

# Taleforståelighet

En enkel modell

**Audun Etnestad**

Master i elektronikk

Oppgaven levert: Juni 2009

Hovedveileder: Lars Magne Lundheim, IET

Biveileder(e): Olav Kvaløy, SINTEF



# Oppgavetekst

Et digitalt kommunikasjonssystem kan karakteriseres ved sin spektraleffektivitet. Denne kan uttrykkes ved bitfeilsannsynlighet eller symbolfeilsannsynlighet som funksjon av signal-støy-forhold.

For diagnose av hørselstap er det gjort undersøkelser om taleforståelighet som funksjon av signal-støy-forhold. Oppgaven går ut på å bruke slike undersøkelser som grunnlag for en kommunikasjonsteoretisk modell for menneskelig talekommunikasjon, og om mulig finne uttrykk for spektraleffektiviteten til et slikt kommunikasjonssystem.

Oppgaven gitt: 23. januar 2009

Hovedveileder: Lars Magne Lundheim, IET



# Sammendrag

Oppgaven har gått ut på å sammenligne symbolfeilsannsynligheten som er målt ved felttesting av HiST-taleaudiometri, med en sterkt forenklet modell for taleforståelighet. Taleaudiometri går ut på å teste taleforståelighet ved forskjellige nivåer eller signal-støy-forhold. HiST-taleaudiometri er en nyutviklet taleaudiometri utviklet av Jon Øygarden ved Høgskolen i Sør-Trøndelag. Testen kan generere 100000 5-ordssetninger ut ifra 50 ord. I testen registreres det hvor mange ord i en setning man oppfatter for å beregne taleforståelighet. Ved hjelp av empiriske data, er det beregnet en gjennomsnittlig symbolfeilsannsynlighet for testen på et utvalg av personer. Denne symbolfeilsannsynligheten er så sammelignet med symbolfeilsannsynligheten til taleforståelighetsmodellen. I forhold til HiST-taleaudiometri, så er det gjort noen forenklinger for modellen: støyen er regnet som hvit, og symbolene er ortogonale. Modellen viste minst symbolfeilsannsynlighet for SNR under ca.  $-3.5$  dB, mens HiST-taleaudiometri, og dermed tale, var bedre enn modellen ved høyere SNR.



# Innhold

<b>1</b>	<b>Innledning</b>	<b>1</b>
1.1	Bakgrunn . . . . .	1
1.2	Rapportstruktur . . . . .	2
<b>2</b>	<b>Taleaudiometri</b>	<b>3</b>
2.1	Taleaudiometri . . . . .	3
2.1.1	Testmateriale . . . . .	4
2.1.2	Presentasjon . . . . .	4
2.1.3	Faktorer som påvirker taleforståelighet . . . . .	5
2.2	HiST-taleaudiometri . . . . .	6
2.2.1	Ordmateriale . . . . .	6
2.2.2	Difonsplitting . . . . .	6
2.2.3	Talestøyen . . . . .	7
2.2.4	Felttest . . . . .	8
2.2.5	Matematisk beskrivelse av 5-ordssetningene . . . . .	8
2.2.6	Symbolfeilsannsynlighet . . . . .	9
<b>3</b>	<b>Modellering av taleforståelighet</b>	<b>11</b>
3.1	AWGN-vektorkanal . . . . .	11
3.1.1	Geometrisk representasjon . . . . .	12
3.1.2	Korrelasjonsmottaker . . . . .	13
3.2	Referanse-systemet . . . . .	15
3.2.1	Symbolfeilsannsynlighet . . . . .	15
3.3	Modell versus målte data . . . . .	17
<b>4</b>	<b>Konklusjon</b>	<b>19</b>
<b>A</b>	<b>Tillegg</b>	<b>21</b>





# Figurer

1.1	Et kommunikasjonssystem . . . . .	1
2.1	Eksempel på scoringskurve. . . . .	5
2.2	Skjøting av ord . . . . .	7
2.3	Eksempel på en sammenskjøtt setning . . . . .	8
2.4	Spektrum for støyen brukt i HiST-taleaudiometri . . . . .	9
2.5	HiST-taleaudiometri kanalmodell . . . . .	9
2.6	Symbolfeilsannsynlighet for HiST-taleaudiometri . . . . .	10
3.1	Blokkdiagram av et digitalt kommunikasjonssystem . . . . .	11
3.2	Geometrisk representasjon av 3 signaler . . . . .	12
3.3	(a) Synthesizer for å generere signalet $s_i(t)$ . (b) Analysator for å generere signal vektoren $\mathbf{s}_i$ . . . . .	13
3.4	Korrelasjonsmottaker . . . . .	15
3.5	Symbolfeilsannsynlighet for modellen . . . . .	17
3.6	Symbolfeilsannsynlighet for korrelasjonsmottakeren kontra SI-funksjonen . . . . .	18
4.1	Forskjell i feilsannsynlighet for et kodet og ukodet system som funksjon av SNR . . . . .	19



# Tabeller

2.1	Ordene som er brukt i 5-ordsetningene. . . . .	6
2.2	Gangen i innspilling og bearbeiding. . . . .	7
A.1	Scoringstabell etter felttest . . . . .	21



# Forkortelser

AWGN Additive White Gaussian Noise

SI Speech Intelligibility

SNR Signal-to-Noise-Ratio

SRT Speech Recognition Threshold

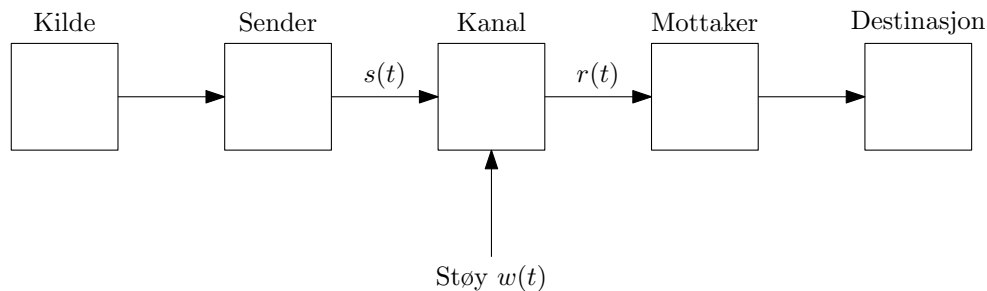


# Kapittel 1

## Innledning

### 1.1 Bakgrunn

I 1948 publiserte Claude E. Shannon artikkelen «A Mathematical Theory of Communication» [1], en banebrytende artikkel som la grunnlaget for informasjonsteorien. Man fikk med informasjonsteorien verktøy slik at man kunne måle effektiviteten av eksisterende kommunikasjonssystemer, og grenser å strekke seg etter for nye.



Figur 1.1: Et kommunikasjonssystem

Shannon delte opp kommunikasjonssystemet i 5 deler som vist i figur 1.1: En informasjonskilde, en sender, en kanal, en mottaker, og til slutt en destinasjon. På samme måte kan en samtale mellom to personer også sees på som et kommunikasjonssystem der de to personene vekselvis er sender og mottaker. I et slikt system er ikke skillene så klare som i f.eks et radiosystem. Hos avsender vil hjernen kunne være både en informasjonskilde og en del av senderen, taleorganet en del av sender og kanal. På mottakersiden vil øret være en del av kanal og mottaker, hjernen en del av mottaker og destinasjon.

Taleaudiometri er en av metodene i audiologi som brukes til å diagnostisere hørselskader og for å evaluere høreapparater. Metoden går ut på at testpersonen får opplept ord eller setninger ved ulike nivåer uten støy, eller med forskjellige signal-støy-forhold hvis man har med støy. Man registrerer så hvor mange av ordene eller hvor mye av teksten testpersonen har oppfattet ved de ulike nivåene eller signal-støy-forholdene. Det nivået som gir en gjenkjennelsesgrad på 50 % kalles høreterskel eller SRT(Speech Recognition Threshold).

