10, 20 og 30 sætninger.

Af figur 5-2 g) fremgår for Brand og Kollmeiers procedure en forholdsvis stor spredning inden for de første 20 sætninger. Denne spredning kan forklares med, at der i proceduren indgår referenceværdien af hældningen af S-kurven repræsenteret ved værdien af s_{tar}. Da referenceværdien er væsentligt større end 5 %/dB, bliver trinene i forhold til en S-kurve med en hældning på 5 %/dB, forholdsvis små. Der skal derfor relativt mange præsentationer til for at opnå en værdi af præsentationsniveauet som er nær værdien af SRT_{50}. For s_{50} lig 10 og 20 %/dB synes MML-proceduren i øvrigt at være marginalt bedre end Brand og Kollmeiers procedure.

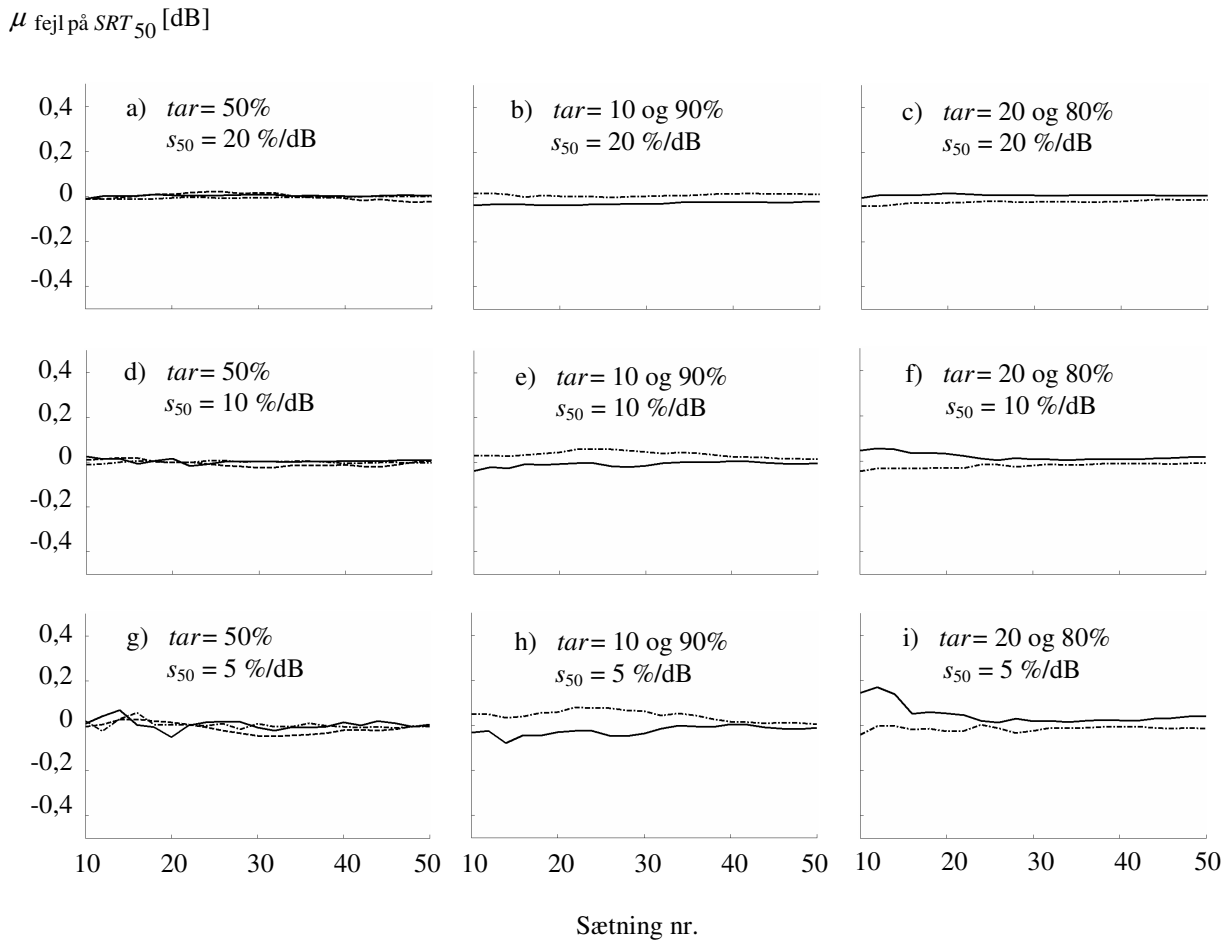
Gennemsnittet af fejlen på s_{50}

Af figur 5-3 fremgår gennemsnittet af fejlen på s_{50}. For både Brand og Kollmeiers procedure og MML-proceduren ses, at gennemsnittet befinder sig over 0 dB, hvorfor der her opstår en uønsket forskydning af bestemmelsen af s_{50}. Denne forskydning er dog mindst for *tar* lig 10 og 90%. Forskydningen af s_{50} ved *tar* lig 50% kan forklares med, at præsentationsniveauerne kun kan variere en smule fra præsentation til præsentation, mens FP'ernes respons kan variere fra nul til fem ord korrekt. På lignende vis kan forskydningen af s_{50} for *tar* lig 20 og 80% samt for *tar* lig 10 og 90% forklares.

Idet forskydningen af s_{50} er størst, når proceduren kun indeholder én værdi af *tar*, bør værdier af s_{50} opnået hermed som skrevet i afsnit 5.3.2 forkastes. At dette er tilfældet, blev også fundet i forbindelse med Pilottest I – se evt. figur C-8, s. 123. Af figur 5-3 ses i øvrigt, at for *tar* lig 50% opnås der med Brand og Kollmeiers procedure en mindre forskydning end hvad, der er tilfældet for MML-proceduren. Dette kan evt. skyldes, at trinene i Brand og Kollmeiers procedure er større end i MML-proceduren. Større trin betyder nemlig, at sætningerne præsenteres over en større del af S-kurvens dynamikområde, hvorved kurvens forløb og herunder hældningen bedre kan bestemmes.

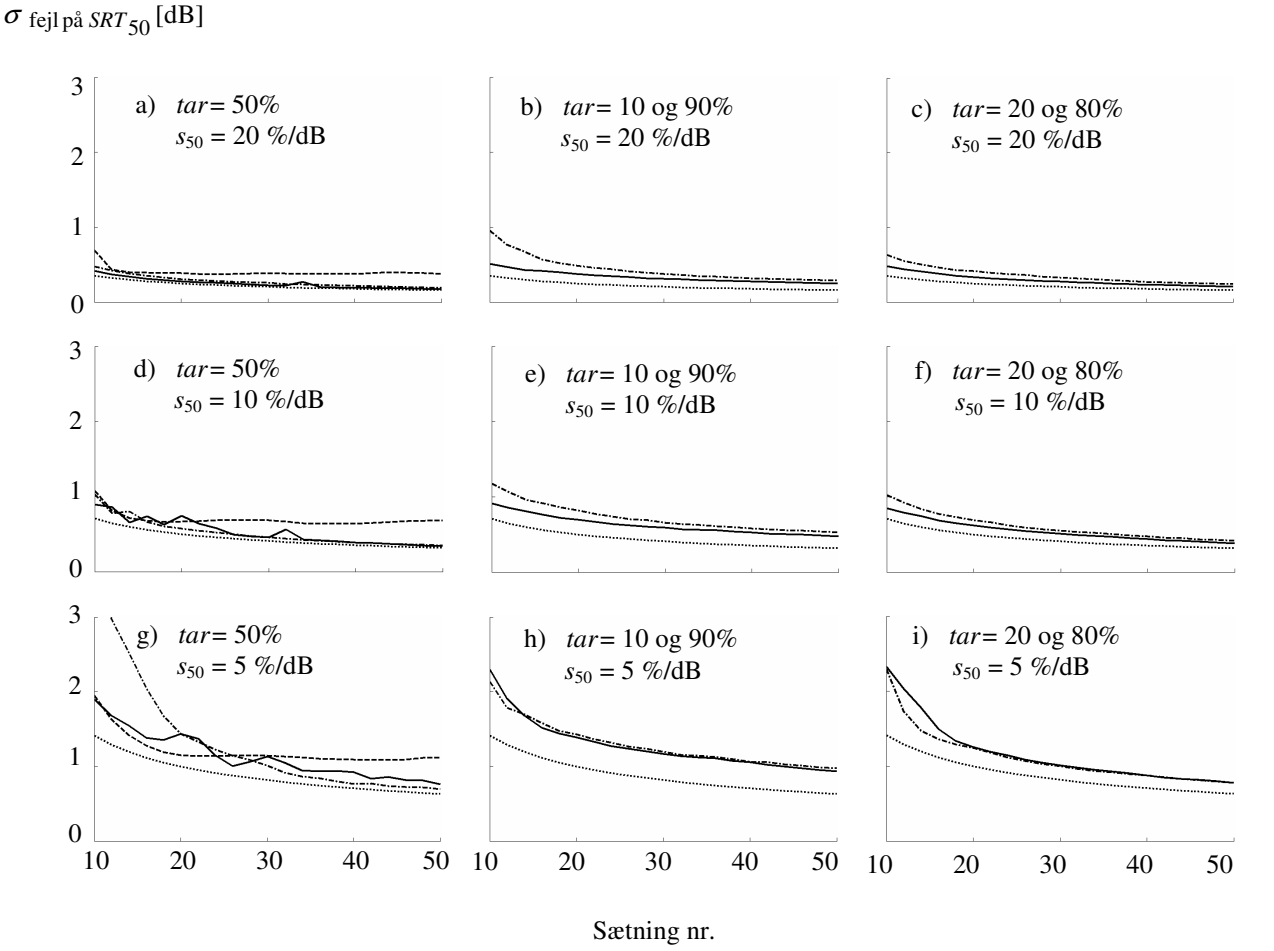
Spredningen af fejlen på s_{50}

Af figur 5-4 fremgår spredningen af fejlen på s_{50}. Som forventet jf. figur 4-3 b), s. 34 ses for både Brand og Kollmeiers procedure og MML-proceduren, at den største spredning opnås for *tar* lig 50%, mens spredningen er mindst for *tar* lig 10 og 90%. Som i figur 5-3 ses, at for *tar* lig 50% opnås der med Brand og Kollmeiers procedure et bedre resultat end hvad, der er tilfældet for MML-proceduren. I øvrigt ses, at spredningen bliver mindre jo flere sætninger, der medtages, hvilket stemmer fint overens med den i Pilottest III fundne re-test variation af s_{50}. Som for re-test variationen af SRT_{50} blev der i pilottesten for både den operatør- og den patientbaserede version ikke fundet nogen statistisk signifikant forskel mellem at bestemme re-test variationen af s_{50} med 10, 20 og 30 sætninger – se evt. figur C-28, s. 159.

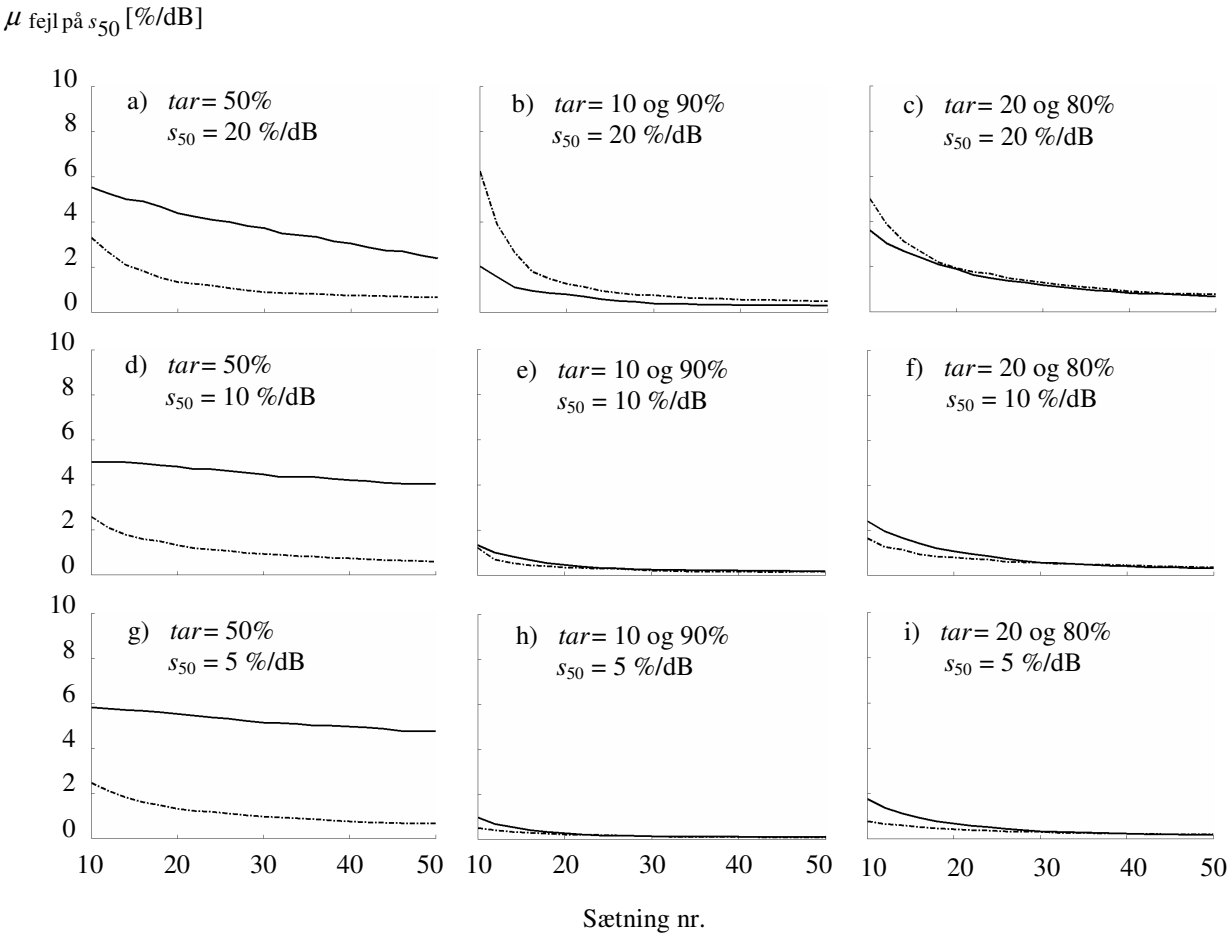


Figur 5-1: Simuleringer – gennemsnittet af fejlen på SRT_{50}. Den stiplede streg refererer til, at værdierne er opnået med Hansen og Ludvigsens procedure, mens værdierne angivet med den stiplede/prikkede streg er opnået med Brand og Kollmeiers procedure. Den fuldtoptrukne streg refererer derimod til, at værdierne er opnået med MML-proceduren. I hver simulering indgik 1.000 fiktive FP'er. Værdien af *tar* angiver ved hvilken eller hvilke sandsynligheder for korrekt svar, sætningerne blev præsenteret, mens værdien af s_{50} refererer til den værdi FP'erne fik tildelt.

