



## **Talegjenkjenning i norske bedrifter**

*En casestudie av effekter og muligheter talegjenkjenning gir norske bedrifter*

**Kaja Hørland Østby og Malin Dahl**

**Veileder: Andreas Haller**

Masterutredning i økonomi og administrasjon

Hovedprofil: Business Analytics

NORGES HANDELSHØYSKOLE

Dette selvstendige arbeidet er gjennomført som ledd i masterstudiet i økonomi- og administrasjon ved Norges Handelshøyskole og godkjent som sådan. Godkjenningen innebærer ikke at Høyskolen eller sensorer inntår for de metoder som er anvendt, resultater som er fremkommet eller konklusjoner som er trukket i arbeidet.

## Forord

Masterutredningen er gjennomført som en del av masterstudiet i økonomi og administrasjon ved Norges Handelshøyskole (NHH), med en spesialisering i Business Analytics. Temaet for masterutredningen er talegjenkjenning som en viktig del av kunstig intelligens, og ble valgt på bakgrunn av interesse og nysgjerrighet for temaet. Temaet er svært dagsaktuelt, og det er stadig flere norske bedrifter som er nysgjerrige på å implementere talegjenkjenning.

Utredningen tar opp viktige utfordringer, muligheter og effekter av implementering av talegjenkjenning i norske bedrifter. Formålet er at masterutredningen skal bidra til økt kunnskap om effekter og muligheter talegjenkjenning gir norske bedrifter. I gjennomføringen av utredningen har vi blant annet gjennomført kvalitative intervjuer av ansatte i ulike bedrifter som er eksperter innenfor talegjenkjenning, og det har gitt oss mulighet til å få fordype oss og få verdifull innsikt i temaet. I denne sammenheng ønsker vi å rette en stor takk til våre informanter som har valgt å sette av tid til å være med på studiet og dele verdifulle erfaringer og kunnskap. Innsikten fra informantene har vært avgjørende for resultatet av utredningen.

Avslutningsvis ønsker vi å rette en spesiell stor takk til vår veileder, Andreas Haller, for gode råd, verdifulle og konstruktive tilbakemeldinger. Vi har satt pris på at du har vært tilgjengelig for samtaler og spørsmål underveis gjennom arbeidet med vår masterutredning.

Norges Handelshøyskole  
Bergen, 01.juni 2022

Kaja Hørland Østby

Malin Dahl

## **Sammendrag**

Denne masterutredningen studerer hvilke effekter og muligheter talegjenkjenning gir for norske bedrifter. Utredningen består av et casestudie og et dokumentstudie, hvor vi gjennom en kvalitativ tilnærming har tilegnet oss verdifull innsikt for å belyse problemstillingen gjennom ulike bedrifter. Vi har også gjennomført et dokumentstudie av en bedrifts interne dokumenter for å analysere effekten av implementering av talegjenkjenning på kundesentre. Utredningen vil i hovedsak være et bidrag til litteraturen når det gjelder effekter og muligheter av talegjenkjenning for norske bedrifter, ettersom det er begrenset med litteratur rundt temaet for det norske språket. I tillegg vil vårt studie bidra til å se på effektivitet i form av lønnsomhet for talegjenkjenning på kundesentre.

Utredningen viser at effekter talegjenkjenning gir for norske bedrifter kan være store kostnadsbesparelser, og at norske bedrifter opplever forbedring i effektivitet. Funn fra dokumentstudiet indikerer en avkastning på investering av talegjenkjenning for bedriften på 60.62%. I tillegg viser utredningen at implementering av talegjenkjenning i norske bedrifter også kan gi andre ikke-kostnads relaterte fordeler, som for eksempel ergonomiske fordeler, bedre kvalitet og informasjonsflyt. Utredningen viser også at effektene og mulighetene talegjenkjenning gir, er avhengig av utviklingen av teknologien og modellene, samt viktigheten av at modellene trenes opp på riktig datamateriell i forhold til hva modellene skal brukes til. Funnene våre indikerer også at mange fremtidige muligheter for norske bedrifter som bruker talegjenkjenning avhenger av hvordan teknologien brukes. De største utfordringene med kvaliteten av talegjenkjenning for det norske språket indikerer funnene våre at er relatert til mangfoldet av dialekter.

# Innholdsfortegnelse

<b>1.0 INTRODUKSJON</b> .....	<b>1</b>
<b>2.0 TEORI</b> .....	<b>3</b>
<b>2.1 TALEGJENKJENNING</b> .....	<b>3</b>
2.1.1 Begrepsforklaring.....	3
2.1.2 Teknisk forklaring.....	4
2.1.3 Historisk perspektiv.....	4
2.1.4 Nåværende perspektiv.....	6
2.1.5 Norsk talegenkjenning.....	8
2.1.6 Talegenkjenning i ulike bedrifter.....	10
2.1.7 Fremtidig perspektiv.....	12
<b>3.0 METODE</b> .....	<b>14</b>
<b>3.1 FORSKNINGSDESIGN</b> .....	<b>14</b>
<b>3.2 DATAINNSAMLING</b> .....	<b>14</b>
3.2.1 Primærdata.....	15
3.2.2 Sekundærdata.....	18
<b>3.3 DATAANALYSE</b> .....	<b>18</b>
<b>4.0 PRESENTASJON AV FUNN</b> .....	<b>19</b>
<b>4.1 DOKUMENTSTUDIE</b> .....	<b>19</b>
4.1.1 Beskrivelse av prosjektet.....	19
4.1.2 Kostnader.....	21
4.1.3 Effekter etter implementering.....	22
<b>4.2 UTFORDRINGER VED BRUK AV TALEGJENKJENNING</b> .....	<b>25</b>
4.2.1 Kvaliteten på talegenkjenning for det norske språket.....	25
4.2.2 Det menneskelige perspektivet kan ikke erstattes.....	27
4.2.3 Nyten av talegenkjenning.....	28
4.2.4 Tilgang på datamateriale.....	29
4.2.5 Endringsvilje.....	30
<b>4.3 FORDELER VED BRUK AV TALEGJENKJENNING</b> .....	<b>31</b>
4.3.1 Effektivitet og reduserte kostnader.....	31
4.3.2 Kvalitet.....	33
4.3.3 Informasjonsflyt.....	34
4.3.4 Ansatte og ergonomi.....	35
<b>4.4 MULIGHETER VED BRUK AV TALEGJENKJENNING</b> .....	<b>36</b>
4.4.1 Innovative muligheter.....	36
4.4.2 Uendelige muligheter.....	36
<b>4.5 FREMTIDEN FOR TALEGJENKJENNING</b> .....	<b>37</b>
<b>5.0 DISKUSJON</b> .....	<b>39</b>
<b>5.1 MULIGHETER FOR KUNDESENTERET</b> .....	<b>39</b>
5.1.1 Bruke transkribering etter en kundesamtale.....	39
5.1.2 Forbedre kunderelasjoner.....	40
<b>5.2 EFFEKTER I ANDRE NORSKE BEDRIFTER</b> .....	<b>41</b>
5.2.1 Helsesektoren.....	42
5.2.2 Undervisningssammenheng.....	43
5.2.3 Andre norske bedrifter.....	43
<b>5.3 MULIGHETER I FREMTIDEN</b> .....	<b>44</b>
5.3.1 Forbedre bruken av tekstanalyse.....	45
5.3.2 Bedre modeller.....	46

<b>5.4 OPPSUMMERING AV RESULTATER</b> .....	47
<b>5.5 IMPLIKASJONER</b> .....	49
<b>6.0 KONKLUSJON</b> .....	51
<b>LITTERATURLISTE</b> .....	53
<b>VEDLEGG</b> .....	56
<b>A1 INTERVJUGUIDE</b> .....	56

## **Figurliste:**

4.1: Illustrasjon av hvordan automatisk talegjenkjenning fungerer på kundesenteret.....	20
4.2: Illustrasjon av plattformen kundesenteret benytter som oversikt over kundesamtaler.....	20
4.3: Illustrasjon av en automatisk transkribert kundesamtale.....	21

## Tabell liste:

3.1: Oversikt over informanter og beskrivelse av rolle.....	16
4.1: Oversikt over kostnader for implementering av talegjenkjenning i bedriften.....	22
4.2: Oversikt over tidsforbruk før og timer spart etter implementering av talegjenkjenning.....	23
4.3: Median årslønn og timelønn kundesentermedarbeider.....	23
4.4: Oversikt over tid- og kostnadsbesparelser i 2021.....	24
4.5: Oversikt over driftskostnader i forbindelse med implementering av talegjenkjenning og totalt spart.....	24
4.6: Oversikt over kostnader i forbindelse med prosjektet, totalt spart og avkastning på investering (ROI).....	25

## 1.0 Introduksjon

De seneste årene har kunstig intelligens vokst i rekordfart, og talegjenkjenning er en viktig del av denne utviklingen. Talegjenkjenning har eksistert siden 80-tallet, men formålet, mulighetene og bruken av teknologien har utviklet seg betraktelig de seneste årene. Dette innebærer økt nysgjerrighet og interesse for talegjenkjenning, både blant privatpersoner og norske bedrifter. Bedrifter er i dag stadig på utkikk etter nye metoder for å effektivisere det administrative arbeidet, og i denne sammenheng vurderer mange bedrifter implementering av talegjenkjenning. Talegjenkjenning kan blant annet brukes til å skrive pasientjournaler automatisk i helsesektoren, men alt tyder på at bruken av talegjenkjenning ikke stopper der. Den økende trenden for talegjenkjenning løfter frem flere utfordringer; Er kvaliteten på norsk talegjenkjenning god nok til profesjonelt bruk? Gir investeringen i teknologien nytteverdi for brukerne, og i så fall på hvilken måte? Implementering av talegjenkjenning i en bedrift er gjerne forbundet med store investeringer, endring i arbeidsmetodikk og ressursforbruk. For at teknologien skal være en verdifull investering for bedriften, er det viktig at den tjener dens formål.

Den økende interessen for talegjenkjenning, samt vår nysgjerrighet rundt effekter og muligheter en slik investering kan gi for norske bedrifter var vår motivasjon for å studere temaet nærmere. Denne oppgaven tar derfor opp følgende forskningsspørsmål:

*“Hvilke effekter og muligheter gir talegjenkjenning for norske bedrifter?”*

Vi studerer det ved å intervjuere eksperter innenfor talegjenkjenning i ulike bransjer og stillinger, for å få deres forståelse av fenomenet. Vi ser også på effekten av talegjenkjenning ved å gjennomføre et dokumentstudie av en bedrifts interne dokumenter, hvor vi analyserer resultater etter implementering av talegjenkjenning på deres kundesentre.

Eksisterende forskning rundt talegjenkjenning for det norske språket argumenterer ofte at talegjenkjenning er utfordrende rent teknisk, spesielt på grunn av mangfoldet av dialekter (Solberg & Ortiz, 2022). Til tross for dette finner vi likevel begrenset litteratur som undersøker effektene av talegjenkjenning i norske bedrifter. Tidligere internasjonale studier viser positive effekter og muligheter ved bruk av talegjenkjenning i spesielt helsesektoren



(Alaiad, 2020). Likevel finner vi den eksisterende litteraturen mangelfull når det gjelder å undersøke om en investering i talegjenkjenning er lønnsomt, dette gjelder spesielt for norske bedrifter. Studiet vårt vil derfor være et bidrag til litteraturen rundt talegjenkjenning for det norske språket, i tillegg til at vi studerer effekten av talegjenkjenning i lys av lønnsomhet. På en annen side, vil studiet være et bidrag til litteraturen i form av et fremtidig perspektiv på muligheter med talegjenkjenning, sett i lys av norske bedrifter.

Utredningen vår viser at bruk av talegjenkjenning for norske bedrifter gir positive effekter i form av sparte kostnader, og at bedrifter som har implementert talegjenkjenning opplever stor forbedring i effektivitet. Funn fra dokumentstudiet indikerer en avkastning på investering av talegjenkjenning for bedriften på 60.62%. Vi har også gjort funn rundt andre positive effekter som ikke er relatert til kostnader, som for eksempel ergonomiske fordeler, høyere kvalitet og bedre informasjonsflyt. Effektene og mulighetene av talegjenkjenning avhenger av hvordan modellene er utviklet, og at datamaterialet modellene er trent opp med er relatert til hva modellene skal benyttes til. Funnene våre indikerer også at mange fremtidige muligheter for norske bedrifter som bruker talegjenkjenning avhenger av hvordan teknologien brukes. Utredningen vår viser også at utfordringer relatert til kvaliteten av talegjenkjenning for det norske språket ofte henger sammen med mangfoldet av dialekter.

Opgaven er strukturert på følgende måte: I kapittel 2 redegjør vi for begrepet talegjenkjenning, og gir en oversikt over talegjenkjenning i et nåværende og historisk perspektiv. Vi presenterer også relevant forskning rundt norsk talegjenkjenning, bruken av det i ulike bedrifter og fremtidig perspektiv på teknologien. I kapittel 3 presenteres metodene som benyttes for å besvare forskningsspørsmålet. Funnene vi har gjort for å belyse effektene og mulighetene for talegjenkjenning er presentert i kapittel 4. I kapittel 5 diskuteres funnene våre i lys av eksisterende litteratur presentert i kapittel 2, samt en diskusjon av oppgavens implikasjoner. Til slutt vil vi i kapittel 6 presentere en konklusjon som skal besvare oppgavens forskningsspørsmål, samt en oversikt over oppgavens begrensninger.

## 2.0 Teori

Siden talegjenkjenning har eksistert over en såpass lang tid og er under en stadig utvikling, har det vært hensiktsmessig å få en oversikt over eksisterende litteratur før vi selv belyser temaet med egne funn.

### 2.1 Talegjenkjenning

#### 2.1.1 Begrepsforklaring

I oppgaven omtaler vi stadig *talegjenkjenning* som et begrep. Dette begrepet omfavner mye, og vi vil derfor gi en oversikt over det mest relevante. Talegjenkjenning som et begrep er det å gjenkjenne menneskelig tale ved hjelp av et system og en maskin. Dette kan også gjerne kalles tale-til-tekst eller automatisk talegjenkjenning, og vi fokuserer på idéen om å gjenkjenne talen for så og få det ned i tekstformat. Hovedpoenget er at menneskelig tale skal registreres, gjenkjennes og identifiseres av datamaskiner for så å kunne anvendes til ulike formål. Et formål kan være å kunne talestyre en maskin ved hjelp av konkrete kommandoer, mens et annet er å kunne skrive ned nøyaktige, digitale tekster og dokumenter uten å bruke et tastatur (Store norske leksikon, 2019). I mange tilfeller kan også hovedformålet være å behandle og analysere denne teksten videre, noe vi omtaler som *tekstanalyse* i denne oppgaven. Tekstanalyse er også selve teknologien vi kaller Natural Language Processing (NLP).

Samtidig referer vi til talegjenkjenning som en konkret programvare og et utviklet system som kan anvendes av personer og som kan implementeres av bedrifter. Hva som innebærer i dette systemet varierer mellom ulike leverandører, andre utviklere av systemet og mellom ulike forskningsgrupper. Disse systemene har ulike navn, som for eksempel Telenor Norge som kaller systemet eller prosjektet sitt for Speech-2-text (S2T), mens Apple kaller sitt system blant annet for Voice Control. Alt dette er talegjenkjenning, men det er med andre ord ulike navn på systemer og modeller som er bygd opp på bakgrunn av ulike formål.

*Transkribering* er et annet begrep vi bruker i forbindelse med talegjenkjenning. Dette er selve prosessen der tale blir til tekst. Transkribering i seg selv er den generelle nedskrivningen, som også kan gjøres manuelt av en person. Ved forskning og systemutvikling kan for eksempel transkribering være den prosessen der ansatte manuelt skriver ned all tekst og eventuelle

lyder fra et lydopptak som senere skal testes opp mot en maskin. Automatisk transkribering er så i de tilfellene der maskinen selv transkriberer denne talen om til tekst. Likevel blir disse begrepene brukt litt om hverandre i dagligtale og derfor naturligvis også under intervjuprosessene våre.