HiST-taleaudiometri er en nyutviklet taleaudiometri utviklet ved Høgskolen i Sør-Trøndelag av Jon Øygarden[2]. Det spesielle med denne testen er at man kan generere 100000 5-ordssetninger

ut i fra 50 ord. Målet med denne oppgaven er å bruke kommunikasjonsteori til å utvikle en enkel modell for taleforståelse, som så kan sammenlignes med målte data fra HiST-taleaudiometri.

## 1.2 Rapportstruktur

Kapittel 2 tar for seg litt om taleaudiometri generelt, og HiST-taleaudiometri spesielt. Kapittel 3 handler om kanaler med additiv hvit gaussisk støy og beskriver en enkel modell for taleforståelighet. Kapittel 4 gir en konklusjon på det hele.



# Taleaudiometri

I dette kapitlet presenteres det litt hva taleaudiometri er, for deretter å se nærmere på HiST-taleaudiometri som danner grunnlaget for oppgaven. Spesielt er det 5-ordsetningene fra HiST-taleaudiometri som skal analyseres.

## 2.1 Taleaudiometri

Materialet i denne delen er i stor del hentet fra [3]. Taleaudiometri er en metode for å diagnostisere hørseltap og evaluere hjelpemidler for dette. Testpersonen får opplest ord eller setninger fra en CD eller en datamaskin ved et bestemt nivå, med eller uten støy. Testpersonen identifiserer hva som blir sagt, og resultatet av hvor mange ord som er oppfattet korrekt blir registrert.

Taleforståelighet (eng. Speech Intelligibility) er definert som

$$SI(SNR) = \frac{N_k(SNR)}{N_t(SNR)} \quad (2.1)$$

der  $N_k(SNR)$  er antall korrekte oppfattede ord i en setning ved et gitt signal-støy-forhold, og  $N_t(SNR)$  er totalt antall ord i setningen.

Det laveste nivået eller SNR hvor testpersonen greier å identifisere ordene 50 % av tiden kalles for “speech recognition threshold” (SRT), eller på norsk høreterskelen.

Hvis støy benyttes i testen, er en mulig måte å generere denne ved å addere mange av de innleste setningene ovenpå hverandre. Alternativt kan man for eksempel måle langtidsspekteret av de innleste setningene, for deretter å forme tilfeldig støy etter dette spekteret.

SI-funksjonen kan som man ser av (2.1) innta en verdi mellom 0 og 1, hvor 1 da er maksimum taleforståelighet,  $SI_{max}$ , for en person med normal hørsel. Hvis en person har alvorlig hørselskade, vil  $SI_{max}$  være lavere enn 1. Dårlig test-materiale vil også kunne føre til lavere taleforståelighet.

Det har vist seg at SI-funksjonen følger en sigmoid-funksjon, det vil si at man ved hjelp av kurvetilpasning kan beskrive taleforståelighet i forhold til nivå eller SNR på formen (2.2)

$$SI(SNR) = \frac{1}{A} \left( 1 + SI_{\max} \frac{A-1}{1 + \exp\left(-\frac{SNR-SNR_{\text{mid}}}{s}\right)} \right) \quad (2.2)$$

hvor  $SNR_{\text{mid}}$  er gitt ved  $SI(SNR_{\text{mid}}) = \frac{1}{2}SI_{\max}$ , og  $s$  er en helnings-parameter. Helningen ved  $SNR_{\text{mid}}$  er gitt ved

$$h = \frac{\partial SI(SNR_{\text{mid}})}{\partial SNR} = \frac{SI_{\max}(A-1)}{4As} \quad (2.3)$$

$A$  er antallet responsalternativer. Hvis man for eksempel bare kan gi responsen ja eller nei, er  $A = 2$ . Det er vanlig at forsøkspersonens alternativer bare er begrenset av sitt ordforråd, da vil  $A$  være tilnærmet uendelig. Formlene (2.2) og (2.3) reduseres da til

$$SI(SNR) = SI_{\max} \frac{1}{1 + \exp\left(-\frac{SNR-SNR_{\text{mid}}}{s}\right)} \quad (2.4)$$

$$h = \frac{SI_{\max}}{4s} \quad (2.5)$$

Ved å kombinere (2.4) og (2.5), og sette  $SI_{\max} = 1$  som gjelder for personer med normal hørsel, kommer man frem til følgende uttrykk

$$SI(SNR) = \frac{1}{1 + \exp(-4h(SNR - SRT))} \quad (2.6)$$

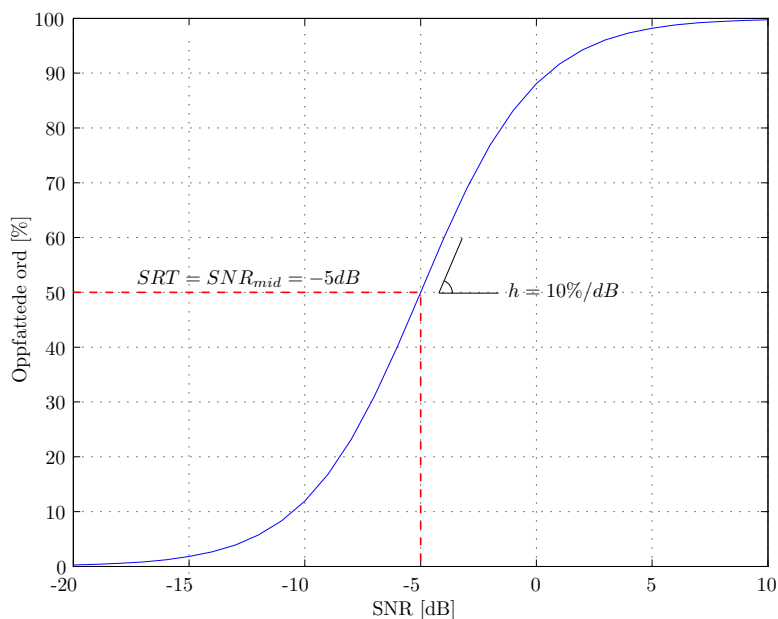
Figur 2.1 viser eksempel på en scoringskurve generert med formel (2.6), med  $SRT = -5$  dB og  $h = 10$  %/dB

### 2.1.1 Testmateriale

Testmaterialet brukt i taleaudiometri kan være “meningsløse stavelser”, som er to konsonanter med en vokal mellom, single ord, eller setninger. Setninger er det som er mest hensiktsmessig da de best representerer en realistisk kommunikasjonssituasjon. For at testen skal gi et nøyaktig og reproducerbart resultat, bør testmaterialet være så homogent som mulig. Det vil si at ordene eller stavelsene i testen bør ha så like scoringskurver som mulig. I tillegg bør distribusjonen av fonemer være representativt for språket som testes. Materialet som skal brukes må testes, evalueres og optimaliseres for å tilfredstille disse kravene.

### 2.1.2 Presentasjon

Presentasjon av materialet kan foregå ved hjelp av hodetelefoner eller i fritt felt ved hjelp av høyttalere. Fritt felt har den fordelen at det er mest naturlig. Bakdelen er at det er vanskeligere å kontrollere parametre i testen, som medfører at reproducerbarheten av testen blir dårligere. Hodetelefoner har den fordelen at det er lett å reproducere testresultater ved at man har bedre kontroll på alle parametre. Støy kan enkelt legges til og romlige forhold kan simuleres ved hjelp av virtuell akustikk.



Figur 2.1: Eksempel på scoringskurve.

### 2.1.3 Faktorer som påvirker taleforståelighet

#### Testmateriale

Forskjellig testmateriale vil gi ulike testresultater. Setninger har brattere taleforståelighetskurve enn enkeltord. Bare standardisert materiale, eller materiale med referanse-skåringsfunksjoner bør brukes.