---------- Hansen og Ludvigsens procedure
-·-·-·- Brand og Kollmeiers procedure
——— MML-proceduren



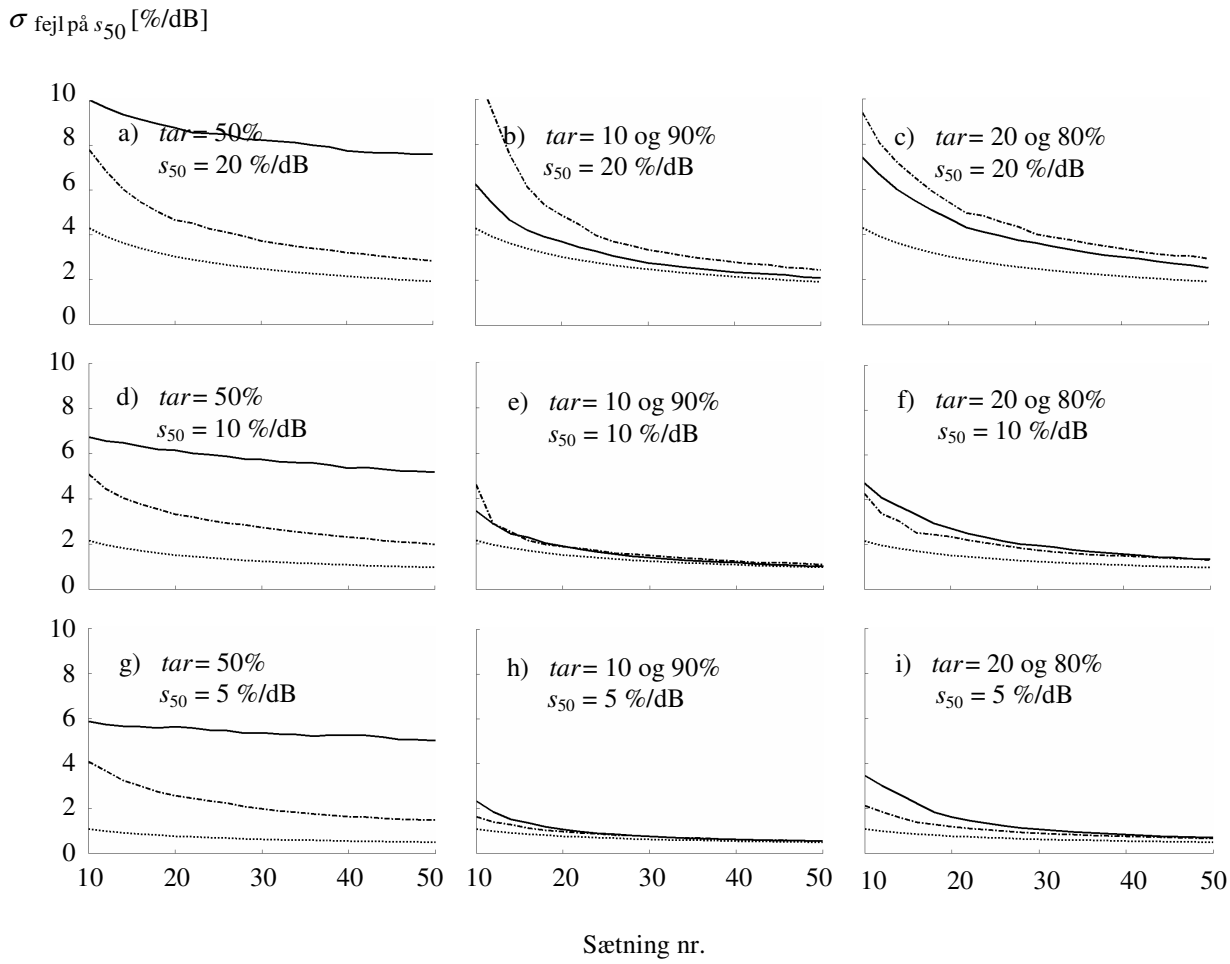
Figur 5-2: Simuleringer – spredningen af fejlen på SRT_{50}. Den stiplede streg refererer til, at værdierne er opnået med Hansen og Ludvigsens procedure, mens værdierne angivet med den stiplede/prikkede streg er opnået med Brand og Kollmeiers procedure. Den fuldtoptrukne streg refererer derimod til, at værdierne er opnået med MML-proceduren. I hver simulering indgik 1.000 fiktive FP'er. Den prikkede streg illustrerer den jf. formel (4-3) mindste og derved bedst opnåelige værdi af den pågældende spredning. Værdien af *tar* angiver ved hvilken eller hvilke sandsynligheder for korrekt svar, sætningerne blev præsenteret, mens værdien af s_{50} refererer til den værdi FP'erne fik tildelt.



Figur 5-3: Simuleringer – gennemsnittet af fejlen på s_{50}. Den stiplede/prikkede streg refererer til, at værdierne er opnået med Brand og Kollmeiers procedure, mens værdierne angivet med den fuldtoptrukne streg er opnået med MML-proceduren. I hver simulering indgik 1.000 fiktive FP'er. Værdien af *tar* angiver ved hvilken eller hvilke sandsynligheder for korrekt svar, sætningerne blev præsenteret, mens værdien af s_{50} refererer til den værdi FP'erne fik tildelt.

Brand og Kollmeiers procedure

MML-proceduren



Figur 5-4: Simuleringer – spredningen af fejlen på s_{50}. Den stiplede/prikkede streg refererer til, at værdierne er opnået med Brand og Kollmeiers procedure, mens værdierne angivet med den fuldtoptrukne streg er opnået med MML-proceduren. I hver simulering indgik 1.000 fiktive FP'er. Den prikkede streg illustrerer den jf. formel (4-8) mindste og derved bedst opnåelige værdi af den pågældende spredning. Værdien af *tar* angiver ved hvilken eller hvilke sandsynligheder for korrekt svar, sætningerne blev præsenteret, mens værdien af s_{50} refererer til den værdi FP'erne fik tildelt.

Brand og Kollmeiers procedure
MML-proceduren
Min. værdi

5.4 Indlæringseffekten

I forbindelse med testafviklingen er der observeret en indlæringseffekt. Denne effekt optræder ved, at værdien af SRT_{50} falder ved på hinanden følgende målinger, indtil et vist antal målinger er opnået. Derefter varierer værdien kun med re-test variationen. For at beskrive indlæringseffekten skelnes mellem to forskellige størrelser, der i den engelsksprogede litteratur kaldes *within-visit learning effect* og *inter-visit learning effect*. Disse størrelser, som her er valgt oversat til *indlæringen inden for en test* og *indlæringen mellem to tests*, beregnes hhv. som forskellen mellem SRT_{50} ved første og sidste måling inden for en test og som forskellen mellem SRT_{50} ved første måling i to forskellige testrunder. I de to følgende afsnit behandles *indlæringen inden for en test* og *indlæringen mellem to tests* for såvel den operatør- som den patientbaserede version.

5.4.1 Indlæringen inden for en test

Med Dantale II materialet er der af Wagener *et al*. (2003) blevet udført en større undersøgelse af *indlæringen inden for en test*. Denne undersøgelse, der blev foretaget med en operatørbaseret afvikling, forløb ved at 60 normalthørende FP'er blev præsenteret for otte på hinanden følgende målinger af SRT_{50}. Hver måling inkluderede 20 sætninger og blev udført i henhold til Brand og Kollmeiers procedure. Ud fra målingerne blev den gennemsnitlige *indlæring inden for en test* bestemt til 2,2 dB – dvs. svarende til forskellen mellem SRT_{50} bestemt ved den første og ottende måling. Samtidigt blev det fundet at SRT_{50} faldt kraftigst inden for de to første målinger – dvs. ved afvikling af de første 40 sætninger. I undersøgelsen konkluderes derfor, at hvis FP'en præsenteres for træningsforløb af 40 sætninger inden den egentlig test afvikles, vil effekten af indlæringen påvirke testresultatet med mindre en 1 dB.

I en undersøgelse foretaget med det tyske materiale, er der fundet en indlæring svarende til de 2,2 dB for det danske materiale (Wagener *et al*., 1999b). Med det svenske materiale er der imidlertid blevet udført flere undersøgelser af indlæringen, der ligeledes viser betydningen af et træningsforløb – se evt. Hagerman (1984), (1997), (2002) og Hagerman & Kinnefors (1995). Ifølge Wagener & Brand (2005) skyldes en betydelig del af den observerede indlæring, at materialer som Dantale II er opbygget af et begrænsede antal ord, hvorfor FP'en under et træningsforløb i en vis grad lærer hvilken type ord, der indgår og derved i takt med, at testen skrider frem kan foretage flere og flere *motiverede gæt*[12]. Hvis dette gør sig gældende, kan indlæring tænkes at være afhængig af, ved hvilket niveau sætningerne præsenteres – dvs. af værdien af *tar*. Er dette tilfældet vil det måske være muligt at reducere antallet af sætninger i et træningsforløb.

På trods af de mange undersøgelser af indlæringen har det ikke været muligt at opspore en undersøgelse af, om der findes en sammenhæng mellem indlæringen og præsentationsniveauet. Dette er derfor blevet undersøgt i Pilottest I ved at præsentere to testgrupper for niveauer ved hhv. 50 og 80% forståelighed. Af testen, der blev afviklet på samme måde som undersøgelsen af Wagener *et al*. (2003), blev der ikke fundet en forskel i størrelsen af indlæringen de to grupper

[12] I tilknytning hertil kan nævnes, at Wagener & Brand kalder en test baseret på et materiale som det danske for en *semi-åben* test. Det gælder selv, når materialet afvikles med en operatørbaseret afvikling – alene grundet materialets opbygning.

imellem – se evt. figur C-5, s. 119 og figur C-6, s. 120. Samtidigt blev størrelsen af indlæringen for begge grupper ikke fundet statistisk signifikant forskellig fra den bestemt af Wagener *et al*. Som for Wagener *et al*. blev det desuden for begge grupper fundet, at hvis en FP præsenteres for et træningsforløb bestående af 40 sætninger synes en efterfølgende måling ikke at være påvirket af effekten af indlæringen. På baggrund heraf vurderes indlæringen primært at bero på en generel tilvænning til testsituationen – herunder at FP'en bliver fortrolig med sætningernes nonsensindhold og bliver fokuseret på at lytte efter sætningerne i støjen. I øvrigt bør det bemærkes, at FP'erne blot præsenteres for de enkelte ord ganske få gange inden indlæringen kan antages at være overstået.

Med en testafvikling som den patientbaserede version med svarmuligheden *ved ikke* har det som skrevet i afsnit 4.2.1 kun være muligt at finde en enkelt undersøgelse. Denne undersøgelse, der blev foretaget af Brand *et al*. (2004)* med det tyske materiale, indbefattede også *indlæringen inden for en test*. Af undersøgelsen blev det konkluderet, at indlæringen bestemt med den patientbaserede afvikling er i samme størrelsesorden som indlæringen bestemt med en operatørbaseret afvikling. For at undersøge dette nærmere indgik der i Pilottest I også en gruppe, som blev præsenteret for den patientbaserede version dog uden muligheden for at svare *ved ikke*. Indlæringen blev her ikke fundet statistisk signifikant forskellig fra den bestemt med hver af de to grupper, der blev præsenteret for den operatørbaserede version. Derved er konklusionen den samme, som den fremført af Brand *et al*.

På baggrund af ovenstående betragtninger for såvel den operatør- som den patientbaserede version er FP'erne i Pilottest III og IV blevet præsenteret for et træningsforløb bestående af 30 sætninger. Dette antal er valgt som et kompromis mellem en minimal påvirkning af indlæringen på det endelige testresultat og en test af en rimelig varighed. At anvende 30 sætninger som træning anbefales i øvrigt også af Hansen & Ludvigsen (2001). Hertil bør nævnes, at jf. Pilottest I tager en afvikling af 30 sætninger i den operatør- og den patientbaserede version hhv. omkring 5 og 10 min. – se evt. figur C-11, s. 127.

5.4.2 Indlæringen mellem to tests

Med en operatørbaserede afvikling med Dantale II materialet er der af Hernvig & Olsen (2005) blevet foretaget en undersøgelse af både *indlæringen inden for en test* og *mellem to tests*. I undersøgelsen indgik 20 hørehæmmede FP'er, som hver blev præsenteret for to testrunder, der var adskilt af en periode på 14 til 43 dage – i snit 27 dage. I hver testrunde blev FP'erne præsenteret for seks på hinanden følgende målinger af SRT_{50}. Hver af disse, hvorunder FP'erne brugte deres eget personlige HA, inkluderede 30 sætninger og blev udført i henhold til Hansen og Ludvigsens procedure. Ud fra målingerne blev den gennemsnitlige indlæring inden for første og anden testrunde bestemt til hhv. 3,2 og 2,2 dB – dvs. svarende til forskellen mellem SRT_{50} bestemt ved den første og sjette måling. Samtidigt blev *indlæringen mellem to tests* bestemt til 1,6 dB – dvs. svarende til forskellen mellem SRT_{50} bestemt ved første måling i hver af de to testrunder.

Sammenlignes indlæringen inden for den første testrunde – dvs. svarende til de 3,2 dB – med den indlæring, der blev bestemt af Wagener *et al*. (2003) fås en forskel på 1 dB. Hernvig & Olsen antyder, at denne forskel måske skyldes, at Wagener *et al*.'s FP'er har været bekendt med

testproceduren – herunder med materialet. At dette skulle være tilfældet, fremgår dog ikke af Wagener *et al*. Imidlertid bør resultaterne fra de to undersøgelser kun sammenlignes varsomt. For det første er målingerne af SRT_{50} foregået på forskellig vis, hvorved indlæringen ikke er bestemt på baggrund af det samme antal sætninger. For det andet indgik der i Wagener *et al*.'s undersøgelse normalthørende FP'er, mens der i Hernvig & Olsens undersøgelse indgik høre-hæmmede – herunder var FP'ernes alder i de to undersøgelser meget forskellig. Disse aspekter vurderes at kunne indvirke på, at der i de to undersøgelser er opnået et forskelligt resultat. Derudover må det tages i betragtning, at forskellen kun er 1 dB og derved måske ligger inden for den variation, der kan forekomme ved tests med et begrænset antal FP'er. Den væsentlige konklusion er dog, at begge undersøgelser pointerer behovet for et træningsforløb. I øvrigt bør nævnes, at der for begge undersøgelser ikke er opgivet nogen spredning af indlæringen. Selvom det derved kan virke fristede, kan der altså ikke blot tages højde for indlæringen ved fast at trække et bestemt antal dB fra det endelige testresultat.

I Pilottest I, hvori der indgik tre forskellige testgrupper, blev *indlæringen mellem to tests* også bestemt. I denne test var de to testrunder adskilt af en periode på 12 til 16 dage – i snit 14 dage. Denne periode er mindre og mere snæver end den i Hernvig & Olsen's undersøgelse. Det bør i øvrigt nævnes, at FP'erne i testen i anden testrunde kun blev præsenteret for en enkelt måling med 20 sætninger, der ligesom målingerne i første testrunde blev udført i henhold til Brand og Kollmeiers procedure. At FP'erne ikke er blevet præsenteret for flere målinger, skyldes alene en afvejning af hvilke faktorer, der ønskes undersøgt i pilottestene. For værdierne af den fundne *indlæring mellem to tests* blev der ikke fundet nogen statistisk signifikant forskel de tre testgrupper imellem – se evt. igen figur C-5, s. 119 og figur C-6, s. 120.