### 2.1.2 Teknisk forklaring

Talegjenkjenning er en del av den kunstige intelligensen vi har tilgjengelig i dagens samfunn, og i likhet med all annen teknologi innenfor dette feltet har talegjenkjenning utviklet seg med årene. Vi skal gå gjennom det historiske perspektivet og ikke minst identifisere det nåværende perspektivet rent teknisk, men innledningsvis vil vi forklare hvordan talegjenkjenning som programvare generelt fungerer i praksis. Dette er på et overfladisk perspektiv teknisk sett, men det er med for å forstå konseptet av teknologien.

Uavhengig av formålet og hva talegjenkjenningen skal brukes til videre, fungerer det slik at man snakker til programvaren gjennom en lydenhet, som for eksempel en mikrofon. Den enheten du snakker til registrerer denne lyden som bølge-filer, og disse filene skal deretter renses for blant annet bakgrunnsstøy. Videre blir disse filtrerte filene brutt ned til såkalte fonemer (Zajechowski, U.å). Et fonem er en lyd, eller en klasse av like lyder som har samme funksjon i et språk eller i en dialekt. Fonemer brukes blant annet for å skille like ord fra hverandre, slik som at for eksempel ordene “hater” og “heter” skilles kun gjennom de to fonemene “a” og “e”. Dessuten kan uttalen på fonemene variere etter språk og dialekt (Store norske leksikon, 2019). For eksempel kan to ulike lyder bety det samme i tekstform, slikt som at en bergenser sier ordet “hater” med skarre-r, mens en østlending heller ruller på r-en. I dette tilfellet er det da ikke to ulike fonemer, men to varianter av samme (Engan et.al, 2018). Felles for disse eksemplene er at fonemene er avgjørende for hva slags betydning ordene, og deretter hele setninger har. Ved å bruke statistisk sannsynlighetsanalyse kan programvaren nå utlede fonemene til komplette setninger (Zajechowski, U.å). På denne måten blir ordene man sier skrevet ned i tekstformat, men med et forbehold om at programvaren har analysert lydene riktig.

### 2.1.3 Historisk perspektiv

For å få en forståelse av bruken og implementeringen av talegjenkjenning mot norske bedrifter, er det relevant å se på det historiske perspektivet. På den måten vil man få en bedre

forståelse av utgangspunktet og formålet til teknologien og hvordan det har utviklet seg til å bli slik vi kjenner det i dag.

I likhet med kunstig intelligens generelt har talegjenkjenning en lang historie med flere bølger av store innovasjoner. Å kunne anvende en maskin som etterligner menneskelig atferd, spesielt evnen til å reagere riktig på vårt daglige talespråk, har fascinert forskere og utviklere i århundre. Utviklingen av talegjenkjenning har beveget seg gjennom siste halvdel av 1900-tallet. Det har gått fra en enkel maskin som klarer å reagere på noen konkrete og bestemte lyder, til et svært omfattende system som reagerer på mer spontan flytende tale og som tar hensyn til variasjon i individuelle språk, dialekter, stemmer og toneleier (Juang et.al, 2014, s.6). I løpet av 80-tallet brukte man disse bestemte lydene som systemet klarte å gjenkjenne til bruk spesielt over telefonen. Dette var for eksempel å kunne be telefonen om å ringe en spesiell person fra en kontaktliste eller fra et sentralbord, eller si et nummer høyt for så å ringe det. Dette kan kategoriseres som regelbaserte modeller under kunstig intelligens, fordi ordene og tallene var registrert som regler og lagt inn manuelt i systemene på forhånd. Det var altså en teknologisk ordbok som oversatte direkte og det gikk derfor ikke an å vike fra reglene eller at programvaren justerte seg selv underveis (Kidawara et.al, 2019, s.5).

Utover 90-tallet kom ulike applikasjoner av talegjenkjenning på banen som ga rom for mer kontinuerlig tale. Det som før bare var kommandoer og enkle, innlærte tall og ord ble nå hele setninger som tilnærmer seg det vi kjenner som maskinlæring av datadrevne modeller i dag. Dette feltet har med andre ord dratt fordeler av fremskritt rundt bruk av stor-data. Man går nå fra å ha skapt et program som kan gjøre enkle oppgaver for deg, til å begynne å utvikle noe som er like raskt som menneske selv og som kan ta komplekse beslutninger. Med dette ser vi at talegjenkjenningen har utviklet seg i takt med datamaskiner og internettutviklingen, og et banebrytende talegjenkjenningssystem kalt skjulte Markov-modeller (HMM) kom derav frem på 2000-tallet. HMM er en modell som fortsatt er populær for ulike talegjenkjenningssystemer nettopp fordi de kan trenes automatisk (Juang et.al, 2014, s.12). Dette kalles uovervåket læring, som for eksempel at modellen registrerer typiske ord eller setninger som brukeren anvender mest og legger så dette inn i systemet. På den andre siden ble denne modellen også brukt i samsvar med overvåket læring. Rent praktisk for talegjenkjenning betyr dette prosessen der en person har transkribert en lydfil fra før av, og måler dette opp mot modellens automatiske transkribering. Her kan personen da gå inn og rette på eventuelle feil og lære modellen til hva som skal transkriberes neste gang. I likhet

med den generelle teknologien for kunstig intelligens er dette en tidkrevende prosess som avhenger av menneskelige ressurser, så det er derfor banebrytende for talegjenkjenning å nærme seg det vi i dagens perspektiv omtaler på som nevralt nettverk og Natural Language Processing (NLP).

#### 2.1.4 Nåværende perspektiv

Siden kunstig intelligens stadig er i utvikling i takt med samfunnet, er det viktig å få en oversikt over det nåværende perspektivet på talegjenkjenning. Dette er for å få en bedre oversikt over hvilken situasjon, både i form av muligheter og begrensninger, som bedrifter som implementerer talegjenkjenning står ovenfor rent teknisk.

NLP omhandler tilnærminger for å behandle, forstå og analysere det menneskelige språket (Hagiwara, 2021, s.19). Dette viser seg å ha en stor betydning for den retningen teknologien er på vei. Måten det fungerer på er at teknologien er utviklet for å kunne simulere så godt som mulig hvordan mennesker forstår tale og deretter reagere. Et typisk vokabular i et slikt system består gjerne av over 60 000 ord, noe som igjen består av tilnærmet uendelig av ordkombinasjoner. Det sier seg selv at det ville tatt svært lang tid for maskinen å lete gjennom alle disse ordene for å “gjette” hvilket ord som ble sagt, så systemet er derfor bygd opp slik at det fokuserer på utnevnte søkeord som gir mening til resten av setningene (Zajechowski, U.å). For å ta et eksempel, så leser man inn setningen “Det er fint vær i dag, det er ser ikke ut som det skal bli noe vind” og som skal automatisk transkriberes. Her kan både “vær” og “vind” være to utnevnte søkeord, for hvis systemet oppfatter ett av disse ordene er det mer sannsynlig at “vind” står sammen med “vær”, enn de tilsvarende like ordene “vinn” og “hver”. Disse to ordene har en helt annen betydning på norsk og ville skapt en uriktig setning, som for eksempel “Det er fint hver i dag”.

I tillegg til denne teknologien består også dagens talegjenkjenningssystemer ofte av dyp læring, som er en del av den tyngre maskinlæringen under NLP. Det er et system som lærer underveis og som stadig er under opplæring. Rent praktisk betyr dette at talegjenkjenningssystemet spesialisere seg og blir god på den personen som bruker det. Her trenes modellen opp på ditt ordforråd og det foregår gjerne gjennom at personen starter med å lese inn en setning på sin naturlige måte (Huang et.al, 2014). Dette kan sammenlignes med installeringen av for eksempel “Siri” eller “Google Assistant” på en ny enhet. Her starter man gjerne med å si “Hei Siri” på forskjellige måter, slik at systemet skal bli kjent med stemmen

din og reagere hver gang du sier noe. For talegjenkjenning som fokuserer mer på nøyaktighet av ord og ikke nødvendigvis hvem som snakker, så kan det være viktigere å være klar over hvilke typiske ord og setningsoppbygninger en person har. På den måten øker sannsynligheten for at det blir automatisk transkribert så fort og nøyaktig som mulig. Det kan for eksempel derfor være aktuelt at maskinen lærer hvilket språk du snakker eller kombinerer i talen din, hvilken setningsoppbygging du foretrekker og ikke minst hvilket stemmeleie du vanligvis ligger på.

For at systemene i utgangspunktet skal kunne lære underveis er modellene avhengig av å ha blitt trent opp av relevant data. I dag har vi store muligheter for å samle inn det vi kaller stor-data og som gir grunnlag for mye variert materiale som modellene kan trenes opp på. Dataen det er snakk om er ulike former av lydopptak og tale, samt gode og utformede ordbøker. Internett har gjort det mulig å tilgjengeliggjøre store mengder data. For eksempel gir internettilgang til en del lydopptak av daglig tale fra ulike sosiale plattformer. Ikke minst har "Siri" og andre lignende applikasjoner vært en rikelig kilde til taledata fra brukerne deres (Huang et.al , 2014).

Typisk for å designe modeller for talegjenkjenning er åpenbart å bruke data som er av interesse og som tjener formålet for systemet. Dersom formålet er å automatisk transkribere et referat fra et møte på Stortinget, bestående av ulike advokater og jurister, vil det være relevant å bruke et system som har et godt juridisk ordforråd. Her kan altså modellen være trent opp på innøvd tale, for eksempel fra tidligere juridiske taler eller opplesninger av noe slag. Dessuten er det desto viktigere å ha modeller som er trent opp på daglig og spontan tale, fordi et møte består gjerne ikke bare av opplesning av manus, men av mer uformell dialog mellom partene. Dersom man bruker modeller som er trent opp på feil data, kalles dette "Mismatch problemet", nemlig at det er for store variasjoner enn hva modellen i utgangspunktet ble trent på (O'Shaughnessy, 2008, s.2). Likevel er det også viktig at disse modellene ikke blir "overtrent" på relevant data. Dersom det meste av datamateriale i dette eksempelet består av en person som helt monotont leser inn juridisk, tunge setninger kan modellen feile totalt i tilfeller der slik formell tale ikke er til stede. For å unngå dette må modellene derfor ideelt sett være testet på variert data, og for å overvåke kvaliteten blir dette testet opp mot treningsdata som er manuelt transkribert (O'Shaughnessy, 2008, s.7).

Siden vi på nåværende tidspunkt har en mengde ulike talegjenkjenningssystemer utviklet av ulike leverandører, er det viktig å påpeke at den største forskjellen mellom disse systemene er dataen den er trent opp på. Det betyr altså at modellene er i utgangspunktet ganske like, men at hvordan de er trent opp skiller de fra nivå av blant annet nøyaktighet, stemmegjenkjenning, støyfiltrering og andre tilegnede egenskaper. Disse egenskapene er tilført etter hvilket marked leverandørene sikter mot, og hva etterspørselen er deretter.

Dette leder videre til de utfordringene som talegjenkjenning står overfor i et nåværende perspektiv. Selv om maskinene skal lære underveis samtidig som man har lett tilgang på store mengder data for å trene de opp, skal det meste i begynnelsen testes opp mot treningsdata. Treningsdata er manuelt transkribert av en person og skal være fasiten i et slikt tilfelle. Et lydopptak på flere timer vil måtte skrives ned manuelt ved hjelp av et tastatur, noe som er svært tidkrevende (Solberg et.al, 2022.)

Det som også er utfordringen til talegjenkjenning generelt er at teknologien behandler ulike lyder isolert. I motsetning til bilder som kan deles inn i farger og dimensjoner, kan ikke lyd deles inn i like mange kategorier. Det vil si at det er mindre kategorier å analysere og bruke for å skape en helhet, men noe lyder kan kategoriseres inn i, er stemmeleie. Vi mennesker er gode på å høre på måten en setning er bygd opp for å forstå konteksten av den. På den måten kan vi både fylle ut ord vi eventuelt ikke hørte, men vi kan også forstå om personen var ironisk, sint eller sa det med største alvor (O'Shaughnessy, 2008, s.1). Per dags dato er dette en stor utfordring for talegjenkjenningsteknologien, og vi skal senere gå inn på hva teorien sier om hvordan disse, og hvilke andre utfordringer som har påvirket ulike bedrifter.

#### 2.1.5 Norsk talegjenkjenning

Bedriftene vi studerer i denne oppgaven er norske og opererer mot det norske markedet og transkriberer og håndterer derfor naturligvis mest det norske språket. Det kan derfor være interessant å få en oversikt over den norske forskningen på talegjenkjenning og ikke minst hva slags utfordringer det norske språket i seg selv står overfor med dagens teknologi.

Automatisk talegjenkjenning er utfordrende på norsk, ettersom språket er ganske mangfoldig. Språket har for det første to offisielle skriftlige standarder, nemlig bokmål og nynorsk. For det andre har språket også mange dialekter, som er ulike leksikalt, grammatisk og fonologisk. Det norske språket har ingen muntlig standard og det brukes også dialekter i alle slags

formelle og offentlige settinger. Det er derfor høyst sannsynlig at man vil bruke sin dialekt, eller foretrekker å bruke den, når man bruker talegjenkjenningssystemet slikt som med de tidligere nevnte applikasjonene “Siri” og “Google Assistant”. Det betyr at norske datasett som skal brukes til trening av modellene krever at taledataen inneholder ulike dialekter, og bør også inkludere transkriberinger i begge skriftlige språk. Standardene for bokmål og nynorsk tillater også mange muligheter; Ord har flere skrivemåter eller bøyningsvarianter, noe som ofte begrunnes i variasjon i dialekt i talespråket. Hvis man tester dette opp mot det vi tidligere beskrev som treningsdata, må altså fasiten vise til forskjellige varianter av samme ord. For eksempel, må modellen klare å håndtere både “veit” og “vet” på samme måte og bygge det riktig opp i en videre setning basert på samme forutsetninger (Solberg et.al, 2022, s.5).

Disse utfordringene står blant annet det norske forskningsprosjektet SCRIBE overfor. SCRIBE er en del av Telenor som vi har vært i kontakt med i forbindelse med studiet vårt. Dette prosjektet har økonomisk støtte fra forskningsrådet med forbehold om en åpen kildekode til deres ferdigtrente modeller (Telenor, 2021). Dette er typisk for forskning på kunstig intelligens fordi det kan være med på å forbedre og utvikle flere systemer. De har derfor i forbindelse med forskningen på talegjenkjenning publisert en god del relevante artikler. SCRIBE sitt mål er derfor å utvikle automatisk transkribering for samtaler mellom flere parter på norsk. Gjennom prosjektet skal det utvikles modeller som er robuste for ubalanser typisk for spontan samtale og modellene skal håndtere turtaking og ikke minst kunne utnytte konteksten i en dialog. Modellene skal i tillegg støtte både bruk av de to norske skriftspråkene, men også bruk av de norske dialekter.

For å nå målet med prosjektet er det derfor nødvendig med forskning og teknologiutvikling innen flere sentrale områder, både når det gjelder språklige universelle utfordringer og utfordringer spesifikt tilknyttet det norske språket. Dette krever godt opplærte modeller og SCRIBE har derfor samarbeidspartnere som gitt de tilgang på mengder av relevant data. Blant annet samarbeider de med Språkbanken på Nasjonalbiblioteket og har fått tilgang på Stortingskorpuset (engelsk: The Norwegian Parliamentary Speech Corpus - NPSC). Dette er materiale som er satt sammen av lydopptak av møter i Stortinget, manuelt transkriberte på bokmål eller nynorsk. Dette er det første offentlig tilgjengelige datasettet som inneholder spontan norsk tale, altså uten manus, og som er designet for å trene automatiske talegjenkjenningssystemer. Fordelen med et slikt datasett er at representantene kommer fra



hele landet og bruker sin egen dialekt, noe som gir større variasjon i opptrening av modellene. Det er likevel viktig å nevne at noen datasett fra Språkbanken inneholder kun manuskript lest data, hvilket innebærer kun planlagt tale og består derfor av mindre nøling, avbrudd og andre lyder typisk for spontan tale. Når innholdet er manusbasert leser gjerne talerne også på bokmål og ikke sin egen dialekt (Solberg et.al, 2022, s.3).