#### Støy og romakustikk

Støy og romklang reduserer taleforståelighet, så ved måling av taleforståelighet i stillhet, må romklang og støy minimeres for å ikke påvirke måleresultat. På den andre siden kan målinger av taleforståelighet i forskjellig støy og romklang gi innblikk i cocktailselskap-effekten (se nedenfor).

#### Cocktailselskap-effekten

Det menneskelige høreselsystem har en imponerende evne til å oppfatte talen fra en bestemt person, selv om denne talen blir maskert av talen til andre mennesker. Denne egenskapen kalles ofte cocktailselskap-effekten. Taleforståeligheten i en situasjon hvor man har mange mennesker som snakker, er påvirket av mange maskeringseffekter: lydtrykk, frekvensspektrum, amplitudemodulasjon, romlig retning, og antallet maskererere. Romlig plassering av taleren og maskerende lydkilder spiller en stor rolle. Binaural høring kan gi en forbedring på 12 dB av SRT i forhold til monaural høring.

#### Nedsatt hørsel

Hørselskade kan føre til økt  $SRT$ , lavere maks taleforståelighet,  $SI_{maks}$ , og en lavere helning på taleforståelighetsfunksjonen. De vanskeligste situasjonene for personer med nedsatt hørsel, er støyete omgivelser med mange konkurrerende lydkilder (Cocktailselskap-effekten). Taleforståelighetstester er derfor et viktig hjelpemiddel for å fastslå hvordan en hørselskade påvirker det daglige livet, og hvor godt et høreapparat fungerer. Taleforståelighet er veldig viktig ved forskning på, og tilpasning av høreapparater.

## 2.2 HiST-taleaudiometri

HiST-taleaudiometri er en norsk taleaudiometritest utviklet av Jon Øygarden ved Høgskolen i Sør-Trøndelag [2]. Testen har følgende talemateriale:

- Enstavelsesord
- Femords setninger på formen: *Navn verb tallord adjektiv substantiv*
- Treords setninger, hvor man bare har kuttet de to første ordene av femords-setningene
- Enstavelse tallord fra Quist-Hanssens innspillinger fra 1960-tallet

Enstavelsesordene anbefales brukt til måling av maksimal taleoppfattelse. Her har man også egne lister tilpasset barn. Femordssetningene og treordssetningene kan brukes til å måle høreterskel og signal-støy-forhold i forskjellige settinger. Enstavelse tallordene er tiltenkt brukt når man tester fremmedspråklige eller andre som har vanskeligheter med ordmaterialet i de andre gruppene. Femords- og treordssetningene i HiST-taleaudiometri er generert ved hjelp av difonsplitting. Det vil si at man generere setningene ved å skjøte sammen ord i vokallyden hvor lyden er tilnærmet stasjonær.

I denne oppgaven er det femordsetningene som har blitt analysert.

### 2.2.1 Ordmateriale

Femordsetningene i testen har alle samme oppbygning: først et navn, deretter et verb, et tallord, et adjektiv, og til slutt et substantiv. Ordene som er valgt ut er inspirert av materialet i de danske og svenske Hageman-testene, og valgt utifra det hensyn at de skal være kjente for forsøkspersonen. Man skal teste taleforståelighet, ikke vokabularet. Tabell 2.1 viser de 50 ordene som er valgt ut til testen.

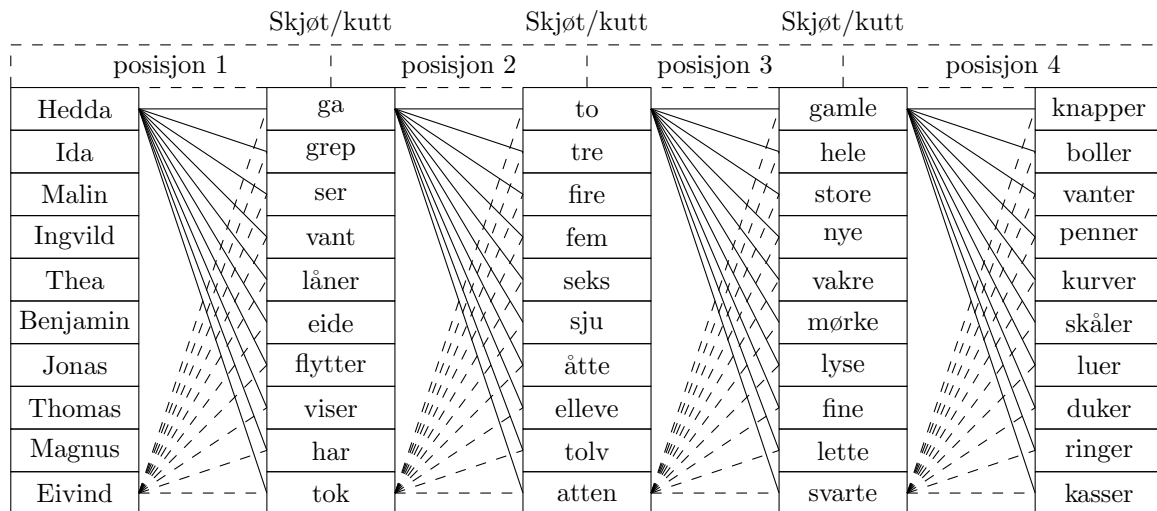
Tabell 2.1: Ordene som er brukt i 5-ordsetningene.

<i>Navn</i>	<i>verb</i>	<i>tallord</i>	<i>adjektiv</i>	<i>substantiv</i>
Hedda	ga	to	gamle	knapper
Ida	grep	tre	hele	boller
Malin	ser	fire	store	vanter
Ingvild	vant	fem	nye	penner
Thea	låner	seks	vakre	kurver
Benjamin	eide	sju	mørke	skåler
Jonas	flytter	åtte	lyse	luer
Thomas	viser	elleve	fine	duker
Magnus	har	tolv	lette	ringer
Eivind	tok	atten	svarte	kasser

### 2.2.2 Difonsplitting

HiST taleaudiometri er laget slik at man kan generere  $10^5$  forskjellige 5-ordssetninger ut ifra de 50 ordene som er valgt til testen. Man kunne realisert dette ved å bare skjøte sammen ordene direkte, men dette fører til veldig unaturlig overganger mellom ordene. Øygarden har løst problemet med unaturlighet ved å benytte difonsplitting. Dette gjøres ved å skjøte sammen

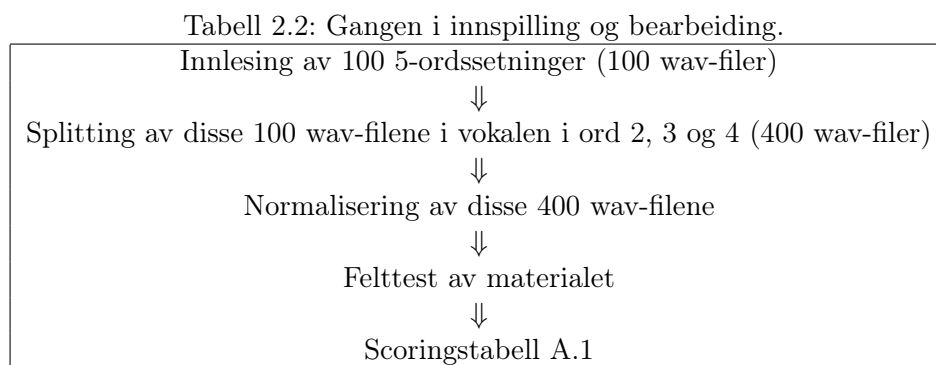
ord i vokallyden, der bølgeformen for ordene er mest stasjonær. Man har splittet i ord 2, 3 og 4 som vist i figur 2.2. For å få med alle overgangene mellom ordene, laget man 100 forskjellige setninger som ble lest inn i hver sin lydfil. Disse filene ble splittet ved hjelp av difonsplitting, der man splitter i en nullgjennomgang i vokalen i ord 2, 3 og 4. Som illustrert i figur 2.2 får man 100 forskjellige overganger mellom de 10 navnene og 10 verbene. Det samme gjelder for verb og tallord, tallord og adjektiv, og adjektiv og substantiv. Dette gir 100 lydfiler for hver av de fire posisjonen vist i figuren, totalt 400 filer.



