

VURDERING
AF
STØJREDUKTIONSS-
ALGORITMER

Philip Juhl

LYNGBY 2004
EKSAMENSPROJEKT
NR. 69

IMM

Trykt af IMM, DTU

Forord

Dette er en gennemgang af en metode til beregning af taleforståelighed. Metoden bliver i dette projekt beskrevet og testet med resultater fra lytteforsøg. Jeg vil gerne takke de forsøgspersoner og vejledere fra *Widex*[®] A/S som har hjulpet mig med nærværende projekt.

Philip Juhl, IMM, 31 August 2004.

Abstract

In this project, an objective model to predict speech intelligibility (SI) for noise reduction (NR) is described. The first part of the model uses Olofsson and Hagermans[11] method for linear separation of speech and noise. To adjust the model previous result[14] from listening exercises were used. The model is tested with different NR implemented in hearing aid (HA) and Matlab[®]. Calculation of the signal-to-distortion-ratio (SDR) and SI was performed. These calculations showed that the Matlab[®] -implementation of a NR using TSVD (Truncated Singular Value Decomposition) was very non-linear. The model could not separate the speech and noise for the NR using TSVD and the predicted SI was not correct.

To verify the models predictions of SI listening exercises was performed. These exercises used two different NR implemented in HA. Statistical test was used to compare the SI for the two NR. The models predictions of the speech intelligibility for the NR, match the result from the exercise. The models prediction was compared to STI (Speech Transmission Index) and PESQ (Perceptual Evaluation of Speech Quality). The predictions from these methods did not matched the result of the listening exercise.

KEYWORDS: noisereduction, hearing aid, speech intelligibility, listening exercise.

Resume

Jeg har i dette projekt opbygget en model til beregning af taleforståelighed for støjreduktionsalgoritmer (SRA). Den første del af modellen anvender Olofsson og Hagermans[11] metode til linear separering af tale og støjsignal. Til fitting af modellen har jeg brugt resultater fra tidligere lytteforsøg [14]. Modellen er blevet afprøvet på forskellige SRA, både SRA implementeret i Matlab[®] og høreapparater med indbygget SRA. På disse SRA har jeg beregnet en værdi for taleforståeligheden og ulineariteten af SRA. Disse beregninger viste at en SRA der brugte TSVD-algoritmen (Truncated singular value decomposition) var meget ulinear. Dette betød at min model til beregning af taleforståelighed ikke virkede ordentligt på denne SRA.

For at verificere om modellens beregning af taleforståelighed er korrekt har jeg udført en række nye lytteforsøg. Ved disse lytteforsøg har jeg testet 2 forskellige SRA fra HA. Forskellene på de 2 SRA stemte overens med resultaterne fra min model. Resultaterne fra lyttetesten er sammenlignet med STI (Speech Transmission Index) og PESQ (Perceptual Evaluation of Speech Quality). Beregningen af taleforståelighed med disse modeller var markant forskellig med resultaterne fra mit lytteforsøg.

KEYWORDS: Støjreduktion, høreapparat, taleforståelighed, lytteforsøg.

Indhold

1	Indledning	11
1.1	Baggrund	11
1.2	Formål	12
1.3	Tidligere projekter	12
1.4	Metode	17
2	Model til bestemmelse af taleforståelighed	19
2.1	Separering af tale og støj efter SRA	19
2.1.1	Bestemmelse af ulinearitet	21
2.2	Psykoakustisk model	23
2.2.1	Spektogram	25
2.2.2	Beregning af SPL	25
2.2.3	Maskering	30
2.2.4	Hørestyrke	32
2.2.5	Beregning af taleforståelighed	34
3	Måling af SRA	37
3.1	Valg af høreapparat	40

3.2	Synkronisering af optagelser	41
3.3	Forbedring af optagelser	42
3.4	Måling af ulineariteten på SRA	43
3.5	Måling af taleforståelighed	46
4	Lyttetest	49
4.1	Resultat fra lyttetest	50
5	Statistiske målinger på resultatet fra lyttetest	53
5.1	Test for normalfordeling	53
5.2	Test for varianshomogenitet	55
5.3	Variansanalyse	57
5.4	Sammenligning af støjreduktioner	58
6	Sammenligning med andre modeller	59
7	Diskussion	63
8	Konklusion	65
A	Beregnet ulinearitet for SRA	69
B	Beregnet taleforståelighed	73
C	Skema til Lyttetest	77

Kapitel 1

Indledning

1.1 Baggrund

Dette projekt undersøger de forhold som personer med høreapparat (forkortet HA) opfatter tale under støjfyldte forhold. Disse personer vil ofte have sværere ved at forstå tale end folk uden HA. Derfor har man udviklet en række støjreduktionsalgoritmer (SRA) til HA som skal forbedre taleforståeligheden. For at bestemme om en given SRA øger taleforståeligheden laver man ofte en række lyttforsøg. For at få et brugbart resultat er det vigtigt at lave mange lyttforsøg og en sådan test kan derfor være meget tidskrævende. Da en enkelt SRA har mange parameter, vil det ofte kræve mange lyttforsøg før de optimale parameter til SRA er fundet. Kvaliteten af en SRA kan bestemmes ud fra forskellige kriterier. Algoritmen kan f. eks gøre støjen mindre ubehagelig at lytte til, forbedre taleforståeligheden eller forvrænge talen. Jeg har valgt at definere taleforståelighed som sandsynligheden for at en forsøgsperson opfatter ordernes fonemer korrekt. En forbedring i taleforståelighed og dæmpning af støj hænger ikke nødvendigvis sammen. Der findes eksempler på SRA der forøger taleforståeligheden men som gør signalet mere ubehageligt at lytte til, ved at tilføje pludselige ubehagelige lyde [15]. Der er også eksempler på SRA som forbedre talesignalets signal-to-noise ratio (SNR) men forvrænger talen så taleforståeligheden mindskes. Mange af de metoder man tidligere har brugt til målinger af forbedret taleforståelighed antager at sammenhængen mellem det originale og det behandlede signal er lineart. De SRA som bruges i HA er ofte adaptive, da de løbende ændrer karakteristik afhængig af om inputsignalet indeholder tale. Antagelsen om at der er linear sammenhæng mellem det oprindelige og det behandlede signal er ikke altid opfyldt. Det er

derfor vanskeligt at beregne taleforståeligheden alene ud fra signalets samlede SNR.

1.2 Formål

Målet med dette projekt er at udvikle en model der for et givent SRA giver en objektiv værdi for taleforståeligheden. Modellen skal bygge på nogle af de tidligere algoritmer som man har brugt til at estimere lyd kvalitet og taleforståelighed. Det er ikke meningen at modellen skal erstatte lytteforsøgene, men skal kunne bruges til at udvælge de SRA der giver det bedste resultat for brugeren mht. taleforståelighed.

1.3 Tidligere projekter

Der har gennem tiden været lavet en lang række undersøgelser omkring taleforståelighed. Jeg vil i dette kapitel komme lidt ind på nogle af de metoder som man har udviklet til objektivt at bestemme lyd kvalitet og taleforståelighed.

Articulations Index

En af de første metoder til beregning af taleforståelighed bygger på Articulations Index(AI) teorien. Denne metode blev udviklet af Fletcher [9] og er senere blevet beskrevet af bl.a. J.B. Allen [1]. Teorien bygger på genkendelsen af cvc ord (constant-vowel-constant) der består af 2 konsonanter og en vokal i midten. Fletchers forsøg viste at genkendelsen af hver af de 3 sproglyde var uafhængige af hinanden [20]. Fletcher lavede også en række lytteforsøg med nonsensord (se kapitel 4) og forskellige low, band og high-pass filtreret støj. Disse forsøg viste at signal-støj-forholdet (SNR) i de forskellige frekvensbånd har forskellig betydning for forståelsen af ordene. Metoden beregner taleforståeligheden ved at se på SNR i forskellige frekvensbånd. I praksis måles SNR efter SRA ved først at sende ren tale igennem og herefter tale og støj. SNR beregnes herefter ud fra de 2 output og vægtes i forhold til AI. Metoden har vist gode resultater [22], [7] når der kun er brugt additiv støj og lineær filtrering. Hvis der bruges en adaptiv SRA (som dem i HA) er det meget van-

skeligt at estimere SNR da SRA bruger forskellige filtre alt efter om signalet indeholder tale eller støj.

Speech Transmission Index

En anden objektiv måde at beregne taleforståelighed er vha. Speech Transmission Index (STI). Denne model er en videreudvikling af AI. STI kan i nogen grad tage højde for en ikke linear transformation samt ulinear ændring i tidsdomænet. Det har dog ikke haft den store succes ved beregning af taleforståelighed ved ulineare SRA. I STI bruges som input til SRA et syntetisk talesignal. Det syntetiske talesignal er for nogle SRA et problem da de bruger taledetektering til at fjerne støj. Med et syntetisk signal vil en taledetektering ikke kunne detektere tale og derfor vil SRA virke anderledes på det syntetiske signal. Der har været en række forsøg på at forbedre metoden så den også har kunne bruge talesignaler [16] men de har ikke kunne bruges tilfredsstillende på ulineare SRA.

Olofsson og Hagermans metode

I mit tidligere projekt *Måling af effektiviteten i støjreduktionsalgoritmer* [14] undersøgte jeg en algoritme udviklet af Olofsson og Hagerman [11]. Denne algoritme estimerer SRA's frekvensrespons for hhv. tale og støj ved at separere tale og støjsignalet efter det har været igennem SRA. Ud fra disse signaler beregnes SRA frekvensrespons for tale og støjsignalet. Ud fra SRA frekvensrespons af tale og støj beregnes et SNR for SRA. I projektet [14] blev algoritmen brugt til at sammenligne effektiviteten for en række SRA. Disse resultater blev sammenlignet med en subjektiv vurdering som blev lavet ved en række lyttetest. En af de SRA jeg brugte er ifølge Hagermans algoritme meget effektiv, hvorimod de subjektive målinger viste en klar forværring af taleforståeligheden. Et problem med metoden er at den måler SNR ud fra hele signalet. Da SRA ofte er adaptive vil SNR ændre sig i signalet.

Talegenkender

Ved transmission af talesignaler over f. eks Internet eller mobil-netværk, bruges der ofte kodningsteknikker til at komprimere talesignalet. Da denne komprimering ikke er tabsfri er det afkodede signal ikke det samme som det

oprindelige signal. Der er derfor udviklet en række metoder til at afgøre hvor stor ændringen i taleforståeligheden er efter komprimering af signalet. Nogle af disse metoder bygger på talegenkendere [13]. Talegenkendere er programmer som er trænet til at kunne oversætte et talesignal til tekst. Der findes i dag en lang række talegenkendere som bruger forskellige teknikker til at genkende ordene i signalet. For at talegenkendere kan bruges i en sådan test, skal programmet have egenskaber der minder om de metoder mennesker bruger til genkendelse af tale. For at beregne forbedringen i taleforståeligheden måles det antal ord som talegenkenderen gentager korrekt før og efter SRA. Forholdet mellem det antal ord som talegenkenderen genkender før og efter komprimeringen er et udtryk for ændringen i taleforståeligheden. Problemet med denne metode er at talegenkendere stadig er meget følsomme overfor støj i talesignalet og måske derfor ikke vil egne sig så godt til forsøg med meget støj.

Psykoakustiske metoder

En anden metode som er blevet brugt til at måle lyd kvaliteten i bl.a. mobiltelefoner er modeller der bygger på *psykoakustiske-modeller*. Disse metoder opstiller en model for hvordan lyden opfanges og behandles i øret [12], [4], [3]. Modellerne bruger et ubehandlet og et behandlet talesignal som input. De 2 signaler indgår i en psykoakustisk-model som beskriver hvordan lyden opfanges og opfattes af en forsøgsperson. En af fordelene ved en psykoakustisk-model er at det også er muligt at 'indbygge' forskellige former for høretab i algoritmen. Modellerne er først og fremmest udviklet og brugt til at bestemme lyd kvaliteten ved talesignaler og der er derfor ikke nogen garanti for at metoden også kan bruges til at estimere taleforståelighed.

Perceptual speech-quality measure(PSQM)

En af disse psykoakustiske-modeller er *Perceptual speech-quality measure* (PSQM) som blev introduceret af Beerends og Stemmerding [4]. Modellen er blevet standardiseret af International Telecommunication Union (ITU) [18] som en standard-model til at måle lyd kvalitet i mobiltelefoner. Lyd kvaliteten måles ved at se på signalet før og efter det har været igennem en "mobil-telefon". PSQM opdeler de 2 signaler i sekvenser på 32 ms. Da sekvenserne fra signalet før og efter bliver sammenlignet, er det vigtigt at de 2 signaler er synkroniseret så talen starter på samme tidspunkt. Justering af de 2 signaler kan

foregå på flere forskellige måder alt efter hvordan signalet er forsinket. Hvis forsinkelsen er tidsuafhængigt invariant kan en krydskorrelation mellem de 2 signaler estimere forsinkelsen. Alternativt kan man bruge en algoritme til detektering af tale for de 2 signaler og justere signalerne ind så talen starter samtidig. Hvis forsinkelsen ikke er konstant over tid kan signalerne deles op i mindre dele (f. eks 32 ms) og for hver del beregne krydskorrelationen mellem de 2 signaler. For at kompensere for en eventuel forstærkning af signalet bliver det behandlede signal skaleret efter signalernes samlede power.

$$S_{global} = \sqrt{\frac{\sum_{startpoint}^{stoppoint} x^2[n]}{\sum_{startpoint}^{stoppoint} y^2[n]}} \quad (1.1)$$

$$y'[n] = S_{global} \cdot y[n]$$

Her er $x[n]$ og $y[n]$ hhv. signalet før og efter. I PSQM har man opbygget en psykoakustisk model, som måler de hørbare ændringer der er på signalet før og efter. Ved at omforme lyden til et tid-frekvens spektrum, får man et groft billede af hvilke frekvenser signalet består af på et givet tidspunkt. Dette modelleres i PSQM ved at beregne et powerspektrum for hver sekvens. Dette powerspektrum omregnes til et hørestyrkespektrum som svarer til hvor kraftigt hvert frekvensområde høres. Til sidst sammenlignes de 2 spektra for signalet før og efter. Ud fra denne sammenligning beregnes en værdi for den hørbare forværring af signalet.

Perceptual evaluation of speech quality (PESQ)

PESQ er en videreudvikling af PSQM og er ligesom PSQM brugt til at estimere talekvalitet i mobiltelefoner. En af de største ændringer er den forbedrede synkroniseringsdel som har vist sig at være et stort problem ved signaler som sendes over et netværk. Hvis det oprindelige og det behandlede signal ikke er synkroniseret korrekt vil PSQM beregne en dårligere værdi for taleforståelighed. For de fleste SRA vil det behandlede signal have en konstant eller en frekvensafhængig tidsforskydning. Denne forbedring i PESQ vil derfor ikke have den store indvirkning når den bruges på SRA. Omregningen fra hørestyrkespektrum til en værdi af forvrængningen er stort set den samme som for PSQM. I PESQ har man valgt at vægte ulineare forstyrrelser højere en lineare forstyrrelser.

Martin Hansens Perceptions model

Ligesom for PSQM bruger Martin Hansens Perceptions model (MHPM) et rent talesignal og et behandlet talesignal. De 2 signaler er som for PSQM blevet synkroniseret m.h.t. talen i signalerne. MHPM bruger i modsætning til PSQM en filterbank til at dele inputsignalet op i 19 bandpass filtreret signaler. I MHPM modelleres den dynamiske kompression og den tidsmæssige tilpasning som der sker inde i øret. Dette modelleres i MHPM v.h.a. feedback-løkker. Hver af de 19 filterede signaler kommer igennem 5 adaption-loops som efterligner den dynamiske kompression og tilpasning som der sker inde i øret. I MHPM er sammenligningen af signalerne en del simple end PSQM. Her bruges en linear korrelationskoefficient for de 2 signaler som mål for lyd kvaliteten. Martin Hansens har i [12] sammenlignet hans MHPM med PSQM. I hans test af de 2 algoritmer klarer PSQM-algoritmen sig en anelse bedre end MHPM.

1.4 Metode

Jeg vil i mit projekt opbygge en psykoakustisk model til at bestemme ændringen i taleforståelighed for en given SRA. Modellen er en videreudvikling af min model fra forrige projekt [14]. Den første del af modellen bruger en modificeret udgave af Hagerman og Olofsson separeringsmodel. Den sidste halvdel af modellen består af en psykoakustisk model der beregner en værdi for taleforståeligheden ud fra den del af talen som ikke er maskeret af støjen. Til justering af modellen vil jeg bruge resultaterne fra mit tidligere lyttforsøg[14]. Til verificering af modellens beregning af taleforståelighed udføres en række nye lyttforsøg. Lyttforsøgene vil igen være med nonsensord til bestemmelse af taleforståelighed. Resultaterne fra disse lyttforsøg bliver med statistiske test sammenlignet. Denne sammenligning skal vise om der er signifikant forskel på SRA og om resultaterne stemmer overens med den beregnede værdi af taleforståeligheden. Tilsidst vil jeg beregne taleforståelighed med 2 andre modeller. Jeg har til denne sammenligning valgt at bruge PESQ og STI.

- Opbygning og implementering af Psykoakustisk model
- Tilpasning af model ud fra tidligere lyttforsøg
- Sammenholde forudsigelser fra min model med lyttforsøg
- Sammenligne min models forudsigelser med PESQ og STI

Kapitel 2

Model til bestemmelse af taleforståelighed

Jeg har valgt at opbygge en metode til bestemmelse af taleforståelighed ud fra nogle af de forskellige metoder som jeg har beskrevet i kapitel 1.3. Modellen vil senere blive evalueret ved at sammenligne den med resultaterne fra lyttetest der måler taleforståelighed. På figur 2.1 ses en skitse af modellen.