Talegjenkjenningssystemer presterer derfor ofte dårligere når de brukes på dialekter, siden modellene ikke har blitt eksponert for det i like stor grad (Nasjonalbiblioteket, 2019).

### 2.1.6 Talegjenkjenning i ulike bedrifter

Som vi har vært inne på ved flere anledninger er talegjenkjenningssystemene utviklet alt etter hvilket formål det skal nytte. Gjennom historien har noen formål blitt flittigere brukt enn andre, spesielt formål for helsesektoren. Her har formålet vært å kunne automatisk transkribere pasientjournaler, og mye av programvarene innenfor talegjenkjenning er derfor i stor grad spesialisert og trent på bakgrunn av dette. Siden vi skal studere flere ulike bedrifter er det derfor relevant å se på teori rundt de mest typiske anvendelsesområdene.

En av hovedgrunnene for hvorfor talegjenkjenning blir kjøpt opp og implementert i bedrifter er potensiale for å forbedre produktiviteten. Med dette menes formålet om å spare ressurser, slik som tid og penger. Blant helsesektoren var talegjenkjenning tidlig ment for å effektivisere journalomløpstiden. Leger må skrive pasientjournaler for hver pasient, og en slik jobb var tildelt sekretærer før talegjenkjenning ble tilgjengelig på mange sykehus. I stedet for at leger snakker inn i en mikrofon som gjøres om til lydopptak som sekretærer manuelt transkriberer, gjøres nå dette automatisk ved hjelp av talegjenkjenningsteknologien. På den måten blir journalene dokumentert forttere og med behov for mindre ressurser (Al.aiad, 2020, s.7).

Hvis man ser på tidligere studier, viser implementeringen av dette systemet betydelige ressursbesparende resultater allerede fra 2002 (Callaway et.al, 2002, s.1). Et annet studie viser faktisk til at journalomløpstiden forbedret seg med 81% ved hjelp av et talegjenkjenningssystem, og at andel rapporter tilgjengelig innen en time forbedret seg med 32% (Koivikko et.al, 2008, s.2). En artikkel fra 2020 har også gått gjennom oppdatert litteratur fra de siste 10 årene som omhandlet nettopp effekten av talegjenkjenningssystemer på produktiviteten i helsesektoren. Funn fra alle disse artiklene identifiserer klart en positiv effekt på avdelingsproduktiviteten på sykehusene, samt lavere behandlingstid for

journalføringen (Alaiad, 2020).

En annen grunn for at talegjenkjenning blant annet blir brukt ved helsesektoren, er av dens ergonomiske fordeler. Det å sitte foran et tastatur gjennom en hel jobbdag kan være belastende for flere deler av kroppen, og ved å kun måtte forholde seg til en mikrofon kan risiko for overbelastning være bemerkelsesverdig lavere (Durling et.al, 2008).

Dette fører videre til et annet bruksområde som inkluderer de fleste bedrifter, nemlig fordelene med talegjenkjenning for de med nedsatt funksjonshemming. Det vil i dette tilfelle være de som fysisk ikke har mulighet til å sitte foran et tastatur å skrive, men også de som ikke har mulighet til å styre digitale verktøy. Det vil både si å kunne bruke en mus, men også å navigere seg i de ulike digitale systemene, som for eksempel å skru av og på en pc eller sende et digitalt dokument videre. Ved hjelp av talegjenkjenning kan slike kommandoer og transkriberinger gi muligheter for å sysselsette flere til ulike arbeidsoppgaver. Dette gir muligheter for blant annet de med motoriske vansker, svaksynte og blinde og ikke minst personer med skrivevansker (Kunnskapsbanken, 2020). En artikkel beskriver hvordan talegjenkjenning også kan brukes til lek og spesielt digitale spill. Formålet her er blant annet å inkludere personer som ikke kan bruke konsoll, men at utfordringen er at talegjenkjenningssystemet ikke kan ta opp det følelsesmessige aspektet (Mustaquim, 2011, s.145).

Et annet utbredt område for bruk av talegjenkjenning er ved skoler og andre pedagogiske institutter. For elever med lærevansker eller utenlandske studenter med et annet språk enn foreleseren/læreren kan det være utfordrende å få inn all informasjon som man kun hører i forelesninger eller møter. I senere tid har også problemer der det er dårlig lyd kvalitet pga begrenset nett-tilgang oppstått. Det kan derfor være til en stor fordel for flere å få dette ned i tekstformat. Alternativet er i mange tilfeller er notater fra foreleser selv eller at man selv skriver ned notater underveis. Dette kan være tidkrevende dersom man ikke behersker språket eller skriften så godt, for ikke å glemme tilfeller der studenter er døve. Et studie basert nettopp på studenter som var døve og studenter med annet morsmål viser at talegjenkjenning har ført til betydelig forbedring i elevenes lese-, skrive- og staveferdigheter (Shadiev et.al, 2014, s.65).

Et annet studie viser dog at det er behov for mer nøyaktighet i talegjenkjenningen dersom

foreleser selv ikke snakker språket bra nok, for da blir det for vanskelig å lese setningene. Et studie sammenlignet resultater for en gruppe før og etter forelesninger med talegjenkjenning og viste resultater av en betydelig forbedret prestasjon i både lekser og prøveresultater. Det er flere studier som har sett på effekten studenter har ved å bruke dette og denne artikkelen konkluderer da med at bruk av talegjenkjenning har forbedret læringsresultatet. Det anbefales derfor å bruke det i situasjoner for folk med lærevansker eller når vanskelighetsgraden øker, men at det også er viktig å komme med en ordrett utskrift som foreleser har korrigert og som de kan bruke som notat senere. Et slikt studie viser også at det er viktig å lære opp elevene og foreleser til å bruke teknologien riktig (Huang et.al, 2015, s 530).

Som nevnt tidligere er det flere utfordringer blant ulike bedrifter sett gjennom talegjenkjenningens nåværende perspektiv. En av hovedoppgavene til talegjenkjenning er nettopp det å tjene sin hensikt, og det kan forstås som at modellen utfører sin egentlige oppgave, og dette kan variere alt etter hva formålet er. I noen tilfeller er det mer interessant å identifisere all bakgrunnsstøy, mens i andre tilfeller er det viktig å nettopp registrere, analysere og dokumentere denne bakgrunnsstøyen. Ved pasientjournaler for eksempel, er formålet klart å få skrevet ned et dokument så nøyaktig, feilfritt og formelt som mulig og bør ikke inneholde nølelyder slikt som “ehm” eller latter. Om talegjenkjenning skal brukes til å automatisk transkribere politiafhør, er det på den andre siden like viktig å få med slike type lyder som ordene selv. Sånne type varierende utfordringer og behov gjør at noen bedrifter og markeder har et bedre utgangspunkt enn andre for å oppnå god effekt ved bruk av talegjenkjenningssystemer (O’Shaughnessy, 2008, s.14).

#### 2.1.7 Fremtidig perspektiv

Når vi nå har sett på utfordringer, både tekniske og praktiske, rundt dagens talegjenkjenningssystemer er det også interessant å se på det fremtidige perspektivet. Ved implementering av talegjenkjenning for bedrifter kan man anta at det er av interesse å se på hvilke muligheter de har for fremtiden. Det kan i beste fall bety at flere bedrifter er villig til å implementere systemer under utvikling, med et håp om at den fremtidige teknologien vil føre til enda bedre resultater.

Et studie viser at det kan være mulig å bruke leppebevegelser for å styrke utfordringen ved ord som ikke blir registrert. Slikt som tidligere nevnt, har lyder færre kategorier enn bilder slik at det kan være vanskeligere å danne et helhetlig bilde. En stor utfordring vil jo derfor

være dersom talegjenkjenningssystemet ikke klarte å registrere et ord og ikke har så mange hjelpemidler for å fylle det ut. Dersom talegjenkjenningssystemet klarer å registrere leppebevegelsene i form av frekvenser, kan dette på mange måter fungere på lik måte som vi mennesker leser lepper for å fylle ut ord (Auforas, 2018, s.6).

Når det gjelder teknologisk utvikling, sies det dog at vi fortsatt kan være minst et par tiår unna å ha virkelig autonome, intelligente kunstige talegjenkjenningssystemer som kommuniserer med oss på en genuint "menneskelignende" måte (Zacjehowski, U.å) Et stort teoretisk spørsmål er nettopp derfor om tilstrekkelig talegjenkjenningssystemer noen gang vil oppnås; For det krever intelligens tolkning på lik linje som selvstyrte biler og det kan til tider oppleves som fatalt ved feil (O'Shaughnessy, 2008). Hva som menes med dette er at dersom en selvstyrt bil skulle gjøre noe feil, er dette livstruende. På samme måte kan pasientjournaler som automatisk har blitt transkribert uriktig og derfor identifisert feil form for sykdom på en pasient også være på lik skala av livstruende.

### **3.0 Metode**

I dette kapitlet presenteres metodene som benyttes for å besvare problemstillingen.

Innledningsvis presenteres valgene som er tatt vedrørende forskningsdesign, datainnsamling og analyse av datamateriale.

#### **3.1 Forskningsdesign**

Talegjenkjenning er et tema som er mye diskutert i eksisterende litteratur, men det er begrenset med forskning og litteratur rundt hvilke effekter og muligheter talegjenkjenning gir for norske bedrifter. På bakgrunn av det, fant vi det hensiktsmessig å benytte oss av et eksplorativt forskningsdesign med en kvalitativ metode ettersom vi ønsket å studere et fenomen nærmere (Saunders et al., 2016, s. 163). Fordelen ved å benytte en kvalitativ forskningsmetode er at metoden har gitt oss mulighet til å undersøke og forstå hvordan fenomenet oppleves fra aktørenes synspunkt (Saunders et al., 2016, s. 165). Vi har valgt å besvare problemstillingen til masteroppgaven gjennom en casestudie, og har benyttet en enkeltcasestudie. Vi har definert Norge som case og fenomenet er talegjenkjenning i ulike bedrifter. I vårt studie har vi benyttet oss av én analyseenhet, ettersom vi ønsker å se på hvilke effekter og muligheter talegjenkjenning gir for norske bedrifter. Vi har innhentet perspektiver fra respondenter fra ulike bedrifter, men hovedformålet vårt er å belyse problemstillingen for ulike bedrifter som én helhet.

Vi fant det hensiktsmessig å kombinere forskningsstrategien casestudie med dokumentstudie, ettersom vi fikk tilgang på interne dokumenter om implementering av talegjenkjenning hos en bedrifts kundesentre. Det var nyttig for oss å gjennomføre dokumentstudie ettersom vi ikke fant tilstrekkelig informasjon fra annen litteratur eller intervjuer (Saunders et al., 2016, s. 183). Dokumentstudiet har gitt oss bedre oversikt og nyttig informasjon som det hadde vært vanskelig for oss å samle inn selv. Ettersom vi benytter oss av eksisterende litteratur og innhenter primærdata fra eksperter på området har vi benyttet oss av en abduktiv forskningstilnærming.

#### **3.2 Datainnsamling**

Datainnsamlingen vi har gjennomført består av både primær- og sekundærdata. Vi har benyttet metodetriangulering som innebærer søk etter likheter mellom ulike kilder, i denne

oppgaven er det metodetriangulering mellom dokumentstudie, intervjuer og eksisterende litteratur.

### 3.2.1 Primærdata

Vi har samlet inn primærdata til forskningen gjennom 8 semistrukturerte intervjuer. I forkant av intervjuene fikk informantene informasjon om tema for intervjuet, og vi utformet en intervjuguide med ulike deltemaer og underspørsmål. Formålet med å innhente egen data var å kunne belyse problemstillingen fra flere innfallsvinkler og forstå hvordan fenomenet oppfattes av ulike informanter, som også er hensikten med et casestudie.

#### **Utvalg av informanter**

Ifølge Johannessen et al. (2016, s. 114) hevder at det viktigste er å skaffe et relevant utvalg av informanter, enn å skaffe mange informanter. Det hevdes også at det bør gjennomføres intervjuer helt til forskeren ikke får noe særlig ny informasjon, men dette kan være for omfattende for en studentoppgave (Johannessen et al., 2011, s.108). Saunders et al. (2016, s. 297) hevder det samme som Johannessen, og mener at for semistrukturerte intervjuer, bør utvalget være på minimum 5 informanter. Selv om vårt utvalg av informanter er begrenset, anser vi vår utvalgsstørrelse som tilstrekkelig ettersom vi ikke fikk tilført noe særlig ny informasjon i våre siste intervjuer, da poengene som ble nevnt av informant allerede var nevnt av tidligere informanter. Vi benyttet en utvalgsstrategi som var basert på et hensiktsmessig utvalg og har benyttet egen dømmekraft til å finne aktuelle informanter som vi anser som verdifulle for å kunne belyse forskningsspørsmålet vårt.

De 8 informantene er blitt anonymisert, og i tabellen nedenfor er informantenes rolle presentert. Våre informanter ble valgt ut for å få sikre at vi fikk ulike perspektiver for å belyse problemstillingen vår, dette var informanter med ulik erfaring, bakgrunn og rolle. Vi mente at det var nødvendig å velge ut informanter i ulike stillinger og organisasjoner, for å få en dypere forståelse av hvilke effekter og muligheter gir talegjenkjenning for norske ulike bedrifter.

**Tabell 3.1:** Oversikt over informanter og beskrivelse av rolle

<b>Informant</b>	<b>Beskrivelse</b>
Informant 1	Rådgiver for et prosjekt for implementering av talegjenkjenning i Telenor og deltar i SCRIBE-prosjektet
Informant 2	Prosjektleder for prosjekt med formål å implementere talegjenkjenning på kundesenter
Informant 3	Produktspesialist Omilon
Informant 4	Forsker innen talegjenkjenning i Telenor og SCRIBE-prosjektet
Informant 5	Foreleser ved universitet som har tatt i bruk talegjenkjenning i undervisning
Informant 6	Rekrutterer for ansatte ved Omilon
Informant 7	Kunderådgiver på kundesenter som har implementert talegjenkjenning
Informant 8	Ledende rolle innenfor prosesser med arbeid rundt taleteknologi i NRK.

**Notater:** Tabellen gir en oversikt over de ulike informantene vi har hatt semistrukturerte intervjuer med, samt hvilken rolle informanten har i dag. Alle informantene er anonymisert, og enkelte bedrifter/organisasjoner er også anonymisert.

Informant 1 er den første informanten som ble intervjuet, og ble intervjuet for å gi oss en bredere forståelse av hvordan Telenor arbeider med talegjenkjenning. Ettersom informanten har jobbet med talegjenkjenning i mange tiår, vil vi betegne informanten som en ekspert på området. Informanten jobber som rådgiver for et prosjekt for implementering av talegjenkjenning i Telenor. Vi fant det også hensiktsmessig å intervjuer en informant som har hatt en sentral rolle i implementeringen av talegjenkjenning på kundesentre i en annen bedrift, og i denne sammenheng ble informant 2 intervjuet. Bedriften startet å arbeide med talegjenkjenning i 2018, og det ble implementert våren 2020. Talegjenkjenningen som har blitt implementert innebærer at alle telefonsamtaler til kundesenteret automatisk blir transkribert, deretter benyttes tekstanalyse til å kategorisere samtalens tema og automatisk skrive et sammendrag på 8 nøkkelord om samtalen. I denne sammenheng fant vi det også

interessant å intervju en kundefrådgiver på et kundesenter som allerede har implementert talegjenkjenning, ettersom det kunne gi oss verdifull innsikt i hvordan det er som ansatt i en bedrift som har implementert talegjenkjenning. Informant 7 ble intervjuet i denne sammenheng, og det må påpekes at informant 7 ikke jobber i samme bedrift som informant 2.