Figur 2.2: Skjøting av ord

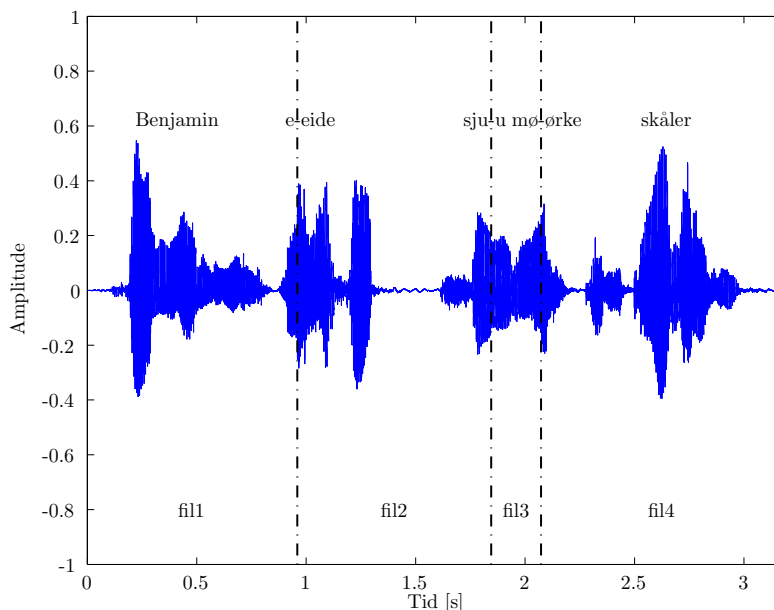
Figur 2.3 viser hvordan en 5-ordssetning er skjøtt sammen i vokalen “e” mellom 1. og 2. fil, “u” mellom 2. og 3. fil, og “ø” mellom 3. og 4. fil.

Tabell 2.2 viser gangen i produksjonen av lydfilene til 5-ordssetningene.



### 2.2.3 Talestøyen

Støyen,  $w(t)$ , som er brukt i HiST-testen er 10.000 tilfeldige 5-ordssetninger som er lagt ovenpå hverandre. Spekteret for støyen er vist i figur 2.4.



Figur 2.3: Eksempel på en sammenskjøtt setning

## 2.2.4 Felttest

For å finne scoringskurver for de forskjellige ordene som er spilt inn, foretok Øygarden en felttest [2]. Fire forskjellige test-sett med 82 setninger ved forskjellige SNR ble laget for denne testen. Antall riktig oppfattede ord ble registrert, ved SNR fra  $-11$  dB til  $3$  dB. Resultatet av felttesten er vist i tabell A.1. Ved hjelp av regresjon tilpasset Øygarden disse data til en SI-funksjon på formen (2.6). Dette har Øygarden gjort for alle ordene, og deretter laget en SI-funksjon som er gjennomsnitt av alle ordene. Denne SI-funksjonen brukes til å finne symbolfeilsannsynligheten.

## 2.2.5 Matematisk beskrivelse av 5-ordssetningene

Bølgeformen for en 5-ordssetning (sammenskjøtt av 4 filer) er gitt ved

$$s_m(t) = \sum_{i=1}^4 g_{m,i}(t - \sum_{j=1}^{i-1} T_{m,j}) \quad (2.7)$$

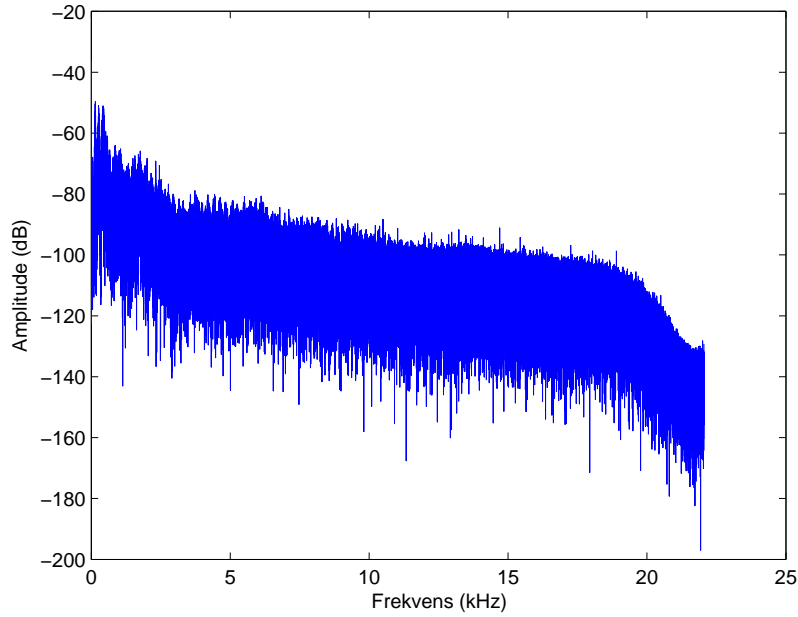
der  $s_m(t)$  er bølgeformen for en av 100000 mulige setninger,  $g_{m,i}(t)$  er en av fire lydfiler som inngår i en setning, og  $T_{m,j}$  er varigheten av en lydfil.

Gjennomsnittseffekt for en lydfil er gitt av

$$\sigma_{g_{m,i}}^2 = \frac{1}{T_{m,i}} \int_0^{T_{m,i}} g_{m,i}^2(t) dt. \quad (2.8)$$

Alle filene er normalisert til samme gjennomsnittseffekt. Gjennomsnittseffekt for en setning er da gitt ved

$$\sigma_{s_m}^2 = \frac{1}{T_m} \int_0^{T_m} \sum_{i=1}^4 g_{m,i}^2(t) dt = \sigma_{g_{m,i}}^2 \quad (2.9)$$



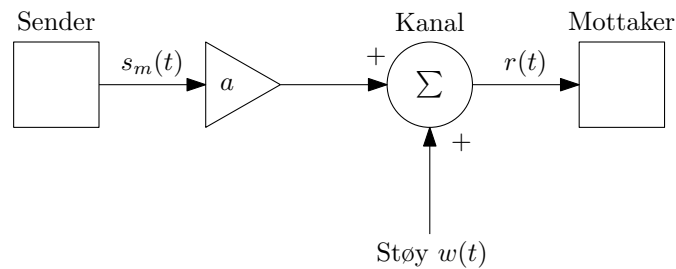
Figur 2.4: Spektrum for støyen brukt i HiST-taleaudiometri

hvor  $T_m = \sum_{i=1}^4 T_{m,i}$  er varigheten av en setning.

Det mottatte signalet er gitt ved

$$r(t) = a \cdot s_m(t) + w(t) \quad (2.10)$$

Figur 2.5 viser kanalmodellen for HiST-taleaudiometri. Mens støyen holdes ved et konstant nivå, varieres amplituden  $a$  på de oppleste setningene.