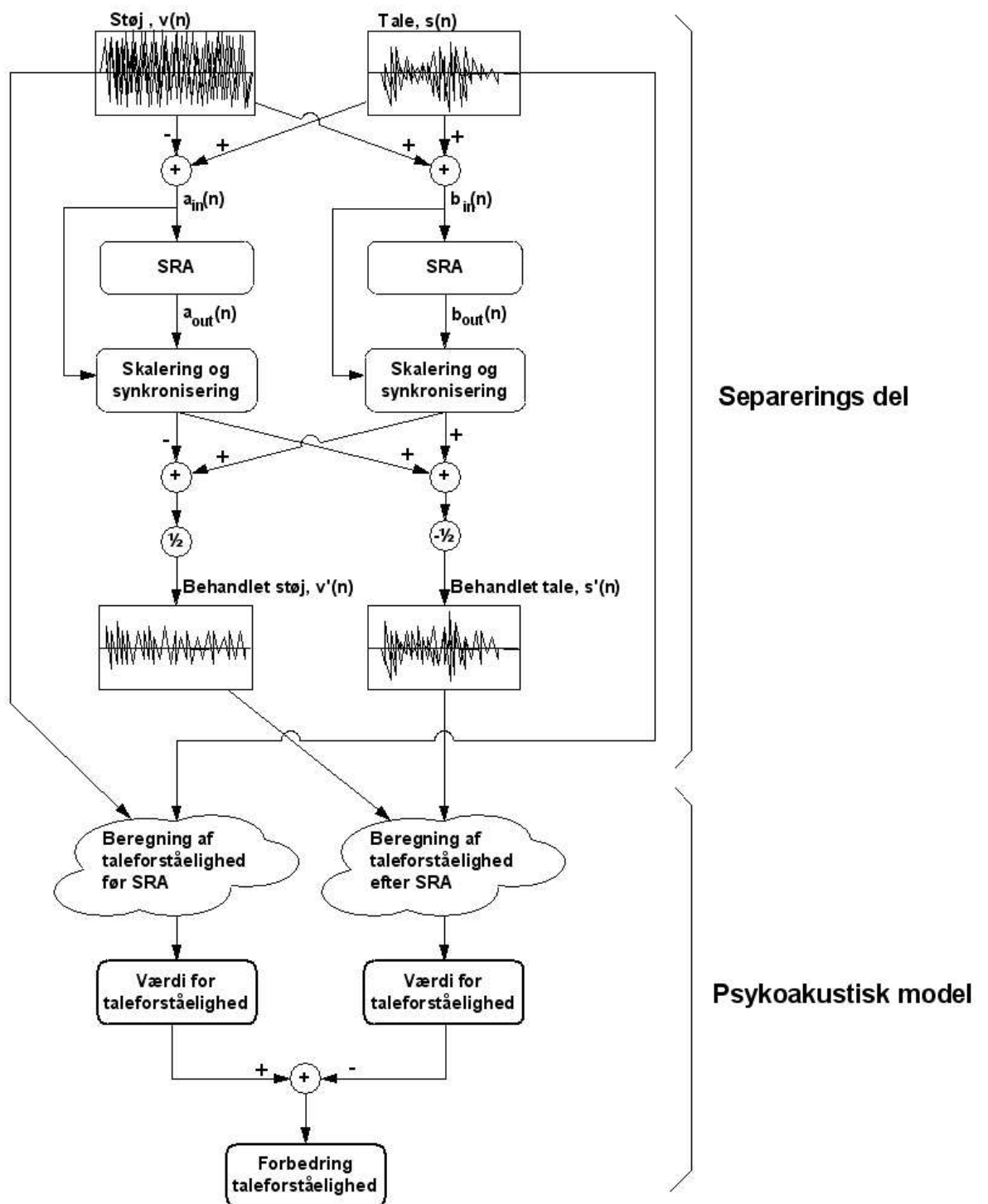
Som det ses på figuren er modellen delt op i 2 dele. Den første del separerer talen og støjen efter SRA. Den anden del af modellen beregner en værdi for taleforståeligheden før og efter SRA. På figuren er kasserne med de skarpe hjørner tale og støjsignaler. I de følgende underkapitler beskrives de enkelte dele af modellen.

2.1 Separering af tale og støj efter SRA

Den første del af algoritmen bruger Olofsson og Hagermans metode til separering af tale og støjsignal efter SRA. For at støjen og talen kan separeres skal det rene talesignal $s[n]$ og støjsignalet $v[n]$ være kendt. Ud fra disse 2 signaler dannes følgende signaler

$$\begin{aligned}a_{in}[n] &= s[n] - v[n] \\ b_{in}[n] &= s[n] + v[n]\end{aligned}$$

Signalerne $a_{in}[n]$ og $b_{in}[n]$ svarer til et støjfuldt signal før en SRA. Da $s[n]$ og $v[n]$ er kendt kan SNR (signal-to-noise ratio) før SRA let beregnes.



Figur 2.1: Skitse af metode til beregning af taleforståelighed

De 2 signaler som er input til den SRA som metoden skal undersøge er

$$\begin{aligned} a_{out} &= SRA(a_{in}) \\ b_{out} &= SRA(b_{in}) \end{aligned}$$

Outputtet fra SRA a_{out} og b_{out} er altså det signal som brugeren vil høre hvis SRA er slået til. Signalerne $a_{out}[n]$ og $b_{out}[n]$ deles op i en linear behandling af talen $s'[n]$ og støjen $v'[n]$ samt en fejl ved denne tilnærmelse $e_1[n]$ og $e_2[n]$.

$$\begin{aligned} a_{out}[n] &= s'[n] - v'[n] + e_1[n] \\ b_{out}[n] &= s'[n] + v'[n] + e_2[n] \end{aligned}$$

Fejlen $e_1[n]$ og $e_2[n]$ kan skyldes en ulinear behandling af inputsignalet eller støj fra SRA. Talen og støjen efter SRA kan herefter separeres på følgende måde

$$s'[n] = \frac{a_{out}[n] + b_{out}[n] - e_1[n] - e_2[n]}{2} \quad (2.1)$$

$$v'[n] = \frac{b_{out}[n] - a_{out}[n] + e_1[n] - e_2[n]}{2} \quad (2.2)$$

Hvis fejlen $e_1[n]$ og $e_2[n]$ er lille og ikke har nogen betydning for taleforståeligheden, kan der ses bort fra disse led og herved få den behandlede tale $s'[n]$ og støj $v'[n]$ ud fra $a_{out}[n]$ og $b_{out}[n]$.

$$\begin{aligned} s'[n] &= \frac{a_{out}[n] + b_{out}[n]}{2} \\ v'[n] &= \frac{b_{out}[n] - a_{out}[n]}{2} \end{aligned}$$

Den behandlede tale og støj kan herefter bruges til bestemmelse af taleforståelighed.

2.1.1 Bestemmelse af ulinearitet

Der findes en række metoder til at estimere den fejl som begås ved at antage at SRA er linear. I [11] og [14] er der angivet 2 metoder til estimering af denne

fejl. Hvis man antager at den ulineare del $e_1[n]$ og $e_2[n]$ er 0 og afspiller de separerede signaler $s'[n]$ og $v'[n]$ kan man lytte efter forvrængninger i tale og støjsignalet. En mere automatisk metode til bestemmelse af størrelsen på $e_1[n]$ og $e_2[n]$ kan laves på følgende måde v.h.a. en Hilbert transformation. Hvis inputsignalet $b_{in}[n]$ bliver Hilbert transformerede fås følgende.

$$\tilde{b}_{in}[n] = \tilde{s}[n] + \tilde{v}[n]$$

Hvor $\tilde{s}[n]$ og $\tilde{v}[n]$ er den Hilbert transformerede af hhv. $s[n]$ og $v[n]$. Ved at bruge $\tilde{b}_{in}[n]$ som input signal til SRA fås følgende output signal

$$\bar{b}_{out}[n] = \tilde{s}'[n] + \tilde{v}'[n] + e_3[n]$$

Outputsignalet $\bar{b}_{out}[n]$ er igen delt op i en linear del $\tilde{s}'[n]$ og $\tilde{v}'[n]$ og en ulinear del $e_3[n]$. Et komplekst signal hvor b_{in} er den reelle del og \tilde{b}_{in} er den imaginære del vil pga. Hilbert transformationen ikke have nogle negative frekvenser. Dette er en af egenskaberne ved en Hilbert transformation.

$$g[n] = b_{in}[n] + j\tilde{b}_{in} = s[n] + v[n] + j\tilde{s}[n] + j\tilde{v}[n]$$

Hvor $j = \sqrt{-1}$. Hvis man på samme måde danner et komplekst signal ud fra b_{out} og \tilde{b}_{out} fås følgende signal

$$h[n] = b_{out}[n] + j\tilde{b}_{out}[n] = s'[n] + v'[n] + e_2[n] + j\tilde{s}'[n] + j\tilde{v}'[n] + je_3[n]$$

Da signalet $g[n]$ ikke indeholder negative frekvenser vil de negative frekvenser for $h[n]$ alene stamme fra $e_2[n]$ og $e_3[n]$. Ved at beregne powerspektrummet for $h[n]$ fås følgende

$$S_h(f) = \begin{cases} S_{e_2}(f) + S_{e_3}(f) & \text{for } f < 0 \\ 4S_s(f) + 4S_v(f) + S_{e_2}(f) + S_{e_3}(f) & \text{for } f > 0 \end{cases} \quad (2.3)$$

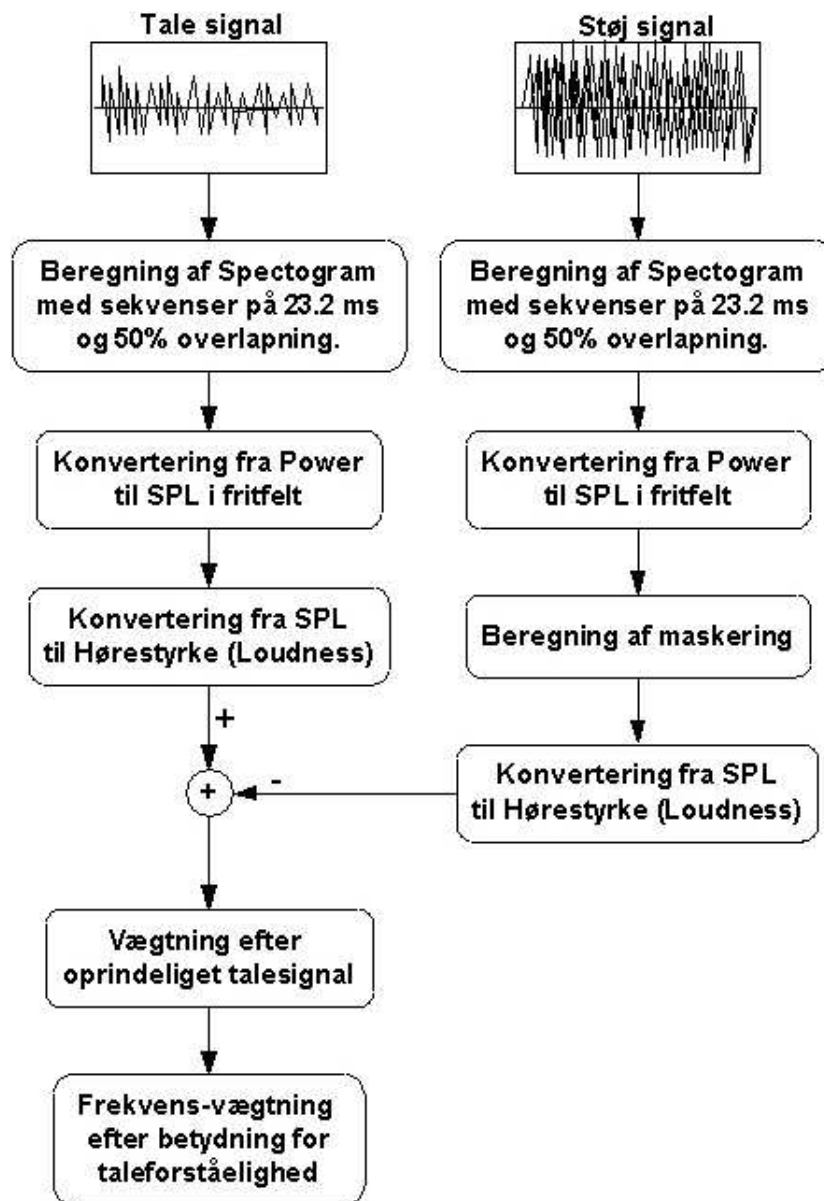
Ved at tage forholdet mellem $S_h(f)$ og $S_h(-f)$ for $f > 0$ vil man få et SDR (signal-to-distortion ratio) for en given frekvens f .

$$SDR(f) = \frac{S_h(f)}{S_h(-f)} \quad \text{for } f > 0 \quad (2.4)$$

Det er med denne metode muligt at sige noget om størrelsen på e_1 og e_2 . Spørgsmålet er så hvor stor en betydning denne ulinearitet har for taleforståeligheden. Dette er et af de uafklarede spørgsmål omkring denne separeringsmetode.

2.2 Psykoakustisk model

Den anden del af modellen består af 2 psykoakustiske modeller til beregning af taleforståelighed. De 2 psykoakustiske modeller for signalerne før og efter SRA er ens. Begge modeller omformer tale og støjsignalet i en række trin. En skitse af den psykoakustiske model kan ses på figur 2.2. Jeg vil i de følgende afsnit beskrive hvordan hvert trin i algoritmen virker.



Figur 2.2: Skitse af psykoakustisk model

2.2.1 Spektrogram

Jeg har valgt at opdele talen og støjen i sekvenser. Fordelen ved denne opdeling er bl.a. at modellen tager højde for de SRA som løbende ændre filterkarakteristik. Jeg har valgt at bruge sekvenser på 1024 samples hvilket cirka svarer til 23.2ms med en samplingfrekvens på 44.1 kHz. For hver sekvens beregnes et Spektral-powerspektrum ($P[i, k]$) på følgende måde

$$\begin{aligned} X[i, k] &= \frac{1}{1024} \sum_{n=nstart_i}^{nstart_i+1023} x[n] \cdot e^{-2\pi j \cdot (k-1) \cdot \frac{n-1}{1024}} \\ \text{for } i &= 0, 1, 2, \dots, N_{sekvens} \\ \text{for } k &= 0, 1, 2, \dots, 1023 \end{aligned} \tag{2.5}$$

Hvor $x[n]$ er signalet ($s[n]$, $v[n]$, $s'[n]$ og $v'[n]$) spektrumet skal beregnes for. For at omregne frekvenserne k til Hz bruges følgende formel

$$f_k = (k/N)F_s$$

hvor N er antallet af samples i sekvensen og F_s er samplingfrekvensen. For at omregne spektrummet til et Spektral-powerspektrum bruges følgende formel

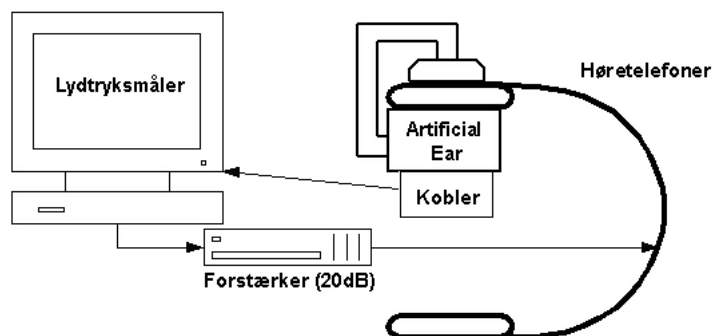
$$P[i, k] = X[i, k] \cdot X^*[i, k]$$

Hvor $X^*[i, k]$ er den komplekskonjugeret af $X[i, k]$.

2.2.2 Beregning af SPL

Ved Sound Pressure Level (SPL) forstås det lydtryk som et signal frembringer. SPL angives normalt i dB hvor 0dB SPL svarer til et lydtryk på $p_0 = 20 \cdot 10^{-6} Pa$. Når et lydtryk angives i SPL er det vigtigt at vide hvor dette er målt. Da øregangens geometri ændrer forstærkning i bestemte frekvensområder vil lydtrykket uden for øregangen (fritfelt) være forskelligt fra lydtrykket inde i øret.

For at beregne et signals SPL i fritfelt har jeg lavet en række målinger på mit system. Systemet består af en PC med Matlab[®] installeret og lydkort. Til lydkortet er tilsluttet et sæt åbne høretelefoner (Beyerdynamic, DT-770).



Figur 2.3: Måling af høretelefonernes frekvenskarakteristik

Målinger har jeg foretaget med en lydtryksmåler som er koblet til høretelefonerne. Jeg har lavet 2 forskellige målinger med lydtryksmåleren. Opstilling til denne måling kan ses på figur 2.3.