For å få et bredere perspektiv på hvordan talegjenkjenning brukes i andre sektorer ble informant 3 intervjuet. Informant 3 er en produktspesialist i Omilon, som er en leverandør for talegjenkjenning. Selskapets hovedfokus og ekspertområde er innenfor helsesektoren, men de har også noen advokater som bruker deres tjenester. Av samme årsak valgte vi også å intervju informant 5 og informant 8. Informant 5 er en foreleser ved et norsk universitet som har tatt i bruk talegjenkjenning i forelesninger. Verktøyet som har blitt brukt har tatt opp det som blir sagt i forelesning, og studentene fått opp teksten direkte på egne dataskjermer i forelesning. Informant 8 har en ledende rolle innenfor prosesser med arbeid rundt taleteknologi i NRK. NRK ble nevnt av flere informanter i tidligere intervjuer, ettersom NRK har et verdifullt arkiv med datamateriale som inneholder bredt utvalg av dialekter. Vi fant det derfor interessant å intervju NRK for å få en dypere innsikt i hvordan de arbeider med taleteknologi, spesielt ettersom NRK ble nevnt i flere sammenhenger av andre informanter.

For å få en mer teknisk innfallsvinkel intervjuet vi informant 4, som er en forsker innen talegjenkjenning i Telenor. Vi fant det hensiktsmessig å snakke med en forsker på temaet ettersom det kan gi oss en dypere forståelse i tekniske utfordringer og fordeler med talegjenkjenning. Informant 6 arbeider med rekruttering av ansatte ved Omilon, og ble valgt ut som informant fordi vi fant det hensiktsmessig å få dypere innsikt i hva en rekrutterer ser etter når en skal ansette mennesker som skal jobbe med talegjenkjenning.

### **Gjennomføring av intervju**

Alle intervjuene ble gjennomført via videosamtaler på grunn av geografisk avstand mellom oss og informantene. Videosamtaler gjorde det mulig å vise frem presentasjoner og systemer, dette sørget for at vi fikk et godt visuelt bilde av samtaleemnene. Vi valgte å ikke ta lydopptak av intervjuene, ettersom vi ønsket å teste ut en modell for tale-til-tekst. Dette sørget også for at vi ikke måtte lagre sensitiv data om informantene, og sparte oss for tid når det kommer til transkribering. Ettersom vi gjennomførte intervjuene sammen, kunne den ene fokusere på å notere og korrigere feil modellen for automatisk tale-til-tekst gjorde, mens den andre fokuserte på å stille spørsmål.



### 3.2.2 Sekundærdata

Sekundærdata er innhentet gjennom litteraturstudie og dokumentstudie. Sekundærdata som er innhentet er offentlig tilgjengelig datamateriale og dokumentstudiet inneholder interne dokumenter fra en bedrift som har implementert talegjennkjennning på kundesenter.

Sekundærdata har vært utredet og publisert for andre formål enn å besvare vår problemstilling, vi har derfor understøttet funn fra vårt litteraturstudie med funn fra informantene og dokumentstudiet.

### 3.3 Dataanalyse

I dette kapitlet vil vi presentere nærmere hvordan analysene av datamaterialet ble gjennomført. Analysen av datamaterialet består av en tematisk analyse. Ettersom vi ikke benyttet lydopptak i intervjuene, var det ingen behov for transkribering i etterkant. Vi valgte vi å bruke en modell for tale-til-tekst, som sørget for at intervjuet ble transkribert automatisk underveis. Vi ville sikre at all informasjon ble gjengitt og oppfattet korrekt, og underveis i intervjuene var en ansvarlig for å korrigere ordfeil. Ettersom vi korrigerer underveis i intervjuet var det som ble sagt friskt i minne vårt, slik at alt informantene sa ble oppfattet korrekt. Ettersom det er en betydelig forskjell mellom skriftlig og muntlig språk, ble det en del ufullstendige setninger og ordfeil, men dette sørget vi for å korrigere underveis (Saunders et al. 2016, s. 572).

Etter hvert intervju ble datamaterialet gjennomgått, hvor vi fremhevet interessante funn. Dette ble gjort individuelt slik at vi i etterkant kunne ha en felles diskusjon, gjennomgang og sammenligning av hverandres fremhevede funn og forståelse. Det ble også diskutert om det forelå temaer hvor det var behov for ytterligere avklaring, samt at hvert intervju dannet grunnlaget for neste intervju. Denne metoden sørget for at ikke viktig informasjon ble utelatt og at vi fikk en grundig og felles forståelse av funnene. Det var enkelte temaer som dominerende hos de fleste informantene, og disse dannet grunnlaget for kodingen. Deretter organiserte vi datamaterialet enda grundigere ved å benytte koding. Det sørget for at vi fikk bedre oversikt over dataen vi hadde samlet inn, og fanget opp gjentakende temaer og mønstre. I neste kapittel, "4. Presentasjon av funn", er alle funnene våre presentert.

## **4.0 Presentasjon av funn**

I dette kapittelet presenteres de ulike funnene vi har gjort gjennom dokumentstudie og basert på datamateriale fra de 8 semistrukturerte intervjuene. Funn vi har gjort fra kvalitative intervjuer presenteres under følgende temaer for talegjenkjenning: utfordringer, fordeler, muligheter og fremtiden. Dette er hovedtemaer som gikk igjen hos mange av informantene våre, og er faktorer som har innvirkning på hvilke effekter og muligheter talegjenkjenning gir for bedrifter.

### **4.1 Dokumentstudie**

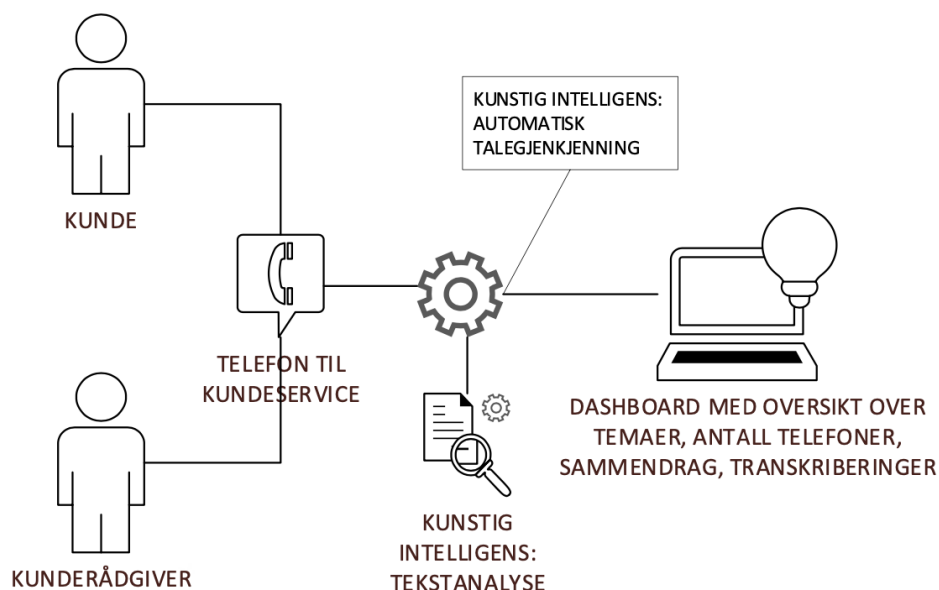
Gjennom studiet vårt ga informant 2 oss tilgang på interne dokumenter som inneholdt rapporter fra deres kundesentre og implementeringsprosjektet. Dette ga oss muligheten til å få dypere innsikt i prosjektet og analysere hvilken effekt implementeringen av talegjenkjenning har hatt for bedriften.

#### **4.1.1 Beskrivelse av prosjektet**

Gjennomgangen av de interne dokumentene ga oss en dypere innsikt i prosjektet som har blitt implementert på kundeservice hos den aktuelle bedriften. Motivasjonen for hele prosjektet var at kunderådgivere bruker mye tid på å kategorisere og skrive sammendrag i etterkant av samtalen med hver kunde, og dette er administrative arbeidsoppgaver og verdifull tid som kan benyttes med kunder i stedet. Kunderådgivere må kategorisere samtaler av flere grunner, blant annet for at bedriften skal forstå hvilke problemer kundene kontakter bedriften for, bidra til å ta riktige valg og forbedre kundeopplevelsen og kundeinnsikten.

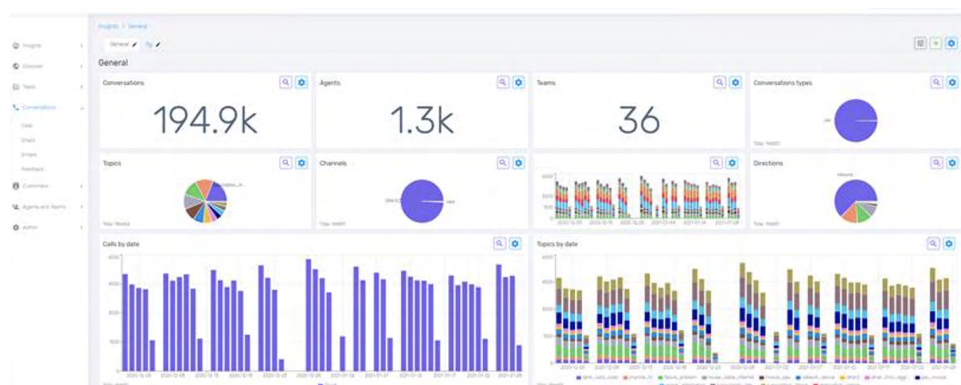
Kunderådgivere må skrive et sammendrag av samtalen for at bedriften skal få en forståelse av hvilke råd som er gitt til kunden og det skal være til hjelp for neste kunderådgiver, dersom kunden ringer tilbake. Ettersom disse arbeidsoppgavene ble sett på som tidstyver og lite effektivt, ønsket bedriften å se på muligheter for forbedring og bestemte seg for å implementere talegjenkjenning. Vi har laget en illustrasjon i figur 4.1 som presenterer kort prosessene i prosjektet som har blitt implementert.

**Figur 4.1:** Illustrasjon av hvordan automatisk talegjenkjenning fungerer på kundesenteret



Som figuren illustrerer, begynner talegjenkjenningssystemet bedriften har implementert først å arbeide når en kunde ringer inn til kundeservice. Samtalen blir automatisk transkribert ved bruk av talegjenkjenning og det blir utført en tekstanalyse av den transkriberte teksten. I stedet for at kunderådgiveren skal bruke tid på å skrive sammendrag og kategorisere samtalen, gjør talegjenkjenningssystemet dette automatisk. Resultatet av talegjenkjenningen blir tilgjengelig i en plattform som gir bedriften oversikt over hvilke temaer kunder ringer inn for, antall samtaler, transkriberinger og sammendrag som består av 8 nøkkelord fra samtalen, en oversikt over plattformen er presentert i figur 4.2.

**Figur 4.2:** Illustrasjon av plattformen kundesenteret benytter som oversikt over kundesamtaler



I plattformen kan brukeren enkelt navigere seg til enkelte samtaler for å lytte og se transkribering, som vist i figur 4.3 nedenfor. Det er svært oversiktlig og enkelt å navigere seg i plattformen, og få et øyeblikksbilde av hva kundene ringer inn for denne uken, eller eventuelt tilbake i tid.

**Figur 4.3:** Illustrasjon av en automatisk transkribert kundesamtale



**Notater:** Figuren illustrerer hvordan det ser ut dersom man ønsker å se transkribering og høre opptak av en konkret kundesamtale. Transkriberingen av samtalen er i denne sammenheng sensurert. Talegjenkjenningen har også identifisert hvem som snakker, slik at det er enkelt å se det i etterkant.

#### 4.1.2 Kostnader

Dokumentstudiet ga oss også innsikt i kostnader bedriften har hatt i forbindelse med implementering av talegjenkjenning på sine kundesentre. Bedriften har til nå implementert talegjenkjenning på tre ulike kundesentre, samt hos en avdeling som jobber med utgående telefonsalg. I 2020 hadde bedriften engangskostnader i forbindelse med implementeringen, og implementerte talegjenkjenning i juli 2020 på ett kundesenter. I 2021 ble talegjenkjenning implementert på kundesenter 2 og 3, samt hos avdelingen som jobber med utgående telefonsalg. Driftskostnadene for 2021 vil derfor naturlig være høyere, ettersom bedriften har implementert talegjenkjenning i flere avdelinger. Tabellen presentert nedenfor gir en oversikt over kostnader.

**Tabell 4.1:** Oversikt over kostnader for implementering av talegjenkjenning i bedriften

	Engangskostnader (2020)	Driftskostnader (2020)	Driftskostnader (2021)
<b>Eksport</b>	600 000 NOK	90 000 NOK	175 000 NOK
<b>Plattform og integrasjoner</b>	390 000 NOK	82 620 NOK	497 900 NOK
<b>Automatisk transkribering (utvikling)</b>	0 NOK	106 380 NOK	338 580 NOK
<b>Automatisk transkribering (kundesenter 1)</b>	0 NOK	227 220 NOK	435 220 NOK
<b>Automatisk transkribering (kundesenter 2)</b>	0 NOK	0 NOK	435 220 NOK
<b>Automatisk transkribering (Kundesenter 3)</b>	0 NOK	0 NOK	77 377 NOK
<b>Automatisk transkribering (Utgående telefonsalg)</b>	0 NOK	0 NOK	106 394 NOK
<b>TOTALE KOSTNADER</b>	<b>990 000 NOK</b>	<b>506 220 NOK</b>	<b>2 065 691 NOK</b>

**Notater:** Talegjenkjenning ble implementert på kundesenter 1 juli 2020, driftskostnadene for 2020 representerer derfor kostnader fra og med juli 2020. Talegjenkjenning ble ikke implementert hos kundesenter 2, 3 og utgående telefonsalg før i 2021. Tabellen gir en oversikt over kostnader rettet til prosjektet, og inkluderer ikke lønnskostnader eller andre interne kostnader hos bedriften.

Driftskostnadene for 2021 vil være mer presise i henhold til hvordan kostnadene vil se ut i fremtiden, ettersom det ble implementert hos alle avdelinger i bedriften som samhandler med kunder i 2021.

#### 4.1.3 Effekter etter implementering

Etter å ha analysert og studert de interne dokumentene ble det tydelig at kundesenteret har hatt stor effekt etter implementering når det gjelder effektivitet. Resultatene av gjennomgangen av de interne dokumentene er presentert i tabell 4.2. Fra tabellen kan man se hva tidsforbruket for kategorisering og sammendrag var før implementering, når arbeidsoppgavene ble gjort manuelt av en kundefrådgiver. I tillegg presenterer tabellen timer spart per måned ved at disse arbeidsoppgavene gjøres automatisk av talegjenkjenningssystemet.

**Tabell 4.2:** Oversikt over tidsforbruk før og timer spart etter implementering av talegjenkjenning

	Arbeidsoppgave	Tidsforbruk
Før implementering	Kategorisering av samtale	13 sekunder per telefon
Før implementering	Sammendrag av samtale	2 minutter per telefon
Etter implementering	Kategorisering av samtale	289 timer spart månedlig
Etter implementering	Sammendrag av samtale	1660 timer spart månedlig

**Notater:** Tabellen presenterer hvor mye tid en kundesentermedarbeider normalt bruker på kategorisering og skrive sammendrag manuelt før implementering av talegjenkjenning, sammenlignet med tid spart etter implementering i 2021.

De interne dokumentene inneholder ikke informasjon om antall telefonsamtaler eller tidsforbruk i 2020, vi vil derfor kun se på effekter av implementering for 2021. Det er uansett 2021 som er det mest relevante året, ettersom bedriften implementerte talegjenkjenning på alle relevante avdelinger i 2021.