Figur 2.5: HiST-taleaudiometri kanalmodell

Signalstøyforholdet for en setning er gitt ved

$$SNR = \frac{a^2 \cdot \sigma_{s_m}^2}{\sigma_w^2} \quad (2.11)$$

## 2.2.6 Symbolfeilsannsynlighet

Resultatene av første felttest er presentert i tabell A.1. Tabellen viser score for hver av de 50 ordene ved SNR fra  $-12$  dB, til  $+3$  dB. Ved hjelp av disse empiriske data og regresjon, har

Øygarden tilpasset disse data til sigmoide-kurver, gitt av formel (2.6). Gjennomsnittlig SRT for alle ordene er  $-5,36$  dB SNR og gjennomsnittlig helning på SI-funksjonen er  $10.6$  %/dB. En symbolfeilsannsynlighetskurve kan da tegnes ved hjelp av formel (2.12).

$$P_e(SNR) = 1 - SI(SNR) \quad (2.12)$$

$$= 1 - \frac{1}{1 + \exp(-4 \cdot 0.106 \cdot (SNR + 5.36))}$$

Den resulterende kurve er presentert i figur 2.6.

I tillegg til symbofeilsannsynligheten regnet ut fra SI-funksjonen, er det også regnet ut et gjennomsnitt for alle ord ved hjelp av data fra felttesten, for å se om kurvetilpasningen stemmer.

Symbolfeilsannsynligheten for et ord er gitt av

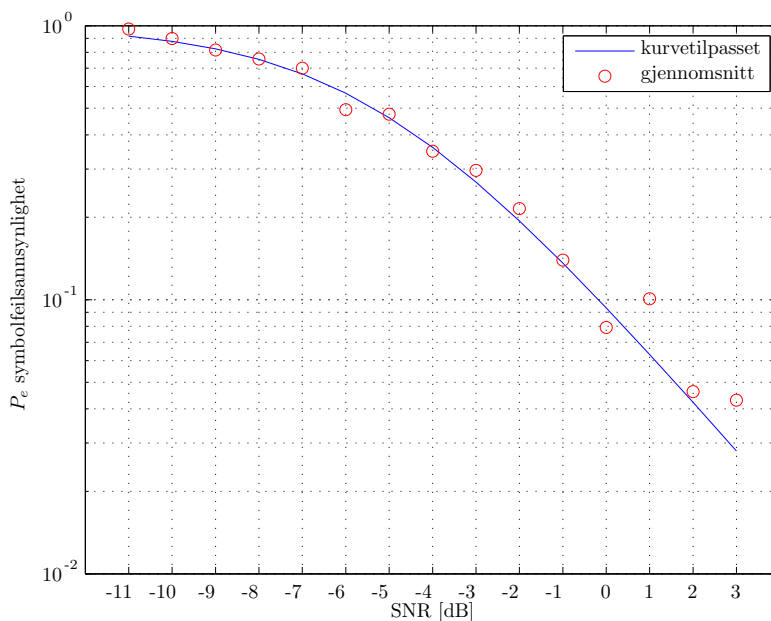
$$P_l(SNR) = 1 - S_l(SNR) \quad (2.13)$$

der  $S_l$  er score for ordet  $l$  ved et gitt  $SNR$  i tabell A.1.

Ved å regne ut gjennomsnittscoren for alle ordene, finner man midlere symbolfeilsannsynlighet

$$\bar{P}_l(SNR) = \frac{1}{50} \sum_{l=1}^{50} P_l(SNR) \quad (2.14)$$

Resultatet av dette er også plottet inn i figur 2.6



Figur 2.6: Symbolfeilsannsynlighet for HiST-taleaudiometri



## Modellering av taleforståelighet

I dette kapitlet skal en enkel modell for taleforståelighet presenteres. Formålet til modellen er å kunne regne ut en symbolfeilsannsynlighet, som kan sammenlignes med symbolfeilsannsynligheten som kommer fra felttesten til HiST-taleaudiometri. Før modellen presenteres, vises litt generell teori om vektorkanaler med hvit gaussisk støy, geometrisk representasjon av signaler og en korrelasjonsmottaker.

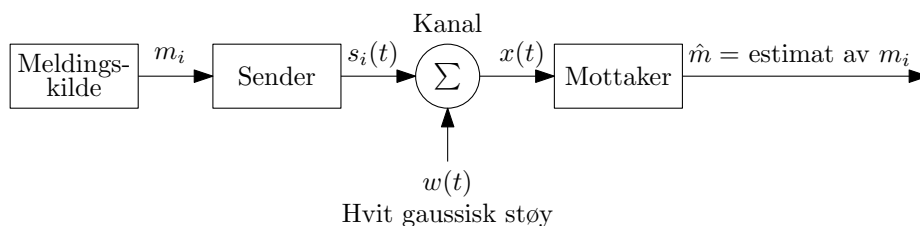
### 3.1 AWGN-vektorkanal

Et kommunikasjonssystem er et system for å kunne overføre meldinger fra et punkt til et annet via en overføringskanal. Overføringskanalen vil kunne være elektrisk, optisk, elektromagnetisk, akustisk, eller til og med mekanisk. Kanalene har ulike bruksområder og ytelser, men felles for dem alle er at de er påvirket av støy. Støy setter begrensning for hvor effektivt man kan få meldingene fra et punkt til et annet.

Formel (3.1) som Shannon utledet i 1948 gir informasjonskapasiteten i bit/s for en kanal dominert av additiv hvit gaussisk støy.  $B$  er kanalens båndbredde mens  $SNR$  er signal/støyforholdet.

$$C = B \cdot \log_2(1 + SNR) \text{ bit/s} \quad (3.1)$$

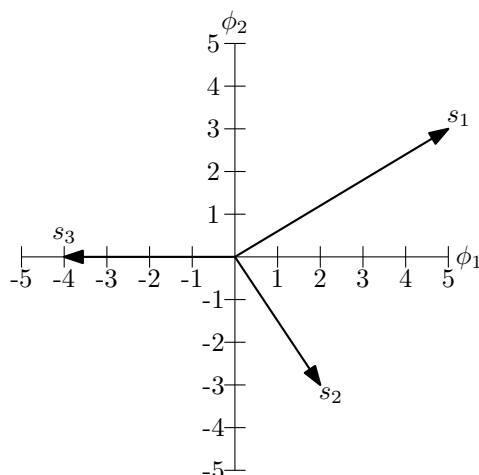
Figur 3.1 viser et digitalt kommunikasjonssystem, som i stor grad kan sammenlignes med HiST-taleaudiometri. Meldingen  $m_i$  er opphavet til bølgeformen  $s_i(t)$  som sendes gjennom et medium hvor bølgeformen blir tilført støyen  $w(t)$ . Mottaker estimerer meldingen  $\hat{m}$ , som forhåpentligvis er lik  $m_i$ .



Figur 3.1: Blokkdiagram av et digitalt kommunikasjonssystem

### 3.1.1 Geometrisk representasjon

Ved å innføre ortonormale tidsfunksjoner som basisfunksjoner, kan et signal skrives som en sum av disse tidsfunksjonene [4]. Med et endelig antall basisfunksjoner kan ekspansjonen representeres i et Euklidsk rom som en vektor. Figur 3.2 viser et eksempel hvor 3 forskjellige signaler er representerte av to basisfunksjoner. Ved hjelp av geometrisk representasjon hadde man fått en elegant metode for å analysere ytelsen til ulike kommunikasjonssystemer.