Den første måling er foretaget med et frekvens sweep mellem [100Hz–10kHz]. Dette sweep bliver genereret af lydtryksmåleren og bliver herefter forstærket af en 20dB forstærker med et meget fladt frekvenskarakteristik. Høretelefonens højttaler sidder fastspændt til et *kunstigt øre* der efterligner frekvenskarakteristikken for et normalt øre. For enden af det kunstige øre sidder en kobler som efterligner øregangen. I kobleren sidder en mikrofon som er forbundet til lydtryksmåleren. Det målte lydtryk er opgivet i dB SPL ved trommehinden. Resultaterne af målingerne kan aflæses på figur 2.4. På figuren for frekvensresponsen ses at det med rimelighed kan antages at forstærkningen på høretelefonens højre og venstre side er ens da den maksimale forskel er på mindre end 4dB. Jeg vil derfor bruge middelværdien som et samlet frekvensrespons for høretelefonerne. Da forstærkningen er angivet i dB har jeg beregnet middelværdien Mix af højre og venstre side på følgende måde.

$$Mix(f) = 10 \cdot \log_{10}\left(\frac{10^{\frac{G_H(f)}{10}} + 10^{\frac{G_V(f)}{10}}}{2}\right)$$

Her er $G_H(f)$ og $G_V(f)$ forstærkningen i dB for hhv. højre og venstre side. Den næste måling jeg foretog på lydsystemet var afspilning af toner med forskellig styrke. Denne måling er foretaget ved at generere en ren tone i Matlab[®], beregne signalets power og afspille den igennem høretelefonerne. Herefter er lydtrykket for tonens frekvens igen målt med lydtryksmåleren i dB SPL ved trommehinden. Tonerne blev generet i Matlab ud fra følgende formel

$$x[n] = A \cdot \cos(2\pi f \cdot \frac{n}{f_s})$$

Hvor f er tonens frekvens, f_s er samplingsfrekvensen og A er tonens amplitude. Ved afspilning af en tone i Matlab skal værdierne ligge i intervallet $[-1, 1]$. Jeg har valgt at afspille alle signaler med en samplingsfrekvens på 44.1kHz. På figur 2.5 ses resultaterne af disse målinger. Her er x-aksen signalets RMS-værdi og y-aksen er det målte lydtryk ved afspilning af tonen. Da målingerne for de forskellige frekvenser alle ligger på parallelle rette linjer vil jeg antage at signalets logaritmiske power værdi og dB SPL er lineart afhængige. En ulempe ved disse målinger er at de er foretaget ved en forholdsvis høj dB SPL. Ved meget lave SPL-værdier vil signalet bruge færre bit og støjen fra kvantisering vil derfor bliver mere tydelig og skabe ulineare forhold. Jeg vil dog se bort fra denne tendens i min model. Ud fra målingerne på figur 2.5 har jeg estimeret følgende ligninger for de 4 rette linjer.

$$\begin{aligned} SPL_{500} &= 19.8 \cdot \log_{10}(RMS) + 77.5 \\ SPL_{1000} &= 19.1 \cdot \log_{10}(RMS) + 78.0 \\ SPL_{5000} &= 19.1 \cdot \log_{10}(RMS) + 82.1 \\ SPL_{7500} &= 19.1 \cdot \log_{10}(RMS) + 80.3 \end{aligned}$$

På figuren ses at de 4 linjer alle har næsten ens hældning. Jeg vil derfor antage at en tones logaritmiske RMS værdi er proportional med den målte dB SPL fra høretelefonerne. Proportionalitetskonstanten har jeg beregnet til følgende

$$\alpha = \frac{19.8 + 19.1 + 19.1 + 19.1}{4} = 19.275$$

Det kan derimod ikke antages at toner med samme RMS-værdi men med forskellig frekvens har samme dB SPL. Dette kan ses ved forskydningen af linjerne. Der er altså en linear sammenhæng mellem en tones power og dB SPL hvis tonens frekvens fastholdes. En sammenligning mellem høretelefonernes frekvenskarakteristik og tonernes dB SPL målt ved 500Hz, 1kHz, 5kHz og 7.5kHz er angivet i tabel 2.1. I tabellen ses at forskellen mellem tonernes dB SPL og frekvenskarakteristikken er næsten konstant. Jeg vil derfor bruge middelværdien af forskellene til at konvertere mellem de 2 forskellige typer

Frekvens (Hz)	dB SPL for testtoner	Målt forstærkning ved frekvenskarakteristik(dB)	db SPL - Forstærkning Δ
500	77.5	20.6	56.9
1000	78.0	21.0	58.0
5000	82.1	24.5	57.6
7500	80.3	23.3	57.0
Gennemsnit			57.4

Tabel 2.1: Sammenhæng mellem høretelefonernes frekvenskarakteristik og toner genereret i Matlab.

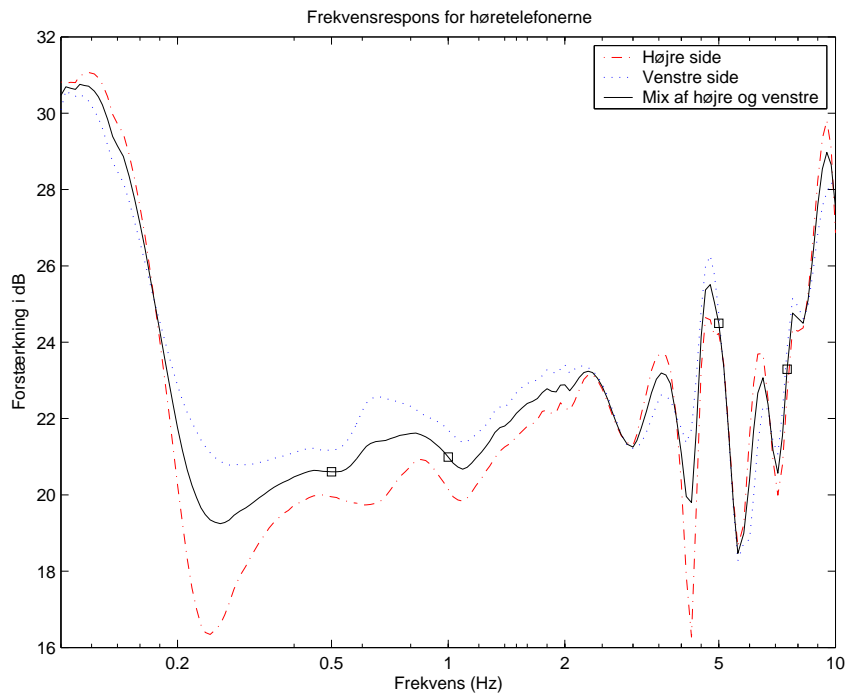
målinger. Ved at trække denne middelværdi fra målingerne af høretelefonernes frekvenskarakteristik kan jeg konvertere en tone med en given frekvens og power til dB SPL.

$$SPL = \alpha \cdot \log_{10}(P) + \beta(Hz) - \Delta$$

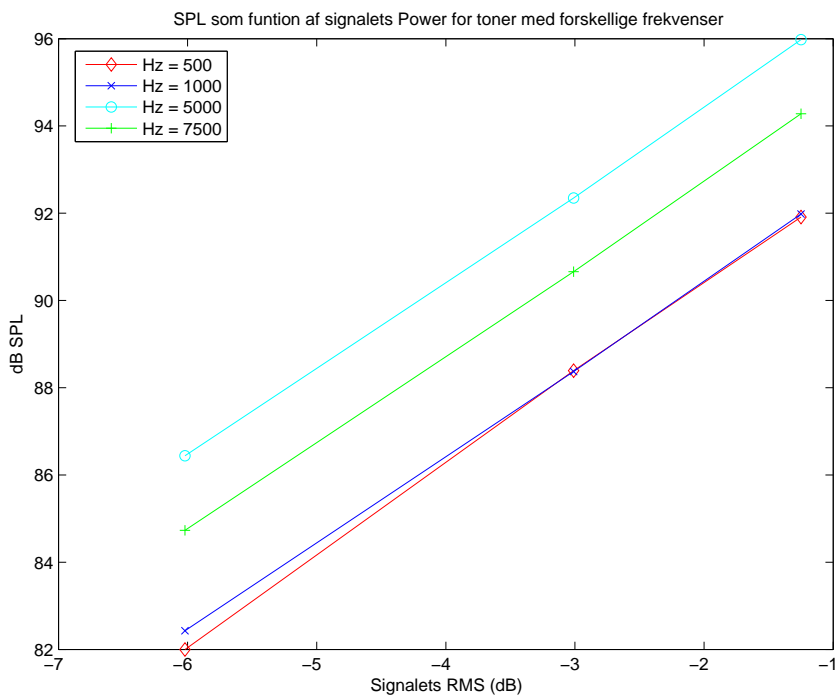
Her er P tonens power og $\beta(Hz)$ er forstærkningen ved tonens frekvens som kan aflæses på figur 2.4. Forskellen mellem frekvenskarakteristik og toners dB SPL er angivet med Δ .

$$X_{SPL}[i, k] = 19.275 \log_{10}(P[i, k]) + \beta\left(\frac{2k}{F_s}\right) - 57.4$$

Med ovenstående ligning konverteres powerspektrummet til et SPL-spektrum hvor $P[i, k]$ er powerspektrummet.



Figur 2.4: Frekvensrespons for høretelefoner for højre og venstre side



Figur 2.5: Sammenhæng mellem et signals RMS-værdi og lydtryk

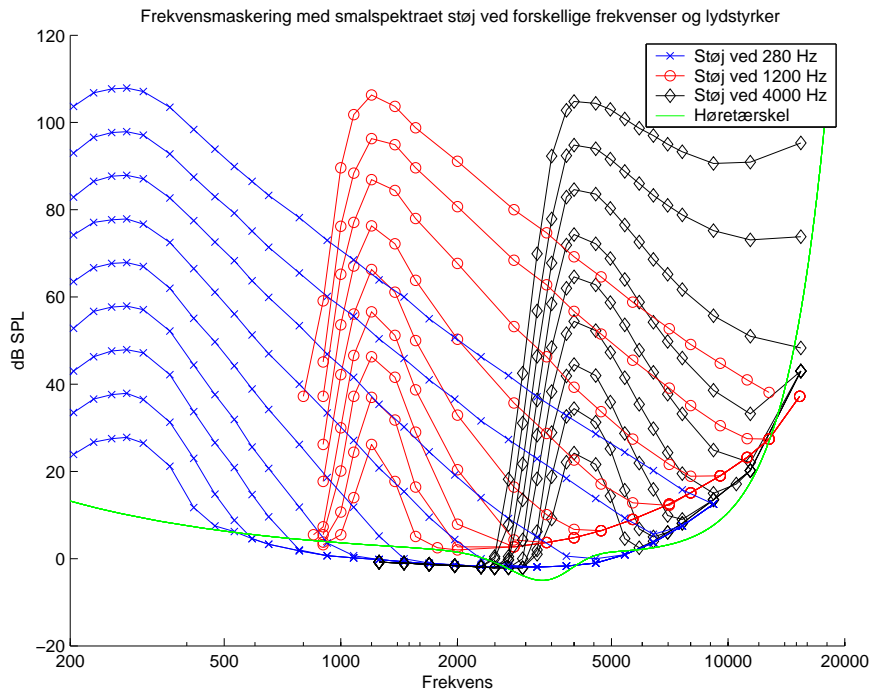
2.2.3 Maskering

Ved maskering forstås at en tone eller støj kaldet *masker*, maskerer en test-tone så dens hørestyrke mindskes eller den ikke kan høres. Maskering udnyttes bl.a. ved komprimering af musik hvor dele af signalet er maskeret og derfor kan fjernes uden at det kan høres. Masker og test-tonen kan have forskellige frekvenser men generelt bliver maskeringen større hvis frekvenserne ligger tæt på hinanden. En *masker* maskerer test-toner med højere frekvens bedre end test-toner med lavere frekvens. Udover maskering af test-toner med forskellige frekvenser kan den også maskere toner før og efter at *masker* er blevet afspillet. Maskering kan således deles op i 3 forskellige typer. Hvis test-tonen kommer før *masker* kaldes det *pre-maskering*. Ved denne type maskering kan der maksimalt være 50 ms mellem test-tonen og *masker*. Pre-maskering har generelt en meget lille effekt på maskeringen. Når test-tonen bliver præsenteret efter *masker* kaldes det *post-maskering*. Forskydningen mellem *masker* og test-tone kan være op til 200ms ved *post-maskering*. Hvis test-tonen og *masker* bliver afspillet samtidig kaldes det *simultan-maskering*. Af de 3 typer maskering er *simultan-maskering* den som maskerer mest. I min model anvendes maskering til at beregne de dele af talen som bliver maskeret af støjen. Maskeringerne bliver kun beregnet ud fra spektrumet af støjen før og efter SRA. I nogle tilfælde kan talen maskere andre dele af talen. Jeg har dog ikke taget denne type maskering af talen med i modellen.

Simultan-maskering

Der er foretaget en række forsøg til bestemmelse af simultan-maskering. Zwicker har i *Psychoacoustics : Facts and models*[23] udført en række lytteforsøg, hvor han har maskeret en test-tone med smalspektret støj. Lytteforsøgene har været udført med maskeringsstøj ved 280, 1200 og 4000Hz og med lydstyrke fra 30 til 110 dB SPL med 10 dB's spring. For hver afspilning af maskeringsstøjen er tærskelværdien for en netop hørbar test-tone bestemt. Resultatet af disse lytteforsøg kan ses på figur 2.6.

På figuren ses det tydeligt hvordan maskeringen fra den smalspektret støj er størst ved frekvenser der er højere end maskeringsstøjens. Dette gælder specielt for maskering ved de lave frekvenser. Samtidig bliver kurverne mere 'flade' i takt med at lydstyrken stiger. Formen på maskeringskurverne afhænger altså både af lydstyrken og maskeringsstøjens frekvens. Jeg har valgt at estimere frekvensmaskeringen ud fra Zwickers målinger. Ved maskeringen



Figur 2.6: Frekvensmaskering med smalspektret støj

med støj med en frekvens og SPL som er forskellig fra Zwicker's målinger beregnes en vægtet middelværdi ud fra de nærmeste frekvensmaskeringskurver. På denne måde er det muligt at estimere en maskeringskurve for støj med en bestemt frekvens og SPL.

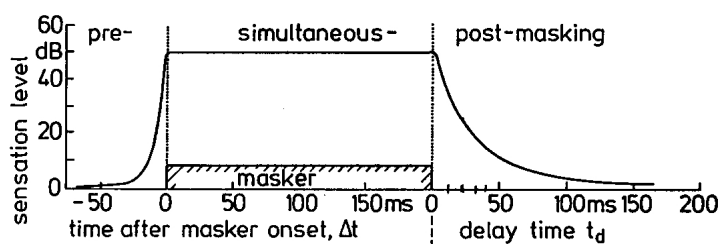
Maskeringskurven for den i 'te sekvens beregnes som maximum af maskeringskurverne for hver af de målte frekvenser. Beregning af simultan-maskering kan således opstilles til følgende ligning

$$X_{Mask_S}[i, k] = \max(Mask_S(i, \frac{2k}{F_s}, 1), Mask_S(i, \frac{2k}{F_s}, 2), Mask_S(i, \frac{2k}{F_s}, 3), \dots)$$

Funktionen $Mask_S(i, \frac{2k}{F_s}, n)$ angiver den maksimale styrke en tone med frekvensen $\frac{2k}{F_s}$ kan maskere. Maskerens styrke i SPL og frekvens er hhv. $X[i, n]$ og $\frac{2k}{F_s}$.

Maskering i tid

Tidsmaskering er mest fremherskende når støjen har nogle korte perioder hvor støjen ikke er hørbar. Ved en konstant støj er der kun maskering i tid lige inden støjen starter og når støjen stopper. I *Psychoacoustics : Facts and models* [23] er der en række figurer der viser hvordan en *masker* typisk maskerer en forsinket test-tone. Figur 2.7 er taget fra denne bog. På figuren



Figur 2.7: Maskering i tid for en støj

ses hvordan en test-tone maskeres ved afspilning af en *masker* på 200-ms. På figuren ses tydeligt at premaskeringen er meget kortere end postmaskeringen. Da premaskeringen er meget kort, har jeg valgt kun at tage post-maskering med i min model. Dette har jeg gjort ved at beregne hvor meget maskeringen aftager for hver 25 ms. Dette svarer nemlig til længde på hver sekvens i mit spektrogram. På figur 2.7 ses at maskeringen halveres i dB for hvert 25 ms. Jeg har derfor i modellen beregnet postmaskeringen som $\frac{1}{2}$ af styrken i forrige sekvens og $\frac{1}{4}$ af styrken fra sekvensen før den. Postmakseringen bliver således beregnet på følgende måde

$$X_{Mask_P}[i, k] = \text{Max}(X_{Mask_S}[i, k], \frac{1}{2}X_{Mask_S}(i-1, k), \frac{1}{4}X_{Mask_S}(i-2, k))$$

Hvor $X_{Mask_S}[i, k]$ er støjens lydstyrken i SPL for den i 'te sekvens og k 'te frekvens.

2.2.4 Hørestyrke

Da toner med samme SPL men med forskellige frekvenser ikke opfattes lige kraftigt har jeg valgt at konvertere fra SPL til *hørestyrke*. Hørestyrke er en

subjektiv værdi for hvor højt en tone opfattes i forhold til en anden tone. Enheden *Sone* er en enhed for hørestyrke, hvor 1 Sone svarer til hørestyrken ved en 1 kHz tone præsenteret ved 40 dB SPL. 0 Sone svarer til høretærsklen for den specifikke tone[23]. Selv om hørestyrke er en subjektiv værdi har lytteforsøg vist at der er en god overensstemmelse (for normalthørende) af hvor kraftigt en given tone opfattes. For at konvertere en sekvens af signalet til et hørestyrke-spektrum har jeg valgt at bruges Zwickers model[23]. Denne model antager en proportionalitet mellem relativ ændring i intensiteten og relativ ændring i *hørestyrke*. Ud fra denne antagelse komme Zwicker frem til følgende formel for beregning af *hørestyrke*.

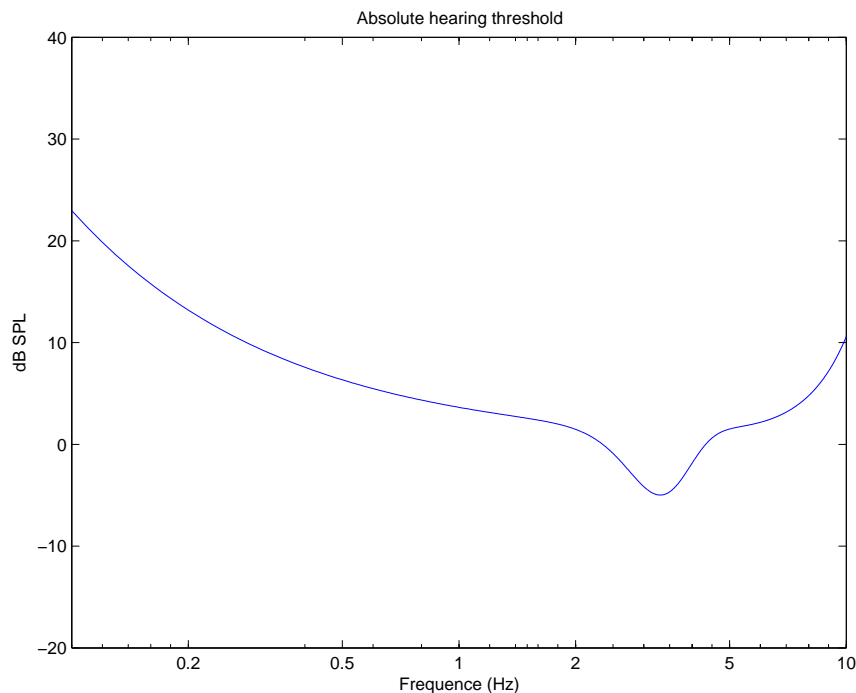
$$N' = 0.08 \left(\frac{E_{TQ}}{E_0} \right)^{0.23} \left[\left(0.5 + \frac{E}{2E_{TQ}} \right)^{0.23} - 1 \right]$$

Her er N' specifik-hørestyrke beregnet i enheden *Sone/Hz*. E_{TQ} er tærskelværdien for intensiteten af en netop hørbar tone for den specifikke frekvens. E_0 er intensiteten som svarer til $10^{-12}W/m^2$ der omtrent svarer til høretærskelen for en tone på 1kHz. For at beregne *hørestyrken* med Zwickers model skal styrken ved en specifik frekvens af et netop hørbart signal, kendes. En sådan høretærskel afhænger i høj grad af forsøgspersonen. Jeg har dog valgt kun at lave lytteforsøg på normalthørende personer og har derfor valgt et standardaudiogram for en normalthørende person. Standardaudiogrammet er estimeret udfra en lang række lytteforsøg, og gennemsnitsværdien af disse målinger beskriver standardaudiogrammet. Jeg har valgt at bruge en ligning fra *Algorithm for extraction of pitch and pitch salience from complex tonal signals*[21], som beskriver et standardaudiogram. På figur 2.8 ses kurven for dette målt i SPL frit felt.