For å se på lønnsomheten av prosjektet fant vi det interessant å se på lønnsomheten i form av tid og ressurser spart, sett i sammenheng med lønnskostnader og kostnader for prosjektet. Bedriften ga oss ikke informasjon om hva årslønnen til deres ansatte på kundesentre er, men vi har tatt utgangspunkt i medianlønn fra SSB, ettersom det er et godt mål på hva som er vanlig å tjene i et gitt yrke (SSB, 2021). Medianlønnen vi har tatt utgangspunkt i er presentert i tabell 4.3.

**Tabell 4.3:** Median årslønn og timelønn kundesentermedarbeider

	Median (2021)
Kundesentermedarbeider (årslønn)	477 000 NOK
Kundesentermedarbeider (timelønn)	244,62 NOK

**Notater:** Tabellen representerer median av årslønn og timelønn for en kundesentermedarbeider (SSB, 2021).

**Tabell 4.4:** Oversikt over tid- og kostnadsbesparelser i 2021

Arbeidsoppgave	Tid spart årlig	Kostnader spart
Kategorisering av samtale	3 468 timer	848 342,16 NOK
Sammendrag av samtale	19 920 timer	4 872 830,40 NOK
<b>Totalt spart</b>	<b>5 721 172,56 NOK</b>	
<b>Årsverk spart</b>	<b>11,99</b>	

**Notater:** Totalt spart for 2021 er beregnet ved å multiplisere median timelønn for 2021 med antall timer spart årlig per arbeidsoppgave, presentert i tabell 4.3.

Analysene vi har gjort når det gjelder kostnadsbesparelser er presentert i tabell 4.4, hvor totalt spart i lønnskostnader tilsvarer årslønnen til 11, nærmere 12 heltidsansatte kunderådgivere. Dette indikerer at implementering av talegjennkjennning på kundesentre kan være svært ressursbesparende ved at manuelle arbeidsoppgaver blir automatisert, noe som innebærer at kostnader blir betraktelig redusert. Det er også hensiktsmessig å se på kostnadsbesparelsene i sammenheng med driftskostnadene for talegjennkjenningsprosjektet. Resultatet av disse analysene er presentert i tabell 4.5.

**Tabell 4.5:** Oversikt over driftskostnader i forbindelse med implementering av talegjennkjennning og totalt spart

	Sum
<b>Totalt spart (2021)</b>	<b>5 721 172,56 NOK</b>
<b>Driftskostnader (2021)</b>	<b>2 065 691,00 NOK</b>
<b>Totalt spart (2021) – driftskostnader (2021)</b>	<b>3 655 481,56 NOK</b>

**Notater:** Tabellen tar utgangspunkt i totalt spart i 2021 i ressurser, fratrukket driftskostnader for 2021.

Etter å ha sett på driftskostnader for 2021 i sammenheng med kostnadsbesparelsene i form av ressurser i 2021 i tabell 4.5, indikerer fortsatt analysene våre at implementering av talegjennkjennning er svært lønnsomt, ressurs- og kostnadsbesparende. For å se på lønnsomheten av hele investeringen av talegjennkjennning for bedriften, har vi valgt å se på avkastning på investering (ROI). Resultatet er presentert i tabell 4.6.

**Tabell 4.6:** Oversikt over kostnader i forbindelse med prosjektet, totalt spart og avkastning på investering (ROI)

	Sum
Engangskostnader (2020)	990 000 NOK
Driftskostnader (2020)	506 220 NOK
Driftskostnader (2021)	2 065 691 NOK
<b>Totale kostnader</b>	<b>3 561 911 NOK</b>
<b>Totalt spart (2021)</b>	<b>5 721 172,56 NOK</b>
<b>Avkastning på investering (ROI)</b>	<b>0,61 / 60,62%</b>

**Notater:** ROI er kalkulert ved å benytte følgende formel  $\frac{\text{netto avkastning}}{\text{investeringskostnad}}$ , hvor netto avkastning representerer totalt spart fratrukket totale kostnader, og investeringskostnad representerer totale kostnader (Visma, u.å).

Tabell 4.6 inkluderer kostnader for hele prosjektet i 2020 og 2021, men kun sparte kostnader for 2021. Dette begrunnes i at vi ikke har noen informasjon om antall telefonsamtaler eller tidsforbruk i 2020, kun kostnader. Likevel er avkastning på investering på 60.62% noe vi anser som svært høyt, og dersom vi hadde hatt informasjon om tidsforbruk og antall telefoner for 2020, hadde avkastning på investeringen vært enda høyere.

## 4.2 utfordringer ved bruk av talegjenkjenning

Gjennom intervju ble det tydelig at temaet *utfordringer* var viktig for informantene, og de påpekte mange utfordringer ved bruk av talegjenkjenning. Det er helt klart at utfordringer med talegjenkjenning er en sentral faktor som har innvirkning på hvilke effekter og muligheter talegjenkjenning gir for norske bedrifter. I dette underkapittelet vil vi presentere de mest sentrale funnene vi har gjort når det gjelder utfordringer ved bruk av talegjenkjenning.

### 4.2.1 Kvaliteten på talegjenkjenning for det norske språket

Flere av informantene hevdet at kvaliteten på talegjenkjenning i dag ikke er bra nok, spesielt når det gjelder det norske språket. Datamaskiner gjør fortsatt feil, og dette kan få alvorlige konsekvenser i enkelte bransjer. Informant 3 ga oss eksempel på dette ved å fortelle om helsesektoren som er selskapets viktigste kunde, hvor leger benytter talegjenkjenning i stedet for å skrive pasientjournaler selv. I denne sammenhengen er det svært viktig at journalen skrives nøyaktig, men det norske språket byr på utfordringer. Her trakk informant 3 frem



negasjoner som en farlig feil talegjenkjenningen kan gjøre, ved at setningen kan få en helt annen betydning. For eksempel “uviktig” i stedet for “viktig”, eller “høyre” i stedet for “venstre”. Ettersom det er kritisk med slike feil i pasientjournaler, kreves det at det settes av mye tid til korrekturlesning av journal slik at eventuelle feil kan korrigeres. Informant 3 nevnte også at en utfordring med det norske språket er at det har mange ord som kan høres svært like ut, og ved å bytte ut den første bokstaven i ordet kan ordet få en helt annen betydning. På en annen side trakk informant 4 frem utfordringen med semantiske ord, dette er ord som har ulike skrivemåter, men samme betydning. Her eksemplifiserte informant 4 ordet “vet” og “veit”, som har samme betydning, men i en transkribering vil muligens ordet “veit” bli feil, på grunn av at fasiten er “vet”. Dette er ord som har samme betydning som alle mennesker vil forstå, men datamaskinen vil telle det som en feil og gi en lavere score når det kommer til treffsikkerheten til den automatiske transkriberingen. I helsesektoren spesielt er det også uenighet om hvordan ord skal skrives, informant 3 forklarte at leger ofte er uenige om en skal bruke det latinske ordet eller fornorsket ord med samme betydning. Dette fører til uenighet mellom leger om pasientjournalen faktisk er skrevet korrekt, selv om betydningen av ordet er det samme. Det ble også nevnt fra flere av informantene at norsk orddeling kan by på utfordringer og kan svekke kvaliteten av den transkriberte teksten.

Flere informanter trakk også frem at det ofte benyttes tenkende og nølende lyder, for eksempel “Ehm” og “Hmm”, disse lydene blir tatt opp og kan påvirke kvaliteten av transkriberingen, på samme måte kan også kvaliteten påvirkes av bakgrunnsstøy. Bakgrunnsstøy ble også trukket frem som en utfordring av flere informanter, da dette gjør det vanskelig å tyde hva som blir sagt. Informant 3 fortalte at ved bruk av talegjenkjenning mot helsesektoren er dette noe de arbeider med, og at de hadde tatt i bruk støykansellering i modellene for å forbedre kvaliteten og unngå transkribering av tenkende lyder. I motsetning trakk informant 4 frem viktigheten av å transkribere nølende og tenkende lyder, spesielt når det gjelder transkribering av barnetale. Dette kan bidra til å forstå transkriberingen av barnet bedre og hva barnet forsøker å si mente informant 4. Talegjenkjenning av barnetale er en utfordring som kom frem i flere intervjuer, flere informanter beskrev dette som vanskelig med bakgrunn i at barn har en ulik uttale, artikulering og det er vanskelig å trekke ut definerte ord. De fleste informantene trakk også frem dialekter som en utfordring, ettersom vi har utrolig mange ulike dialekter i Norge. Flere hevdet at talegjenkjenningen helt klart er best på østlandsdialekt, noe som kan begrunnes i at modellene er mest trent på denne type

datamateriale. Informant 4 trakk også frem at kvaliteten på talegjenkjenningen blir redusert dersom det snakkes med utenlandske aksenter.

Alle informantene hevdet at kvaliteten på har en lang vei å gå, men hva talegjenkjenningen skal brukes til i etterkant er faktoren som er avgjørende for viktigheten av kvaliteten. I denne sammenheng nevnte informant 8 at presisjonen til deres modeller er målt til rundt 60-70% nøyaktighet, som ofte holder når journalistene bruker det internt til å strukturere datamateriale, ettersom journalistene kan identifisere mangler og høre på klippet hvor transkriberingen er mangelfull i etterkant. På en annen side, hvis resultatet av talegjenkjenningen brukes i offentlig sammenheng og skal være tilgjengelig for kunder, må kvaliteten bedres slik at materialet får verdi for kunden. Dette ga informant 8 oss et eksempel på ved å forklare at de skal ta i bruk automatisk transkribering av podcaster, slik at mennesker som er døve eller har nedsatt hørsel kan ta en del av trenden med podcaster. En modell som har en presisjon på 60-70% holder ikke i denne sammenhengen, og informant 8 påpekte at presisjonen til modellen må forbedres betraktelig før de kan tilby det til kundene sine.

#### 4.2.2 Det menneskelige perspektivet kan ikke erstattes

Flere av informantene påpekte at en utfordring med talegjenkjenning i dag er at det aldri vil kunne erstatte mennesket fullstendig. Informant 1 hevdet i denne sammenheng at dersom en kunde velger å ringe inn til kundeservice i stedet ved å finne svaret på nett, så ønsker kunden å snakke med et menneske og ikke en datamaskin. Vi gjorde samme funn hos informant 7 som påpekte at kundene som ringer inn til kundesenteret er svært takknemlige for at det fortsatt er mulig å snakke med mennesker, dette er noe informant 7 hevdet at kundene ofte gir tilbakemelding på.

Informant 2 forklarte at deres modell for talegjenkjenning automatisk genererer nøkkelord for hver samtale på kundeservice, slik at kunderådgiveren ikke behøver å skrive sammendrag etter samtalen. Informanten forklarte at de har gjort undersøkelser på sine kundesentre rundt dette, og har sett at omtrent halvparten av sammendragene må skrives manuelt allikevel etter implementering av talegjenkjenning. Dette begrunnet informant 2 med at det kunne være kompliserte saker, eller rådgivernes eget ønske. Det samme hevdet også informant 7 som jobber på et kundesenter som har implementert talegjenkjenning, hvor informant 7 påpekte at

sammendragene som skrives i etterkant ikke alltid holder. Informanten forklarte at i enkelte kundesaker må kunderådgiverne fortsatt skrive egne sammendrag, slik at det skal være forståelig for neste kunderådgiver dersom kunden ringer tilbake igjen. Informanten påpekte at det ofte kommer frem ulike ting i kundesamtaler som er viktig at er notert til neste rådgiver, og da holder det ikke at modellen for talegjenkjenning genererer enkelte nøkkelord om samtalen.

Informant 2 påpekte også at en utfordring med talegjenkjenning som gjør at det menneskelige perspektivet ikke kan erstattes er kunder som er på vei ut av bedriften. Informanten beskrev videre at kunderådgivere identifiserer kunder som er på vei ut av bedriften i samtaler og blir belønnet med bonus om de forhindrer kunden i å forlate bedriften. Informanten hevdet i denne sammenheng at de har en modell som kan analysere dette i stedet, men kunderådgiveren kan føle seg overkjørt og bedriften ser fortsatt behovet for det menneskelige perspektivet når det gjelder kunder som holder på å forlate bedriften.

Informant 4 påpekte også at modellene ikke kan forstå konteksten av samtalen på samme måte som et menneske, en modell vil ikke forstå for eksempel ironi. Det er hensiktsmessig å forstå konteksten i samtalen dersom modellen ikke oppfatter alle ord, for da vil et menneske mest sannsynlig forstå hvilket ord som mangler, men ettersom modellen ikke forstår konteksten, vil det være utfordrende for modellen.

#### 4.2.3 Nyttien av talegjenkjenning

Flere av informantene påpekte utfordringer når det kommer til hvordan nytten av talegjenkjenning begrenses i enkelte tilfeller. Informant 2 påpekte at nytten blir begrenset på deres kundesentre ettersom transkriberingen og tekstanalysen ikke blir tilgjengelig umiddelbart, men senest neste morgen. I denne sammenheng beskrev informanten at det hadde vært svært mange muligheter til å utnytte seg av transkriberingen og tekstanalysen dersom det hadde blitt tilgjengelig umiddelbart. Informant 2 trakk frem eksempelet om at dersom det er mye kø en dag, kan man umiddelbart se hvorfor mange ringer inn ved å se på hvilke nøkkelord som går igjen. Det ble også trukket frem at dersom kunden ringer inn igjen samme dag, vil ikke den nye kunderådgiveren se at kunden allerede har vært i kontakt med kundesenteret samme dag. Vi fant samme funn hos informant 5, som også påpekte at det en utfordring som begrenser nytten av talegjenkjenningen er at transkriberingen tok for lang tid. Informant 5 benyttet en modell som automatisk transkriberte det informanten sa i forelesning,

men informanten påpekte at transkriberingen som skulle være direkte, var veldig forsinket. Dette ble beskrevet som en utfordring ettersom hensikten var at studentene skulle slippe å notere i forelesninger og mange spørsmål fra studentene kom for sent. Sitat informant 5: *“Det tok like lang, om ikke lenger tid, som å notere for hånd. For så å vente på dette og ta imot spørsmål”*. Det ble i denne sammenhengen påpekt av informanten at opplevelsen av talegjenkjenning ikke var like effektivt som tiltenkt. Informant 5 hevdet også at en annen utfordring som begrenser nytten er at mange studenter ønsket å sortere og organisere notatene på sin egen måte, noe som gjorde at studentene hang etter under selve forelesningen, og tilbakemelding er at mange studenter bruker mye tid på det i etterkant av forelesningene.

Det ble også nevnt av flere informanter at det er utfordrende å finne en god og riktig metode for å gjennomføre tekstanalyse av det som er transkribert, dette er også en utfordring som bidrar til å begrense nytten av talegjenkjenning. Informant 2 påpekte at deres kundesenter har forsøkt mange ulike modeller, og det er viktig at tekstanalysen resulterer i et sammendrag av samtalen. Informanten beskrev at deres modell utfører en tekstanalyse og identifiserer 8 nøkkelord om samtalen som skal erstatte et manuelt skrevet sammendrag, samt identifiserer et tema for samtalen, men dette holder nødvendigvis ikke i alle tilfeller. I denne sammenhengen påpekte også informant 4 at det er svært utfordrende å lage en modell som lykkes med å treffe riktig tema etter tekstanalyse, dette kan være ekstra utfordrende i kundesamtaler på kundeservice hvor ordene kunden sier ikke er beskrivende nok for å identifisere temaet på samtalen. Videre påpekte også informant 4 at en utfordring som begrenser nytten av talegjenkjenning er at transkriberingen tar svært lang tid, normalt kan man bruke en time på fem minutter med tale. Informanten nevnte også at det kan mangle ord i ordboken til modellen, og da må ordene legges inn manuelt. Det oppleves derfor som svært tidkrevende å bygge opp en modell som en kan dra nytte av.