På nær testgruppen som blev præsenteret for den operatørbaserede version med niveauer ved 50% forståelighed, blev værdierne af *indlæring mellem to tests* ikke fundet statistisk signifikant forskellig fra den tilsvarende værdi bestemt af Hernvig & Olsen. I forbindelse hermed kan nævnes, at for hver af de tre grupper blev den fundne *indlæring inden for en test* sammenlignet med de 3,2 dB bestemt af Hernvig & Olsen i første testrunde. Her blev det igen fundet, at på nær testgruppen som blev præsenteret for den operatørbaserede version med niveauer ved 50%, blev værdierne af denne indlæring ikke fundet statistisk signifikant forskellig fra værdien 3,2 dB. Af pilottesten blev det konkluderet af selv, hvis en FP 14 dage forinden er blevet godt trænet, synes et træningsforløb ikke er at kunne undværes.

5.5 Sammenfatning

I forbindelse med en adaptiv testafvikling er der flere parametre, som skal fastsættes. Mht. præsentationsformen kan det være en god idé at variere niveauet at talesignalet, mens støjniveauet holdes konstant, idet der derved sikres en veldefineret baggrundsstøj. Vælges et kontinuert støjsignal opnås en mere realistisk lyttesituation og derved en mere valid test end hvad der er tilfældet, hvis støjen er afbrudt mellem sætningspræsentationerne. Det signal, hvis niveau holdes konstant, skal have et for FP'en behageligt hørbart niveau. Samtidigt bør præsentationsniveauet af den første sætning vælges således, at den enkelte FP som min. kan høre et af de afpillede ord. Ændringen af præsentationsniveauet ud fra FP'ens respons kan foregå ved at anvende en af de tre omtalte afviklingsprocedurer. Hansen og Ludvigsens procedure, som er designet til en ope-

ratørbaseret afvikling, er forholdsvis simpel, hvorfor en test baseret på denne procedure derved kan afvikles manuelt. I modsætning til denne procedure kan Brand og Kollmeiers procedure og MML-proceduren afvikles således, at der både kan bestemmes en værdi af SRT_{50} og s_{50}. Ønskes begge værdier bliver re-test variationen af SRT_{50} imidlertid større end, hvis kun SRT_{50} ønskes bestemt. Inden en egentlig test afvikles bør FP'en præsenteres for et træningsforløb. Dette gælder uafhængigt af responsmetoden og om FP'en tidligere er blevet præsenteret for materialet.

6 Alternativer til Dantale II-støjen

I stedet for at anvende det støjsignal, der er genereret i forbindelse med udviklingen af Dantale II materialet, kan det overvejes at anvende et andet signal. Dantale II-støjen kan eksempelvis udskiftes med et støjsignal hidrørende fra en af de tidligere danske tests – det kunne være sig støjen fra Dantale I eller den såkaldte Gunnar Nu-støj fra GS-listerne. For at øge sammenligneligheden med tests på andre sprog kan der alternativt anvendes et standardiseret støjsignal. Et meget oplagt valgt vil hertil være en af de ni ICRA-støje – disse vil blive behandlet nedenfor. Imidlertid minder disse støjsignaler ligesom Dantale II-støjen ved lytning ikke meget om virkelig tale, hvorfor der i forbindelse med dette projekt er genereret et forslag til et nyt støjsignal. Dette signal, der tilsvarende Dantale II-støjen er genereret ud fra materialets testsætninger, vil ligeledes blive behandlet. Som en bibemærkning kan det ganske kort nævnes, at EHIMA (The European Hearing Instrument Manufacturers Association) som en del af projektet ISMADHA (International Standards for Measuring Advanced Digital Hearing Aids) i skrivende stund arbejder på at udvikle et støjsignal bestående af talelyde fra otte forskellige sprog – heriblandt engelsk, spansk, arabisk og kinetisk (Gert Ravn – DELTA, pers. medd. marts 2007). Da signalet endnu ikke er offentligt lanceret vil det imidlertid ikke yderligere blive berørt i denne rapport.

6.1 ICRA-støj

ICRA-støj er en samling af ni forskellige støjsignaler. Disse støjsignaler, som offentligt blev lanceret omkring år 2000, findes på den såkaldte ICRA-cd. Da denne cd er til fri kopiering, er signalerne at finde på den til rapporten vedlagte cd-rom. Signalerne er designet således, at de kan anvendes til tekniske målinger af digitale HA'er og som mulige baggrundsstøje i forbindelse med tale i støj-tests. Formålet med udviklingen af signalerne var at etablere internationale standardiserede støjsignaler for at øge sammenligneligheden af tests på tværs af forskellige undersøgelser og herunder sprog. Hvert af de ni støjsignaler er kendetegnet ved at have et langtidsspektrum, der er sammenligneligt med det for almen tale og en modulation svarende til en eller flere samtidige talere.

Selvom intet af disse ni signaler er blevet anvendt som baggrundsstøj i forbindelse med de i dette projekt udførte pilottests, er de et meget oplagt valgt. I det følgende fremgår derfor en kort beskrivelse baseret på Dreschler *et al*. (2001) af signalernes karakteristika og måden, hvorpå de er genereret. For at illustrere signalernes anvendelighed i forbindelse med taletests refereres derefter en tidligere undersøgelse foretaget med det tyske materiale. Denne undersøgelse resulterede i to nye støjsignaler, som af Wagener *et al*. (2006) er konstrueret ud fra den såkaldte *icra5*. Disse modificerede signaler, som også behandles, findes ligeledes som de oprindelige støjsignaler på den vedlagte cd-rom.

6.1.1 Karakteristika

Hver af de ni ICRA-støje repræsenterer enten en mandlig taler, en kvindelig taler eller kombinationer deraf. Signalerne indeholder derved forskellige modulationsgrader og har tilmed fået tildelt forskellige stemmestyrker – normal, lettere hævet og kraftig stemmestyrke. Genereringen er foretaget således, at det ikke er muligt at skelne enkelte ord. Da signalerne derved ikke indeholder forståelige segmenter, kan de med fordel anvendes på tværs af forskellige sprog. Oprindeligt fik hvert støjsignal tildelt en betegnelse refererende til det pågældende signals karakteristika. I stedet for disse betegnelser anvendes imidlertid ofte en nummeret betegnelse, hvor de enkelte numre henviser til signalernes placering på ICRA-cd'en – se eksempelvis van Toor & Verschuure (2002), Wagener & Brand (2005) og Wagener *et al*. (2006). Frem for de oprindelige anvendes de nummererede betegnelser derfor i det følgende.

Af tabel 6-1 fremgår en oversigt over, hvad der adskiller de ni ICRA-støje. Af tabellen ses, at *Icra1-3* svarer til umoduleret, kontinuert støj repræsenteret med en mandsstemme ved forskellige stemmestyrker. Hensigten med disse tre signaler er, at de skal simulere støjen fra mange samtidige talere. *Icra4* og *5* repræsenterer en enkelt taler – hhv. en kvinde- og en mandsstemme – med normal stemmestyrke. *Icra6* repræsenterer derimod to samtidige talere – en mands- og en kvindestemme – med samme stemmestyrke. *Icra7-9* simulerer med forskellige stemmestyrker seks samtidige talere (eng.: Babble noise) – tre mands- og tre kvindestemmer. For hvert af de tre signaler er niveauet af to af mands- og to af kvindestemmerne sænket 6 dB i forhold til niveauet af de to øvrige talere. Dette er gjort for at simulere, at de seks talere befinder sig ved to forskellige afstande fra lytteren.

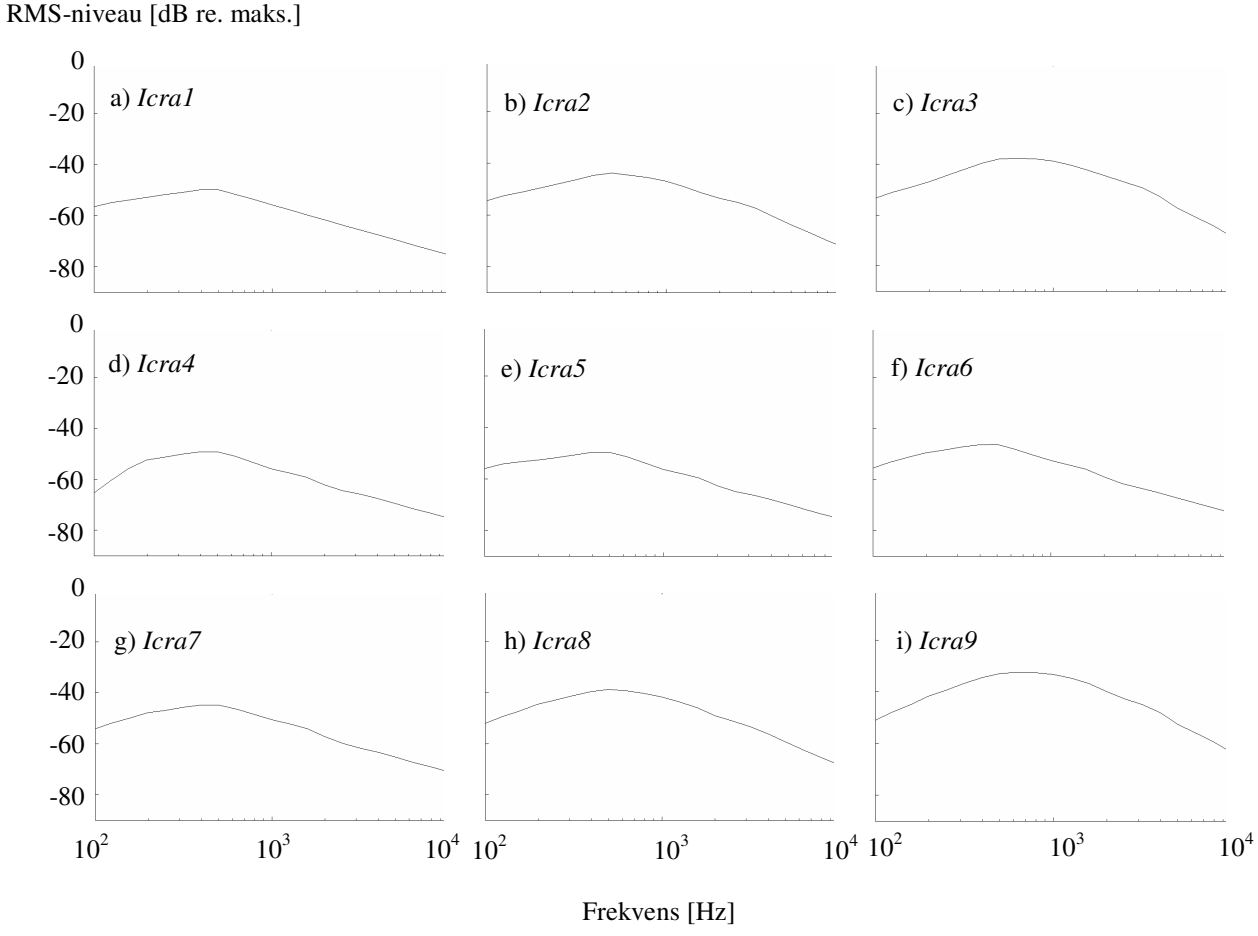
Tabel 6-1: Skematisk oversigt over kendetegn ved hver af de ni ICRA-støje. Betegnelserne er valgt ud fra signalernes placering på ICRA-cd'en.

(Forsimplet gengivelse af Dreschler *et al*., 2001 – Tabel 1).

Betegnelse	**Modulationsgrad**	**Køn**	**Stemmestyrke**
Icra1	Umoduleret	Mand	Normal
Icra2	Umoduleret	Mand	Lettere hævet
Icra3	Umoduleret	Mand	Kraftig
Icra4	En taler	Kvinde	Normal
Icra5	En taler	Mand	Normal
Icra6	To talere	Blandet	Normal
Icra7	Seks talere	Blandet	Normal
Icra8	Seks talere	Blandet	Lettere hævet
icra9	Seks talere	Blandet	Kraftig

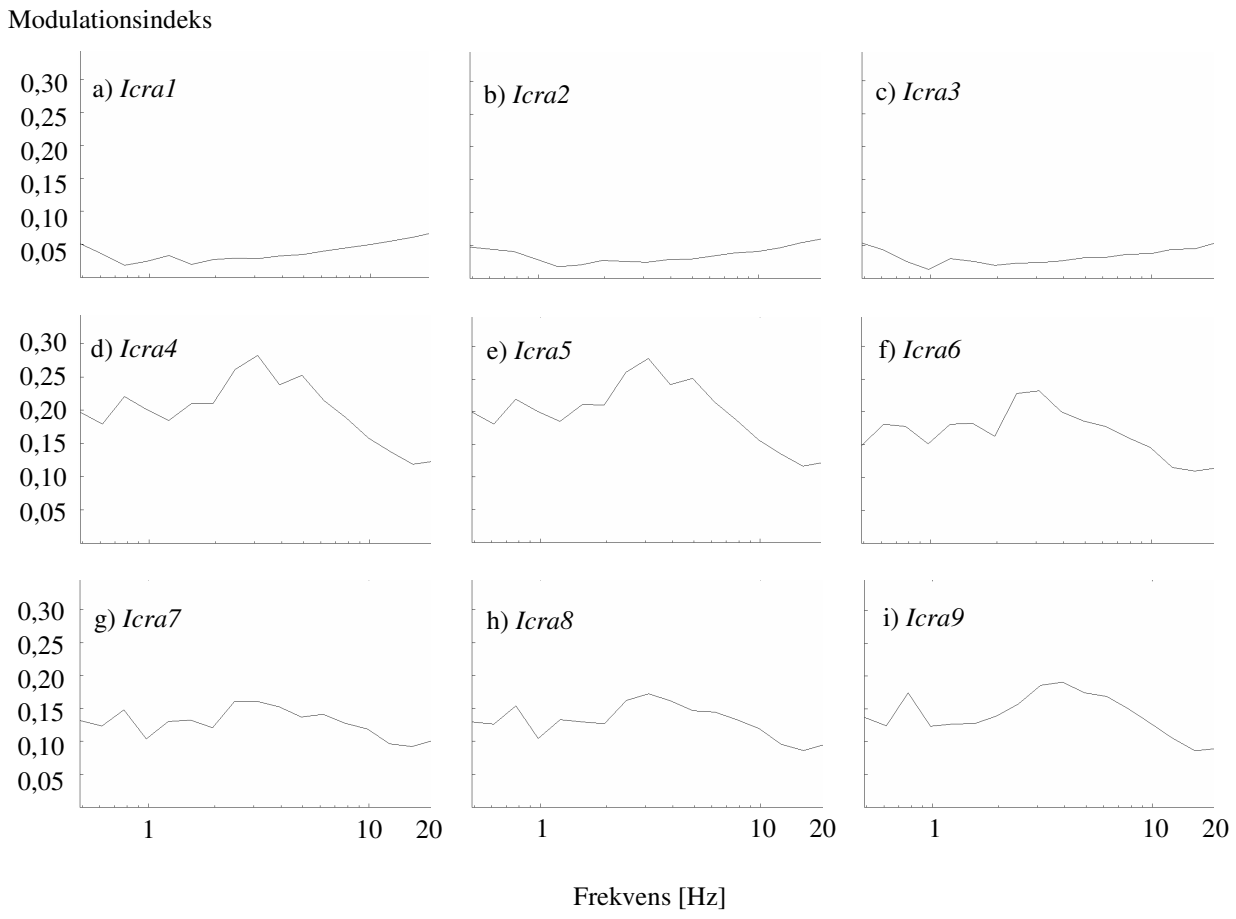
Generelt bør det bemærkes, at der ikke er taget højde for, at frekvensindholdet af flere samtidige talere i et rum er afhængigt af, i hvilke afstande talerne befinder sig fra rummets begrænsningsflader. Denne afhængighed skyldes, at lavfrekvent lyd i højere grad reflekteres fra begrænsningsfladerne, hvorved disse lyde ikke dæmpes i samme grad som højfrekvent lyd. Hvis tanken er, at Dantale II-støjen skal simulere flere samtidige talere, gør den samme problematik sig imidlertid også gældende her. Et ændret frekvensspektrum i forhold til testsætningernes kan dog have betydning for den maskerende effekt.