Ved hjælp af ovenstående standardaudiogram og de målinger som er foretaget på mit lydssystem (se figur 3.1) har jeg konverteret signalernes spektrum fra SPL til hørestyrke. Denne konvertering kan opskrives på følgende måde

$$X[i, k]_{Loud} = 0.08 \left(\frac{AHL(\frac{2k}{F_s})}{AHL(1000)} \right)^{0.23} \left[\left(0.5 + \frac{X_{SPL}[i, k]}{2AHL(1000)} \right)^{0.23} \right]$$

Hvor $X_{SPL}[i, k]$ er det oprindelige spektrum i SPL og $X[i, k]_{Loud}$ er spektrummet målt i enheden *Sone*. Til konverteringen bruges funktionen $AHL(\frac{2k}{F_s})$ der angiver en normalthørenes høretærskel for en given frekvens k .



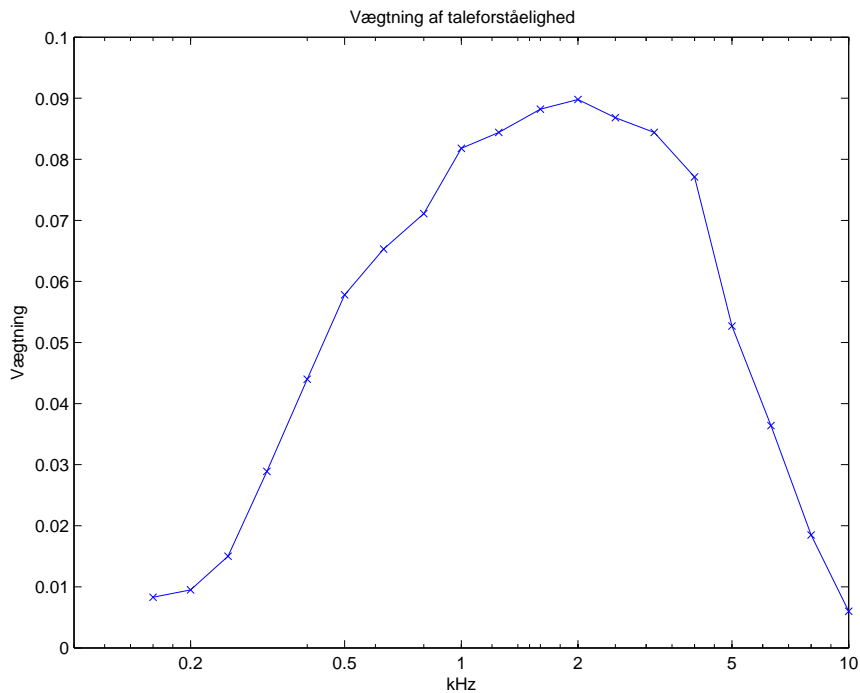
Figur 2.8: Skitse af standardaudiogrammet for en normalt-hørende

2.2.5 Beregning af taleforståelighed

Ved hjælp af ovenstående beregninger har jeg transformeret tale og støjsignalet til et *loudness*-spektrum. For at beregne taleforståeligheden før og efter SRA har jeg forsøgt at beregne de dele af talen som ikke er maskeret af støjen. Ved at trække støjens spektrum fra talens spektrum opnås et slags *hørbar tale*-spektrum $TF[i, k]$ der udtrykker hvilke dele af talen som kan høres.

$$TF[i, k] = \begin{cases} 0 & \text{for } X_{Speech}[i, k] < X_{Noise}[i, k] \\ X_{Speech}[i, k] - X_{Noise}[i, k] & \text{else} \end{cases}$$

De steder hvor støjen er kraftigere end talen sættes til 0, da talen her er maskeret af støjen. For at få dette spektrum vægtet til et enkelt tal har jeg anvendt forskellige metoder. Mit mål er at den beregnede værdi for taleforståelighed stemmer overens med resultatet fra mine lytteforsøg[14]. Fra tidligere undersøgelser er det vist at ikke alle frekvenser er lige vigtige for taleforståeligheden. Derfor har jeg valgt at beregne en vægtet middelværdi for hver sekvens efter opgivne værdier fra ANSI[17]. Vægtningen af frekvenserne kan ses på figur 2.9.



Figur 2.9: Vægtning af taleforståelighed

På denne måde opnås en værdi for hver sekvens. Den vægtet middelværdi for hver sekvens beregnes på følgende måde

$$TF_{sekvens}[i] = \frac{1}{N_k} \sum_{k=1}^{N_k} FV\left(\frac{2k}{F_s}\right) TF[i, k]$$

Hvor N_k er antallet af frekvenser i det *hørbar tale*-spektrum og $FV\left(\frac{2k}{F_s}\right)$ er betydningen af taleforståelighed ved frekvensen $\frac{2k}{F_s}$.

Da det ikke er alle sekvenser af det *hørbare tale*-spektrum som er lige vigtige for taleforståeligheden har jeg valgt at vægte sekvenserne efter det oprindelige tale-spektrum $X_s[i, k]$. Det oprindelige tale-spektrum er først normaliseret.

$$X_{norm}[i, k] = \frac{X_s[i, k]}{\sum_{i=1}^{N_i} \sum_{k=1}^{N_k} X_s[i, k]}$$

Herefter er der beregnet en vægtning $SV[i]$ af hver sekvens på følgende måde.

$$SV[i] = \sqrt{\frac{1}{N_k} \sum_{k=1}^{N_k} X_{norm}^2[i, k]}$$

Vægtningen $SV[i]$ afhænger af niveauet på det oprindelige tale-spektrums. På denne måde bliver de sekvenser som indeholder tale (høj RMS-værdi) vægtet højere end dem der ikke indeholder tale. Beregningen af taleforståelighed TF_{SRA} for en signal før eller efter SRA beregnes på følgende måde

$$TF_{SRA} = \frac{1}{N_i} \sum_{i=1}^{N_i} SV(i) TF_{sekvens}[i]$$

For at beregne om en given SRA giver en forbedring eller forværring af taleforståeligheden har jeg beregnet en værdi for taleforståeligheden før og efter SRA. Herefter tages forskellen på taleforståeligheden før og efter SRA.

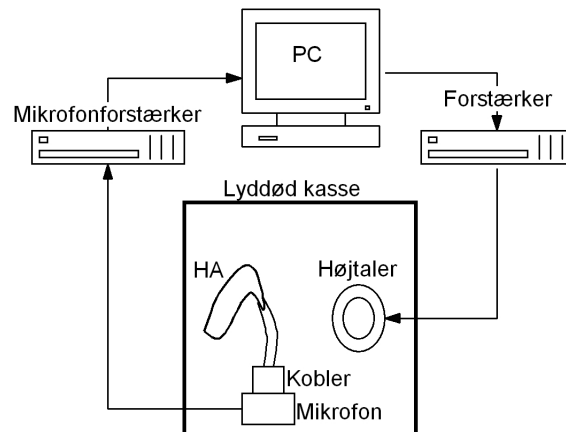
$$TF_{Samlet} = TF_{SRA_{fr}} - TF_{SRA_{efter}}$$

Hvis dette tal er negativ betyder det en forværring i taleforståeligheden. Til at vurdere om ovenstående beregning af taleforståeligheden stemmer overens med målingerne fra mit lyttforsøg, har jeg brugt modellen på de samme SRA som dem fra lyttforsøget. På figur B.1(a) ses den beregnet taleforståelighed for 5 sætninger. Signalerne er de samme som dem jeg har brugt i mit forrige lyttforsøg. På figuren ses at der er en fin overensstemmelse mellem resultaterne fra lyttforsøget og min model. Taleforståeligheden for SRA ligger korrekt under taleforståeligheden uden nogen SRA, og desuden giver den kraftige SRA (*enhanced*) en dårligere taleforståelighed end *soft*. Dette er i overensstemmelse med resultaterne fra mit tidligere lyttforsøg. Jeg har derfor valgt at bruge denne metode til beregning af taleforståeligheden på de forskellige SRA.

Kapitel 3

Måling af SRA

Jeg har valgt at teste min metode med SRA fra forskellige digitale høreapparater. Da jeg ikke har de algoritmer som høreapparaterne bruger, har jeg lavet en opstilling hvor det fysiske høreapparat indgår. Forsøgsopstillingen bruges til at måle output fra høreapparatet når input er kendt. På figur 3.1 ses den opstilling som jeg har brugt til mine målinger. Fra Computerens



Figur 3.1: Opstilling til måling af SRA i høreapparat

lydkort afspilles de lydsekvenser som man ønsker at måle på. Lydsekvensen kommer herefter igennem en effektforstærker som er tilsluttet en højttalerenhed i den lyddøde boks. Med en lydtryksmåler er lydtrykket målt til at ligge mellem 60 og 70 dBC SPL når lydsekvenserne afspilles.

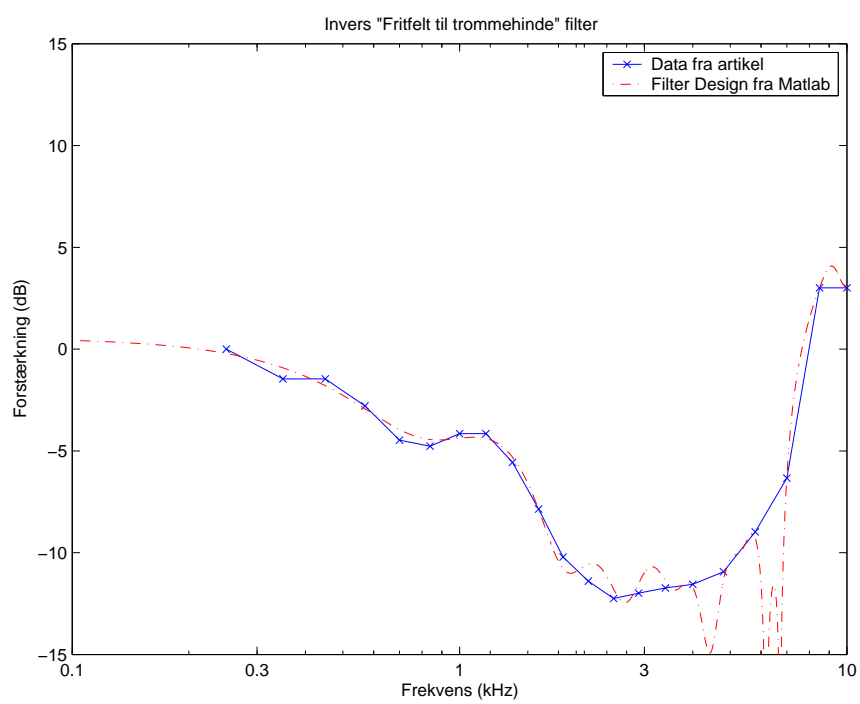
Dette lydtryk er målt ved HA mikrofonindgang. Da HA er af typen BTE (Behind the ear) er lydtrykket ikke helt det samme som lydtrykket ved øre-

indgangen. Jeg har valgt at se bort fra denne forskel. Højtalerenheden har en meget flad frekvensgang i frekvensområdet 100 Hz til 10 kHz. Den lyd-dødeboks (model TBS 25 fra Interacoustics) dæmper støj fra omgivelserne samtidig med at der fra 100Hz til 10kHz er næsten lyddødt. Inde i boksen sidder det høreapparat som skal testes. Høreapparaterne er alle blevet tilpasset med deres tilhørende fitting-program. Efter at høreapparatet er blevet programmeret er det tilsluttet en G.R.A.S. kobler. En kobler efterligner øregangen ind til trommehinden.

Efter kobleren sidder en mikrofon som måler lydtrykket. Dette svarer til lydtrykket ved trommehinden. Mikrofonen er tilsluttet en mikrofonforstærker som er sat til linear forstærkning og som sender signalet videre til lydkortets signalindgang. Med denne opstilling er det altså muligt at afspille lyd til høreapparatet og optage output.

Da optagelserne er foretaget med en kobler svarer det til at signalet har været igennem øregangen. Derfor lyder det anderledes hvis man afspiller det i et fritfelt (f. eks åbne høretelefoner). Dette skyldes at signalet allerede har været igennem en kobler/øregang. For at tage højde for dette har jeg brugt et invers *fritfelt til øregangs* filter. Jeg har brugt de data som er opgivet i artiklen *Transfer Function and Correction Factors Used in Hearing Aid Evaluation and Research*[5]. Her har man for 100 forsøgspersoner estimeret en filter-funktion af fritfelt til øregang. På figur 3.2 ses det inverse filter og de estimeret data taget fra artiklen.

Jeg har valgt at estimerer et filter ud fra de målte data med Matlab® -funktionen *Firls*. Denne funktion estimerer et FIR filter v.h.a. mindste kvadraters metode. Efter at optagelserne er foretaget bliver det filtreret med mit invers *fritfelt til øregangs* filter, for på denne måde at modvirke den transformation der sker når der måles med en kobler. På figur 3.2 ses at filteret afviger fra målepunkterne ved 4.5kHz og omkring 6kHz. Dette kan skyldes at mit FIR filter har for høj orden.



Figur 3.2: Fritfelt til øregangs-filter

3.1 Valg af høreapparat

Jeg har brugt følgende høreapparater til at teste modellen på

- Senso Diva SD-9M , *Widex*[®]
- Canta7 BTE , *GN ReSound*
- Natura[™] 2SE , *Sonic innovations*

Jeg har desuden brugt følgende Matlab[®] -modeller af SRA

- SRA i Senso Diva SD-9M
- *Truncated Singular Value Decomposition* (TSVD)¹

Matlab[®] -modellerne har den store fordel at et input altid giver det samme output, man er på denne måde fri for den støj der kommer fra måleudstyr og HA. TSVD-algoritmen er ikke en som er anvendt i HA da den ikke har en særlig god lyd kvalitet. Matlab[®] -modellen af Senso Diva SD-9M kan ikke direkte sammenlignes med støjreduktionen i HA da der også er mange andre funktioner der påvirker signalet i HA.

Høreapparaterne har mange forskellige indstillingsmuligheder. Det er derfor ikke muligt at sætte de forskellige høreapparater op efter en standard metode. Jeg har dog forsøgt at tilpasse høreapparaterne efter følgende principper. En del af høreapparaterne har flere mikrofoner og derved mulighed for at lave en form for retningsbestemt støjreduktion. Hvis dette er tilfældet har jeg sat høreapparatet til ikke at bruge retningsbestemt støjreduktion. Støjreduktioner som sidder i HA har alle mulighed for at indstille hvor aggressivt SRA skal være. Jeg har for alle SRA brugt 3 forskellige indstillinger, nemlig *soft*, *normal* og *enhanced*. De 3 HA kan desuden slå støjreduktionen fra og Senso Diva SD-9M kan sættes i en såkaldt *linear mode* hvilket betyder at støjreduktionen og kompressoren er slået fra.

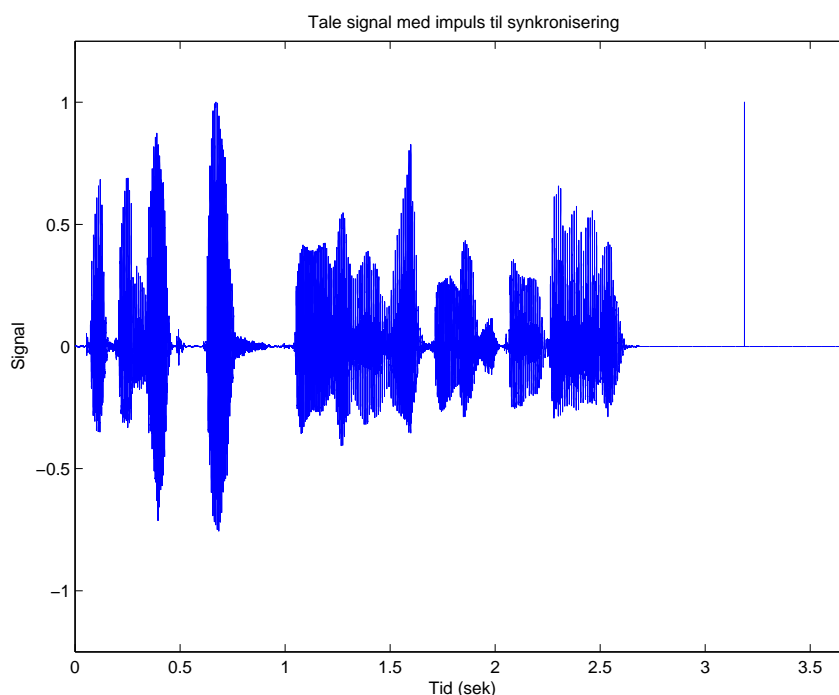
Jeg har valgt at bruge et 60dB fladt audiogram, da dette betyder at høreapparatet skal forstærke inputsignalet og måske derved øge støjreduktionens effekt. En person med et sådan høretab har en høretærskel som ligger konstant 60 dB SPL over høretærsklen for en normalthørende. En person med et sådan høretab har en moderat til middelsvær hørenedsættelse

Alle tilpasningerne er lavet i programmet *Noah[™] 3* hvor høreapparaternes tilpasningsprogrammer er installeret.