#### 4.2.4 Tilgang på datamateriale

Flere av informantene påpekte at en stor utfordring med talegjenkjenning er tilgangen på datamateriale. Det hevdes fra flere av informantene at en god og solid modell for talegjenkjenning og tekstanalyse krever trening med riktig datamateriale. Gjennom flere intervjuer kom det frem at datamateriale fra Språkbanken var mest brukt, dette er datamateriale som er tilgjengelig for alle. Informant 4 hevdet at utfordringen med dette datamaterialet er at det inneholder mest formelt språk og mye datamateriale hvor mennesker

leser fra manus, det kan føre til at mennesker undertrykker dialekten sin fordi de leser rett fra et manus på enten bokmål eller nynorsk og det er lite eller ingen spontan tale.

Informant 3 påpekte i denne sammenheng at det har vært utfordrende å få tak i datamateriale fra helsesektoren, på grunn av sensitiv pasientinformasjon. Etersom bedriften er spesialisert mot helsesektoren, har det vært kritisk å trene modellene på medisinsk språk. På en annen side påpekte informant 4 at det har vært utfordrende å få tak i datamateriale med spontan tale, for å forbedre kvaliteten til modellene for talegjenkjenning hevdet informanten at det er kritisk at modellene trenes med datamateriale hvor det snakkes fritt uten manus med alle tenkende og nølende lyder og hvor dialekter ikke blir undertrykt. Videre hevdet informant 4 at årsaken til at det er vanskelig å få tak i denne type datamateriale, er at mennesker som regel ikke samtykker til taleopptak av fri tale. Informanten påpekte videre i denne sammenheng at de supplerer forskningen med datamateriale fra Språkbanken og data fra Stortinget, men datamaterialene inneholder lite spontan tale. Likevel oppleves det som en utfordring å få tilgang på datamateriale som inneholder hele samtaler med spontan tale.

I motsetning til de andre informantene påpekte informant 8 at de har tilgang på mye datamateriale, men NRK bruker svært mye tid på å hente ut datamateriale til universiteter og andre institusjoner de samarbeider med. Det trekkes frem som en utfordring og jobber med å finne en løsning for en mer effektiv deling av datamateriale. Informant 4 påpekte også at det var stort potensiale i å bruke data fra NRK TV og NRK Radio, ettersom datamateriale inneholder mangfold av dialekter og det er en blanding av formell tale, dagligtale og spontan tale.

#### 4.2.5 Endringsvilje

Flere av informantene trakk også frem endringsvilje hos de ansatte i organisasjonen som en utfordring. I denne sammenheng påpekte informant 3 at selv om enkelte leger har brukt mikrofon i mange år, hvor lydopptaket har blitt sendt til en sekretær som transkriberer manuelt, har det vært enkelte leger som har vært skeptiske til å gå over til talegjenkjenning med automatisk transkribering. Informant 2 påpekte også at det har vært enkelte kunderådgivere på kundesentrene som ikke har ønsket å stoppe å skrive sammendrag i etterkant av samtalene. Etter implementeringen av talegjenkjenning og tekstanalyse hos kundesentrene var hensikten at kunderådgiverne ikke skulle bruke tid på å skrive sammendrag etter samtalene, men det ønsket rådgiverne å fortsette med. Informanten

forklarte at det tok tid for de ansatte å akseptere, og flere mener fortsatt at ikke alle sammendrag kan skrives av en datamaskin.

I denne sammenheng beskrev informant 7 opplevelsen av rollen som kunderådgiver etter implementering av talegjenkjenning på kundesenteret hvor informanten er ansatt. Sitat informant 7: *“Tidligere var det rom for å ta et lite pust i bakken mellom samtalene på kundesenteret, ettersom man skulle skrive et sammendrag av samtalen i etterkant. Nå får man ikke rom for å nullstille seg før man er rett inn i en ny samtale med en kunde”*. Informanten beskrev dette som utfordrende ettersom at det kan påvirke samtalen med neste kunde, ved at man muligens bringer med seg følelser fra forrige kundesamtale inn i en ny. Dette påpekte informanten at kan oppleves som en dårlig kundeopplevelse.

Lignende utfordringer ble også beskrevet av informant 6, ettersom informanten opplever at utfordringen for Omilon er å gi sykehus og ansatte forståelse av hvordan og i hvor stor grad talegjenkjenning kan utgjøre en forskjell i deres hverdag. Videre påpekte informanten endringsprosesser gjennom flere år på sykehus, som kan påvirke dette. Sitat informant 6: *“Omilon opplever også at mange i helsesektoren, spesielt på sykehusene, i mange år har vært gjennom mange større og mindre endringsprosesser, og er lei av endring. Dette gjør det tidvis vanskelig å overtale personell om flere endringer, men samtidig ser man at dagens teknologiskifte har en svært positiv innvirkning på hvordan personell tilnærmer seg digitale virkemidler og verktøy til forskjell for ti år siden.”*

### **4.3 Fordeler ved bruk av talegjenkjenning**

Gjennom intervju ble det tydelig at temaet *fordeler* var viktig for informantene, og de påpekte mange fordeler ved bruk av talegjenkjenning. Fordeler med talegjenkjenning er en sentral faktor som har innvirkning på hvilke effekter og muligheter talegjenkjenning gir for norske bedrifter. I dette underkapittelet vil vi presentere de mest sentrale funnene vi har gjort når det gjelder fordeler ved bruk av talegjenkjenning.

#### **4.3.1 Effektivitet og reduserte kostnader**

Det er tydelig at effektivitet er en tydelig fordel med talegjenkjenning, ettersom det ble nevnt av alle informantene våre. Informant 3 påpekte i denne sammenhengen at leger har mulighet til å ha flere pasienter i løpet av en dag, noe som også innebærer at det er mulighet for å tjene

mer penger. Det ble også trukket frem at det blir mer effektivitet i organisasjonen ettersom sekretæren, som tidligere kun har transkribert lydopptak til pasientjournaler, har mulighet til å gjøre andre oppgaver og bistå på andre områder i organisasjonen. Informant 3 påpekte også at dette ikke bare gjelder for helsesektoren, men effekten har de også sett i andre bransjer som for eksempel advokater, journalister, politi og for studenter. De fleste informantene påpekte hvor tidkrevende det er å transkribere tekst manuelt, og spesielt i helsesektoren og NRK hvor de er avhengige av transkribering av lydopptak. Talegjenkjenning kan dermed bidra til reduserte kostnader i form av ressursbesparelse ved transkribering. Informant 3 påpekte i denne sammenheng at talegjenkjenningen skriver 4 ganger raskere enn hva som er mulig på tastaturet.

Informant 2 påpekte også at en av de viktigste fordelene med talegjenkjenning er effektivitet. Informanten ga eksempel på dette ved at kundefrådgivere sparer mye tid, etter hver samtale må kundefrågiveren kategorisere samtalen og deretter skrive et kort sammendrag av samtalen manuelt. Etter implementering av talegjenkjenning og tekstanalyse på kundesenteret, er det nå mulig for kundefrådgivere å ta ny samtale umiddelbart og bruke mindre tid på administrative oppgaver. Transkribering, kategorisering og sammendrag av samtalen blir gjort automatisk av en modell.

Informant 7 påpekte at etter implementering var det mulig å bruke mindre tid på administrative oppgaver og mer tid til kunder, noe som igjen bidrar til mindre kø og mer effektivitet. Informant 8 hevdet også at det er store fordeler ved å slippe å transkribere manuelt ettersom det krever mye tid og ressurser, ettersom bedriften er avhengige av å ha materiale transkribert til tekst for å kunne strukturere datamaterialet. Informanten påpekte videre at når bedriften sparer ressurser, innebærer det også reduserte kostnader.

Informant 3 trakk også frem at deres løsning innebærer ingen stor engangssum, men at kundene betaler per bruker månedlig. Informanten beskrev at kundene deres derfor har lavere terskel for å teste ut talegjenkjenning, ettersom kundene tar en mindre risiko enn ved en stor investering. Videre forklarte informanten at dette gir kundene mulighet til å få bekreftet selv at implementering av talegjenkjenning innebærer reduserte kostnader og sparte ressurser. NRK har krav på seg fra Medietilsynet om at alt lydinnhold som publiseres skal være transkribert, påpekte informant 8. Dette innebærer veldig mange timer med manuell transkribering, og talegjenkjenning bidrar til å løse denne tidstyven og gir ansatte muligheten

til å bruke tiden på noe annet. Ettersom spesielt podcaster har blitt populært de senere årene, blir en stor andel av kundene til NRK ekskludert fra å lytte til podcaster på grunn av at de har nedsatt hørsel eller er døve, påpekte informant 8.

#### 4.3.2 Kvalitet

Kvalitet er en fordel med talegjenkjenning som ble trukket frem av mange av informantene. I denne sammenheng hevdet informant 3 at kvaliteten i helsesektoren øker ved at det aldri vil oppnås skrivefeil. Modellen for talegjenkjenning har en definert ordbok, og vil derfor aldri transkribere et konkret ord med skrivefeil. Ettersom medisinsk språk inneholder mange latinske ord, er dette et argument for at kvaliteten i arbeidet til legene øker. Det ble også påpekt av informant 3 at mange utenlandske leger behersker det norske språket godt muntlig, men ikke skriftlig, og med talegjenkjenning blir ikke rettskrivning en utfordring lenger. Informanten påpekte også at mange leger benytter talegjenkjenning aktivt for å kommunisere bedre med pasienter under konsultasjonen. Det innebærer at legen snakker høyt og pasienten kan bekrefte at det som blir skrevet ned er riktig, som igjen bidrar til å øke kvaliteten i arbeidet.

Informant 2 påpekte i denne sammenheng at kvaliteten er bedret etter implementering av talegjenkjenning fordi kundesenteret får muligheten til å forstå kunden bedre. Videre forklarte informanten at dette gir bedriften mulighet til å få en dypere kundeinnsikt, gi en bedre kundeopplevelse, samt forbedre tilbakemeldinger, markedsføringskampanjer og produktlanseringer. Informanten hevdet at dette er noe de har sett gjennom resultater etter implementering, hvor kundetilfredsheten har økt betraktelig. Verktøyene og datamaterialet brukes også i opplæring og videreutvikling av ansatte, som er med på å gi de ansatte en bredere og dypere forståelse og forbedrer kvaliteten i arbeidet. De samme funnene gjorde vi hos informant 7, som påpekte at kunderådgiverne måles av ledelsen på effektivitet og kundetilfredshet. Informanten forklarte videre at egen effektivitet og kundetilfredshet har bedret seg betraktelig etter implementering. Informant 7 påpekte også at resultatet av transkriberingen er blitt benyttet til opplæring og videreutvikling, og informanten mente at det hadde vært veldig lærerikt å lese transkribering av egen kundesamtale.

I undervisningssammenheng bidrar talegjenkjenning til at kvaliteten øker betraktelig når enkelte studenter følger forelesningen digitalt, ettersom det ofte innebærer dårlig kvalitet på lyd og bilde, hevdet informant 5. Talegjenkjenning bidrar til at disse studentene får tilgang på



automatisk transkribert tekst under forelesningen, noe som innebærer at studentene ikke går glipp av noe selv om lyden svikter. Informanten fortalte videre at engelske studenter har gitt tilbakemelding på at talegjenkjenningen bidrar til at forelesningene blir mer forståelig for studentene i skriftlig form. Informanten påpekte også at det er en stor fordel for studentene at vanskelige ord og begreper blir automatisk transkribert, ettersom studentene ofte bruker mye tid på det selv i forelesning og sliter med å følge med samtidig. Dette hevdet informanten er spesielt for studier med mange vanskelige ord og begreper, for eksempel innenfor helsevitenskap hvor informanten selv underviser. Informant 5 påpekte at i tilfeller hvor transkriberingen ikke har blitt tilgjengelig for studentene direkte, har informanten korrigert transkriberingene i etterkant og sendt resultatet til studentene. Dette hevdet informanten at studentene hadde gitt gode tilbakemeldinger på, og flere studenter mente at det var mer utbytte av forelesningen ved å slippe å notere selv. Det ble også påpekt av informant 5 at talegjenkjenningen har bidratt til at forelesere som ikke har norsk som morsmål eller har dårlig engelsk uttalelse, har hatt muligheten til å tilby studentene forelesningen transkribert uten skrivefeil. Dette sikrer at studentene ikke har gått glipp av viktig informasjon. Informanten påpekte også at studenter med generelle lærevansker har fått bedret kvaliteten på lærings situasjonen ettersom de nå kan lese det foreleser sier høyt. Informant 2 hevdet at kvaliteten på samtale på kundesenteret har økt, både fordi kunderådgivere har fått bedre opplæring gjennom talegjenkjenning og fordi det er blitt mindre kø ettersom kunderådgiverne bruker mindre tid på administrative oppgaver, noe som innebærer mer fornøyde kunder og høyere kvalitet på hver samtale.

#### 4.3.3 Informasjonsflyt

I mange intervjuer ble det også påpekt at talegjenkjenning bidrar til bedre informasjonsflyt innad i bedriften. Informant 3 fra Omilon ga oss et godt eksempel på dette, ettersom praksisen til leger som ikke bruker talegjenkjenning normalt er å spille inn lydopptak som transkriberes manuelt av en sekretær. Videre påpekte informanten at sekretæren ofte har flere lydopptak som skal transkriberes, og det kan danne seg kø. Dersom pasienten har blitt videresendt til en annen avdeling, er det hensiktsmessig at ny avdeling har journalen tilgjengelig umiddelbart, noe som ikke alltid er tilfelle ved manuell transkribering. Informanten påpekte også at talegjenkjenning gjør det enklere med vaktskifte på sykehus, slik at informasjonen fra den ene vekten automatisk er oppdatert til vaktskifte, og ingen lydopptak ligger i kø for manuell transkribering. Informanten hevdet at en bedret informasjonsflyt bidrar en forbedret pasientsikkerhet og kvalitet.

Lignende funn gjorde vi også hos informant 7, som hevdet at det ved flere tilfeller har hendt at kunderådgivere på kundesenteret har glemt å skrive sammendrag av samtalen med en kunde. Informanten forklarte at dersom samme kunde ringer inn igjen for samme problem, er det nyttig å kunne lese et sammendrag av forrige samtale eller eventuelt annen viktig informasjon. Talegjenkjenning sikrer at sammendrag ikke blir glemt, og gir den neste kunderådgiveren mulighet til å forstå problemet bedre og sikrer en bedre informasjonsflyt på kundesenteret. Informanten påpekte også at det blir bedre informasjonsflyt mellom kunderådgivere og ledelsen, ettersom hver samtale blir kategorisert etter tema kunden ringer inn for, blir ledelsen oppmerksom på problemer kunder ringer inn for. Dersom det er enkelte temaer som går igjen, kan ledelsen sette inn tiltak for å forbedre tjenesten på det området. Informant 2 påpekte også nytten av å få oversikt over temaer kundene ringer inn for. Informanten presenterte plattformen de benytter for oss, og der kunne informanten navigere seg til en ordsky med de mest populære temaene den siste uken. I denne sammenheng hevdet informanten at man ofte så at dersom tiltak ble satt inn fra ledelsen, forsvant temaet fra ordskyen etter en liten stund ettersom kunder ikke ringte inn for det aktuelle problemet lenger.

#### 4.3.4 Ansatte og ergonomi

Flere av informantene påpekte også at en viktig fordel med talegjenkjenning er mer fornøyde ansatte. Først og fremst hevdet informant 3 at talegjenkjenning gir ergonomiske fordeler, i denne sammenheng trakk informanten frem at brukere av talegjenkjenning unngår kjente skader i rygg og arm som mange får ved å jobbe mye med tastatur og datamus. Informanten trakk også frem at talegjenkjenning også kan benyttes som et tilbud for uføretrygdete som for eksempel har mistet hendene sine, å komme tilbake på jobb. Det ble påpekt at Omilon sin talegjenkjenning også fungerer med bruk av kommandoer, ved at brukeren for eksempel kan si “åpne nytt dokument” eller lignende.