Af figur 6-1 fremgår frekvensspektret for hver af de ni ICRA-støje – dvs. langtidsspektrene. Spektrene er beregnet med 1/3 oktavbånds filtre på nøjagtig samme måde som det for Dantale II-støj – se evt. afsnit 3.2.2. Som for spektret for Dantale II-støjen er frekvensen angivet fra 100 Hz til 10 kHz, mens RMS-niveauet er opgivet relativt til det maksimale niveau i computeren. Af figuren ses, at spektrene ændres en smule ved de forskellige stemmestyrker og at spektret for kvindestemmen svarende til figur 6-1 d) indeholder mindre lavfrekvent energi end hvad, der er tilfældet for det tilsvarende spektrum for en mandsstemme svarende til figur 6-1 e). Samtidigt ses, at spektrene generelt er stort set ens de ni støjsignaler imellem. Tilsvarende spektre opgivet i 1/1 oktavbånd fremgår i øvrigt af Dreschler *et al*. (2001 – Figur 2).



Figur 6-1: Langtidsspekteret for hver af de ni ICRA-støje – samtlige spektre er beregnet med 1/3 oktavbånd. RMS-niveauet er opgivet relativt til det maksimale niveau i computeren.

For nemmere at differentiere mellem signalerne er modulationsindekset for hver af disse beregnet med opløsning på 1/3 oktav i frekvensområdet fra 0,5 til 20 Hz. De beregnede indekser fremgår af figur 6-2 og er bestemt på tilsvarende vis som det for Dantale II-støjen – se igen evt. afsnit 3.2.2. Som det ses af figuren falder modulationen, jo flere samtidige talere støjsignalet simulerer, hvilket var forventeligt. De tre umodulerede støjsignaler, der fremgår af figur 6-2 a) til c), er principielt sammensat af uendelig mange samtidige talere, hvorved modulationen stort set er lig nul. I Dreschler *et al*. (2001 – Figur 3) fremgår tilsvarende modulationsindekser – disse er angivet fra 0,5 til 32 Hz. Til sammenligning kan nævnes, at beregningen af Dreschler *et al*. er foretaget med en opløsning på 1/1 oktav, hvorved skaleringen af ordinaten er en anden og detaljeringsgraden er mindre end den i figur 6-2.



Figur 6-2: Modulationsindekset for hver af de ni ICRA-støje med en opløsning på 1/3 oktav i frekvensområdet fra 0,5 til 20 Hz.

Som for Dantale II-støjen er der for modulationen af hver af de ni ICRA-støje desuden beregnet et samlet modulationsindeks for den lavfrekvente modulation – dvs. op til 50 Hz. Disse fremgår af tabel 6-2, hvoraf det ligeledes kan sluttes, at modulationen falder, desto flere samtidige talere støjsignalet simulerer.

Tabel 6-2: For hver af de ni ICRA-støje fremgår det samlede modulationsindeks for den lavfrekvente modulation – dvs. op til 50 Hz.

Icra1	*Icra2*	*Icra3*	*Icra4*	*Icra5*	*Icra6*	*Icra7*	*Icra8*	*Icra9*
0,26	0,22	0,23	0,91	0,91	0,77	0,59	0,59	0,63

Som en ekstra bemærkning kan nævnes, af der i skrivende stund er ved at blive udviklet en ny samling af lydsignaler kaldet ICRA 2 (Søren Westermann – Widex, pers. medd. feb. 2007). Disse signaler, som skal simulere naturligt forekommende lyde, er imidlertid ikke tiltænkt som mulige baggrundsstøje i forbindelse med tale i støjtests, men derimod til evaluering af lydkvalitet og måling af HA'ers opførsel i forskellige lyttesituationer. På nuværende tidspunkt er der optaget i alt 47 forskellige lyde. Disse er ikke endeligt kalibrerede, hvorfor de ikke endnu er klar til brug.

6.1.2 Generering

Ud fra et indspillet talesignal er støjsignalerne genereret vha. en digital signalbehandling, hvis fremgangsmåde beskrives i det følgende. Beskrivelsen er her valgt angivet som en punktopstilling, hvor de enkelte punkter henviser til numrene i figur 6-3. Forløbet af genereringen kan opdeles i to dele. Først modificeres det indspillede talesignal, hvorved dets frekvensspektrum minder om det for hvidstøj, mens den oprindelige modulation bibeholdes (punkt 1 til 6). Dernæst justeres frekvensspektret, hvorved det hhv. modsvarer det for en almen mands- og kvindestemme ved forskellige stemmestyrker (punkt 7 til 10). Ved at kombinere disse genererede signaler på passede vis, er de ni støjsignaler, der er listet i tabel 6-1, frembragt. Følgende er baseret på Dreschler *et al*. (2001), hvorfor der for en mere specifik beskrivelse henvises hertil.

1) Støjsignalerne blev genereret ud fra en indspilning af en engelsk tekst. Teksten omhandlede tallære og blev oplæst af en kvinde.

2) Det optagede talesignal blev digitaliseret med en samplingsfrekvens på 44,1 kHz, hvorefter signalet blev filtreret. Ved denne filtrering blev signalet opdelt i tre frekvensbånd. Opdelingen i netop tre bånd blev valgt for, at det resulterende signal skulle minde om tale uden, at de enkelte ord kan skelnes.

3) Signalerne i hvert bånd blev modificeret vha. en såkaldt Schroders proces. Ved denne proces blev hver sample tildelt en sandsynlighed på 50% for at vende fortegnet. Derved blev fortegnet vendt på ca. halvdelen af alle samples, mens fortegnet på de resterende

blev bibeholdt uforandret. Ved at anvende denne fremgangsmåde opnåede hvert af de tre modificerede signaler den samme modulationsgrad som den for det oprindelige indspillede talesignal. Signalerne var imidlertid uforståelige og havde et fladt (hvidt) frekvensspektrum.

4) Signalerne blev derefter filtreret endnu en gang. Dette blev gjort således, at de resulterende signaler kun indeholdte frekvenskomposanter i det samme område, som de tre umodificerede signaler – dvs. som signalerne i punkt 2.

5) Amplituden af hvert af de tre filtrerede signaler blev justeret, hvorved RMS-niveauet af hvert signal blev det samme som før filtreringen foretaget i punkt 4.

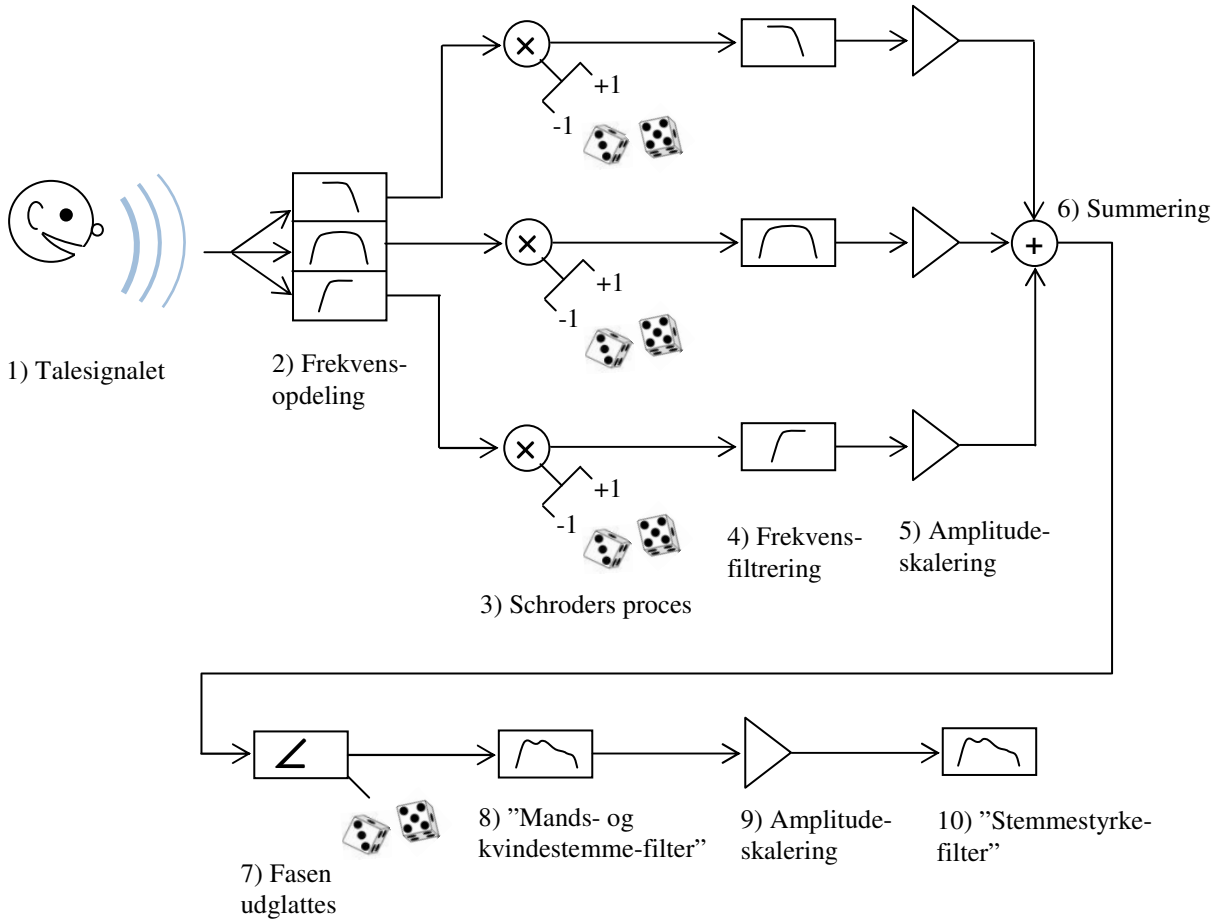
6) De tre bånd blev adderet.

7) Da det resulterede signal havde en ubehagelig skrattende lyd blev fasen udglattet ved først at konvertere signalet fra tids- til frekvensdomænet, hvor selve udglatningen blev udført ved at randomisere fasen. Derefter blev signalet konverteret tilbage til tidsdomænet, hvor signalets enkelte segmenter blev adderet med et overlap.

8) Det samlede signal blev derefter filtreret på to forskellige måder, hvorved der blev opnået to forskellige signaler, hvis frekvensspektre hhv. modsvarer det for en mands- og kvindestemme. De heraf opnåede spektre er i overensstemmelse med LTASS, som er middel langtidsspektret for gennemsnitlig tale på tværs af forskellige sprog

9) Amplituden af de to støjsignaler blev justeret, hvorved på forhånd bestemte RMS-niveauer blev opnået.

10) Slutteligt blev signalerne filtreret således, at deres spektrum modsvarer det for normal, lettere hævet og kraftig stemmestyrke.



Figur 6-3: Figuren illustrerer, hvordan de ni ICRA-støje er genereret. For en beskrivelse af de enkelte angivne punkter henvises til hovedteksten.

(Modificeret gengivelse af Dreschler *et al*., 2001 – Figur 1).

6.1.3 Anvendelse i forbindelse med taletests

Med det tyske materiale er det af Wagener & Brand (2005) blevet undersøgt, om det støjsignal, der oprindeligt er genereret til dette materiale – den såkaldte *olnoise*, kan erstattes med *icra1*. Støjsignalet *olnoise*, der ligesom *icra1* er umoduleret, er genereret på tilsvarende vis som Dantale II-støjen ved at sammenkopiere de tyske testsætninger. Undersøgelsen, hvori der indgik ti normalthørende og ti hørehæmmede FP'er, viste for begge grupper af FP'er ingen statistisk signifikant forskel i bestemmelsen af hverken SRT_{50} eller s_{50} ved anvendelse af de to forskellige støjsignaler. Da langtidsspekteret for forskellige sprog stort set er ens – se evt. afsnit 3.2.2, konkluderer Wagener & Brand, at standardiseret støj som *icra1* vil kunne anvendes på tværs af tests i forskellige lande. Derved indikeres, at det ikke er nødvendigt at generere et specifikt støjsignal til et givet talemateriale.

I samme undersøgelse indgik også tests med to modulerede støjsignaler *icra5* og *icra7* – disse repræsenterer hhv. en mandlig taler og seks samtidige talere. Undersøgelsen viste, at de normalhørende opnåede statistisk signifikant lavere værdier af både SRT_{50} og s_{50} med de to modulerede end med de to umodulerede støjsignaler. De laveste værdier blev imidlertid opnået med *icra5*. Forskellen mellem SRT_{50} opnået med de umodulerede signaler og *icra5* var på omkring 15 dB. Dette kan skyldes, at den pågældende støj indeholder pauser – såkaldte stilleintervaller – på helt på til 2 sek. Disse intervaller kan falde sammen med store dele af talesignalet, hvorved FP'en lettere kan afgive et korrekt svar end, hvis stilleintervallerne ikke havde været til stede.

Testene viste imidlertid også, at det ikke var alle hørehæmmede, der kunne drage nytte af disse intervaller, hvorved testen kan betragtes som værende sensitiv. Hertil kan nævnes, at nogle af de hørehæmmede opnåede en værdi af SRT_{50} med de modulerede signaler, der var op til 10 dB bedre end værdien opnået med de umodulerede signaler, mens andre imidlertid opnåede værdier, der var sammenlignelige. På baggrund heraf blev det af Wagener & Brand konkluderet, at en kraftig moduleret støj som *icra5* er ideel til at differentiere mellem personer med forskellige grader af hørenedsættelser. Da ønsket med en taletest som oftest er at simulere en lyttesituation, hvor der er flere samtidige talere, synes moduleret støj desuden at bidrage til at gøre testen mere valid. Som bibemærkning kan nævnes, at moduleret støj tillige er bedre til at teste, hvorvidt støjreduktionen og kompressionen i digitale HA'er er designet hensigtsmæssigt – for en undersøgelse af sidstnævnte henvises til van Toor & Verschuure (2002).