¹Denne Matlab[®] -model er blevet udviklet og implementeret i et projekt fra DTU

3.2 Synkronisering af optagelser

Da modellen bruger 2 optagelser til separering af tale og støj er det vigtigt at optagelserne er synkroniseret. For at synkronisere signalerne har jeg indsat et impulssignal i det afspillet signal. Impulssignalet består af $\frac{1}{2}$ sekund stilhed efterfulgt af een sampleværdi som er sat til det maksimale og herefter $\frac{1}{2}$ sekund stilhed igen. På figur 3.3 ses et talesignal hvor impulssignalet er indsat tilsidst i signalet. For at undersøge om optagelsen og afspilningen



Figur 3.3: Talesignal med impuls til synkronisering

foregår med den korrekte frekvens har jeg afspillet et testsignal som består af 5 impulser. Den første impuls bruges til at synkroniser optagelsen og afspilningen efter. Når den første impuls er synkroniseret skal de resterende 4 impulser også være synkroniseret. Hvis dette ikke er tilfældet er afspilning og optagelsens samplefrekvens forskellig. Jeg har valgt at afspille og optage med en samplefrekvens på 44.1kHz. Mit første test lydkort var et *SoundBlaster® Live®* med dette lydkort havde optagelsen 8-9 samples for meget pr. sekund. For at forbedre dette udskiftet jeg lydkortet med et *SoundBlaster® Audiogy® 2 NX* her var forskellen på inputsignalet og optagelsen 200 samples pr. sekund. Med det nye lydkort var fejlen altså endnu større. Herefter udførte jeg samme test med et *ECHO® Gina 24/96*. Med dette lydkort var der ikke

mistet nogen samples. *ECHO*[®] Gina 24/96 var altså klart det bedste lydkort i denne test, og jeg har derfor valgt at bruge dette til alle mine optagelser.

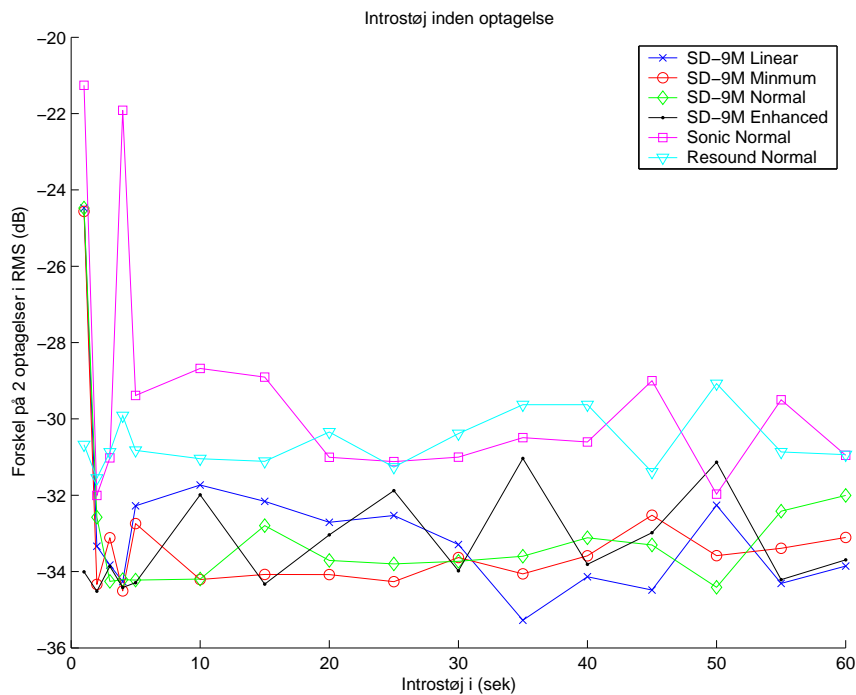
3.3 Forbedring af optagelser

Efter at have lyttet til det separerede tale og støj, var det klart at separeringen ikke var god nok. Ved aflytning af den separerede støj efter den havde været igennem høreapparatet indeholdt signalet en del af talesignalet. Det separerede talesignal indeholdt ligeledes en del af støjsignalet. Jeg har derfor undersøgt hvor denne fejl stammer fra og hvordan den kan minimeres.

For at undersøge hvor stor fejlen er har jeg lavet 2 optagelser $y_1(n)$ og $y_2(n)$ med det samme inputsignal $x(n)$ og synkroniseret de 2 optagelser med hinanden v.h.a. et impulssignal. Jeg har herefter taget forskellen på de 2 signaler $\delta(n)$ og undersøgt dette signal. For at få et tal for hvor ens de 2 optagelser ($y_1(n)$ og $y_2(n)$) er på hinanden tages RMS-værdien på $\delta(n)$. Hvis RMS-værdien på $\delta(n)$ er tæt på 0 er de 2 optagelser næsten ens.

En fejlkilde ved en sådan metode er at høreapparaterne i høj grad afhænger af tidligere input. Høreapparaterne bruger det tidligere input til at justere på bl.a. kompressoren og støjreduktionen. Dette betyder at når der udføres 2 optagelser lige efter hinanden, vil høreapparaterne opføre sig anderledes ved den 2. optagelse da det tidligere input her er den første optagelse. Derfor er det vigtigt at der inden hver optagelse afspilles en sekvens med f. eks støj, så høreapparaterne har det samme tidligere input inden selve optagelsen begynder. For at undersøge hvor meget støj der skal afspilles før en optagelse har jeg udført en række forsøg med varierende længde støj inden optagelsen. På figur 3.4 er RMS-værdien for $\delta(n)$ som funktion af hvor lang tid der er afspillet støj inden optagelsen. Testen er lavet for de 3 HA sat i forskellige tilstande. På figuren ses at der ved en introstøj på under 5 sek. er en stor forskel på de 2 optagelser. Der skal altså minimum være 5 sek. mellem de 2 optagelser.

Da de fleste SRA kræver en hvis *indkørselstid* før de virker optimalt har jeg testet hvor lang denne indkørselstid skal være. Dette er gjort ved at afspille et signal med ren hvid støj i 1 minut. Outputtet fra HA har jeg optaget og delt op i sekvenser på 1 sekund. For hver af disse sekvenser har jeg beregnet RMS-værdien. På figur 3.5 har jeg plottet RMS-værdierne for outputtet af HA som funktion af tiden. På figuren ses tydeligt at der er en indkørselstid før HA dæmper støjen. Den langsomme dæmpning skyldes støjreduktionens indkørselstid, og man kan på figuren se at den kan være helt op til 30 sek.

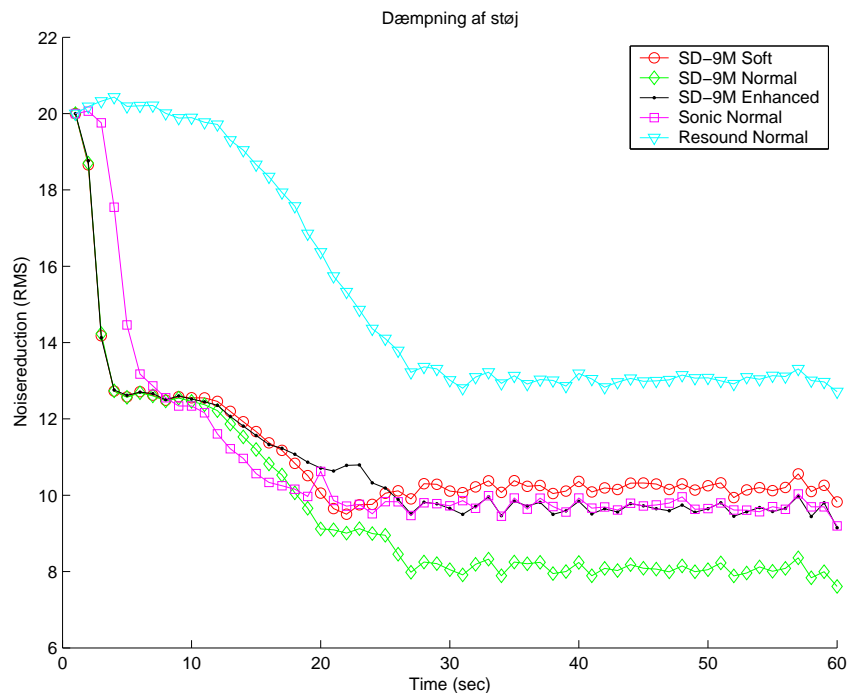


Figur 3.4: Optagelser med samme inputsignal

For at få SRA til at virke optimalt har jeg derfor valgt at afspille 30 sek. støj inden hver optagelse.

3.4 Måling af ulineariteten på SRA

Da behandlingen af signalerne i HA er ulinear har jeg forsøgt at bestemme den fejl der begås ved at antage at systemet er lineart. Ulineariteten på HA skyldes ikke alene SRA men gælder hele behandlingen af signalet. For hver SRA har jeg målt ulineariteten for 5 testsignaler. De 5 testsignaler er taget fra lytteforsøget og indeholder hver en nonsenssætning (se kapitel 4) hvor der er pålagt hvid støj. På disse optagelser har jeg beregnet SDR ved hjælp af ligning 2.4. Til Hilberttransformationen af signalerne har jeg brugt Matlab[®]'s indbygget *Hilbert*²-funktion. Jeg har delt min test op i 2 typer. Den første type består af de SRA som er simuleret i Matlab[®], nemlig Matlab[®]-modellen af Senso Diva SD-9M og SRA der bruger TSVD. Målingerne på disse SRA kan ses på figur A.1 i bilag A. Den anden type SRA som der er målt på er dem der sidder i HA, Senso Diva SD-9M, Canta7 BTE og Natura[™] 2SE.

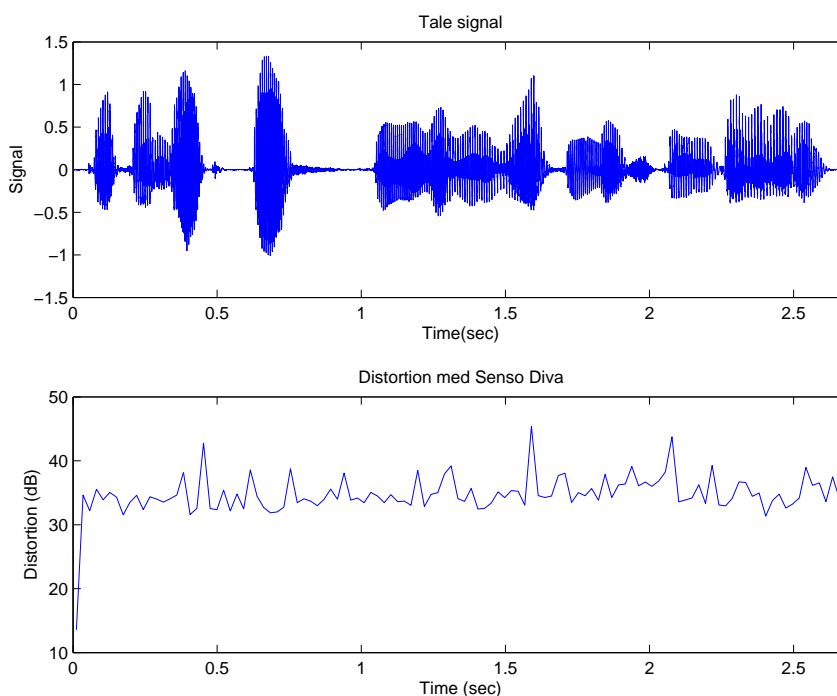


Figur 3.5: Indkørselstid for forskellige SRA

Resultatet af denne test kan ses på figur A.2 i bilag A. For at få et enkelt tal for ulineariteten har jeg for hver SRA beregnet en vægtet middelværdi af SDR. Jeg har valgt at vægte efter betydningen af taleforståelighed, så de frekvenser der har størst betydning for taleforståeligheden vægtes højest. På figur 2.9 ses vægtningen af de forskellige frekvenser. På figur A.1 i bilag A ses tydeligt at ulineariteten for TSVD-algoritmen er meget større end Matlab[®]-modellen af Senso Diva SD-9M. Hvis der lyttes på den separerede tale og støj kan der for TSVD tydeligt høres at talen og støjen ikke er separeret særligt godt. Separeringen af tale og støj for Senso Diva SD-9M er som forventet fin og der kan ikke høres tale i det separeret støjsignal. På figuren ses det ligeledes at der ikke er særlig stor forskel på om støjreduktionen er *soft*, *standard* eller *enhanced*. Dog er det som forventet den kraftige støjreduktion *enhanced* der er den mest ulineare af de 3 typer. For HA modellen har jeg ligeledes målt ulineariteten. For HA modellerne er ulineariteten lidt højere end Matlab[®]-modellen af Senso Diva SD-9M, men meget lavere end TSVD. For Senso Diva SD-9M fra *Widex*[®] ses at ulineariteten er størst når HA har slået SRA fra. Dette er lidt uventet, da SRA ikke har nogen indflydelse på signalet. Den målte SDR må derfor skyldes andre funktioner i HA som f. eks. kompressoren. I *linear*-mode er både kompressoren og SRA slået fra, og

den målte ulineariteten i denne mode skyldes derfor ikke disse funktioner. Den målte SDR i *linear*-mode er da også den højeste for dette HA. For de 2 andre HA fra *GN ReSound* og *Sonic innovations* er den målte ulinearitet lidt højere end for Senso Diva SD-9M . I tabel 3.1 har jeg for alle SRA beregnet en talværdi for deres SDR.

Jeg har desuden beregnet ulineariteten for hver sekvens af signalet. Størrelsen af disse sekvenser er den samme som ved beregningen af signalets spektrogram (23.2 ms). Dette har jeg gjort for at undersøge om ulineariteten er større i de sekvenser der indeholder tale. Hvis ulineariteten er størst i de områder hvor der er tale kan det have stor betydning for taleforståeligheden. Jeg har for hver sekvens beregnet SDR på samme måde som i forrige afsnit. På figur 3.6 er den beregnede ulinearitet (SDR) for Senso Diva SD-9M . SRA er i dette forsøg sat til *standard* og inputsignalet er den første sætning fra mit lytteforsøg. Den øverste del af figuren er det oprindelige talesignal. Den nederste



Figur 3.6: Beregnet ulinearitet på sekvenserne

del er den beregnet SDR for signalet. På figuren er SDR næsten konstant og den ser ikke ud til at være større for de sekvenser der indeholder tale.

3.5 Måling af taleforståelighed

Til beregning af taleforståeligheden med min model har jeg brugt de samme sætninger som dem fra lyttetesten. Da modellen beregner en værdi for taleforståeligheden for hver enkelt sætning, har jeg noteret middelværdien af de beregnede værdier for taleforståeligheden. Jeg har igen valgt at dele SRA op i 2 typer, de SRA som er implementeret i Matlab[®] og dem som er implementeret i et HA.

SRA	Beregnet taleforståelighed	Målt SDR
Ingen SRA	0	91.5 dB
Matlab-model af Senso Diva SD-9M , <i>soft</i>	-0.49	48.0 dB
Matlab-model af Senso Diva SD-9M , <i>standard</i>	-0.88	42.9 dB
Matlab-model af Senso Diva SD-9M , <i>enhanced</i>	-0.86	41.5 dB
Matlab-model af TSVD, <i>soft</i>	-0.05	2.5 dB
Matlab-model af TSVD, <i>standard</i>	-1.20	2.0 dB
Matlab-model af TSVD, <i>enhanced</i>	1.70	1.7 dB

Tabel 3.1: Beregnet taleforståelighed og ulinearitet på Matlab-modeller

I tabel 3.1 ses den beregnede taleforståelighed og SDR for de 2 SRA der er implementeret i Matlab[®]. Den beregnede taleforståelighed for Matlab[®]-modellen af Senso Diva SD-9M er i overensstemmelse med de resultater jeg fik fra mit forrige lytteforsøg [14]. På figuren ses at de 3 indstillinger *soft*, *normal* og *enhanced* alle ligger under den beregnede taleforståelighed for ingen SRA. Derudover er forskellen imellem de 3 indstillinger også i overensstemmelse med resultaterne fra mit forrige lytteprojekt. Den beregnede taleforståelighed for TSVD-algoritmen viser en stor forbedring af taleforståeligheden. Dette vil sandsynligvis ikke kunne vises ved et lytteforsøg. Grunden til at TSVD-algoritmen viser en forbedring af taleforståeligheden må findes i ulineariteten. TSVD-algoritmen er meget ulinear hvilket bevirker at den behandlede tale og støj ikke er godt separeret. Modellen beregner derfor taleforståeligheden efter SRA på nogle dårligt separerede signaler. Værdien af taleforståeligheden efter SRA bliver derfor forkert.

Den beregnede taleforståelighed for de 3 HA fra *Widex[®]*, *GN ReSound* og *Sonic innovations* kan ses i tabel 3.2. Indstillingen ”uden SRA” er målinger hvor SRA er slået fra i HA.

Den beregnede taleforståelighed er meget forskellig for de 3 HA. Der er desuden heller ikke en generel sammenhæng på forbedringen/forværringen og

SRA	Beregnet taleforståelighed	Målt SDR
<i>Widex</i> [®] , Senso Diva SD-9M , Linear Mode	Ej beregnet	40.0 dB
<i>Widex</i> [®] , Senso Diva SD-9M , uden SRA	-0.20	36.2 dB
<i>Widex</i> [®] , Senso Diva SD-9M , <i>soft</i>	-1.00	39.9 dB
<i>Widex</i> [®] , Senso Diva SD-9M , <i>standard</i>	-1.50	37.8 dB
<i>Widex</i> [®] , Senso Diva SD-9M , <i>enhanced</i>	-1.80	37.6 dB
<i>GN ReSound</i> , Canta7 BTE , uden SRA	-2.70	27.9 dB
<i>GN ReSound</i> , Canta7 BTE , <i>soft</i>	-3.00	28.2 dB
<i>GN ReSound</i> , Canta7 BTE , <i>standard</i>	-2.90	25.6 dB
<i>GN ReSound</i> , Canta7 BTE , <i>enhanced</i>	-3.50	27.7 dB
<i>Sonic innovations</i> , Natura [™] 2SE , uden SRA	-3.40	27.9 dB
<i>Sonic innovations</i> , Natura [™] 2SE , <i>soft</i>	-3.10	29.0 dB
<i>Sonic innovations</i> , Natura [™] 2SE , <i>standard</i>	-1.10	28.8 dB
<i>Sonic innovations</i> , Natura [™] 2SE , <i>enhanced</i>	-1.50	26.1 dB

Tabel 3.2: Beregnet taleforståelighed og ulinearitet på HA

indstillingen på SRA. Generelt kan det dog ses at ingen af SRA giver en beregnet taleforståelighed som er bedre end ingen SRA. For Senso Diva SD-9M bliver den beregnede taleforståelighed værre når SRA bliver kraftigere. Denne tendens er der også lidt for Matlab[®] -modellen af Senso Diva SD-9M . For *GN ReSound* giver modellen en lavere værdi for taleforståeligheden med SRA end uden SRA, hvorimod det for *Sonic innovations* er omvendt.