Informant 2 påpekte at talegjenkjenning har bidratt til at de ansattes tilfredshet har økt, dette begrunnet informanten med at de ansatte sparer tid og at de påvirkes av fornøyde kunder. De ansatte kan bruke mer tid på å skape gode kundeopplevelser, i stedet for å gjøre administrative arbeidsoppgaver. Vi gjorde det samme funnet hos flere informanter, og mange av funnene tyder på at ansatte blir mer tilfredse i arbeidssituasjonen når de får redusert mengde ensformig arbeid.

#### 4.4 Muligheter ved bruk av talegjenkjenning

I flere av intervjuene tydelig at bruk av talegjenkjenning også innebærer mange store *muligheter*, og informantene ga oss mange eksempler på hvilke muligheter talegjenkjenning bidrar til. I dette underkapittelet vil vi presentere de mest sentrale funnene vi har gjort når det gjelder muligheter ved bruk av talegjenkjenning.

##### 4.4.1 Innovative muligheter

Det ble trukket frem av flere av informantene at talegjenkjenning gir muligheter for å være innovative. I denne sammenheng påpekte informant 2 at talegjenkjenningen gir bedriften mulighet til å utnytte dataanalyse på en innovativ måte. Automatisk transkribering av alle samtaler, samt tekstanalyse i etterkant gir bedriften verdifull datamateriale med uendelige muligheter. Informant 2 påpekte at dette gir bedriften en unik mulighet til å differensiere seg fra konkurrentene sine. Informanten påpekte videre at ettersom bedriften nå har tilgang til verdifullt datamateriale, gir det muligheter til å benytte datamaterialet til salgsmuligheter, spesielt når det gjelder utgående samtaler. Det ble også påpekt at datamaterialet kan benyttes til å analysere bedriftens interne prosesser og identifisere hvor hyppig kundene blir kontaktet av organisasjonen. Informanten hevdet at det gir bedriften mulighet til å endre interne prosesser for en forbedring, samt at datamaterialet kan utnyttes til å lage mer effektive kampanjer og produktlanseringer. Flere informanter hevdet også at kombinasjonen av talegjenkjenning og tekstanalyse gir bedriften unike muligheter til å lage analyser, rapporter og modeller for å videreutvikle seg og forstå kunden på et dypere nivå.

##### 4.4.2 Uendelige muligheter

Funn fra flere intervjuer indikerte at talegjenkjenning har uendelige muligheter og det er muligheter for å bruke det i de fleste bransjer. Flere informanter påpekte blant annet at politiet har et eget prosjekt for talegjenkjenning. I denne sammenheng forklarte informant 8 at NRK har et samarbeid med Oslo Politidistrikt hvor formålet er å transkribere barnetale. NRK bidrar med innhold fra NRK Super, slik at politiet har datamateriale for å trene modellene sine ettersom transkribering av barnetale kan være utfordrende. Videre fortalte informanten at NRK har svært mye datamateriale med bredt utvalg av dialekter i arkivet sitt, og at de samarbeider med universiteter og kommersielle bedrifter for å bidra til videreutvikling av talegjenkjenning. Informant 4 påpekte også at talegjenkjenning gir muligheter for barn og voksne å bli bedre på uttale av andre språk. Videre fortalte

informanten at det er en applikasjon ved navn “Duolingo” som bruker talegjenkjenning for å forbedre barn og voksnes uttale i ulike språk. Denne applikasjonen gir muligheter for å bruke talegjenkjenning til språklæring.

#### **4.5 Fremtiden for talegjenkjenning**

Temaet *fremtiden* for talegjenkjenning var også et viktig tema i flere av intervjuene. Fremtiden for talegjenkjenning er en viktig faktor å studere for å belyse hvilke effekter og muligheter talegjenkjenning gir for norske bedrifter. I dette underkapittelet vil vi presentere de mest sentrale funnene vi har gjort når det gjelder fremtiden ved bruk av talegjenkjenning.

I denne sammenheng hevdet informant 3 at Omilon ser for seg en større kundegruppe innenfor andre sektorer i fremtiden, ettersom 90% av deres kunder i dag er innenfor helsesektoren. Informanten påpekte videre at bedriften har behov for å trene modellene sine på mer formell og norsk dagligtale med et bredt utvalg av dialekter for å gjøre talegjenkjenningen effektiv innenfor andre sektorer. Omilon så også for seg å ta i bruk Natural language processing i fremtiden, slik at det kan brukes for å kvalitetssikre journalene, hvor modellen for eksempel automatisk spør legen “Høyre eller venstre?”. Informanten fortalte videre at alle pasienter må registreres med ulike koder for diagnoser og sykdommer, i denne sammenheng har bedriften en ambisjon om å utnytte tekstanalyse til å automatisk registrere kode ut fra teksten skrevet i journal, noe som sparer legen for mye tid. Det samme påpekte også informant 6, hvor fremtidige planer for Omilon er å utvikle integrasjoner og tjenester som kan gjøre medisinsk koding enklere for de ulike sykehusene, dette kan gi Omilon gevinst gjennom høyere volum av brukere av deres løsning.

Informant 2 påpekte at de så for seg å utnytte talegjenkjenning og tekstanalyse til å identifisere salgsmuligheter med automatiske algoritmer. Målet innebærer at selskapet kan få økte inntekter ettersom salgsmuligheter som normalt ikke ville vært oppdaget, oppdages av automatiske algoritmer. På den andre siden har informant 4 ambisjoner om å forbedre talegjenkjenning på et generelt grunnlag, slik at talegjenkjenningen kan brukes av alle. Informanten fortalte også om ambisjoner om å forbedre chatbots og trene opp roboter som kan ta over resepsjonsoppgaver, som for eksempel å booke møter. Informant 4 påpekte at teknologien begynner å bli så bra at det kan bli skummelt dersom teknologien kommer i feil hender, i denne sammenheng trekker informanten frem at den kan brukes i falske nyheter

eller i identitetstyveri. Informanten trakk også frem behovet for å forbedre voice assistants i fremtiden, slik at tjenestene kan brukes av alle mennesker med ulike dialekter og utenlandske aksenter.

Informant 8 påpekte at fremtiden for talegjenkjenning for NRK er å fortsette med utvikling av talegjenkjenning som er åpent og tilgjengelig for alle. NRK har en ambisjon om at alle bedrifter og universiteter som arbeider med talegjenkjenning kan bidra til videreutvikling, og ikke drive med silovirksomhet. Informanten påpekte videre at de også ønsker å bruke talegjenkjenning og natural language processing til å håndtere store datamengder og koble sammen arkivinnhold med nytt datamateriale. Informanten eksemplifiserte dette med at journalister kan automatisk koble nytt datamateriale om en hendelse med arkivinnhold fra lignende hendelse på samme sted tidligere. Ambisjonen er at talegjenkjenningen og tekstanalysen skal trekke ut en kontekst, og benytte kontekstforståelse til å koble sammen arkivinnhold.

## 5.0 Diskusjon

I dette kapitlet skal vi diskutere funnene våre opp mot problemstillingen, samt identifisere funnenes begrensninger. Like viktig vil vi også påpeke studiets implikasjoner. Funnene våre har i seg selv identifisert en god del effekter og muligheter talegjenkjenning gir bedriftene, men ved å sammenligne bedriftene mot hverandre og inkludere litteratur vil vi kunne spekulere rundt flere aspekter.

### 5.1 Muligheter for kundesenteret

I kapitlet under funn viser vi til konkrete utregninger vi selv har utført basert på bedriftens interne dokumenter. Denne analysen bekrefter overfor kundesenteret at en investering i talegjenkjenning er lønnsomt, og at engangskostnadene for investeringen og driftskostnadene er lavere enn sparte kostnader i form av ressurser. Dokumentstudiet gir oss også muligheten til å diskutere om teknologien kan brukes på andre måter for å øke effektiviteten og derav lønnsomheten.

#### 5.1.1 Bruke transkribering etter en kundesamtale

I tråd med funn fra andre informanter kan man spekulere i om tekstanalysen og bruk av talegjenkjenningssystemet kan anvendes på en annen måte for å gjøre det om ikke enda mer lønnsomt og effektivt. Siden flere kunderådgiverne må inn etter en samtale og skrive om sammendragene for at notatene skal bli mer forståelig, er det åpenbart at en forbedring her vil kunne spare enda mer tid per samtale. Et slikt notat fra en samtale med en kunde er nærliggende likt det en lege skriver inn i en pasientjournal etter et pasientbesøk. Slikt som informant 3 forklarte, har leger full oversikt over hva som blir skrevet ned i sammendragene og ikke minst mulighet til å rette på dokumentet. I dette tilfellet ville altså en kunderådgiver kunne legge på med kunden og etterpå snakke inn en oppsummering i det samme systemet. Her vil man naturligvis inkludere de mest relevante nøkkelordene fra samtalen som tekstanalysen kunne plukket opp. Hovedgrunnen for at dette kan være en forbedring er tiden spart ved å ikke måtte skrive inn et sammendrag manuelt på et tastatur. Om det hadde blitt mer lønnsomt og effektivt kan åpenbart beregnes på lik linje som vi har gjort i kapittel 4.1.3. Dersom rundt halvparten av kunderådgivere går inn og skriver manuelle sammendrag kan man lett regne ut gjennomsnittstiden dette tar opp mot tiden det tar å automatisk transkribere det inn.

I tillegg kan også tiden for kategoriseringen av samtalen forbedres ved at nøkkelordene nå blir mer nøyaktige. Som informant 2 forklarte skrev de fleste kunderådgivere manuelle sammendrag i mer kompliserte saker. Det er trolig disse sakene som også vil ha utfordringer rundt tekstanalysen ved å kunne identifisere de 8 mest relevante nøkkelordene og kategorisere deretter. Dersom sammendraget hadde blitt automatisk transkribert i samme system kan man se for seg at analysen blir enklere dersom de relevante nøkkelordene befinner seg både i samtalen og sammendraget. I tillegg til lønnsomheten kan man også spekulere i om plattformen de bruker for å få oversikt over kundesamtalene vil kunne bli enda mer nøyaktig. Dette kan sammenlignes med Omilons forklaring av en forbedret informasjonsflyt i helsesektoren; Temaer (Topics by date) og samtaletype (Conversation Types) oppdateres fortere og ikke minst kan dette påvirke kundesenterets køsystem i en positiv forstand. For dersom sammendragene forbedrer nøkkelordene og øker informasjonsflyten vil det også være enklere å identifisere grunner for at kunder ringer inn og derav kunne kontrollere køen bedre, noe som i utgangspunktet er et av målene for implementeringen. Som et tilleggsargument for å bruke talegjenkjenning også i etterkant av kundesamtalene, er muligheten for å begrense utfordringer rundt det menneskelige perspektivet. For det første vil en slik ordning gjøre at man ikke er totalt avhengig av teknologien, men at et menneske faktisk kan formulere et sammendrag som retter opp eventuelle teknologiske misforståelser, slikt som informant 4 forklarte som ironi, følelser og ufullstendige setninger. I tillegg vil det kanskje også forbedre endringsviljen ved at kunderådgiveren nå får et større *“rom for å nullstille seg før man er rett inn i en ny samtale med en kunde”* som informant 7 beskrev.

### 5.1.2 Forbedre kunderelasjoner

Som litteraturen beskriver er nok talegjenkjenningsteknologien langt unna en genuint “menneskelignende” kunstig intelligens, noe som i stor grad beskriver hvorfor blant annet kunder ønsker å komme i kontakt med “ekte” kunderådgivere når de først tar kontakt med leverandøren. Det som dog er interessant med våre funn er at bruk av talegjenkjenning på kundesenteret kanskje faktisk har en motsatt effekt, nemlig at den forbedrer kunderelasjoner i den grad at det oppfattes mer personlig og menneskenært. Informant 2 påpekte at transkriberingen og tekstanalysen per dags dato ikke blir tilgjengelig med en gang, men senest neste morgen. Dette er noe som på sikt skal oppdateres umiddelbart, og vi ønsker å illustrere med eksempel hvordan talegjenkjenning kan forbedre kunderelasjoner når dette først er tilgjengelig. Vi kan ta et eksempel der en kunde ringer inn for andre gang på samme

dag med et problem som ikke ble løst første gangen. Siden det ikke er lang tid mellom samtalene har ikke kundefrådgiver rukket å skrive noe manuelt sammendrag og er også opptatt i en annen samtale. Alternativt kan det også være et tilfelle likt det informant 7 nevnte om kundefrådgivere som glemmer å skrive sammendrag og at det derfor ikke er noe relevant logg av en kunde som ringer inn igjen. Men på grunn av talegjenkjenningssystemet og plattformen de bruker kan kundefrågiveren nå blant annet tidlig registrere det at kunden har vært i kontakt tidligere samme dag, hvem vedkommende var i kontakt med, hva problemet gjaldt og andre eventuelle personlige detaljer. I første omgang kan også kundefrågiveren være klar over hva kunden mest sannsynlig ringer inn for, fordi systemet har registrert et gjentagende problem for andre tilsvarende kunder denne dagen. Med andre ord gir teknologien og en forbedret omløpshastighet muligheten til å forbedre kundefrågiveren og gjøre den mer personlig.

Ser man på dette eksempelet med et fremtidig perspektiv kan kundefrågiveren også forbedres i takt med teknologiens utvikling, for eksempel det informant 2 nevnte rundt bruk av dataanalyse for salgsmuligheter. En mulighet her vil jo være å identifisere kundenes personlige behov, ønsker, egenskaper og muligheter og som igjen kan forsterke kundefrågiveren. Det er også relevant å nevne at tekst- og dataanalysen også har potensiale til å brukes for å identifisere kundetilfredsheten blant kunder som tar kontakt, for eksempel ved å identifisere positive, og negative ladede ord fra en transkribert samtale eller et sammendrag. Dette kan være en måte å måle effekten av teknologien, på lik linje som å kalkulere lønnsomheten i kroner.

## **5.2 Effekter i andre norske bedrifter**

Dokumentstudie og videre eventuelle muligheter for kundesenteret kan også være med på å argumentere for effekter som blant annet økt lønnsomhet i de andre norske bedriftene vi studerer. Ettersom vi ikke har noe talldata på disse bedriftene vil det derfor være mest relevant å identifisere de nåværende effektene og diskutere muligheter for å kunne kalkulere lønnsomheten med interne, gitte tall.



### 5.2.1 Helsesektoren

Som litteraturen beskriver gjennom 20 år så har talegjenkjenning økt effektiviteten i helsesektoren og det er interessant om våre funn støtter oppunder dette. Like interessant er det også å se på om funnene kan kalkuleres på noe måte. Informant 3 nevnte at tjenesten de tilbyr innebærer ingen store investeringer, men kun enhetskostnader i forhold til hvor mange som bruker tjenesten. Vi fikk ikke noen konkrete tall av informanten, men det skal ikke være noen store faste summer for helsesektoren ved å bruke talegjenkjenning. I denne sammenheng ble det også påpekt at det er en åpenbar forbedring av omløpstider for blant annet pasientjournaler ved hjelp av talegjenkjenning. De samme funnene gjorde vi i eksisterende litteratur, som viste at bruk av talegjenkjenningssystem gjorde at andel rapporter tilgjengelig innen time forbedret seg med 32%. Dette fører naturligvis til at helsesektoren kan få flere pasienter inn daglig i de forskjellige avdelingene. Som informant 3 påstod så skyldes dette at pasientjournalene skrives 4 ganger raskere enn tidligere og at ingen pasientjournaler blir sittende fast i sekretærkø på grunn av for eksempel sykdom. I tillegg behøver ikke pasienten å stå i like lang kø dersom de blir henvist videre i påvente om at den nye avdelingen skal få oppdaterte pasientjournaler. Med disse forholdene kan man med andre ord regne ut hvor mange flere pasienter man får inn i snitt hver dag og deretter de økte inntektene for helsesektoren.