Ved anvendelse af moduleret frem for umoduleret støj bør der rettes opmærksomhed på, at re-test variationen kan blive påvirket i en sådan grad, at testen ikke længere kan betragtes som værende reliabel. Selvom re-test variationen i Wagener & Brand's undersøgelse blev målt større for de modulerede end for de umodulerede signaler, viste dette sig imidlertid ikke at være tilfældet. Det bør hertil huskes, at en test ikke kan være sensitiv uden, at den også er reliabel. At tests med moduleret frem for umoduleret støj er medvirkende til bedre at skelne forskellige grader af hørenedsættelser, konkluderes bl.a. også af Gustafsson & Arlinger (1994), Hagerman (2002) og Summers & Molis (2004). For ikke at påvirke reliabiliteten unødigt er der af Hagerman (2002) blevet anvendt et moduleret støjsignal, som var tilpasset hvert enkelt ord således, at disse blev maskeret ens, hver gang de blev præsenteret.

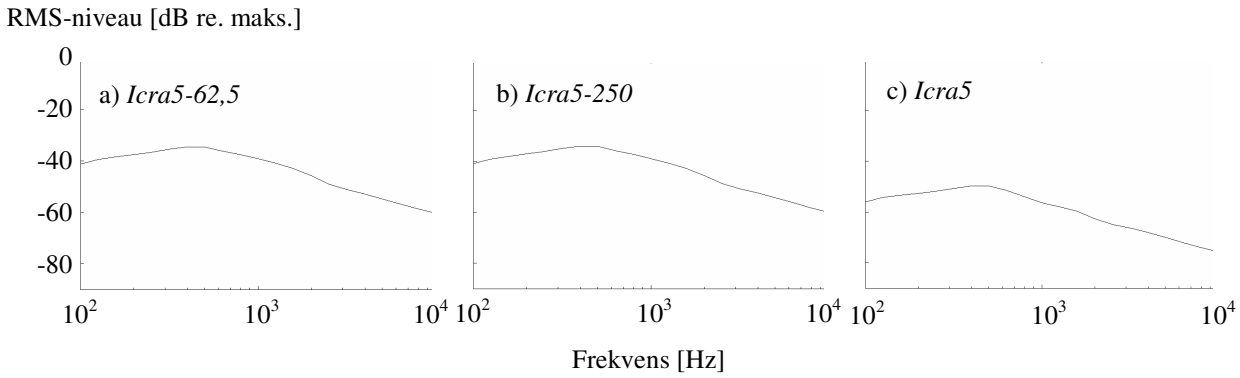
6.1.4 Modificering af *icra5*

I forrige afsnit blev det refereret, at Wagener og Brand (2005) ud fra en undersøgelse med både normalthørende og hørehæmmede slutter, at anvendelse af en kraftigt moduleret støj som *icra5* resulterer i en mere sensitiv test end hvad, der er tilfældet, hvis der anvendes en støj, som kun er svagt moduleret. På trods heraf er de lange stilleintervaller i *icra5* jf. Wagener og Brand ikke hensigtsmæssige. Derfor er der af Wagener *et al*. (2006) blevet konstrueret to støjsignaler på basis af *icra5* kaldet *icra5-62,5* og *icra5-250*, hvor varigheden af stilleintervallerne hhv. er reduceret til maks. 62,5 og 250 ms.

Af Wagener *et al*. blev der udført en sammenlignende undersøgelse af disse modificerede signaler og den oprindelige *icra5*. Af undersøgelsen, hvori der indgik ti hørehæmmede FP'er, blev det konkluderet, at SRT_{50} bestemt med *icra5* primært er afhængig af SRT_{50} bestemt uden en kontrolleret baggrundsstøj grundet de lange stilleintervaller. Værdien af SRT_{50} bestemt med

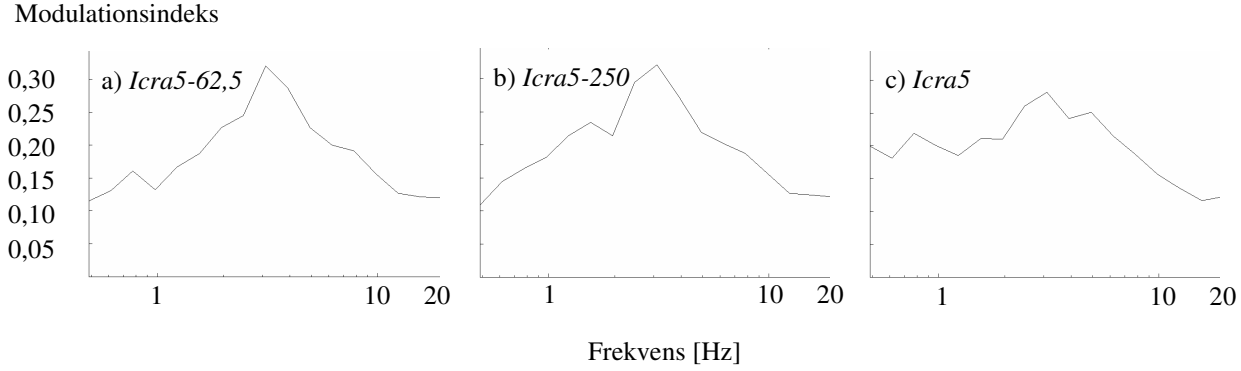
de to modificerede signaler er derimod både afhængig af SRT_{50} bestemt uden støj og SRT_{50} bestemt med en umoduleret støj som eksempelvis *icra1* – førstnævnte gælder selvom stilleintervallerne i et *icra5-62,5* kun var maks. 62,5 ms. Ønskes et mål for en FP's evne til at opfatte tale i støj, anbefales det derfor af Wagener *et al.* at anvende et af de to modificerede støjsignaler. Som skrevet i afsnit 3.1.3 er de tyske testsætninger indtalt af en mand, mens der for det danske er anvendt en kvindelig taler. I stedet for at tage udgangspunkt i *icra5* bør derfor i forbindelse med tests med det danske materiale eksempelvis anvendes *icra4*. Dette gælder, idet frekvensspektret for en mands- og kvindestemme ikke er helt ens, hvorved dette vil kunne influere på maskeringen af testsætningerne.

Efter kontakt til Kirsten Wagener er de to modificerede signaler via hjemmesiden for Carl von Ossietsky Universität i Oldenburg blevet gjort offentligt tilgængelige. Som for de oprindelige ICRA-støje er der for hvert af de to modificerede signaler foretaget en frekvens- og modulationsanalyse. Af figur 6-4 fremgår frekvensspektret – dvs. langtidsspektret – for både *icra5-62,5* og *icra5-250*. For let at kunne sammenligne disse med den oprindelige *icra5* er spektret for denne også medtaget. Som det ses, påvirkes selve formen af spektrene ikke, når varigheden af stilleintervallerne reduceres. Til gengæld bliver RMS-niveauet højere, hvilket var forventeligt. Ved bl.a. en sammenlignende test med disse støjsignaler bør der tages højde herfor.



Figur 6-4: Langtidsspekteret for de to modificerede signaler – hhv. *icra5-62,5* og *icra5-250* – og for den oprindelige *icra5*. De tre spektre er beregnet med 1/3 oktavbånd og RMS-niveauet er opgivet relativt til det maksimale niveau i computeren. Bemærk; delfigur c) er identisk med figur 6-1 e).

Modulationsindekset for hvert af de to modificerede signaler og den oprindelige *icra5* fremgår af figur 6-5. Indeksene er i overensstemmelse med tilsvarende indekser angivet i Wagener *et al.* (2006 – Figur 1), som dog er opgivet i 1/1 oktavbånd. Som det ses af figuren, reduceres den lavfrekvente modulation, når varigheden af stilleintervallerne reduceres.



Figur 6-5: Modulationsindekset for hvert de to modificerede signaler – hhv. *icra5-62,5* og *icra5-250* – og for den oprindelige *icra5*. Som for de øvrige behandlede støjsignaler er modulationsindekserne opgivet med en opløsning på 1/3 oktav i frekvensområdet fra 0,5 til 20 Hz. Bemærk; delfigur c) er identisk med figur 6-2 e).

Af tabel 6-3 fremgår for hvert af de to modificerede signaler og den oprindelige *icra5* det samlet modulationsindeks for den lavfrekvente modulation – dvs. op til 50 Hz. Som ovenfor kan det heraf sluttes, at når varigheden af stilleintervallerne reduceres, mindskes modulationsindekset.

Tabel 6-3: For de to modificerede signaler – hhv. *icra5-62,5* og *icra5-250* – og for den oprindelige *icra5* fremgår det samlede modulationsindeks for den lavfrekvente modulation – dvs. op til 50 Hz.

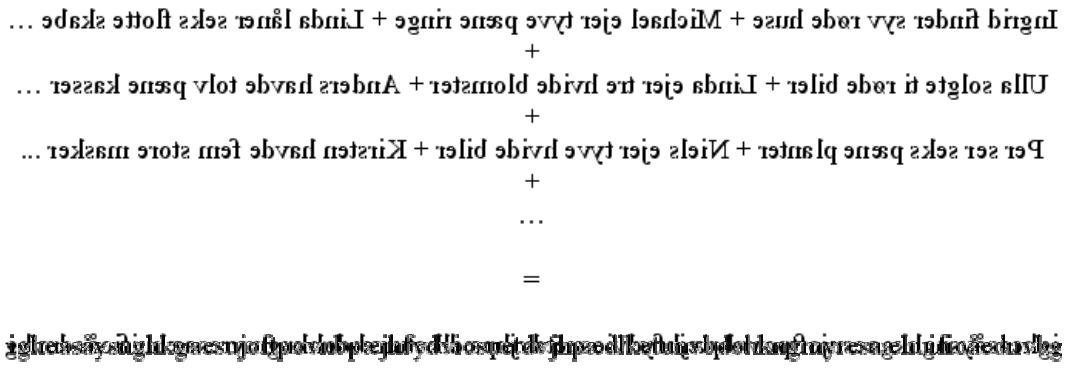
Icra5-62,5	*Icra5-250*	*Icra5*
0,84	0,88	0,91

6.2 Forslag til nyt støjsignal

Selvom ICRA-støjene har et frekvensspektrum nær det for almen tale og en modulation svarende til en eller flere samtidige talere, lyder signalerne ligesom Dantale II-støjen ikke som egentlig tale. Der er derfor i forbindelse med dette projekt genereret et forslag til et nyt støjsignal. Dette signal vurderes at kunne bidrage til en test, der er bedre til at simulere en lyttesituation ved et snakkesaligt selskab end hvad, der er tilfældet for eksempelvis Dantale II-støjen – dvs. testen bliver mere valid. Signalet findes som de øvrige behandlede støjsignaler på den vedlagte cd-rom. I det følgende fremgår en beskrivelse af, hvordan signalet er genereret samt en sammenligning med Dantale II-støjen.

6.2.1 Generering

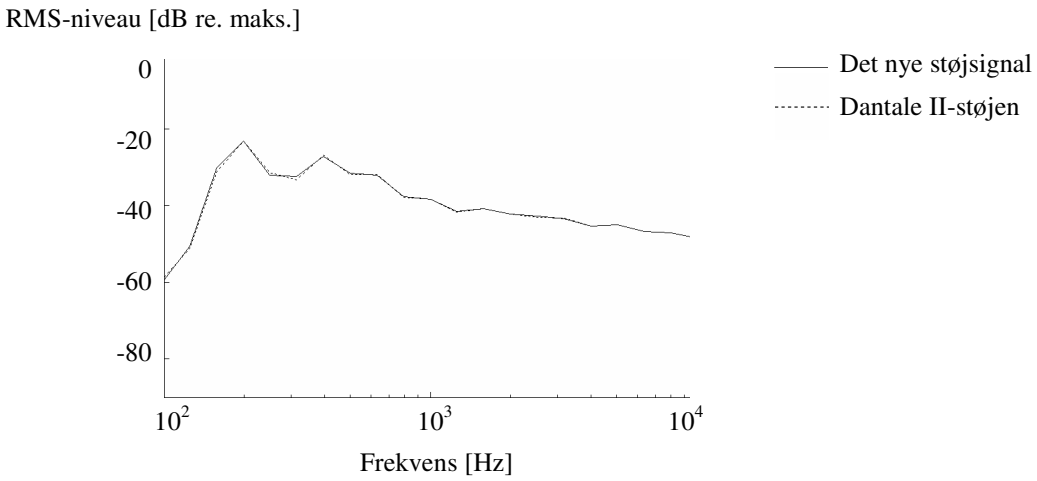
Ved genereringen af det nye støjsignal er hver enkel af materialets i alt 160 testsætninger blevet afspillet baglæns – inspirationen hertil er hentet i Hygge *et al*. (1992). De *bagvendte* sætninger blev derpå klippet sammen uden pause ti og ti – svarende til hele lister. Idet materialet består af 16 lister, opstod der derved 16 sekvenser af ca. 22 sek.'s varighed. Disse blev efterfølgende afspillet oveni hinanden således, at der opstod én sammenhængende sekvens – dette er forsøgt illustreret i figur 6-6. Da sekvensen var forholdsvis kort, blev den gentaget indtil dens varighed var 150 sek. – dvs. 2,5 min. – svarende til længden af Dantale II-støjen. For at sikre et ens frekvensindholdet for det nye støjsignal og Dantale II-støjen blev frekvensspektret for begge støjsignaler bestemt. Derefter blev det nye støjsignal filteret således, at dets frekvensspektrum svarede til det for Dantale II-støjen, hvorved RMS-niveauet af de to støjsignaler er ens. Det bør hertil huskes, at Dantale II-støjen afviger en anelse fra middel langtidsspektret for testsætningerne – se evt. afsnit 3.2.2.



Figur 6-6: Illustration af hvordan det det nye støjsignal er genereret. Først blev samtlige af materialets testsætninger afspillet baglæns. Inden for hver liste blev disse *bagvendte* sætninger klippet sammen – dette er forsøgt vist for de tre første lister med de i figuren tre øverste linjer. Derved opstod 16 sekvenser som blev lagt sammen ved at afspille disse oveni hinanden, hvorved det nye støjsignal fremkom. Sammenkædningen af sætningerne samt sammenlægningen af de heraf kommende sekvenser er angivet med et additionstegn.