Kapitel 4

Lyttetest

Til at udføre min lyttetest, har jeg valgt at bruge såkaldte *nonsenssætninger* fra projektet *Properties of Intelligibility Test Materials* [2]. Disse består af en bære-sætning hvor der er indsat et nonsensord, f. eks *bjald*. Nonsensordet er altid indsat som det 4. ord i sætningen. Et eksempel på en nonsenssætning kunne være: ”Han kunne dog *bjald* ikke finde bilen”. Her er bære-sætningen ”Han kunne dog ikke finde bilen” og nonsensordet er *bjald*. Forsøgspersonen skal prøve at finde og gentage nonsensordet sådan som han/hun har opfattet det. Da der ikke er nogen sammenhæng mellem nonsensordet og bære-sætningen kan forsøgspersonen ikke gætte nonsensordet ud fra resten af sætningen. Dette betyder at det selv ved høje signal støjforhold er svært at gentage nonsensordet. Da støjreduktionerne i HA kræver et forholdsvist højt SNR før de giver en forbedring i signalet, er det en fordel at bruge nonsenssætninger. For at kunne bruge nonsenssætningerne i et lyttforsøg er det vigtigt at lyttetesten giver en score mellem 10% og 90 %. I mit forrige projekt [14] lå scoren mellem 45% og 85%. Da jeg i denne lyttetest har valgt samme type støj og SNR som i mit forrige projekt vil jeg forvente at scoren ved lyttetesten er i samme størrelsesorden.

Lyttetesten er delt op i 4 dele. Den første del består af en introduktion så forsøgspersonerne vænner sig til støjen og talen. Introduktionsdelen består af 5 testsætninger hvor der ikke er brugt nogen SRA. Efter hver testsætning er afspillet skal forsøgspersonen prøve at gentage nonsensordet. På denne måde testes om forsøgspersonen har forstået opgaven korrekt og om nonsensordet bliver udtalt tydeligt af forsøgspersonen. Under introduktionsdelen kan forsøgspersonen ligeledes stille spørgsmål til testen. Introduktionsdelen er til for at minimere en eventuel indlæringsseffekt som testen måtte have.

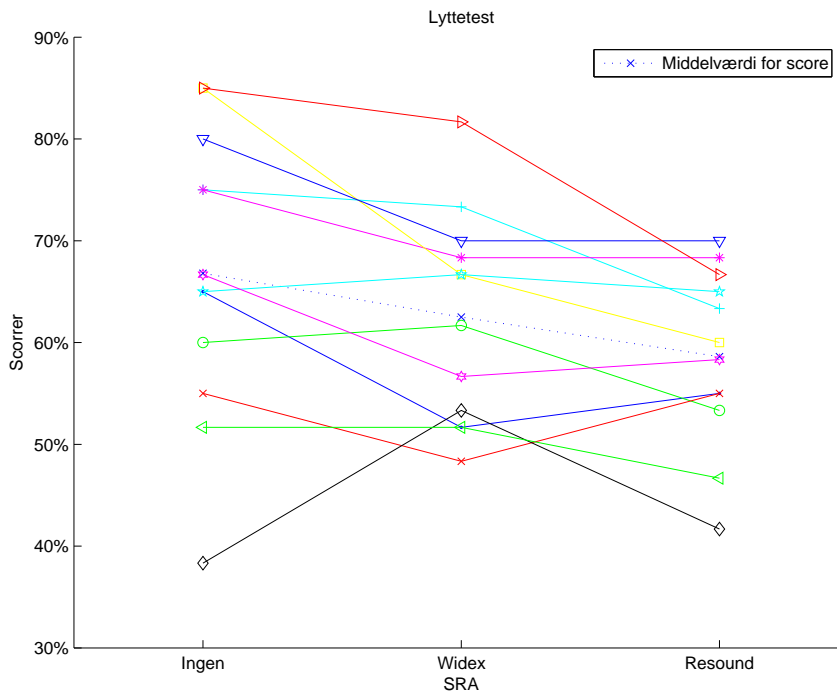
Efter introduktionsdelen kommer der 3 deltest som hver består af 20 non-

senssætninger. De 3 deltest er blevet behandlet af hver deres SRA. De SRA som jeg har valgt at bruge er Senso Diva SD-9M fra *Widex*[®] med *standard* SRA, Canta7 BTE fra *GN ReSound* med *enhanced* SRA og ingen SRA. Den sidste er altså ikke en SRA, men svarer til det signal som HA får som input. Jeg har valgt at teste de 2 SRA fra *Widex*[®] og *GN ReSound* da der ifølge modellen er forskel på taleforståeligheden. I tabel 3.2 ses at den beregnet taleforståelighed for *Widex*[®] og *GN ReSound* er hhv. -1.50 og -3.50. Der bør derfor være en stor sandsynlighed for at jeg med lyttetesten kan finde en signifikant forskel mellem de 2 SRA.

Rækkefølgen for sætningerne er den samme for hver lyttetest, mens den SRA som der er brugt på hver af de 3 deltest er forskellig. Dette er gjort for at minimere en eventuel forskel på sværhedsgraden af nonsenssætningerne i hver deltest. For at gøre lyttetesten mere effektiv har jeg delt hvert nonsensord op i 3 fonemer. Scoren for et nonsensord afgøres ved hvor mange fonemer forsøgspersonen har udtalt korrekt. Et nonsensord som *bjald* har som initial fonem *bj*, medialt fonem *a* og som finalt fonem *ld*. Forsøgspersonen får 1 point for hvert fonem der udtales korrekt. Det er på denne måde muligt at score mellem 0 og 3 point på hvert nonsensord. Kan forsøgspersonen ikke finde nonsensordet gives der 0 point for dette ord. Testskemaet som jeg har brugt til at notere scoren på kan ses i bilag C. Det kan i øvrigt bemærkes at nogle af nonsensordene er rigtige ord som f. eks *tak*. Nonsensordene giver dog ingen mening i den bæresætning hvor de er indsat.

4.1 Resultat fra lyttetest

Jeg har kørt min lyttetest på 12 forskellige forsøgspersoner. De fleste af forsøgspersonerne er fra samme afdeling i *Widex*[®], som hvor forsøgene er udført. Disse forsøgspersoner har alle før været med i lignende lyttforsøg. De har derfor en svag tendens til at have en lidt højere score end de andre forsøgspersoner der er brugt. Fordelen ved at bruge forsøgspersoner med denne erfaring er at variansen på målingerne ofte er mindre, da de alle har den samme træning i lyttetest. På figur 4.1 ses den score som forsøgspersonerne har fået. Scoren er angivet i % af den maksimale score som er 60 point for hver deltest. På figuren har jeg delt resultaterne op efter hvilken SRA der er brugt i deltesten. På figuren ses at der er stor forskel på hvor meget forsøgspersonerne scorer ved hver SRA. Der er dog en tendens til at linjerne er parallelle hvilket tyder på at scoren for de 3 deltest er personafhængig. Der er også



Figur 4.1: Resultat fra lyttetest, opdelt efter SRA

en af forsøgspersonerne som har en markant lavere score ved ingen SRA end ved de 2 SRA. Forsøgspersonen med denne måling har ikke tidligere været med i nogle lytteforsøg og dette er nok også årsagen til at scoren generelt ligger under middelværdien. Scoren for hver deltest ligger dog alle mellem 85% og 35%. Dette betyder at "sværhedsgraden" ved testen er perfekt da ingen af forsøgspersonerne har ligget i yderenderne af skalaen. Havde SNR været lavere ville SRA ikke havde ændret så meget ved signalet, og havde SNR været højere, ville nogle af forsøgspersonernes score ligge tæt på 100%. På plottet kan desuden ses en generel tendens til at scoren er højest for deltesten uden SRA, og derefter kommer *Widex*[®] SRA og den laveste score har *GN ReSound*. Dette er i fin overensstemmelse med målingerne fra min Matlab[®]-model. For at bekræfte om denne tendens er signifikant, har jeg lavet en række statistiske test på mine målinger.

Kapitel 5

Statistiske målinger på resultatet fra lyttetest

Til at undersøge om der er signifikant forskel på målingerne bruges statistiske test. Jeg har valgt at forkaste en 0-hypotese hvis teststørrelsen ligger i det kritiske område for et 5%-signifikantsniveau. Det vil sige at der er 5% chance for at jeg har forkastet en sand 0-hypotese. Ved de test hvor jeg ikke kan forkaste min 0-hypotese har jeg beregnet testens styrke. Styrken af en given test fortæller hvor stor sandsynligheden er for at en acceptere 0-hypotese er sand. Styrken af en test giver således kun mening hvis jeg vælger at acceptere min 0-hypotesen. Til beregning af styrken på de forskellige test har jeg valgt programmet G*Power [8]. Styrken af testen beregnes normalt før lyttetesten udføres. Normalt ønskes en styrke på minimum 70% hvis 0-hypotesen accepteres, og man kan ud fra dette beregne antallet af forsøgspersoner der som minimum skal bruges for at opnå en sådan styrke.

5.1 Test for normalfordeling

Mange af de statistiske test som jeg har brugt, antager at observationerne er normalfordelt. For at undersøge denne antagelse har jeg brugt en χ^2 -test. Ved denne test er min 0-hypotese at observationerne stammer fra samme fordeling. Ved antagelsen om at observationerne er normalfordelt har jeg estimeret normalfordelingens middelværdi og varians. Jeg har valgt at dele observationerne op efter hvilken af de 3 deltest og efter hvilken SRA de har være behandlet med. En observation indgår således i 2 grupper nemlig efter

Gruppe	Deltest 1	Deltest 2	Deltest 3	Ingen SRA	<i>Widex</i> [®]	<i>GN ReSound</i>
0%, 25%	4	3	4	3	4	2
25%, 50%	1	3	2	4	2	4
50%, 75%	3	4	3	2	3	2
75%, 100%	4	2	3	3	3	4
Teststørrelse	2	0.67	0.67	0.67	0.67	1.33
Signifikans-niveau	36.79%	71.65%	71.65%	71.65%	71.65%	51.34%
Styrke	22.55%	10.36%	10.36%	10.36%	10.36%	16.27%

Tabel 5.1: Opdeling af observationer i 4 grupper

deltest (deltest 1, 2 og 3) og SRA (Ingen, *Widex*[®] og *GN ReSound*). Dette har jeg gjort da jeg både ønsker at undersøge om observationerne sorteret efter deltest og SRA kan antages at være normalfordelt. Jeg har valgt at dele observationerne ind i 4 grupper. Dette er gjort ud fra en tommelfingerregel der siger at antallet af grupper cirka skal være $2n^{2/5}$ hvor n er antallet af observationer. Da jeg kun har 12 observationer fra hver deltest, har jeg således valgt 4 grupper. For at få den størst mulige styrke i testen er det forventede antal for hver gruppe $\frac{1}{4}$ af antallet af observationer. Jeg har for den enkelte deltest beregnet grænseværdierne for de 4 grupper ud fra 4 fraktiler i normalfordelingen. Den første gruppe indeholder således de observationer som er mindre end 25%-fraktilen, anden gruppen indeholder observationer som ligger mellem 25% og 50%-fraktilen o.s.v. I tabel 5.1 ses hvorledes de forskellige observationer er fordelt. Da jeg har 12 observationer for hver deltest, og har valgt 4 grupper som hver indeholder $\frac{1}{4}$ af normalfordelingen, er 3 det forventet antal observationer pr. gruppe. Ud fra tabellen beregnes teststørrelsen Z på følgende måde

$$Z = \sum_{i=1}^4 \frac{(X_i - 3)^2}{3} \quad (5.1)$$

Hvor 3 er det forventede antal observationer og X_i er antallet af observationer for den i 'te gruppe. I tabellen ses teststørrelsen for hver deltest og SRA.

For at undersøge min 0-hypotese skal teststørrelserne sammenlignes med fraktilen i en χ^2 -fordeling. Da jeg har opdelt observationerne i 4 grupper, og har estimeret 2 værdier (middelværdi og varians) ud fra observationer, skal χ^2 -fordelingen have $4 - 2 = 2$ frihedsgrader. Jeg har for hver af de 6 teststørrelse undersøgt signifikansniveauet ved χ^2 -testen. Disse værdier ses i nederste række i tabellen. Da ingen af signifikansniveauerne er under 36%, kan det ikke

afvises at observationerne er normalfordelt. Styrken af en χ^2 -test er normalt ikke særlig stor. Jeg har dog ikke kunne finde noget alternativ til denne test. Dette betyder at selv om jeg ikke kan afvise at observationerne er normalfordelt, er det ingen garanti for at de er det. For at beregne styrken for min χ^2 -test har jeg beregnet følgende værdier

$$w = \sqrt{\sum_{i=1}^4 \frac{p_0(i) - p_1(i)}{p_0(i)}}$$

Hvor $p_0(i)$ er sandsynligheden for at en observation tilhøre den i 'te gruppe hvis 0-hypotesen er sand. I mit tilfælde har jeg valgt at hver gruppe skal have samme sandsynlighed derfor er $p_0(i) = 0.25$. Værdierne for $p_1(i)$ er beregnet sandsynligheder ud fra antallet af observationer der tilhører den i 'te gruppe. For at beregne styrken af min χ^2 -test har jeg desuden brugt følgende værdier.

Signifikansniveau	$\alpha = 5\%$
Antal samples	$n = 12$
Frihedsgrader	$Fd = 2$

Med ovenstående værdier indtastet i G*Power har jeg beregnet styrken af hver test, disse værdier kan ses nederst i tabellen 5.1. Som det ses af testen er styrken af χ^2 -testen meget lille, og det er derfor ikke sikkert at min 0-hypotese kan accepteres. For at øge styrken af testen kan antallet af observationer øges. Med programmet G*Power har jeg beregnet at med 150 observationer (og alt andet lige) vil styrken være omkring 70%. Der skal altså markant flere observationer til for at min 0-hypotese med sikkerhed kan accepteres. Jeg vil dog alligevel antage at mine observationer er normalfordelte.

5.2 Test for varianshomogenitet

For at bruge variansanalyse til undersøgelse af mine observationer kræves at hver deltest er normalfordelt og har samme varians. Til undersøgelse af varianshomogenitet har jeg valgt at bruge Bartlett's test. Først opdeles observationerne X efter hvilken af de 3 deltest de tilhører.

$$X_{ij} \quad j = 1, \dots, 12, \quad i = 1, 2, 3$$

med

$$\begin{aligned} E(X_{ij}) &= \mu_i \\ V(X_{ij}) &= \sigma_i^2 \end{aligned}$$

Herefter opstilles min 0-hypotese om varianshomogenitet

$$\begin{aligned} H_0 & : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \\ H_1 & : \exists i, j (\sigma_i^2 \neq \sigma_j^2) \end{aligned}$$

Herefter beregnes teststørrelsen Z på følgende måde

$$Z = f \log_e s^2 - \sum_{i=1}^3 f_i \log_e s_{1i}^2$$

hvor

$$\begin{aligned} f_i & = 12 - 1 \\ f & = \sum_{i=1}^3 (12 - 1) = 33 \\ s_{1i}^2 & = \frac{1}{12 - 1} \sum_{j=1}^{12} (X_{ij} - \bar{X}_i)^2 \\ s^2 & = \frac{1}{f} \sum_{i=1}^3 f_i s_{1i}^2 \end{aligned}$$

Ud fra mine observationer har jeg beregnet teststørrelsen Z til følgende

$$Z = 3.12$$

Ved sammenligning af teststørrelsen med $\chi^2(2)$ accepteres min 0-hypotese på et 21%-signifikantsniveau. Jeg kan altså ikke afvise min 0-hypotese, at deltest 1, 2 og 3 har samme varians.

Jeg har lavet samme test for forsøgspersonerne, for at finde ud af om det kan antages at forsøgspersonerne har samme varians. Testen køre efter samme model som for undersøgelsen af deltesten. Med mine observationer får jeg en teststørrelse Z på følgende

$$Z = 13.69$$

Ved sammenligning af teststørrelsen med $\chi^2(11)$ accepteres min 0-hypotese på et 25%-signifikantsniveau. Jeg kan altså ikke afvise min 0-hypotese, at observationerne fra hver forsøgsperson har samme varians.

Desværre er testens power ikke særlig stor, så der kan altså stadig godt være en forskel i varianserne som testen ikke kan detektere.

5.3 Variansanalyse

For at undersøge om der er en generel forskel på scorens middelværdi for de 3 deltest og for forsøgspersonerne, har jeg valgt at lave en tosidet variansanalyse. For at kunne bruge denne test, kræves følgende om observationerne (X_{ij}).

- Stokastisk uafhængige
- Normalt fordelte
- Samme varians (σ^2)
- Additiv struktur for middelværdierne $E(X_{ij}) = \mu_{ij} = \mu + \alpha_i + \beta_j$

Det antages at observationerne er stokastisk uafhængige da der er forskel på den rækkefølge forsøgspersonerne har fået de 3 deltest. Jeg har i de 2 forrige afsnit undersøgt om punkt 2 og 3 er opfyldt. Jeg har i forrige afsnit ikke kunne afvise at observationerne opfyldte de 2 punkter. Antagelsen om additiv struktur af middelværdierne, kræver at der er flere observationer af hver X_{ij} . Dette vil kræve at forsøgspersonen kørte den samme deltest igennem med den samme SRA, flere gange. Dette er ikke muligt i praksis, da forsøgspersonen da allerede kender nonsensordene i deltesten, og observationerne ville således ikke længere være stokastisk uafhængige.