Figur 3.2: Geometrisk representasjon av 3 signaler

$M$  energisignaler eller symboler kan representers som en lineær kombinasjon av  $N$  ortonormale basisfunksjoner.

$$s_i(t) = \sum_{j=1}^N s_{ij} \phi_j(t) \quad \begin{cases} 0 \leq t \leq T \\ i \in \{1, 2, \dots, M\} \end{cases} \quad (3.2)$$

Koeffisientene for  $s_i$  finnes ved

$$s_{ij} = \int_0^T s_i(t) \phi_j(t) dt \quad \begin{cases} i \in \{1, 2, \dots, M\} \\ j \in \{1, 2, \dots, N\} \end{cases} \quad (3.3)$$

som forutsetter ortonormalitet, altså at alle funksjonene er ortogonale og har energi lik 1. Det vil si

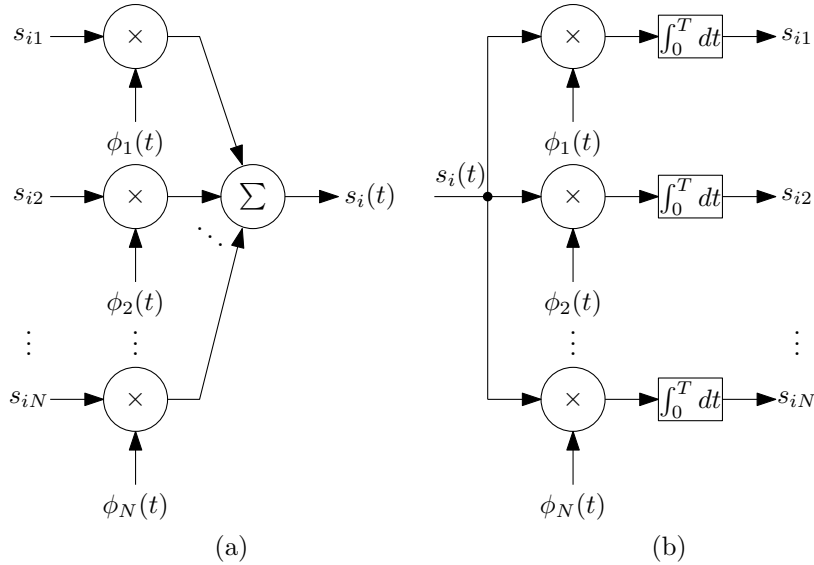
$$\int_0^T \phi_i(t) \phi_j(t) dt = \delta_{ij} = \begin{cases} 1 & \text{hvis } i = j \\ 0 & \text{hvis } i \neq j \end{cases} \quad (3.4)$$

Koeffisientene  $\{s_{ij}\}_{j=1}^N$  kan ses på som en  $N$ -dimensjonal vektor,  $\mathbf{s}_i$ .

$$\mathbf{s}_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{iN} \end{bmatrix} \quad (3.5)$$

Et hvert signal er eksakt beskrevet av signalvektoren  $\mathbf{s}_i$ .

Figur 3.3a viser hvordan man genererer et signal,  $s_i(t)$ , ved hjelp av ortonormale basisfunksjoner, mens figur 3.3b viser hvordan man gjenvinner signalvektoren  $\mathbf{s}_i$ .



Figur 3.3: (a) Synthesizer for å generere signalet  $s_i(t)$ . (b) Analysator for å generere signalvektoren  $\mathbf{s}_i$

### 3.1.2 Korrelasjonsmottaker

Analysatoren i figur 3.3b er det man kaller en korrelasjonsmottaker. Multiplikatoren og integratoren utgjør korrelatoren, som beregner indreproduktet mellom mottatt signal og en basisfunksjon. Utgangene på korrelatorene vil være koeffisientene  $s_{ij}$  som til sammen utgjør signalvektoren  $\mathbf{s}_i$ . I tillegg får man eventuell støy som adderes i kanalen.

Indreprodukt mellom to reelle tidsfunksjoner er gitt av

$$\langle f(t), g(t) \rangle = \int_0^T f(t)g(t)dt \quad (3.6)$$

Hvis nå  $s_i(t)$  overføres over en additiv hvit gaussisk kanal, som vist i figur 3.1, vil man få addert støy til signalet. Det mottatte signalet er da gitt av

$$x(t) = s_i(t) + w(t) \quad \begin{cases} 0 \leq t \leq T \\ i \in \{1, 2, \dots, M\} \end{cases} \quad (3.7)$$

hvor  $w(t)$  er en hvit gaussisk prosess. Utgangen av korrelator  $j$  er gitt av

$$\begin{aligned}
x_j &= \int_0^T x(t)\phi_j(t)dt & (3.8) \\
&= s_{ij} + w_j, \quad j \in \{1, 2, \dots, N\}
\end{aligned}$$

der  $s_{ij}$  er gitt av (3.3), og

$$w_j = \int_0^T w(t)\phi_j(t)dt \quad (3.9)$$

Når  $w(t)$  er en hvit gaussisk prosess med null middelvei er middelveien og varians for  $x_j$

$$\mu_{X_j} = s_{ij} \quad (3.10)$$

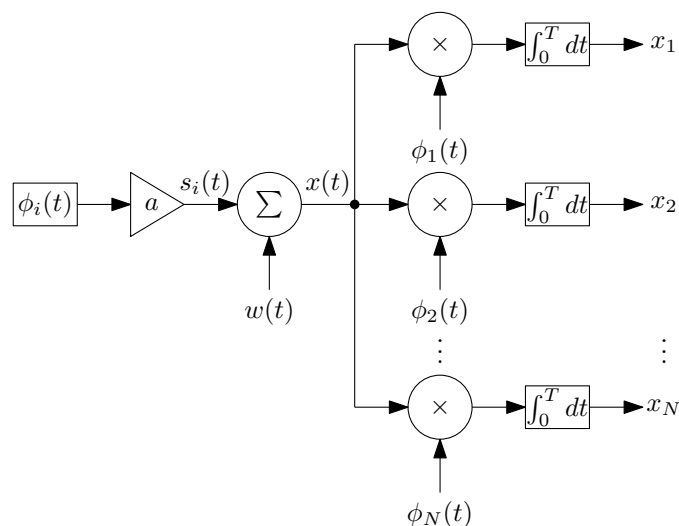
$$\sigma_{X_j}^2 = \frac{N_0}{2} \quad (3.11)$$

hvor  $N_0$  er støyeffektthet.

Ser man på alle utgangene under ett, får man en normalfordelt observasjonsvektor  $\mathbf{x}$ . Hvis de sendte symbolene har lik sannsynlighet, velger man i en Maximum Likelihood mottaker det symbolet som har en vektor som har minst avstand til den observerte vektoren.

## 3.2 Referanse-systemet

Systemet som brukes som modell for taleforståelighet, består av en sender som har like mange ortonormale basisfunksjoner som meldinger. Disse basisfunksjonene kan skaleres med faktoren  $a$ . Dette betyr at alle meldinger er ortogonale og har energien  $a^2$ . Figur 3.4 viser strukturen på systemet.



Figur 3.4: Korrelasjonsmottaker

Siden alle symbolene består av en skalert basisfunksjon, er det sendte signalet gitt av

$$\mathbf{s}_i = a \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (3.12)$$

hvor posisjonen til 1-tallet er gitt av indeksen  $i$ .

Mottatt signal er gitt av

$$\mathbf{x} = \mathbf{s}_i + \mathbf{w} \quad (3.13)$$

hvor  $\mathbf{w}$  er en vektor med støysamples fra en hvit gaussisk prosess.

### 3.2.1 Symbolfeilsannsynlighet

For å avgjøre hvilket symbol som er sendt, beregner man indreproduktet mellom det mottatte signalet og alle basisfunksjoner. Det indreproduktet som gir størst verdi vil være det riktige symbolet. Feil har man da hvis det riktige symbolet  $\phi_0$  har mindre indreprodukt med  $x$ , enn indreproduktet til  $x$  og et hvilket som helst av de andre symbolene  $\phi_i$ .