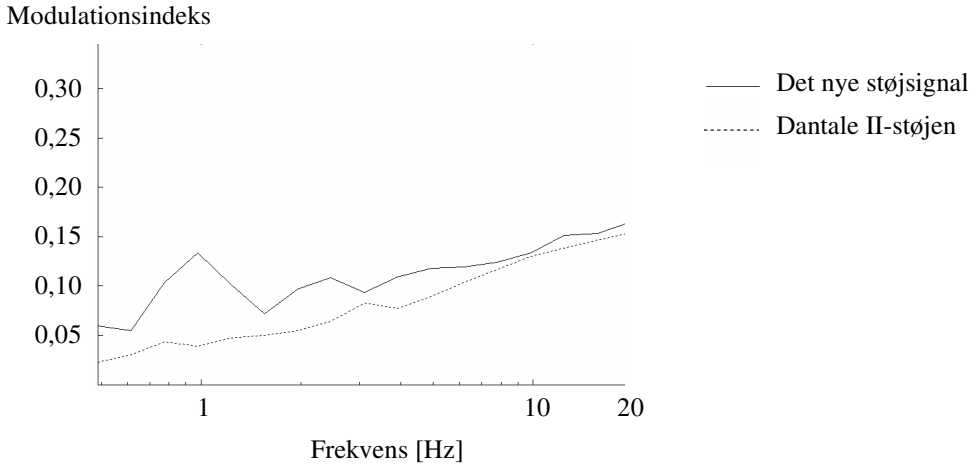
6.2.2 Det nye støjsignal vs. Dantale II-støjen

Ligesom for de forrige behandlede støjsignaler er der for det nye støjsignal blev foretaget en frekvens- og modulationsanalyse. Af figur 6-7 fremgår frekvensspektret – dvs. langtidsspektret – for både det nye støjsignal og for Dantale II-støjen. Som det ses er de to spektre identiske, hvilket skyldes den i forrige afsnit skrevne filtrering. Da begge signaler har samme langtids-spektrum maskerer de rent frekvensmæssigt testsætningerne ens.



Figur 6-7: Langtidsspekteret af det nye støjsignal og Dantale II-støjen – disse er hhv. angivet med en fuldtoptrukken og prikket linje. Begge spektre er beregnet med 1/3 oktavbånd og RMS-niveauet er opgivet relativt til det maksimale niveau i computeren. Bemærk; spektret for Dantale II-støjen er identisk med det i figur 3-4, s. 25.

Af figur 6-8 fremgår modulationsindekset for det nye støjsignal og Dantale II-støjen. Som det ses er det nye støjsignal en anelse mere moduleret end Dantale II-støjen – det er dog langt fra så moduleret som eksempelvis *icra5*. Som for Dantale II-støjen er der for modulationen beregnet et samlet modulationsindeks for den lavfrekvente modulation – dvs. op til 50 Hz. Dette samlede indeks blev for det nye støjsignal bestemt til 0,43. Til sammenligning blev det tilsvarende indeks for Dantale II-støjen bestemt til 0,36 – se evt. afsnit 3.2.2.



Figur 6-8: Modulationsindekset for det nye støjsignal og Dantale II-støjen – disse er hhv. angivet med en fuldtop-trukken og prikket linje. Indekset er opgivet med en opløsning på 1/3 oktav i frekvensområdet fra 0,5 til 20 Hz Bemærk; spektret for Dantale II-støjen er identisk med i figur 3-5, s. 27.

I Pilottest IV, hvori der indgik hørehæmmede FP'er, blev det nye støjsignal testet op imod Dantale II-støjen. Dette blev gjort ved at sammenligne værdier af SRT_{50} og s_{50} opnået med hhv. det nye støjsignal og Dantale II-støjen. Testen viste, at FP'erne generelt havde sværere ved at sortere det nye støjsignal fra sætningerne end Dantale II-støjen, hvorved FP'erne opnåede højere værdier af SRT_{50} med det nye end med den oprindelige støj – se evt. figur D-4, s. 172. I undersøgelsen foretaget af Wagener & Brand (2005), der er refereret i afsnit 6.1.3, blev der opnået de højeste værdier af SRT_{50} med de umodulerede signaler frem for med de to modulerede signaler. I Pilottest IV blev der altså opnået det omvendte. Dette kan forklares ud fra, at det nye støjsignal lyder mere som egentlig tale end hvad, der er tilfældet for Dantale II-støjen, hvorfor det nye støjsignal på trods af den lidt kraftigere modulation er sværere af sortere fra testsætningerne.

Af testen blev det fundet, at begge signaler er lige gode (/dårlige) til at differentiere mellem de implicerede FP'er, hvorfor testens sensitivitet synes at være upåvirket – se evt. igen figur D-4, s. 172. Samtidigt blev værdier af såvel SRT_{50} som s_{50} opnået ved hvert af de to støjsignaler sammenlignet med FP'ernes gennemsnitlige høretærskler ved 500 Hz, 1, 2 og 4 kHz. For en beskrivelse af de heraf opnåede resultater henvises til Bilag D, afsnit D.7 – herunder til figur D-8, s. 174 og D-9, s. 175. I forbindelse med testen blev det i øvrigt også vurderet at størrelsen af indlæringseffekten ikke synes at være påvirket af indførelsen af det nye støjsignal – se evt. figur D-11, s. 177. Imidlertid er det ikke blevet undersøgt, om det nye støjsignal evt. resulterer i en større re-test variation end den bestemt med Dantale II-støjen, hvorfor dette bør undersøges, hvis det nye signal ønskes anvendt i et endeligt testprogram.

6.3 Sammenfatning

Som alternativ til Dantale II-støjen er der flere muligheder. Hvis der ønskes et standardiseret støjsignal er et oplagt valg en af de ni ICRA-støje. Hvert af disse støjsignaler har samme langtidsspektrum som det for almen tale og en modulation svarende til en eller flere samtidige talere. En undersøgelse med det tyske materiale har vist, at kraftig moduleret støj som *icra5* frem for umoduleret støj er bedre til at differentiere mellem forskellige hørenedsættelser. Da *icra5* indeholder forholdsvis lange stilleintervaller, er der af Wagener *et al*. (2006) imidlertid blevet konstrueret to tilsvarende støjsignaler, hvor varigheden af stilleintervallerne er reduceret til hhv. 62,5 og 250 ms. Imidlertid minder ICRA-støjene ligesom Dantale II-støjen ved lytning ikke meget om virkelig tale, hvorfor der i forbindelse med dette projekt er lavet et forslag til et nyt støjsignal. Dette signal er tilsvarende Dantale II-støjen genereret ud fra materialets testsætninger – dog er testsætningerne her afspillet baglæns. Mens langtidsspektret for disse to støjsignaler er ens, er det nye støjsignal en anelse mere moduleret end Dantale II-støjen. I forbindelse med Pilottest IV blev de to signaler sammenlignet. Resultatet heraf viste, at FP'erne generelt opnåede højere værdier af SRT_{50} med det nye støjsignal end med Dantale II-støjen og at de to signaler var lige gode (/dårlige) til differentiere mellem de implicerede FP'er.

7 Afslutning

Afslutningsvist opridses, hvad der i forbindelse med dette projekt er blevet undersøgt. Derefter følger en kort diskussion og perspektivering.

7.1 Konklusion

I et testprogram som det udviklet af DELTA indgår mange forskellige parametre, hvis valg der bør tages stilling til. Programmet, som er opbygget omkring talematerialet Dantale II, består af to forskellige versioner – den operatør- og den patientbaserede version. I forbindelse med sidstnævnte kan det overvejes, om denne evt. skal afvikles med en åben frem for med en lukket responsmetode. En åben responsmetode svarer til måden, hvorpå FP'en i den operatørbaserede version afgiver sit respons, mens en lukket svarer til en *x*AFC-test. Vælges testen afviklet som en *x*AFC-test bør der ikke forventes resultater, der er sammenlignelige med de, der kan opnås med den operatørbaserede version. Dette gælder, idet FP'en i en test, der afvikles med en lukket responsmetode har en vis sandsynlighed for ud fra svarmulighederne at gætte det korrekte svar – dvs. både via *rene* og *motiverede gæt*.

Omfanget af *rene gæt* er omvendt proportionelt med antallet af svarmuligheder. Det samme formodes at være tilfældet for *motiverede gæt*. Da der kun kan tages højde for de *rene gæt* og derved ikke for de *motiverede*, bør der medtages så mange svarmuligheder som muligt. Hvis svarmulighederne skal være ord, der er indeholdt i talematerialet, svarer det til ti svarmuligheder for hvert afspillet ord. I øvrigt kan det vises, at jo flere svarmuligheder der medtages, desto mindre bliver den teoretisk mindste spredning af såvel SRT_{50} som s_{50}, hvilket er af betydning for den forventede størrelse af re-test variationen og herunder for testens samlede reliabilitet. I Pilottest II blev det i øvrigt fundet, at materialets homogenitet ikke er afhængig af, om testen afvikles med den operatørbaserede version eller med en 10AFC-test. Vælges færre svarmuligheder end der er ord indeholdt i materialet og vælges de enkelte ord gennem testen præsenteret med forskellige svarmuligheder kan der være fare for, at selv samme ord kan få varierende sværhedsgrad.

Ud over antallet af svarmuligheder bør det også overvejes, om FP'en skal have mulighed for at svare *ved ikke*. Ved at medtage denne svarmulighed undgås eventuelle frustrationer hos FP'en over, at skulle afgive et ord som svar til noget, der ikke er hørt. Imidlertid kan forskellige FP'er gøre forskelligt brug af svarmuligheden, hvorfor FP'er med selv samme taleopfattelsesevne kan opnå et forskelligt testresultat. Vælges denne svarmulighed, vurderes det derfor som en god idé, at konvertere alle *ved ikke*-svar til en $1/x$ korrekte svar, hvorved en evt. forskellig brug af svarmuligheden i en vis grad udlignes. I forbindelse hermed bør det huskes, at det i Pilottest II blev fundet, at de FP'er, som havde mulighed for at svare *ved ikke*, generelt foretog færre *motiverede gæt* end de FP'er der ikke havde denne svarmulighed. Ved at indføre svarmuligheden *ved ikke* blev der med den operatør- og den patientbaserede version derved opnået testresultater, som var sammenlignelige. For den patientbaserede version bør det dog nævnes, at hvad enten denne afvikles med en åben eller lukket responsmetode, kan der drages tvivl om, hvorvidt en sådan test

kan vinde indpas i den kliniske praksis. Det gælder jf. spørgeskema-undersøgelsen, idet selve kontakten med patienten er et centralt element i den kliniske behandling.

I forbindelse med testafviklingen skal det afgøres, om det under testen enten er tale- eller støjniveauet, der skal varieres og om støjen skal være kontinuert eller afbrudt. For at sikre en veldefineret baggrundsstøj bør niveauet af talesignalet imidlertid varieres, mens støjniveauet holdes konstant. Støjen bør samtidigt være på et niveau, hvorved selv en FP med meget dårlig hørelse kan høre signalet – niveauet kan evt. sættes individuelt. I tilknytning hertil bør præsentationsniveauet af den første sætning sættes således, at den enkelte FP som min. kan høre et af de afpillede ord. Ud over ovenstående skal det bestemmes, hvordan præsentationsniveauet ud fra FP'ens respons skal ændres. Hansen og Ludvigsens procedure, som er designet til afvikling med en åben responsmetode, giver kun en værdi af SRT_{50}, men til gengæld kan en test baseret på denne procedure afvikles manuelt. For at afvikle en test baseret på enten Brand og Kollmeiers procedure eller MML-proceduren bør der imidlertid anvendes en computer. Fordelen ved de to sidstnævnte procedurer er, at de kan anvendes således, at der både bestemmes en værdi af SRT_{50} og s_{50}.

Ønskes både en værdi af SRT_{50} og s_{50} bliver re-test variationen af SRT_{50} imidlertid større end, hvis kun SRT_{50} ønskes bestemt. Hvordan testresultatet angives har altså betydning for testens reliabilitet, hvilket blev observeret ved både computersimuleringer og ved at sammenligne resultater opnået med Pilottest I og III. Re-test variationen er dog også under påvirkning af, hvor lang testen er – dvs. af hvor mange sætninger FP'en præsenteres for. Der bør derfor foretages et kompromis mellem en rimelig re-test variation og en test af en rimelig varighed.

I tilknytning til re-test variationen bør det vurderes til hvilke formål, testen kan og ønskes anvendt – herunder om der ønskes en værdi af s_{50}. Ønskes testen anvendt i forbindelse med udvælgelse/tilpasning af HA'er bør det undersøges, om den forventede forskel mellem eksempelvis to forskellige HA'er er større end den fundne re-test variation. Hvis ikke, vil testen ikke kunne anvendes til dette formål. Selvom en test ikke kan vise en forskel, er det dog ikke ensbetydende med, at en FP ikke kan opleve en forskel mellem eksempelvis to forskellige HA-typer eller -indstillinger. Hertil bør nævnes, at re-test variationen af SRT_{50} forventes at være højere for hørehæmmede end for normalthørende, mens det omvendte forventes at gøre sig gældende for re-test variationen af s_{50}. Samtidigt kan det nævnes, at der opnås en test som er mere sensitiv ved at opgive SRT_{80} frem for SRT_{50} – dog bliver re-test variationen større.

Inden en egentlig test afvikles, skal der tages højde for, at der ved afvikling af et talemateriale som Dantale II forekommer en ikke ubetydelig indlæringseffekt – dette gælder også selvom FP'en tidligere er blevet præsenteret for materialet. I Pilottest I blev det fundet, at indlæringseffekten i den operatørbaserede version ikke er afhængig af præsentationsniveauet, og at den har samme størrelsesorden for både den operatørbaserede version og for en 10AFC-test. Som et kompromis mellem en minimal påvirkning af indlæringen på det endelige testresultat og testens samlede varighed anbefales det, at anvende 30 sætninger som træning.

I Dantale II materialet er der indbefattet et taleformet støjsignal, der kun er svagt moduleret og ikke lyder som egentlig tale. Flere undersøgelser har vist, at moduleret frem for umoduleret støj er medvirkende til bedre at kunne skelne forskellige grader af hørenedsættelser, hvorved en test med moduleret frem for umoduleret støj bliver mere sensitiv. Ved at anvende et mere moduleret støjsignal end Dantale II-støjen kan re-test variationen dog blive større og derved

dårligere. I forbindelse med dette projekt er der genereret et forslag til et nye taleformet støjsignal, der ligesom Dantale II-støjen er genereret ud fra testsætningerne. Sammenlignet med Dantale II-støjen er det nye støjsignal en anelse mere moduleret og lyder mere som egentlig tale, hvorved testen bliver mere valid. I Pilottest IV blev der imidlertid ikke observeret en større sensitivitet ved at anvende det nye støjsignal frem for Dantale II-støjen. Til gengæld opnåede FP'erne generelt højere og derved dårligere værdier af SRT_{50} med det nye støjsignal end med Dantale II-støjen, hvilket netop kan forklares med, at det nye støjsignal lyder mere som tale.

7.2 Diskussion

Ved udførelse af en taletest kan det generelt debatteres, hvorvidt testen reelt afspejler en FP's kommunikationshandicap. Det skyldes, at i dagligdagen kan taleopfattelsen betragtes som et bisensorisk fænomen, idet den ikke alene er baseret på det auditive signal, men også på viden om samtaleemnet og den visuelle støtte som mundaflæsning, gestikulation og ansigtsmimik giver anledning til. Nogle personer vil derfor måske kunne opnå et testresultat, som afspejler en dårligere evne til at opfatte tale end den, som de selv oplever. Det bør i øvrigt nævnes, at ved en audiovisuel indspilning af et talemateriale er det sjældent, at indtaleren bruger gestikulation og ansigtsmimik, hvorfor stort set samme problematik tillige gør sig gældende.