Den tosidet variansanalyse skal undersøge følgende hypoteser

$$\begin{aligned} H_{01} &: \alpha_1 = \dots = \alpha_3 = 0 \quad \text{mod} \quad H_{11} : \exists i(\alpha_i \neq 0) \\ H_{02} &: \beta_1 = \dots = \beta_{12} = 0 \quad \text{mod} \quad H_{12} : \exists j(\beta_j \neq 0) \end{aligned}$$

Her er α_i og β_j den additive middelværdi fra hhv. deltest i og forsøgsperson nr j . Hvis min H_{01} hypotese kan antages vil jeg antage at der ikke er forskel på sværhedsgraden af de 3 deltest. For at undersøge de 2 hypoteser har jeg opstillet et variansanalysekema for testen som kan ses i tabel 5.2. Til beregning af de forskellige værdier har jeg brugt formlerne fra *En introduktion til statistik*[6].

Ved sammenligning af de beregnede test-størrelser fra tabel 5.2 og opslag i tabel over fraktiler i F-fordelingen fås følgende

$$\begin{aligned} F(2, 22)_{1-0.35} &= 1.10 > 0.90 \\ F(11, 22)_{1-0.001} &= 4.7 < 7.06 \end{aligned}$$

Variation	SAK	f	Test
Mellem Deltest	$SAK_4 = 0.0077$	2	0.90
Mellem Forsøgspersoner	$SAK_3 = 0.3312$	11	7.06
Vekselvirkning	$SAK_2 = 0.0938$	22	
Total	$SAK_0 = 0.4327$	35	

Tabel 5.2: Variansanalysekema

Da 0-hypotese forkastes hvis teststørrelsen er større end værdien i F-fordelingen, forkastes således H_{02} hypotese på et 0.1%-signifikantsniveau. Derimod kan H_{01} ikke forkastes på et 35% signifikantsniveau. Det kan derfor ikke afvises at der er signifikant forskel på sværhedsgraden af de 3 deltest. Der er derimod signifikant forskel på scoren for de 12 forsøgspersoner.

5.4 Sammenligning af støjreduktioner

Jeg har i forrige afsnit ikke kunne vise en forskel i sværhedsgraden på hver deltest. Jeg vil derfor antage at det ikke har nogen signifikant betydning for observationerne om en given SRA er brugt på deltest 1, 2 eller 3. Da der er signifikant forskel på forsøgspersonernes score, har jeg valgt at bruge en parret t-test til at sammenligne de forskellige SRA. I tabel 5.3 ses resultaterne af en parret t-test. Navnene i de 2 første søjler er de 2 SRA der sammenlignes. Den sidste er signifikantsniveauet for om SRA 1 er bedre end SRA 2. Jo mindre signifikantsniveauet er, desto større er sandsynligheden for at SRA 2 er bedre end SRA 1. Fra tabellen ses at der på et 5% signifikantsniveau kan forkastes at SRA 1 er bedre end SRA 2. Dette betyder at jeg ud fra mine observationer må konkludere at *Ingen* har den højeste score og *GN ReSound* har den laveste score. Dette stemmer også overens med resultaterne fra min Matlab[®] -model.

SRA 1	SRA 2	Signifikantsniveau for at 1 er bedre end 2
<i>Widex[®]</i>	<i>Ingen</i>	2.24%
<i>GN ReSound</i>	<i>Widex[®]</i>	3.30%
<i>GN ReSound</i>	<i>Ingen</i>	0.07%

Tabel 5.3: Sammenligning af SRA ved en parret t-test

Kapitel 6

Sammenligning med andre modeller

Det er svært at sammenligne resultaterne af min model med andre modeller. Modellerne har forskellige forudsætninger og resultaterne kan ikke sammenlignes direkte. De 2 modeller som jeg har valgt at bruge er STI og PESQ. Disse 2 modeller bruges til at beregne hhv. taleforståelighed og lyd kvalitet. Jeg har testet disse metoder på Canta7 BTE fra *GN ReSound* med *enhanced* SRA og Senso Diva SD-9M fra *Widex*[®] med *standard* SRA. Dette er de samme SRA som dem jeg brugte i lyttetesten, og indstillingen af HA er også den samme.

Speech Transmission Index

Til måling af STI er programmet ”Speech Transmission Index” [10] brugt. Dette program gør det muligt at måle STI v.h.a. lyd kortet fra Pc’en. Til måling af STI på HA har jeg brugt opstillingen fra figur 3.1.

Programmet afspiller STI-signalet som er et frekvensmoduleret støjsignal. STI-signalet gør i denne test det ud for et talesignal. Det er altså ikke muligt at måle taleforståeligheden for en given sætning.

For alligevel at kunne sammenligne resultaterne fra STI testen med min test, har jeg blandet STI-signalet med et støjsignal. Støjsignalet er det samme som det der er brugt i lyttetesten og SNR = 0dB. Da STI-signalet ikke har lange pauser uden lyd har jeg beregnet SNR ud fra hele signalets RMS-værdi.

For at beregne STI-værdien for de 2 SRA Senso Diva SD-9M og Canta7

BTE har jeg afspillet signalet gennem HA. Jeg har indstillet lydstyrken så det afspillede signal er cirka 70dBC SPL målt i fritfelt ved mikrofonindgangen til HA.

Ud fra optagelserne beregner STI-programmet et SNR i oktav-bånd fra 125Hz til 8kHz samt en samlet værdi for Speech transmission index. Værdien af STI er et tal mellem 0 og 1 hvor 0 svarer til at alt informationen fra talen er væk og 1 at alt informationen fra talen er bevaret. STI undersøger altså hvor meget af informationen SRA bevarer i et talesignal. Sammenhængen mellem STI-værdien og scoren for min lyttetest er ikke mulig at fastslå da den afhænger af hvilken type lyttetest der er brugt. Men det kan antages at en højere STI-værdi svarer til en højere score i lyttetesten. I tabel 6.2 ses de beregnede STI-værdier samt SNR i de forskellige frekvensbånd. Ud fra

HA	STI	125Hz	250Hz	500Hz	1kHz	2kHz	4kHz	8kHz
Ingen SRA	0.62	-9.88	-6.78	-3.61	3.59	8.38	12.16	15.00
Canta7 BTE	0.60	-7.84	-5.15	-2.48	3.61	6.49	9.94	11.82
Senso Diva SD-9M	0.70	-8.65	-6.84	-2.37	7.17	14.82	15.00	14.94

Tabel 6.1: Resultater fra STI-program

ovenstående målinger ses at Senso Diva SD-9M klarer sig bedst da den har den højeste værdi for STI og Canta7 BTE og ingen SRA giver næsten samme STI-værdi. Dette er forskellig fra resultaterne af min lyttetest.

PESQ

Ved PESQ metoden sammenlignes 2 signaler med hinanden og PESQ giver en værdi for hvor meget det behandlede signal er forvrænget i forhold til det oprindelige signal. Ved sammenligning af en optagelse med tale og støj før og efter en given SRA giver PESQ en værdi for hvor meget signalet (tale og støj) er forvrænget. Dette er dog ikke så brugbart da en forvrængning af støjen sjældent har nogen negativ effekt på taleforståeligheden. Derfor har jeg valgt at sammenligne det oprindelige talesignal uden støj med de behandlede signaler med både tale og støj. Til beregning af PESQ har jeg brugt et program som er implementeret af ITU-T[19]. Dette program tager som input de 2 lyd-signaler og beregner PESQ ud fra disse. Jeg har valgt at beregne PESQ-værdien for 5 nonsenssætninger med støj. Disse 5 sætninger er de samme som dem der er brugt ved lyttetesten. I tabel 6.2 ses PESQ-værdierne for de 5 testsætninger.

HA	PESQ 1	PESQ 2	PESQ 3	PESQ 4	PESQ 5	Middel
Ingen SRA	1.312	1.372	1.333	1.195	1.350	1.312
Canta7 BTE	1.357	1.401	1.370	1.240	1.430	1.196
Senso Diva SD-9M	1.091	1.308	1.291	0.967	1.325	1.360

Tabel 6.2: Resultater fra PESQ-program

Forskellen i PESQ-værdien for de 2 SRA og ingen SRA er meget lille og det kan derfor være svært at konkludere hvilken SRA der er bedst. I tabellen ses at *Widex*[®] har den højeste middelværdi og ingen SRA har det laveste gennemsnit. Dette er forskelligt fra resultaterne af lyttetesten. Det er derfor ikke umiddelbart muligt at bruge PESQ til beregning af taleforståelighed for disse SRA.

Kapitel 7

Diskussion

Min models forudsigelser er blevet sammenlignet med observationerne fra lytteforsøg. For en del af de statistiske test har jeg accepteret 0-hypotesen, selv om testen har en meget lille styrke. For at øge styrken i disse test skal der laves flere lytteforsøg.

Modellen er kun blevet brugt til beregning af taleforståelighed hvor der er brugt hvid støj. Det kunne være interessant at teste modellens forudsigelse af taleforståelighed med andre typer støj. Da modellen bruger post- og simultan-maskering vil taleforståeligheden måske også kunne beregnes for støjtyper som f.eks. partynoise. Det kunne specielt være interessant at undersøge om modellen kan beregne taleforståeligheden for SRA i et HA med forskellige indstillinger. Modellen kan herved være med til at finde de optimale indstillinger m.h.t. taleforståelighed. Problemet med et sådan forsøg er at ændringen i taleforståeligheden ikke er særlig stor. Det kræver derfor mange lytteforsøg for at detektere en forskel i taleforståeligheden.

Modellen har været brugt til at forudsige taleforståeligheden ved afspilning af signaler i åbne høretelefoner. Før modellen kan bruges i praksis skal modellen kunne beregne taleforståelighed for HA der sidder i øret. Dette burde dog ikke være et stort problem da der findes filtre der kan tage højde for bl.a. øregangsresonansen. Et andet problem med modellen er at den ikke er testet på forsøgspersoner med høretab. Dette kræver at modellen ændres på en række forskellige punkter, så den også kan tilpasses til et bestemt høretab. En hørehæmmet vil opfatte en tone med en anden hørestyrke end en normalhørende. Konverteringen til hørestyrke skal derfor tilpasses til personens høretab.

Til separering af tale og støjsignal anvendes Olofssens og Hagermans metode. Da behandlingen af signalet i SRA ikke er linear, vil separeringen indeholde

fejl. Der findes dog andre metoder der kan bruges til separering af signaler efter en ulinear transformation. Disse metoder vil måske i højere grad kunne separere tale og støjsignalet så beregning af taleforståeligheden bliver mere præcis. Beregningen af SDR for SRA har vist at der for TSVD-algoritmen var en god overensstemmelse mellem SDR og hvor godt talen og støjen var separeret. En bestemmelse af hvor stor betydning den målte ulinearitet har på taleforståeligheden ville være et oplagt projekt at arbejde videre med.

Kapitel 8

Konklusion

Der findes mange forskellige metoder til bestemmelse af lyd kvalitet og taleforståelighed. Modellen beskrevet i nærværende projekt estimerer taleforståelighed ved SRA i f. eks HA. Modellen kræver at behandlingen af signaler fra SRA er tæt på linear. Den målte SDR for målinger på HA stemmer godt overens med tidligere resultater[11]. Den målte ulinearitet på optagelser fra HA var ikke meget højere end ulinearitet ved simulerede SRA. En måling af hvor stor betydning ulineariteten har for taleforståeligheden, er dog ikke fastlagt.

Ingen af de valgte SRA gav en forbedring af taleforståelighed. Disse resultater er verificeret med lyttetest. Selv om målingerne på SRA tyder på en forværring af taleforståeligheden, kan resultatet for høreapparatsbrugere godt være anderledes. Målingerne og lyttforsøgene bygger nemlig på optagelser fra HA. Høreapparatsbrugere vil høre signalet på en anden måde da det ikke har været igennem noget måleudstyr.

Ved sammenligning med STI og PESQ var der ingen af disse metoder der stemte overens med resultaterne fra lyttforsøgene. Disse metoder gav begge SRA fra Senso Diva SD-9M den højeste score. Det skal dog bemærkes at sammenligningen med STI og PESQ er meget svær, da begge metoder bruger et lidt andet input. Min model har desuden været tilpasset til beregning af taleforståelighed for netop denne slags lyttforsøg.

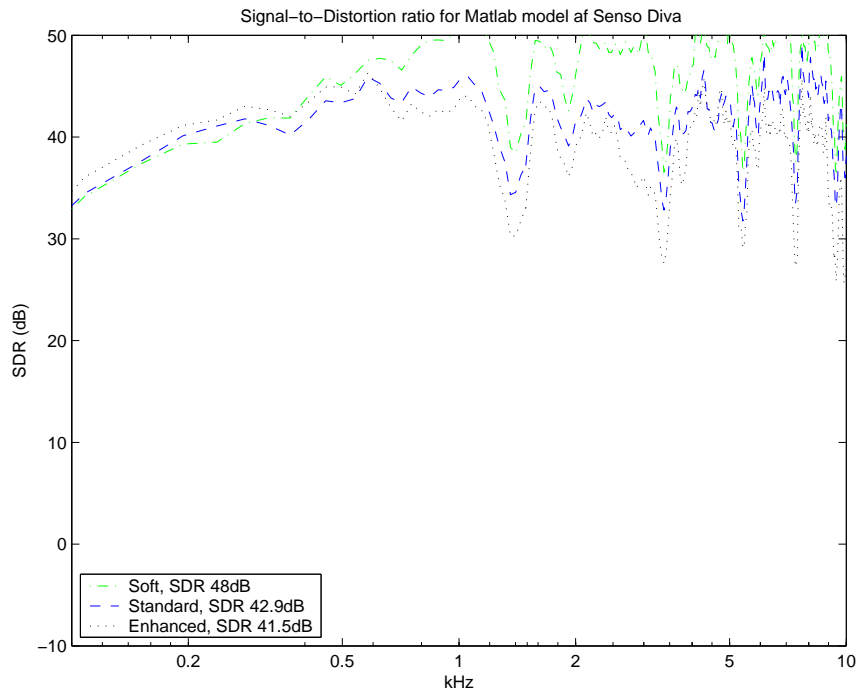
Litteratur

- [1] J. B. Allen. How do humans process and recognize speech? *IEEE Transactions on speech and audio processing*, 2:567–577, October 1994.
- [2] J. Barfod. Properties of intelligibility test materials. Technical report, The Acoustics Laboratory, Technical University of Denmark, 1976.
- [3] J. G. Beerends and A. P. Hekstra. Perceptual evaluation of speech quality the new itu standard for end-to-end speech quality assessment. *J. Audio Eng. Soc.*, 50:765–778, October 2002.
- [4] J. G. Beerends and J. A. Stemerdink. A perceptual speech-quality measure based on a psychoacoustic sound representation. *J. Audio Eng. Soc.*, 42:115–123, 1994.
- [5] R. A. Bentler and C. V. Pavlovic. Transfer function and correction factors used in hearing aid evaluation and research. *Ear And Hearing*, 10:58–63, January 1989.
- [6] K. Conradsen. *En introduktion til statistik*. Informatic and Mathematical Modelling, Technical University of Denmark, 1999.
- [7] J. R. Dubno and A. B. Schafer. Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners. *J. Acoust. Soc. Am.*, 1992.
- [8] Franz Faul and Edgar Erdfelder. *How to Use G*Power*, 1997. <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>.
- [9] H. Fletcher. *Speech and Hearing in communication*. ASA, 1953.
- [10] Lexington Center/School for the Deaf. *Speech Transmission Index, Version 3.1*, 2003. <http://www.hearingresearch.org/STI.htm>.
- [11] B. Hagerman and Å. Olofsson. A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. Technical report, Unit of Technical and Clinical Neuroscience, Karolinska Institutet, 2003.
- [12] Martin Hansen. *Assessment and prediction of speech transmission quality with an auditory processing model*. PhD thesis, Vom Fachbereich Physik der Universität Oldenburg, 1998.
- [13] W. Jiang and H. Schulzrinne. Speech recognition performance as an

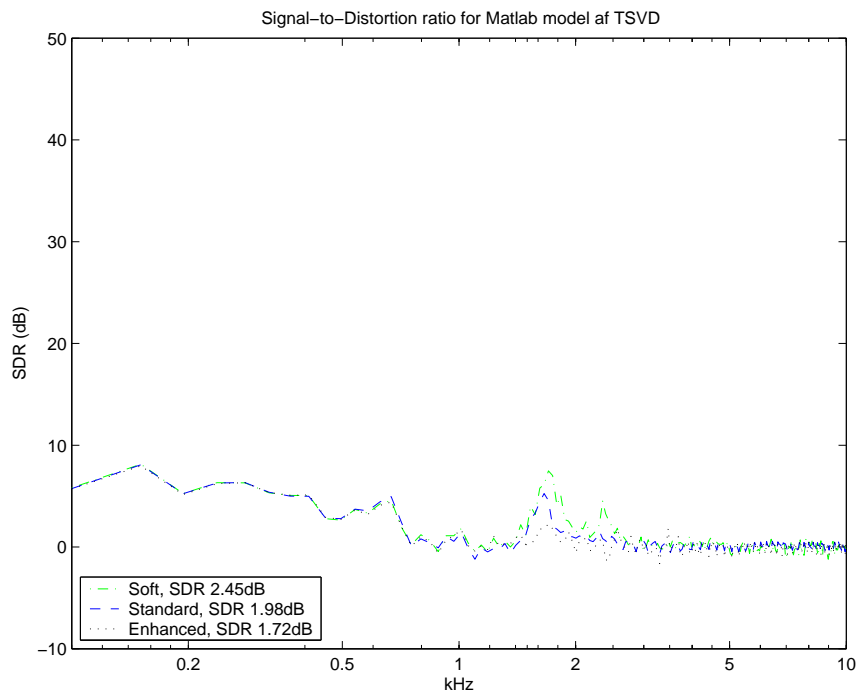
- effective perceived quality predictor. Technical report, Columbia University, Department of Computer Science, 2001.
- [14] Philip Juhl. Måling af effektiviteten i støjreduktionsalgoritmer. Technical report, Informatic and Mathematical Modelling, Technical University of Denmark, 2003.
- [15] H. Levitt. Noise reduction in hearing aids: An overview. *Journal of Rehabilitation Research and Development*, 38, January 2001.
- [16] C. Ludvigsen. The use of objective methods to predict the intelligibility of processed speech, in recent developments in hearing instrument technology. Technical report, Proc. 15th Danavox Symposium 81-94, 1993.
- [17] Acoustical Society of America. Methods for calculation of the speech intelligibility index. Technical report, American National Standards Institute, Inc., 1997.
- [18] ITU-T Telecommunication Standardization Sector of ITU. Objective quality measurement of telephoneband (300-3400hz) speech codecs. Technical report, International Telecommunication Union, 1998.
- [19] ITU-T Telecommunication Standardization Sector of ITU. Perceptual evaluation of speech quality(pesq). Technical report, International Telecommunication Union, 2001.
- [20] Herman J.M. Steeneken. The measurement of speech intelligibility. *TNO Human Factors, Soesterberg, the Netherlands*.
- [21] E. Terhard, G. Stoll, and M. Seewann. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J. Acoust. Soc. Am.*, 71:679–688, March 1982.
- [22] P. M. Zurek and L. A. Delhorne. Constant reception in noise by listeners with mild and moderate sensorineural hearing impairment. *J. Acoust. Soc. Am.*, 1987.
- [23] E. Zwicker. *Psychoacoustics : Facts and models*. Springer, 2. ed edition, 1999.