Sannsynlighet for feil

$$\begin{aligned}
 P_e &= P(\langle \mathbf{x}, \phi_0 \rangle < \langle \mathbf{x}, \phi_i \rangle) & (3.14) \\
 &= P(\langle \mathbf{s}_0 + \mathbf{w}, \phi_0 \rangle < \langle \mathbf{s}_0 + \mathbf{w}, \phi_i \rangle) \\
 &= P(\langle a\phi_0 + \mathbf{w}, \phi_0 \rangle < \langle a\phi_0 + \mathbf{w}, \phi_i \rangle) \\
 &= P(\overbrace{a\langle \phi_0, \phi_0 \rangle}^1 + \langle \mathbf{w}, \phi_0 \rangle < a\overbrace{\langle \phi_0, \phi_i \rangle}^0 + \langle \mathbf{w}, \phi_i \rangle) \\
 &= P(\langle \mathbf{w}, \phi_0 \rangle - \langle \mathbf{w}, \phi_i \rangle < -a) \\
 &= P(w < -a) \\
 &= P(w > a)
 \end{aligned}$$

Støyen  $w$  er normalfordelt, med et middel på 0 og varians på  $\sigma^2$ . Ved hjelp av Q-funksjonen (3.15), som brukes til å beregne arealet under halen på en normalfordeling, kan en regne ut sannsynligheten for feil.

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt, \quad x \geq 0 \quad (3.15)$$

Referert til figur 3.4, er effekten på det sendte signalet

$$\begin{aligned}
 \sigma_s^2 &= \frac{1}{T} \int_0^T s_i^2(t) dt & (3.16) \\
 &= a^2 \frac{1}{T} \overbrace{\int_0^T \phi_i^2(t) dt}^1 \\
 &= \frac{a^2}{T}
 \end{aligned}$$

Støyeffekten

$$\sigma_w^2 = \frac{N_0}{2} \cdot 2B = N_0 B \quad (3.17)$$

Sammenhengen mellom båndbredde og symbolengde for en Nyquist-kanal er

$$B = \frac{1}{2T} \quad (3.18)$$

Vi får da følgende uttrykk for signal-støy-forholdet

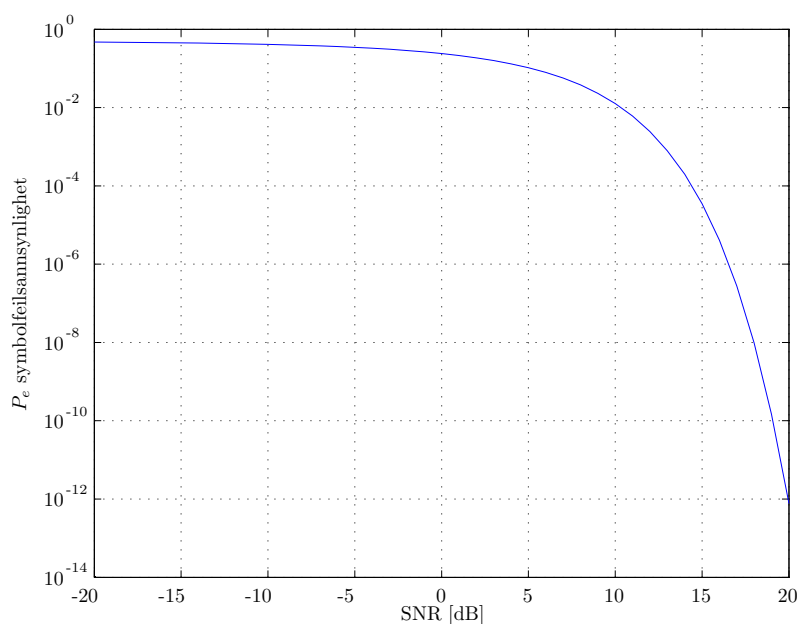
$$SNR = \frac{\sigma_s^2}{\sigma_w^2} = \frac{a^2/T}{B \cdot N_0} = \frac{2 \cdot B \cdot a^2}{B \cdot N_0} = \frac{2 \cdot a^2}{N_0} \quad (3.19)$$

Uttrykket for symbolfeilsannsynlighet som funksjon av SNR blir da

$$\begin{aligned} P_e &= Q\left(\frac{a}{\sqrt{2\sigma^2}}\right) \\ &= Q\left(\frac{a}{\sqrt{N_o}}\right) \\ &= Q\left(\sqrt{\frac{SNR}{2}}\right) \end{aligned} \tag{3.20}$$

Funksjon (3.20) er plottet i figur 3.5.

Figur 3.5 viser for symbolfeilsannsynligheten for modellen. Ytelsen er styrt av normalfordelingen, som fører til at man får det karakteristiske ”fossefallet“ ved ca 10dB.

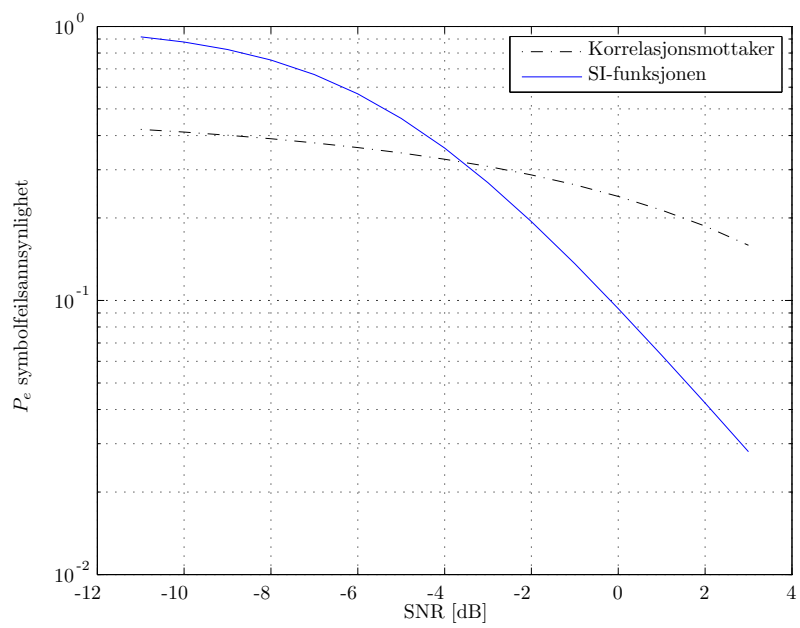


Figur 3.5: Symbolfeilsannsynlighet for modellen

### 3.3 Modell versus målte data

Figur 3.6 viser symbolfeilsannsynligheten til modellen, kontra den gjennomsnittlige symbolfeilsannsynlighet funnet i Øygaardens felttest. Grafen er kun plottet mellom  $-11$  dB og  $+3$  dB, da data for  $-12$  dB fra felttesten var for upålitelig.

Grafen viser at modellen gjør det best ved lave SNR, mens HiST-talaudiometri, og dermed mennesketale, gjør det beste fra ca.  $-3.5$  dB og oppover.



Figur 3.6: Symbolfeilsannsynlighet for korrelasjonsmottakeren kontra SI-funksjonen

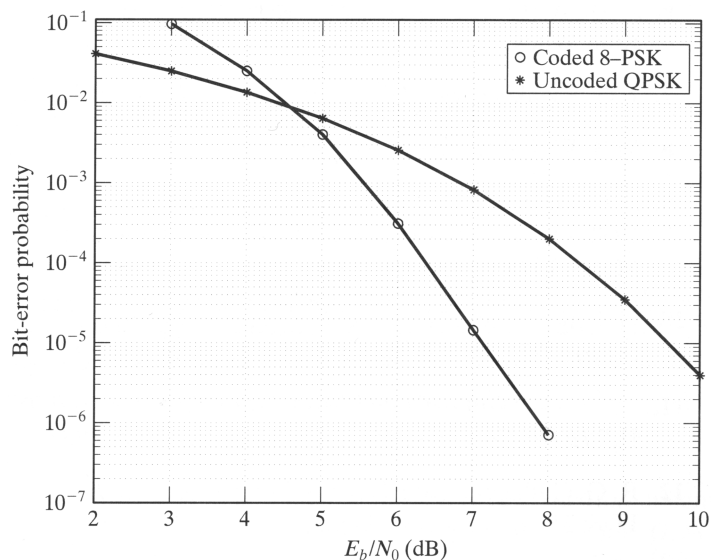


## Konklusjon

Symbolfeilsannsynligheten for HiST-taleaudiometri og taleforståelighetsmodellen for seg selv, er ikke uventede. Det er også intuitivt at korrelasjonsmottakeren vil gjøre det bedre enn det menneskelige øret, siden modellen for taleforståelighet benytter seg av ortogonale symboler. Talespråket består derimot ikke av ortogonale symboler, dette vil gjøre det vanskeligere å skille symboler.