7.3 Perspektivering

I forbindelse med videreudvikling af testprogrammet kunne det være interessant at foretage en undersøgelse, hvori resultater opnået med Dantale I og II materialet sammenholdes. Desuden kunne der evt. udføres en undersøgelse, som sammenholder en test afviklet med Dantale II med den danske HINT-test. I et fremtidigt perspektiv kunne testprogrammet i øvrigt tænkes udvidet således, at det også indeholder en testafvikling tiltænkt børn, hvor længden af testsætningerne er reduceret til at bestå af tre frem for fem ord. Derudover kunne programmet også indbefatte andre typer af tests – herunder tests til bestemmelse af høretærsklen for rentoner. Fordelen ved et sådant testbatteri er, at det vil kunne øge sammenligneligheden af testresultater opnået på tværs af forskellige høreklinikker.

Det kunne ligeledes være interessant, om der kunne opbygges et nyt talemateriale – evt. kaldet Dantale III – og derved en test, der kan fortælle ved hvilke frekvenser, FP'en har svært ved at opfatte tale. Hvis en sådan test skal afvikles som en *x*AFC-test bør de forskellige svarmuligheder udformes som minimalpar. Opbygges materialet således, at det ikke er bestående af sætninger, men af enkeltord, kunne ordene vælges som navneord. Derved ville testen kunne afvikles ved, at FP'en præsenteres for billeder frem for ord, hvorved målgruppen formodes at kunne udvides. Hvorvidt der i det danske sprog findes nok navneord, som ligner hinanden, bør dog undersøges. Alternativt kunne der anvendes nonsensord, hvorved en test med dette materiale vil kunne anvendes internationalt – dog vil benævnelsen Dantale III nok ikke være passende.

Forkortelsesliste

I rapporten optræder følgende forkortelser;

AFC	Alternative Forced Choice
B&K	Brüel & Kjær
CL	Carl Ludvigsen
CVC	Consonant-Vowel-Consonant
DELTA	Dansk, Elektronik, Lys og Akustik
DL	Discrimation Loss
DS	Discrimation Score
DTU	Danmarks Tekniske Universitet
ECTS	European Credit Transfer and Accumulation System
EHIMA	The European Hearing Instrument Manufacturers Association
FIR	Finite Impulse Response
FP	Forsøgsperson
GS	Gentofte Standard-listerne
GUIDE	Graphical User Interface Development Environment
HA	Høreapparat
HC	Hørecentral-listerne
HearCom	Hearing in the Communication Society
HINT	Hearing In Noise Test
HL	Hearing Level – niveauet i dB over den standardiserede høretærskel
ICRA	International Collegium of Rehabilitative Audiology
IIR	Infinite Impulse Response
ISMADHA	International Standards for Measuring Advanced Digital Hearing Aids
LTASS	Long-Term Average Speech Spectrum
MCL	Most Comfortable Level
MLE	Maksimum-likelihood-estimat
MML	Method of Maximum Likelihood
NATASHA	Network and Tools for the Assessment of Speech/Language and Hearing Ability
OPE	Den operatørbaserede version
PATm	Den patientbaserede version med svarmuligheden ved ikke
PATu	Den patientbaserede version uden svarmuligheden ved ikke
PDC	Prolog Development Center
PTA	Pure-tone average
RMS	Root Mean Square
Sig.	Signifikansniveau
SNR	Signal to Noise Ratio
SPL	Sound Pressure Level – niveau i dB over referencen 20 µPa.
SPSS	Statistical Product and Service Solution
SRT	Speech Reception Threshold
TAL	Teknisk Audiologisk Laboratorium

Symbolliste

I forbindelse med diverse formler er følgende symboler anvendt;

a , b	Konstanter som indgår i Brand og Kollmeiers procedure
c	Antal ord FP'en svarer korrekt
$c_{gæt}$	Antallet af ord FP'en via *rene gæt* svarer korrekt
ch	Mål for hvor chancetagen FP'en er
$f(i)$	Funktion som indgår i Brand og Kollmeiers procedure
i	Antallet af reverseringer
j	Antallet af uafhængige elementer i en sætning
k	Tæller i likelihood-funktionen
m	Antal sætninger eller præsentationsniveauer
n	Antallet af ord FP'en præsenteres for
$n_{gæt}$	Antallet af ord FP'en ikke har hørt
N	Refererer til listenumre
$N_{n_{gæt},x\text{AFC}}$	Antallet af mulige udfald for at gætte $c_{gæt}$ ord korrekt
p_{Like}	Likelihood-funktion
p_{ord}	Sandsynligheden for korrekt gengivelse af et enkelt ord
p_{OPE}	Sandsynligheden for korrekt svar i den operatørbaserede version
$p_{\text{sætning}}$	Sandsynligheden for korrekt gengivelse af en hel sætning
p_{xAFC}	Sandsynligheden for korrekt svar i en *x*AFC test
$prev$	Diskriminationsværdien – dvs. andelen af ord FP'en svarer korrekt
$P_{n_{gæt},x\text{AFC}}$	Sandsynligheden for via *rene gæt* at gætte et bestemt antal ord korrekt
s_{tar}	Hældningen af S-kurven ved tar
s_{50}	Hældningen af S-kurven ved SRT_{50}
$s_{50,\text{ord}}$	Den gennemsnitlige hældning af ordenes S-kurver ved SRT_{50}
SNR	Præsentationsniveauet
SRT_{th}	Taletærsklen opgivet ved th
tar	Sandsynlighed for korrekt svar hvorved sætningerne præsenteres
th	Sandsynlighed for forståelighed hvorved taletærsklen opgives
x	Antallet af svarmuligheder
X	Stokastisk variabel
ΔSNR	Ændring af præsentationsniveauet – dvs. trinstørrelsen
μ	Gennemsnitsværdi
σ	Spredning
$\sigma_{SRT_{50,\text{ord}}}$	Spredningen af ordenes SRT_{50}

Litteraturliste

Nedenfor fremgår den litteratur, der er blevet anvendt i forbindelse med udarbejdelsen af denne rapport. Imidlertid forekommer der et vist overlap den opgivne litteratur imellem. Det skyldes, at fire af de opgivne tidsskiftsartikler er baseret på mindre dele af to større ph.d.-afhandlinger. For at undgå unødige dobbelt-henvisninger, er der i selve hovedrapporten valgt kun at henvise til artiklerne, hvor dette har været muligt. Nedenfor i parentes fremgår, hvilke afhandlinger de pågældende artikler er baseret på. Eksempelvis er artiklen af Brand & Kollmeier (**2002**) baseret på afhandlingen af Brand (**2000** – Kap. 2), hvorfor der af de to referencer fremgår de samme konklusioner. Generelt er beskrivelserne af de emner, som artiklerne omhandler, mere udførlige i afhandlingerne. For at hjælpe læseren findes der en kopi af den litteratur, der har været offentligt tilgængelig via Internettet, på den til rapporten vedlagte cd-rom – herunder de to netop omtalte ph.d.-afhandlinger. Derudover er rapporten fra det i *forordet* nævnte litteraturstudie også at finde på cd'en. For at markere hvilken litteratur det i øvrigt drejer sig om, er der i forbindelse med referencer fast gennem rapporten angivet en asterisk (*) – det samme gør sig gældende nedenfor. Det bør bemærkes, at de anvendte standarder er angivet til slut i litteraturlisten.

Aczel, A. D. (**1996**): *Complete Business Statistics*, Irwin McGraw-Hill, International Edition, 3. udgave.

Ardenkjær-Madsen, R. & Josvassen, J. L. (**2001**): *Forslag til en ny dansk sætnings-baseret audiovisuel taleaudiometrisk test i støj* – kandidatafhandling, Københavns Universitet, Institut for Almen Anvendt Sprogvidenskab.

Boothroyd, A. & Nittrouer, S. (**1988**): *Mathematical Treatment of Context Effects in Phoneme and Word Recognition*, J. Acoust. Soc. Am, 84 (1), s. 101-114.

Brand, T. (**2000**)*: *Analysis and Optimization of Psychophysical Procedures in Audiology* – ph.d.-afhandling, Carl von Ossietsky Universität Oldenburg, Fysisk Institut.

Brand, T. & Kollmeier, B. (**2002**): *Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests*, J. Acoust. Soc. Am, 111 (6), s. 2801-2810.
(Denne artikel er baseret på Brand, T. (**2000**) – Kap. 2).

Brand, T.; Wittkop, T.; Wagener, K. C. & Kollmeier, B. (**2004**)*: *Vergleich von Oldenburger Satztest und Freiburger Wörtertest als geschlossene Versionen*, Deutsche Gesellschaft für Audiologie (DGA).

Byrne, D.; Dillon, H. & Tran, K. (**1994**): *An international comparison of long-term average speech spectra*, J. Acoust. Soc. Am, 96 (4), s. 2108-2120.

Danielsen, K.; Elberling, C. & Salomon, G. (**1971**): *Experimental Monosyllabel Discrimination Test for Evaluation of Hearing Aids*, Scand. Audiol., suppl. 1, s. 45-48.

DELTA (**2003**)*: *Sammenligning af Dantale DVD i forhold til CD*, TAL.

Dillon, H. & Ching, T. (**1995**): *What Makes a Good Speech Test?* i Plant, G & Spens, K. E. (eds.), *Profound Deafness and Speech Commuication*, Whurr Publishers Ltd., London, s. 305-348.

Dreschler, W. A.; Verschuure, H.; Ludvigsen, C. & Westermann, S. (**2001**): *ICRA Noises: Artificial Noise Signals with Speech-like Spectral and Temporal Properties for Hearing Instrument Assessment*, Audiology, 40 (3), s. 148-157.

Elberling, C.; Ludvigsen, C. & Lyregaard, P. E. (**1988**): *Dantale – Compact Disc*, Teknisk rapport, Laboratoriet for Akustik, DTH (nu DTU).

Elberling, C.; Ludvigsen, C. & Lyregaard, P. E. (**1989**): *Dantale: A New Danish Speech Material*, Scand. Audiol. 18 (3), s. 169-175.

Elbro, C; Moller, S & Nielsen, E. M. (**1995**): *Functional Reading Difficulties in Denmark – A Study of Adult Reading of Common Texts*, Reading and Writing 7 (3), s. 257-276.

Evans, P. (**1997**): *Speech audiometry for differential diagnosis*, i Martin (**1997**), s. 131-149.

Ewertsen, H. W. (**1973**): *Operation Helen – Auditive, Visual & Audio-visual Perception of Speech*, Chapter 1: Perception of questions by normal subjects, Statens Hørecenter, Bispebjerg Hospital.

Gelfand, S. A. (**2001**): *Essentials of Audiology*, Thieme, 2. udgave.

Goodman, J. W. (**2000**): *Statistical Optics*, Wiley Inter-Science, 1.udgave.

Gustafsson, H. Å. & Arlinger, S. D. (**1994**): *Masking of speech by amplitude-modulated noise*, J. Acoust. Soc. Am, 95 (1), s. 518-529.

Hagerman, B. (**1984**): *Some Aspect of Methodology in Speech Audiometry – Studies of reliability, computer simulations and development of a new speech material for measuring speech reception threshold in noise.* Scand. Audiol., suppl. 21.

Hagerman, B. (**1997**): *Attempts to Develop an Efficient Speech Test in Fully Modulated Noise*, Scand. Audiol., 26, s. 93-98.

Hagerman, B. (**2002**): *Speech Recognition Threshold in Slightly and Fully Modulated Noise for Hearing-impaired Subjects*, International Journal of Audiology, 41, s. 321-329.

Hagerman, B. & Kinnefors, C. (**1995**): *Efficient Adaptive Methods for Measuring Speech Reception Threshold in Quiet and in Noise*, Scand. Audiol, 24, s. 71-77.

Hansen, M. & Ludvigsen, C. (**2001**): *Dantale II – Danske Hagerman sætninger*, Danske taleaudiometri-materialer, Værløse.

Hernvig, L. H. & Olsen, S. Ø. (**2005**): *Learning effect when using the Danish Hagerman sentences (Dantale II) to determine speech reception threshold*, International Journal of Audiology, 44 (9), s. 509-512.

Hüche, E. (**1988**): *Digital Signal Behandling*, Teknisk Forlag A/S, 1. udgave.

Hygge, S.; Ronnberg, G. J.; Larsby, B. & Arlinger, S. (**1992**): *Normal-Hearing and Hearing-impaired Subjects Ability to Just Follow Conversation in Competing Speech, Reversed Speech, and Noise Backgrounds*, Journal of Speech and Hearing Research, 35 (1), s. 208-215.

Keidser, G. (**1991**): *Reference-data for Dantale (normalthørende)*, Laboratoriet for Akustik ved DTH (nu DTU), Publikation nr. 40.

Keidser, G. (**1993**): *Normative Data in Quiet and in Noise for "Dantale" – A Danish Speech Material*, Scand. Audiol. 22, s. 231-236.

Keidser, G. (**1994a**): *Articulation Index Transfer Functions for 'Dantale' (200 Monosyllabic Words)*, Scand. Audiol. 23 (1), s. 75-77.

Keidser G. (**1994b**): *Forsøg med '4AFC'*, ODIN projektrapport (ad hoc I-IV) del 32, 33, 34 og 37, DTH (nu DTU).

Kollmeier, B. & Wesselkamp, M. (**1997**): *Development and Evaluation of a German sentence test for objective and subjective speech intelligibility assessment*, J. Acoust. Soc. Am, 102 (4), s. 1412-1421.

Ludvigsen, C. (**1974**): *Construction and Evaluation of an audio-visual test (The Helen Test)*, Scand. Audiol. Suppl. 4, s. 67-74.

Ludvigsen, C. (**1981**): *Operation Helen – Auditive, Visual & Audio-visual Perception of Speech*, Chapter 2: Perception of questions by normal-hearing and hearing-impaired persons in quiet and in noisy environments, Audiologisk afd., Bispebjerg Hospital.

Ludvigsen, C. (**2005**): *Indførelse af videnskabeligt baseret taleaudiometri i Danmark: Fra uformel hviskeprøve til computerstyret test*, Dansk Akustisk Selskab, Dansk Akustik – 1955-2005, Jubilæumsskrift, s. 150-160.

Martin, M. (**1997**): *Speech Audiometri*, Whurr Publishers Ltd, 2. udgave, San Diego, USA.

Nielsen, H. & Carver, K. (**1997**): *Håndbog i audiologiske tests*, Special-pædagogisk forlag.

Nielsen, J. B. (**2005**): *Construction of a Danish HINT*, 21. Danavox Symposium, s. 573-583.

Nielsen, K. O. & Parving, A. (**1986**): *Undersøgelse af ordmaterialer i klinisk taleaudiometri*, Ugeskrift for Læger, 138. årgang nr. 8, Den almindelige danske lægeforening, s. 441- 444.

Nilsson, M.; Soli, S. D. & Sullivan, J. A. (**1994**): *Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise*, J. Acoust. Soc. Am, 95 (2), s. 1085-1094.