Bilag A

Beregnet ulinearitet for SRA

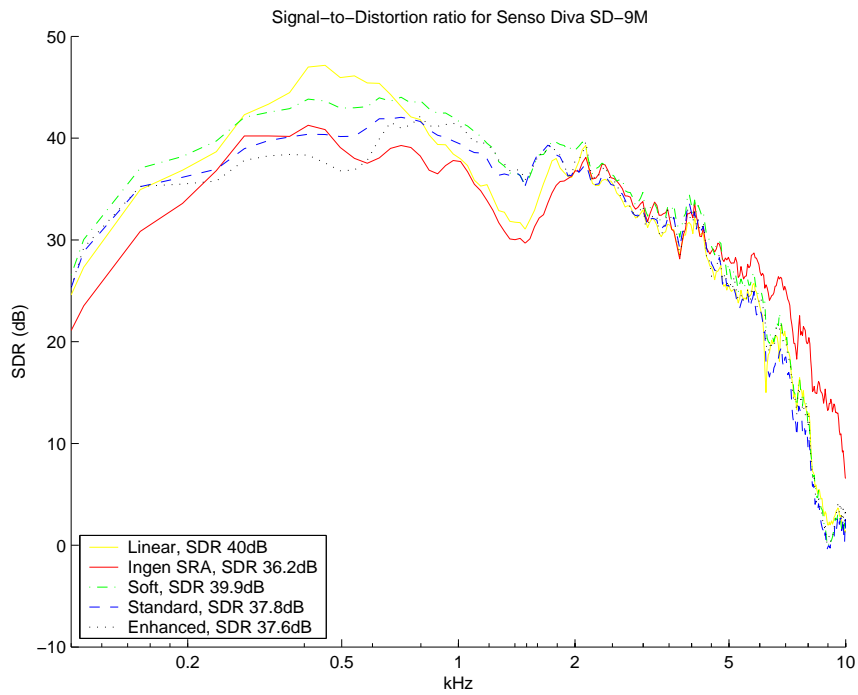


(a) Matlab model af Senso Diva SD-9M

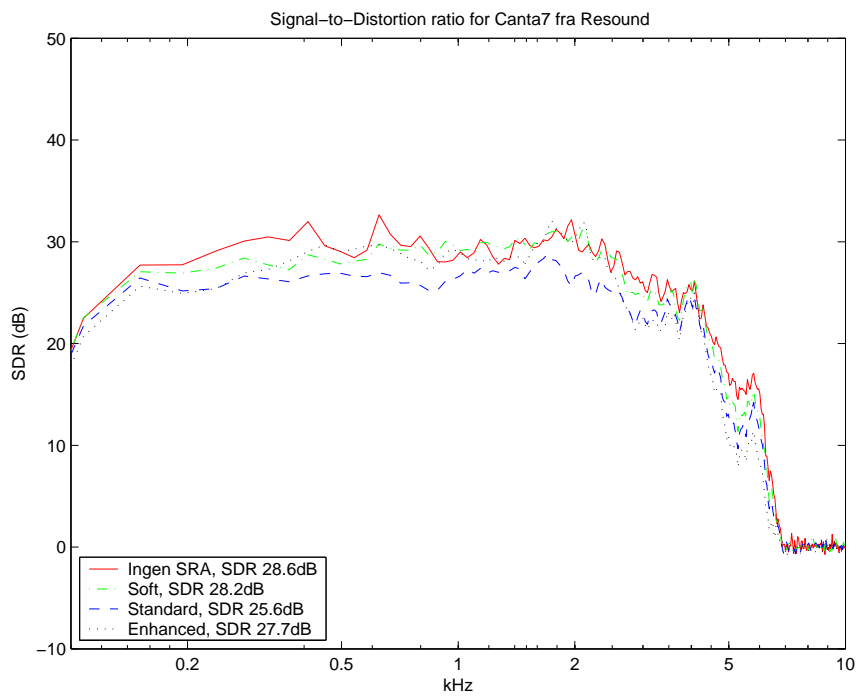


(b) Matlab model af TSVD

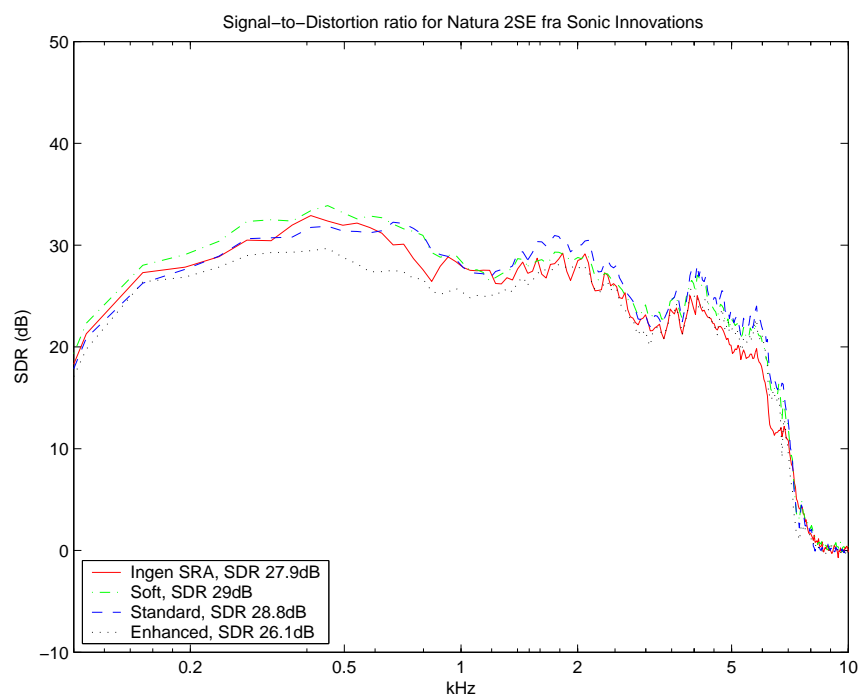
Figur A.1: Målt SDR med Matlab modeller



(a) Senso Diva SD-9M fra Widex



(b) Canta7 BTE fra Resound

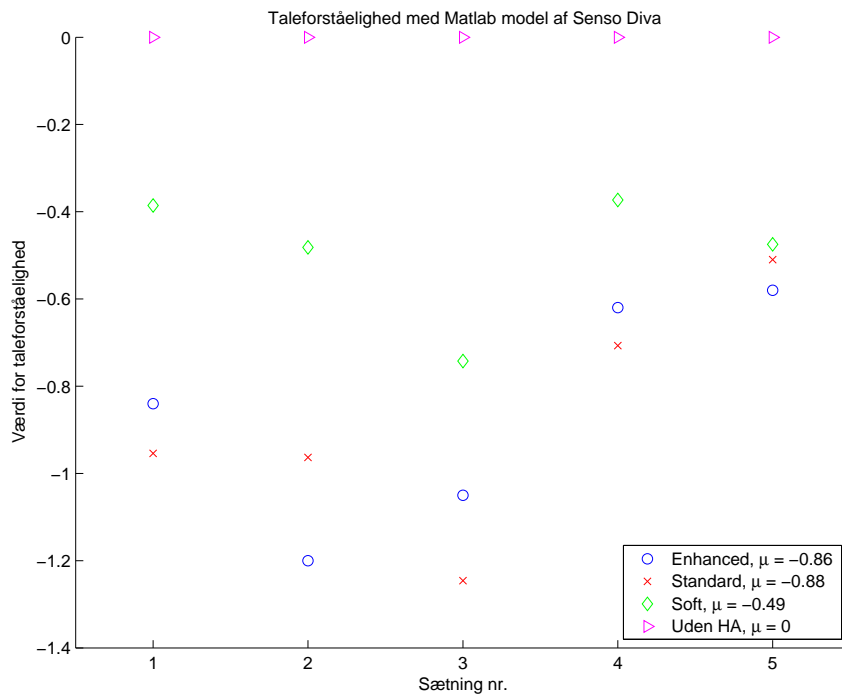


(c) Natura 2SE fra Sonic

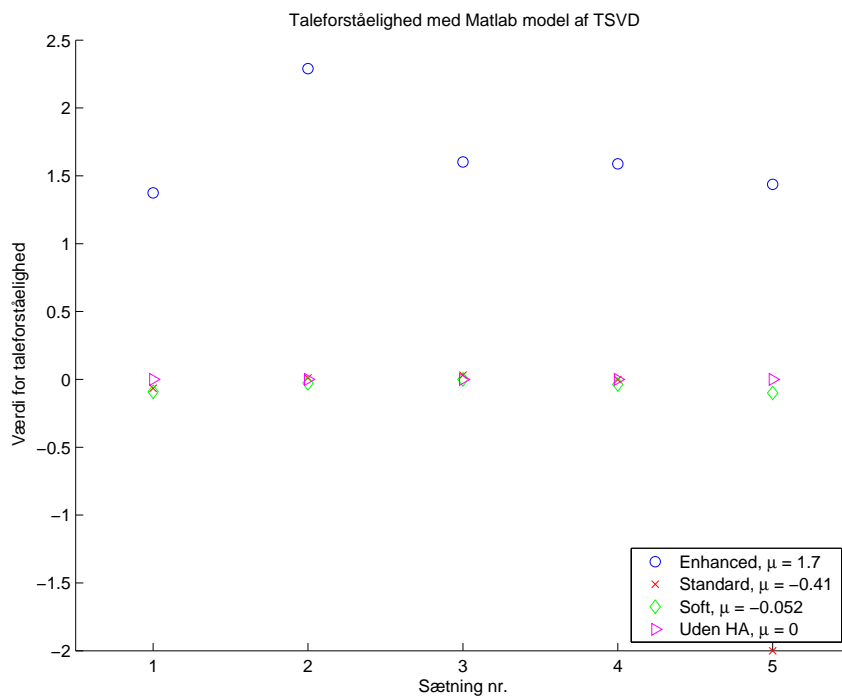
Figur A.2: Målt SDR med SRA fra HA

Bilag B

Beregnet taleforståelighed

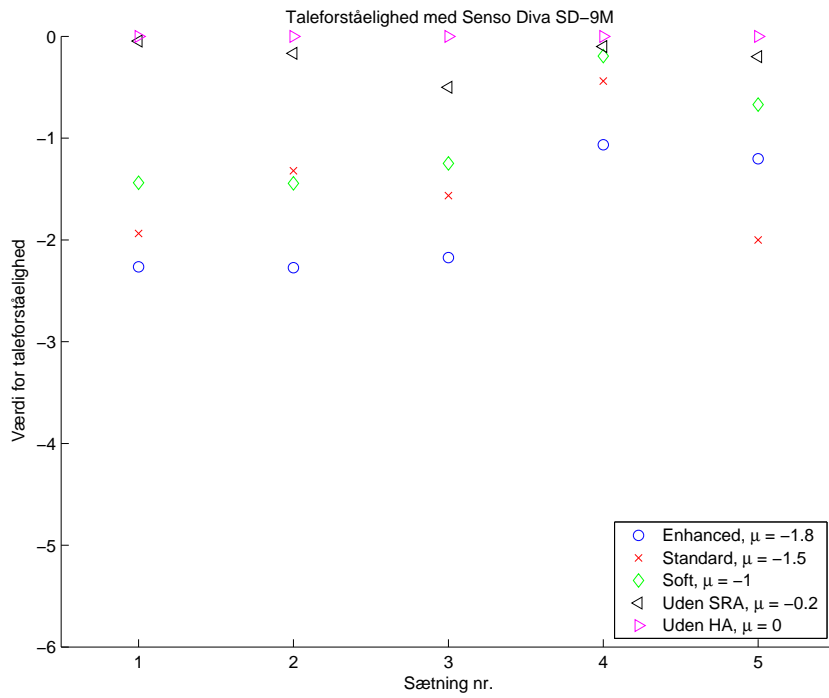


(a) Matlab model af Senso Diva SD-9M

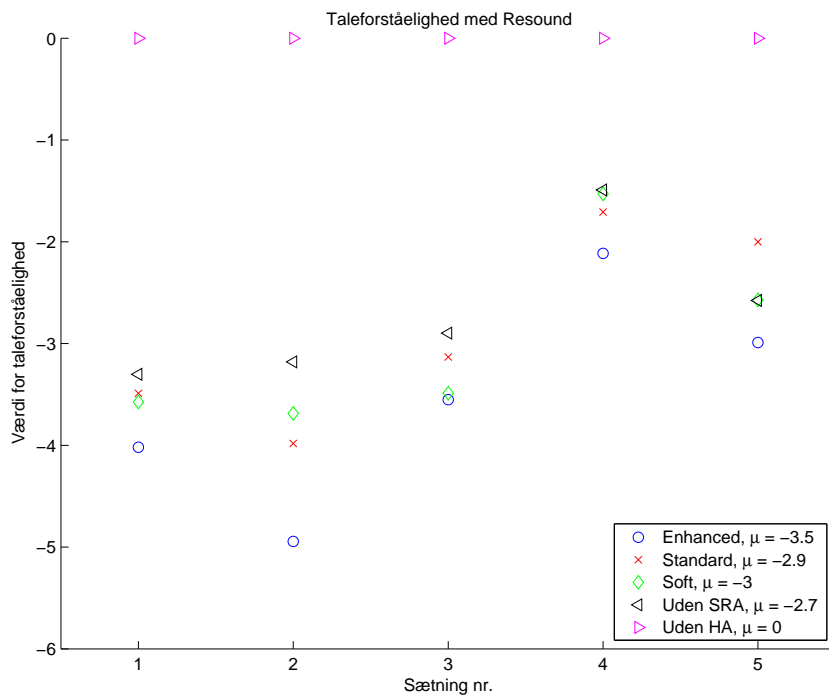


(b) Matlab model af TSVD

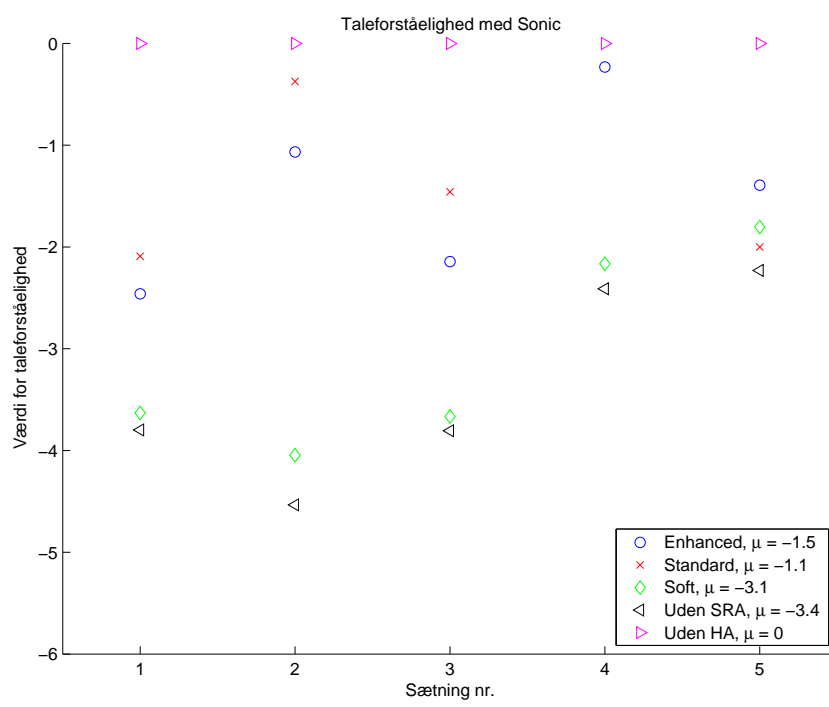
Figur B.1: Beregnet taleforståelighed for Matlab modeller af SRA



(a) Senso Diva SD-9M fra Widex



(b) Canta7 BTE fra Resound



(c) Natura 2SE fra Sonic

Figur B.2: Beregnet taleforståelighed for HA med SRA

Bilag C

Skema til Lyttetest

Testskema for Lyttetest (Version 2)

Information om testen

Testnr.	Test Person	TestType

Test 1

Nr	Ord	Point	Nr	Ord	Point	Nr	Ord	Point	Nr	Ord	Point
1	husk		6	jån		11	len		16	tung	
2	tak		7	hik		12	fus		17	risp	
3	bøf		8	pusp		13	pæst		18	stasp	
4	såf		9	vil		14	pun		19	gav	
5	gæg		10	ryv		15	pan		20	tusp	

Test 2

Nr	Ord	Point	Nr	Ord	Point	Nr	Ord	Point	Nr	Ord	Point
1	bong		6	mon		11	mug		16	høv	
2	båd		7	nøl		12	pon		17	ryb	
3	tar		8	kak		13	jyl		18	fisp	
4	vån		9	nøst		14	hung		19	hem	
5	fæs		10	mys		15	stel		20	måv	

Test 3

Nr	Ord	Point	Nr	Ord	Point	Nr	Ord	Point	Nr	Ord	Point
1	tal		6	skof		11	raf		16	pam	
2	nær		7	såb		12	jad		17	bev	
3	gyd		8	fæd		13	nung		18	næs	
4	sosk		9	dåf		14	lost		19	dør	
5	hom		10	hyn		15	søn		20	såsp	