Det uventede resultatet er at øret og tale, ved et signalstøyforhold på mer enn ca.  $-3.5$  dB, har en lavere symbolfeilsannsynlighet enn korrelasjonsmottakeren.

En mulig forklaring på at HiST-taleaudiometri, og da tale-kanalen, gjør det bedre enn referansesystemet ved høyere SNR, kan man finne ved å se til informasjonsteori og da spesielt kodingsteori. Ved å sammenligne figur 3.6, og 4.1 som er hentet fra [5], ser man likhet i de kryssende grafene. Figur 4.1 viser at ved å innføre redundans i et signal ved hjelp av koding, vil det kunne gi en lavere feilsannsynlighet ved et gitt SNR. Men, ved lavt nok SNR vil det ukodede signalet gjøre det best.



Figur 4.1: Forskjell i feilsannsynlighet for et kodet og ukodet system som funksjon av SNR

Referansesystemet er en veldig enkel modell som ikke innhar noe koding eller redundans av noe form. Talespråket derimot inneholder redundans. Et godt eksempel på at tale inneholder redundans er telefonkanalen. En telefonkanal har en båndbredde fra 300 til 3400 Hz, mens talsignalet i seg selv har et spekter som strekker seg over hele det hørbare området. Likevel er det mulig å føre en normal samtale via en telefonkanal. Grunnen til dette er redundans og at mesteparten av informasjon ligger ved de lavere frekvenser. Konsonantene “s” og “f” kan være vanskelige å skille på en telefonkanal da de inneholder mye energi ved høyere frekvenser, men de lar seg skille ved hjelp av kontekst.

Det kan også tenkes at man ved lave SNR bruker mer mental kapasitet til å oppfatte ordene, og derfor har mindre kapasitet til å analysere setningen og ordenes mening, og derfor ikke så lett greier å forutse hvilke ord som kommer. Ordene blir på denne måten mer overraskende: man rekker ikke å dekode det forrige ordet før det kommer ett nytt.

# Tillegg

Tabell A.1 viser data fra Øygardens første felttest. Den viser gjennomsnittlig taleforståelighet i % for ordene ved forskjellige SNR. Det er også oppgitt antallet personer som er testet ved hver SNR.

Tabell A.1: Scoringstabell etter felttest

Antall-målt	30	18	33	26	30	18	33	26	30	18	33	26	30	18	33	26
SNR[dB]	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3
Hedda	0	0	18	4	87	67	52	38	37	89	64	100	97	100	91	96
Ida	0	0	15	8	23	39	76	8	53	100	100	65	90	100	100	100
Malin	0	0	33	31	7	44	79	88	50	94	73	92	97	100	97	100
Ingvild	0	0	6	4	20	28	42	4	43	11	48	85	57	28	82	77
Thea	0	0	24	8	10	22	73	85	80	78	82	81	97	100	100	100
Benjamin	0	0	6	15	7	11	85	62	57	72	82	96	100	100	100	100
Jonas	23	6	6	4	17	39	24	46	87	94	91	100	93	94	100	100
Thomas	3	33	3	58	60	67	88	65	73	100	97	88	100	100	100	92
Magnus	0	0	6	81	90	56	82	92	80	94	97	96	100	100	100	100
Eivind	0	0	3	4	3	28	15	31	33	78	76	54	80	83	88	92
ga	3	0	0	0	0	0	0	46	50	39	76	96	33	100	94	77
grep	0	0	0	0	7	28	27	38	30	50	55	69	50	89	94	69
ser	3	0	0	0	27	50	42	23	87	28	39	96	100	100	94	100
vant	0	0	21	0	50	0	36	8	67	50	82	81	97	89	94	100
låner	0	0	0	0	0	11	30	4	60	17	24	88	100	89	97	96
eide	0	0	0	38	7	11	39	19	30	50	82	96	100	56	100	92
flytter	0	0	6	0	13	0	3	0	27	33	73	65	73	83	82	100
viser	0	0	3	4	33	28	64	62	60	89	73	100	87	100	100	100
har	7	0	27	42	17	33	33	58	80	78	85	88	100	83	97	100
tok	0	0	3	15	7	0	27	23	10	78	79	65	100	89	97	100
to	0	0	15	4	7	11	3	23	80	44	76	81	93	67	85	100
tre	0	0	6	8	20	0	79	69	77	89	85	88	97	94	97	100
fire	3	0	18	50	0	44	48	27	83	89	82	96	100	94	91	96
fem	0	22	3	19	17	50	55	65	87	67	70	92	100	89	97	100
seks	70	28	67	73	60	89	97	88	93	100	94	92	97	100	91	100
sju	17	0	6	0	70	72	76	85	73	78	100	96	100	94	100	100
åtte	0	0	0	8	13	11	30	27	50	50	73	88	100	100	97	96
elleve	0	0	6	4	7	6	24	73	77	61	85	85	97	72	100	92
tolv	3	0	0	4	10	28	64	8	57	89	100	92	87	89	100	100
atten	0	0	15	35	3	0	73	27	27	89	55	81	97	100	97	92
gamle	0	0	9	12	27	72	21	85	77	33	100	88	100	100	100	100
hele	0	0	3	0	3	17	36	27	63	61	48	81	93	61	79	85
store	0	0	9	8	7	56	79	73	100	89	91	100	93	94	100	100
nye	3	0	6	8	10	11	61	81	90	67	94	73	100	83	100	100
vakre	0	0	3	27	33	56	73	50	87	56	100	69	83	50	100	96
mørke	3	0	9	0	10	11	12	23	73	56	18	77	100	83	82	100
lyse	0	0	3	15	13	11	76	69	87	89	79	73	100	94	100	100
fine	0	0	9	15	23	6	27	96	70	61	88	92	93	100	97	96
lette	0	6	0	0	7	6	12	31	17	28	79	81	63	83	91	92
svarte	10	0	15	65	53	78	67	96	80	89	91	92	100	100	97	88
knapper	3	0	42	23	23	50	82	88	97	94	85	81	87	100	100	85
boller	13	0	15	4	60	17	61	69	57	100	97	92	100	94	100	100
vanter	13	0	3	38	30	39	82	88	97	94	91	96	100	100	97	100
penner	0	0	3	19	3	0	85	85	93	100	79	96	100	89	88	96
kurver	0	0	6	4	27	6	15	23	33	56	64	81	93	100	97	100
skåler	3	0	18	46	70	28	94	100	93	100	100	96	100	100	94	96
luer	0	0	6	8	0	6	48	35	20	44	67	81	93	89	88	96
duker	0	0	3	0	3	22	12	65	60	72	82	88	97	100	100	96
ringer	0	0	12	31	60	56	58	46	73	72	79	77	90	100	100	92
kasser	27	39	21	77	67	78	64	100	93	78	94	96	100	94	97	100



# Bibliografi

- [1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
- [2] J. Øygarden, *Norwegian Speech Audiometry*. PhD thesis, Norwegian University of Science and Technology, 2009.
- [3] B. M. B. Kollmeier, T. Brand, *Springer Handbook of Speech Processing*, ch. 4. Perception of Speech and Sound, pp. 61–82. Springer, 2008.
- [4] S. Haykin, *Communication Systems*, ch. 5, pp. 309–343. John Wiley & Sons, 2001.
- [5] D. J. C. Shu Lin, *Error Control Coding*. Pearson Education International, 2004.