Olsen, S. Ø. (**1996**): *Evaluation of the List Numerals in the Danish Speech Audiometry Material*, Scand. Audiol. 25 (2), s. 103-107.

O'Regan, J. K. & Humbert, R. (**1989**): *Estimating psychometric functions in forced-choice situations: Significant biases found in threshold and slope estimations when small samples are used*, Perception & Psychophysics, 46 (5), s. 434-442.

Parving, A.; Barford, J. & Salomon, G. (**1976**): *Taleaudiometri klinisk og socialt – en sammenlignende undersøgelse med forskellige taleaudiometriske materialer*, Ugeskrift for Læger, 138. årgang nr. 4, Den almindelige danske lægeforening, s. 237- 240.

Pedersen, E. R. (**2006**)*: *Psykoakustik – et specialeforberedende litteraturstudie*, Syddansk Universitet Odense, Institut for Sensorer, Signaler og Elektroteknik.

Plomp, R. & Mimpen, A.M. (**1979**): *Improving the Reliability of Testing the Speech Reception Threshold for Sentences*, Audiology 18, s. 43-52.

Poulsen, T. (**1993**): *Taleforståelighed (note nr. 3111)*. Laboratoriet for Akustik, DTH (nu DTU).

Poulsen, T. (**2001**): *Psykoakustiske målemetoder (note nr. 3108)*. Version 3.3, Institut for Akustisk Teknologi DTU.

Røjskjær, C. E. (**1952**): *Monaural Speech Audiometry* – doktorafhandling. Københavns Universitet. Poul Søndergaards Bogtrykkeri.

Salomon, G.; Parving, A. & Danielsen, K. (**1985**): *Definitioner og gradueringer af høreskader og kommunikationshandicap til brug ved mediko-legale kompensationer*, Ugeskrift for Læger, 147. årgang nr. 8, Den almindelige danske lægeforening, s. 685- 691.

Smits, C. & Houtgast, T. (**2006**): *Measurements and calculations on the simple up-down adaptive procedure for speech-in-noise tests*, J. Acoust. Soc. Am, 120 (3), s. 1608-1621.

Summers, V. & Molis, M. R. (**2004**): *Speech Recognition in Fluctuating and Continuous Maskers: Effects of Hearing Loss and Presentation Level*, Journal of Speech, Language, and Hearing Research, 47, s. 245-256.

van Toor, T. & Verschuure, H. (**2002**): *Effects of high-frequency emphasis and compression time constants on speech intelligibility in noise*, International Journal of Audiology, 41 (7), s. 379-394.

Wagener, K. C. (**2003**)*: *Factors Influencing Sentence Intelligibility In Noise* – ph.d.-afhandling, Carl von Ossietsky Universität Oldenburg, Fysisk Institut.

Wagener, K. C. (**2005**): *Paediatric speech intelligibility tests in noise*, 21. Danavox Symposium, s. 449- 463.

Wagener, K. C. & Brand, T. (**2005**): *Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters*, International Journal of Audiology, 44 (3), s. 144-156.
(Denne artikel er baseret på Wagener, K. C. (**2003**) – Kap. 5).

Wagener, K. C.; Brand, T. & Kollmeier, B. (**1999a**): *Entwicklung und Evaluation eines Satztests für die deutsche Sprache – Teil II: Optimierung des Oldenburger Satztests,* Zeitschrift für Audiologie, 38 (2), s. 44-56.

Wagener, K. C.; Brand, T. & Kollmeier, B. (**1999b**): *Entwicklung und Evaluation eines Satztests für die deutsche Sprache – Teil III: Evaluation des Oldenburger Satztests,* Zeitschrift für Audiologie, 38 (3), s. 86-95.

Wagener, K. C.; Brand, T. & Kollmeier, B. (**2006**): *The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners*, International Journal of Audiology, 45 (1), s. 26-33.
(Denne artikel er baseret på Wagener, K. C. (**2003**) – Kap. 6).

Wagener, K. C.; Josvassen, J. L. & Ardenkjær, R. (**2003**): *Design, optimization and evaluation of a Danish sentence test in noise,* International Journal of Audiology, 42 (1), s. 10-17.
(Denne artikel er baseret på Wagener, K. C. (**2003**) – Kap. 2).

Wagener, K. C.; Kühnel, V. & Kollmeier, B. (**1999c**): *Entwicklung und Evaluation eines Satztests für die deutsche Sprache – Teil I: Design des Oldenburger Satztests*, Zeitschrift für Audiologie, 38 (1), s. 4-15.

Anvendte standarder

Dansk Standard (**1991**): *DS/ISO 8253-1. Akustik – Audiometri – Del 1: Toneaudiometri*, 1. udgave.

Dansk Standard (**1993**): *DS/ISO 8253-2. Akustik – Audiometri – Del 2: Frifeltsaudiometri*, 1. udgave.

Dansk Standard (**1998**): *DS/EN ISO 8253-3. Akustik – Audiometri – Del 3: Taleaudiometri*, 1. udgave.

Dansk Standard (**2000**): *DS/EN ISO 389-1. Akustik – Nulpunktsreference ved kalibrering af audiometre – Del 1: Høretærskler for rene toner ved brug af supraaurale øretelefoner*, 1. udgave.

Dansk Standard (**2000**): *DS/EN ISO 7029. Akustik – Statistisk fordeling af høretærskler som funktion af alder*, 1. udgave.

Deutsche Norm (**1996**): *IEC 61260. Bandfilter für Oktaven und Bruchteile von Oktaven*, 1.udgave.

International Standard (**1993**): *IEC 645-2. Audiometers – Part 2: Equipment for speech audiometry*, 1. udgave.

Stikordsregister

A

Afviklingsniveauet 47
Afviklingsprocedurer..... 48

B

Basisliste..... 19
Bernoulli-forsøg 30;35;53
Brand og Kollmeiers procedure..... 49

C

Chancetagen 39
CL-listerne..... 13
Computersimuleringer..... 52
CVC..... 12

D

Dantale I 13
Dantale II..... 17
Dantale II-støjen..... 24
Den operatørbaserede version 28
Den patientbaserede version..... 34
Det nye støjsignal 74
DL-værdi 12

F

Fiktiv FP 53
Fonetisk balancering..... 12;13;23;20
Frekvensspektrum 25; 66; 75

G

GS-listerne..... 12
Gunnar Nu-støj 13

H

Hansen og Ludvigsens procedure..... 48
HC-listerne 12
Helen-materialet 15
HINT-test..... 15

I

ICRA 2 68
ICRA-støj 64
Indlæringen inden for en test..... 60
Indlæringen mellem to tests..... 61
Indlæringseffekten..... 60
Indspilningsteknik 20

J

j-faktoren..... 18

K

Koartikulationen..... 20

L

Langtidsspektrum..... 25;66;75
Logistisk funktion 29;37
LTASS..... 26

M

Maksimum-likelihood-estimat 50
MCL 13
Minimalpar..... 14
MML-procedure..... 51
Modulationsindekset 26;67;75
Motiverede gæt..... 35

O

Oldenburger Satztest 17
Optimeringsproces 22
Ordscoring..... 18

P

Præsentationsformen 45

R

Referenceværdier 29
Reliabilitet..... 10
Rene gæt..... 35
Responsmetode 28

S

Segmentering..... 21
Sensitivitet..... 10
S-kurve 29;38
Spondæ..... 12

T

Talegenkendelse..... 43
Taletærsklen 30

V

Validitet..... 10

X

*x*AFC-test 34

Appendiks De danske Hagerman sætninger

Nedenfor fremgår testsætningerne i talematerialet Dantale II – disse kaldes undertiden de danske Hagerman sætninger. Sætningerne er inddelt i 16 lister og indeholder hver fem ord tilhørende forskellige ordklasser, som angives i følgende rækkefølge; egennavn, udsagnsord, talord, tillægsord og navneord.

Liste 1

Ingrid	finder	syv	røde	huse
Michael	ejer	tyve	pæne	ringe
Linda	låner	seks	flotte	skabe
Ulla	får	fjorten	hvide	jakker
Niels	solgte	ti	store	masker
Henning	ser	ni	smukke	planter
Anders	vandt	otte	sjove	gaver
Kirsten	købte	tre	nye	blomster
Per	valgte	tolv	fine	biler
Birgit	havde	fem	gamle	kasser

Liste 2

Ulla	solgte	ti	røde	biler
Linda	ejer	tre	hvide	blomster
Anders	havde	tolv	pæne	kasser
Michael	valgte	otte	fine	masker
Niels	vandt	fjorten	store	jakker
Birgit	finder	ni	nye	ringe
Henning	låner	syv	sjove	skabe
Ingrid	ser	fem	flotte	planter
Kirsten	købte	seks	smukke	gaver
Per	får	tyve	gamle	huse

Liste 3

Per	ser	seks	pæne	planter
Niels	ejer	tyve	hvide	biler
Kirsten	havde	fem	store	masker
Ulla	finder	otte	nye	jakker
Linda	vandt	ni	røde	skabe
Michael	får	ti	flotte	gaver
Henning	valgte	tre	gamle	ringe
Anders	købte	syv	fine	kasser
Ingrid	låner	tolv	sjove	huse
Birgit	solgte	fjorten	smukke	blomster

Liste 4

Anders	vandt	tolv	store	kasser
Per	købte	ti	fine	biler
Ulla	ejer	syv	røde	jakker
Michael	havde	fem	nye	planter
Niels	solgte	tre	smukke	blomster
Linda	valgte	ni	hvide	skabe
Birgit	finder	tyve	pæne	huse
Ingrid	låner	otte	gamle	masker
Henning	får	seks	flotte	ringe
Kirsten	ser	fjorten	sjove	gaver

Liste 5

Henning	låner	syv	hvide	jakker
Anders	finder	fem	gamle	skabe
Linda	havde	ti	fine	gaver
Kirsten	solgte	seks	sjove	ringe
Michael	får	tolv	røde	huse
Birgit	ser	tre	store	blomster
Niels	købte	tyve	pæne	planter
Per	valgte	ni	flotte	masker
Ulla	vandt	otte	smukke	biler
Ingrid	ejer	fjorten	nye	kasser

Liste 6

Henning	havde	fjorten	sjove	blomster
Kirsten	ser	syv	store	huse
Linda	finder	fem	nye	masker
Ulla	købte	tyve	hvide	jakker
Michael	får	tolv	smukke	planter
Anders	ejer	seks	flotte	biler
Niels	låner	ti	fine	gaver
Per	solgte	ni	gamle	ringe
Birgit	valgte	tre	røde	skabe
Ingrid	vandt	otte	pæne	kasser

Liste 7

Anders	finder	seks	store	gaver
Michael	ser	tyve	gamle	blomster
Kirsten	vandt	tolv	flotte	ringe
Ingrid	valgte	otte	hvide	planter
Birgit	får	fem	fine	masker
Ulla	låner	syv	sjove	skabe
Henning	havde	tre	nye	huse
Linda	solgte	fjorten	røde	kasser
Niels	ejer	ti	smukke	biler
Per	købte	ni	pæne	jakker

Liste 8

Niels	ser	ti	fine	jakker
Per	får	tolv	store	kasser
Michael	vandt	tre	nye	skabe
Anders	ejer	tyve	flotte	gaver
Henning	låner	fem	smukke	ringe
Linda	havde	seks	sjove	masker
Kirsten	købte	fjorten	pæne	planter
Ulla	finder	otte	gamle	blomster
Ingrid	valgte	ni	røde	biler
Birgit	solgte	syv	hvide	huse

Liste 9

Linda	havde	ti	fine	blomster
Ulla	låner	tre	flotte	ringe
Per	finder	fem	store	masker
Michael	vandt	syv	smukke	biler
Ingrid	valgte	fjorten	sjove	jakker
Kirsten	får	otte	hvide	kasser
Henning	solgte	tolv	gamle	gaver
Niels	ejer	tyve	pæne	huse
Anders	købte	ni	nye	planter
Birgit	ser	seks	røde	skabe

Liste 10

Niels	finder	tyve	smukke	skabe
Henning	valgte	tolv	store	kasser
Linda	låner	fjorten	gamle	biler
Birgit	vandt	fem	fine	huse
Ingrid	købte	ni	hvide	planter
Anders	ser	seks	røde	jakker
Michael	får	tre	nye	masker
Kirsten	solgte	syv	sjove	blomster
Ulla	havde	otte	pæne	gaver
Per	ejer	ti	flotte	ringe

Liste 11

Linda	solgte	fjorten	flotte	huse
Niels	vandt	ni	hvide	blomster
Ulla	havde	tolv	røde	jakker
Kirsten	købte	seks	sjove	biler
Per	ejer	syv	store	kasser
Ingrid	får	otte	smukke	masker
Anders	finder	tre	pæne	gaver
Birgit	ser	ti	gamle	skabe
Hanning	låner	tyve	nye	planter
Michael	valgte	fem	fine	ringe

Liste 12

Ulla	ejer	fem	røde	jakker
Birgit	får	tre	store	planter
Linda	solgte	otte	flotte	huse
Michael	havde	fjorten	fine	kasser
Kirsten	ser	ni	pæne	ringe
Niels	finder	tyve	gamle	masker
Anders	valgte	seks	sjove	gaver
Henning	låner	syv	smukke	skabe
Ingrid	købte	tolv	hvide	biler
Per	vandt	ti	nye	blomster

Liste 13

Linda	ejer	fjorten	hvide	jakker
Henning	havde	ni	nye	gaver
Ulla	købte	ti	store	biler
Per	finder	fem	fine	huse
Ingrid	ser	tre	smukke	masker
Niels	vandt	seks	flotte	blomster
Anders	får	syv	pæne	planter
Birgit	valgte	tolv	gamle	skabe
Michael	solgte	tyve	sjove	kasser
Kirsten	låner	otte	røde	ringe

Liste 14

Per	valgte	otte	pæne	masker
Michael	ejer	seks	nye	huse
Linda	solgte	fem	store	kasser
Niels	købte	fjorten	gamle	jakker
Anders	vandt	tre	fine	blomster
Birgit	ser	tolv	røde	skabe
Kirsten	får	ni	smukke	ringe
Ingrid	låner	ti	flotte	planter
Henning	havde	syv	hvide	biler
Ulla	finder	tyve	sjove	gaver

Liste 15

Per	får	otte	flotte	blomster
Kirsten	ejer	syv	fine	ringe
Linda	havde	fem	smukke	huse
Ingrid	valgte	seks	nye	biler
Niels	ser	ni	pæne	skabe
Michael	solgte	fjorten	gamle	planter
Ulla	finder	tre	store	masker
Henning	låner	tyve	hvide	kasser
Anders	købte	tolv	sjove	jakker
Birgit	vandt	ti	røde	gaver

Liste 16

Henning	vandt	syv	flotte	skabe
Per	ser	otte	røde	jakker
Niels	ejer	tre	sjove	kasser
Kirsten	havde	tolv	store	ringe
Ulla	får	tyve	pæne	planter
Anders	finder	ni	smukke	gaver
Linda	købte	fjorten	nye	huse
Ingrid	låner	fem	hvide	biler
Birgit	valgte	seks	fine	blomster
Michael	solgte	ti	gamle	masker