I tillegg til dette kan helsesektoren også regne ut sparte kostnader i form av sekretærlønn. Ettersom manuell transkribering er svært tidskrevende, vil talegjenkjenning brukt av legene selv bidra til at sekretæren nå kan bruke ressurser på noe annet lønnsomt for bedriften. En gjennomsnittslønn for en sekretær er normalt 454 560,- årlig (SSB, 2021). Det vil si at bedriften kan ha sparte kostnader på 454 560,- for hver sekretær som er ansatt for å bare transkribere lydfiler til pasientjournaler. Med alle overnevnte forhold kan man med andre ord regne ut samme kalkulasjoner som ved dokumentstudie. I motsetning til kundesenteret kan man her registrere hvilke avdelinger eller sykehus som vil dra nytte av å forbedre en omløpstid, og da om denne forbedringen blir lønnsomt opp mot enhetskostnaden for å implementere talegjenkjenningssystemet i dette tilfellet. På den måten kan dette føre til en ganske sikker investering før selve implementeringen, men også gi muligheten for å teste ut systemet og kalkulere lønnsomheten før det blir tatt i bruk.

### 5.2.2 Undervisningssammenheng

For institutter som bruker talegjenkjenning ved undervisning, beskriver det meste av funnene våre til at kvaliteten på undervisningen og andre studieforhold forbedrer seg. Likevel er det interessant å diskutere om man også her kan forbedre lønnsomheten ved å anvende systemet riktig og til de riktige formål. Med det mener vi tilfeller der forelesere bruker mindre tid på å forberede seg før undervisningen eller bruk av etterarbeid for notater; Dersom transkriberingen under forelesningen erstatter forberedte notater eller dokumenter som foreleser må lage i ettertid. Dette er åpenbart konkrete timer spart og derav sparte lønnskostnader for instituttet. Dette beviser at talegjenkjenning også kan føre til sparte kostnader og ikke nødvendigvis bare blir sett på som en investering for å øke kvaliteten for undervisningen.

Dersom lærere kan implementere talegjenkjenning på samme måte som Omilon gjør i helsesektoren vil det også være mulig å identifisere hvilke tilfeller der det faktisk er lønnsomt å bruke talegjenkjenning og når det har størst effekt. Som informant 3 beskriver så vil man i helsesektoren kunne betale for per bruker av systemet månedlig, og på den måten vil ikke hele bedriften måtte implementere et stort system med tilsvarende store kostnader. Som informant 5 forklarte så er det ikke alle lærere, studenter, fag og konkrete forelesninger som vil ha noe nytte av å bruke talegjenkjenning. Kan man derfor diskutere om talegjenkjenning heller kunne blitt brukt i spesifikke sammenhenger? For å ta et eksempel fra funnene kunne man gitt utenlandske studenter tilgang på talegjenkjenning for deres forelesninger for å kunne følge bedre med både språklig og eventuelle digitale utfordringer. Litteraturen nevner også bedre muligheter for personer med lese- og skrivevansker. I tillegg kunne også tungt faglige undervisninger ha talegjenkjenning tilgjengelig for alle studenter slik at stoffet ble enklere å forstå. På den måten kan talegjenkjenning både ha en stor effekt for de som faktisk har bruk for det, samtidig som det ikke vil føre til unødvendige store kostnader.

### 5.2.3 Andre norske bedrifter

Idéen om at talegjenkjenningssystemet kan implementeres for noen relevante brukere og ikke for en hel bedrift med store kostnader gjør at man kan diskutere muligheter for å forbedre lønnsomheten for andre norske bedrifter utenom våre informanter. For eksempel beskrev både informant 1 og 8 bruken av talegjenkjenning for politiavhør og gjerne i samsvar med barn. I dette eksemplet vil jo talegjenkjenning gi en stor effekt fordi det kan være lettere å

forstå barn når ord blir transkribert. Ikke minst kan dette også brukes for personer med utfordrende uttale rent språklig. Én ting er at det forbedrer selve avhøret, men det vil også være nærliggende å tro at det sparer tid ved at man ikke behøver store ressurser på å transkribere så nøyaktig som mulig i etterkant av et avhør. Et annet eksempel er det informant 3 nevnte om bruk av talegjenkjenning i advokatvirksomheter. Her vil det være lett å kunne kalkulere spart tid ved å avskaffe sekretærer på lik linje som i helsesektoren.

I disse overnevnte eksemplene kan man også diskutere om omløpstiden for alle dokumenter ville forbedret seg betraktelig. Litteraturen og funnene våre viser stadig til forbedring i bedrifter der dokumentene både transkriberes fortere, men også at de digitaliseres. Dersom notater fra forelesninger, avhørsrapporter, saksdokumenter fra advokater eller oppdatert kundeinformasjon blir raskere tilgjengelig er det stor sannsynlighet for at bedrifter nyter godt av det. Som et tilleggsargument kan man spekulere i om talegjenkjenning også åpner opp for muligheter til å frigjøre flere arbeidsressurser og spare lønnskostnader, slik som tilfellet med sekretærer i helsesektoren.

Avslutningsvis ønsker vi også å nevne mulighetene talegjenkjenning gir bedrifter for å inkludere flere ansatte. Informant 3 beskrev den positive fordelene talegjenkjenning har rent ergonomisk for de som bruker det. Det faktum at leger kan forhindre å pådra seg typiske skader ved å skrive manuelt kan jo relateres til alle yrkestitler der mus og tastatur anvendes. Dette er et interessant tilleggsperspektiv fordi det ser mer på mulighetene en bedrift har for å gjøre arbeidsplassen mer attraktivt og at flere ansatte kan jobbe der over lengre perioder. Ikke minst gir det mulighet for at blant annet svært faglig kompetente personer, men som i utgangspunktet ikke har mulighet til å bruke tastatur, faktisk kan ansettes på grunn av talegjenkjenningsteknologien. Dette kan også relateres til utenlandske ansatte som er avhengig av direkte oversettelse eller annen transkribert skrift fordi de for eksempel ikke behersker språket eller befinner seg geografisk et annet sted. Med tanke på dette er det jo derfor ikke til å unngå å nevne at dette igjen kan være kostnadsbesparende og øke kvaliteten for enhver norsk bedrift.

### **5.3 Muligheter i fremtiden**

Talegjenkjenning kan også gi muligheter for norske bedrifter i fremtiden. I forrige avsnitt diskuterte vi om det er mulig å bruke den nåværende teknologien annerledes i fremtiden for å

blant annet spare kostnader. Men ser vi derimot på litteraturens historiske perspektiv, kan man kanskje ganske bastant konkludere med at teknologiutviklingen har store fordeler for vår problemstilling. Det er derfor like interessant å spekulere i hva slags effekt den fremtidige teknologiske utviklingen vil kunne gi for bedriftene. Litteraturen og funnene våre har mye til felles rundt det fremtidige perspektivet på talegjenkjenning og studiet vårt kan derfor spekulere i ulike scenarier på bakgrunn av dette.

### 5.3.1 Forbedre bruken av tekstanalyse

Mye av mulighetene for talegjenkjenning virker å ligge i tekstanalysen og annet bruk av den automatiske transkriberingen i ettertid. Dette henger i tråd med litteraturen som beskriver NLP som teknologisk dominerende i fremtiden. Informant 2 nevnte det å bruke datamateriale til å hevde seg på markedet. I tillegg til å bruke informasjonen til å forbedre kunderelasjoner er det også åpenbart verdifull informasjon der man kan identifisere salgsmuligheter, som for eksempel å foreslå å oppgradere personlige abonnement eller tilleggstjenester. Det kan jo diskuteres om noe lignende kan brukes i andre bedrifter også. Et eksempel er om alt fra en forelesning blir automatisk transkribert- både det informant 5 selv sier, men også hva studentene diskuterer eller spør om. Dersom man hadde utført noe tekstanalyse på denne dataen kan det hende at lærere hadde fått en raskere og bedre oversikt over hva slags spørsmål som går igjen fra studentene og justert undervisningen deretter. I tillegg kan man også lettere få en oversikt over hva slags eksempler, temaer eller scenarier som er tatt opp i forelesninger og bruke det som bevis for pensumgrunnlag ved eksamen. Informant 8 nevner jo at de ønsker å bruke teknologien for å kunne koble nye saker med arkivinnhold for å få større sammenhenger. Vi finner det også sannsynlig at noe slikt ville gitt verdi både for politiavhør, advokatvirksomheter og også i undervisningssammenhenger.

Informant 3 beskrev et ønskelig fremtidig perspektiv der teksten ble kvalitetssikret ved å få opp ulike forslag. Dette var blant annet at modellen automatisk spør legen om avgjørende ord slikt som om pasienten hadde vondt i “høyre” eller “venstre” bein, men også om legen mente at det var “uviktig” eller “viktig” informasjon. En slik teknologi kan muligens også ha stor verdi for de andre bedriftene og informantene. Ser man også i sammenheng med litteraturen, kan teknologiutviklingen i form av å registrere leppebevegelser forbedre den generelle kvaliteten. En slik forbedring i kvalitet kan sikre at man unngår det litteraturen beskriver som livstruende feil ved bruk av talegjenkjenning.

Kvalitetsforbedringen kunne også muligens hjulpet på utfordringen om sammendrag og nøkkelord for kundesenteret; Hvis man ser for seg et scenario der kunderådgiveren automatisk fikk opp forslag til nøkkelord eller forslag til kategorisering av samtale direkte etter kundesamtalen og ville derav fått bedre kvalitet på notatene og ville også sluppet å føre det inn selv. Dette kan sammenlignes med det informant 3 beskrev som en fremtidig mulighet for helsesektoren, der tekstanalyse benyttes for å generere en kode for å registrere pasienter. Man kan også se for seg at dette kunne blitt brukt i forbindelse med direkte oversettelse av sendinger slikt som i NRK for at setningene ville blitt bedre i tilfeller der ord i utgangspunktet var uforståelig. Et enda større perspektiv er å se på det i undervisningssammenheng, der tekstanalysen kunne identifisert eventuelle feil som foreleser sier eller identifisert temaer som blir tatt opp og som vanligvis bør brukes bedre tid på.

### 5.3.2 Bedre modeller

En stor del av både litteraturen og våre funn viser til det spesielle for talegjenkjenning i Norge og for norske bedrifter, nemlig bruk av dialekter. Jo mer trent modellene som de norske bedriftene bruker er på norske dialekter, jo bedre ville alle de ovennevnte effektene og mulighetene være. Modellene vil rett og slett bli mer nøyaktig og det er ingen av bedriftene som ikke ville ha dratt fordeler av bedre trente modeller på dialekter, rett og slett fordi man kan anta at alle bedriftene har ansatte, kunder, pasienter, studenter eller lærere som har forskjellige dialekter. Litteraturen nevner jo også det juridiske rundt våre to skriftspråk og at man i mange tilfeller skal ha mulighet til å velge om teksten skal automatisk transkriberes på bokmål eller nynorsk. På den måten kommer man ikke unna det faktumet at talegjenkjenning ville blitt mer tilgjengelig generelt om alle modellene var like gode på begge skriftspråkene og på alle dialekter. For å ta et eksempel fra undervisningssammenheng ser man for seg en utenlandsk student som sitter i et annet land og hører på en foreleser fra Molde. Studenten forstår skriftlig bokmål, men har ikke mulighet til å følge med på muntlig norsk med Molde-dialekt. Dersom forelesningen i dette tilfellet ble automatisk transkribert til perfekt bokmål, som for eksempel at “I” (som betyr “jeg” i Molde) hadde blitt oversatt til “Jeg” er det ingen tvil om at verdien ville økt betraktelig.

Dette er det litteraturen forklarer som NLP, og det er flere aspekter ved denne teknologien som kan gjøre det enklere for en del av de nåværende utfordringene til de norske bedriftene. Informant 3 nevnte mulighetene for å spesialisere seg på hver bruker. Det at modellene er

trent opp på én foreleser og da vedkommendes typiske setningsoppbygninger eller dialekt vil naturligvis forbedre transkriberingen.

Et interessant perspektiv er også det litteraturen ser på som muligheter for å inkludere mer av det følelsesmessige perspektivet. Dersom teknologien utvikler seg til å kunne forstå mer av blant annet ironi, sinne eller generelt tolke blant annet nølelyder i bedre forstand vil mulighetene trolig bli enda større. Det vil da kanskje bli lettere å transkribere podcaster, radio og annet for NRK fordi teksten kan inkludere tegn slikt som utropstegn eller forklarende ord for døve slikt som "... Roper med gråt i halsen". I tillegg kan dette naturligvis være svært verdifullt for politiavhør. Som litteraturen beskriver, er det også verdifullt for spillutvikling og kan sammenlignes med det informant 4 blant annet forklarte rundt bruken av talegjenkjenning for fremtidig språklæring og appen "Duolingo".

Avslutningsvis er det viktig å påpeke at jo flere bedrifter som bruker teknologien og som implementerer talegjenkjenning, jo mer data vil også naturligvis være tilgjengelig for modellene. Informant 4 uttrykte i flere tilfeller etterspørselen etter data fra blant annet flere dialekter og generell spontan tale. Informant 3 etterspurte også mer faglig språk for andre bedrifter, slik som for eksempel advokatvirksomheter. Informant 4 kunne også fra et forskningsperspektiv se for seg at talegjenkjenning kunne bli brukt mer i lek, spill og ikke minst digitale plattformer for å lære seg språk og dialekter. Dette i seg selv vil jo kanskje føre til mer data om spontan tale og også da forminske tilfeller med overtrente modeller for konkrete bedrifter. I tråd med dette er det også essensielt å påpeke viktigheten som litteraturen og blant annet NRK og SCRIBE beskriver om bruken av open source. Det å ha tilgjengelige og åpne koder vil generelt kunne forbedre bruk av talegjenkjenning for norske bedrifter og oppsummert da gi både flere effekter og muligheter for alle.

## **5.4 Oppsummering av resultater**

Resultatene fra vårt studie viser at den største og mest relevante effekten av talegjenkjenning for norske bedrifter er økt lønnsomhet. Dette er spesifikt knyttet til kundesenteret og de kalkulerede tallene som viser til sparte kostnader. Måten vi har kombinert tall fra interne dokumenter og tilgjengelig offentlige tall, har gjort det mulig å spekulere i hvordan andre norske bedrifter også kan kalkulere hvor stor effekt talegjenkjenning kan ha på deres

lønnsomhet. Dette er interessant fordi mange av våre funn fra informantene har ikke i utgangspunktet fokusert på direkte lønnsomhet i form av økte inntekter eller sparte kostnader.

De andre type effektene som talegjenkjenning nå gir for norske bedrifter kategoriserer vi som ikke-kostnads relaterte fordeler. Dette er blant annet de ergonomiske fordelene som gjør det mulig å unngå skader ved bruk av mus og tastatur samtidig. I tillegg gir det også en positiv effekt for studenter der kvaliteten på undervisningen øker. Flere av bedriftene har også ved flere anledninger påpekt en bedre informasjonsflyt ved bruk av teknologien. Dette skyldes både at dokumenter kommer inn raskere i systemene, men ikke minst at systemer blir digitalisert og samlet på én plattform.

Resultatene åpner opp for mange fremtidige muligheter for norske bedrifter som bruker talegjenkjenning, og vi spekulerer i at dette avhenger av hvordan teknologien brukes. I første omgang vil mulighetene være større dersom norske bedrifter kan kjøpe talegjenkjenning per bruker og derfor unngå store enhetskostnader eller betale for at teknologien er tilgjengelig i unødvendige tilfeller. Resultatene viser også at muligheten kan gi enda mer lønnsomhet for kundesenteret og for undervisning dersom de bruker transkriberingen på lik linje som en pasientjournal. I tillegg kan gode kunderelasjoner og salgsmuligheter forbedre lønnsomheten ved å bruke transkribert kundeinformasjon. Dersom talegjenkjenning blir tilgjengelig for de med ergonomiske, språklige eller geografiske utfordringer kan bedriftene øke inkludering og muligheter for ansatte, kunder, lyttere og studenter.

I tillegg til riktig bruk av talegjenkjenningsteknologien, viser også resultatene at mulighetene er større sammen med den fremtidige teknologiske utviklingen. Dersom modellene trenes opp på større både kvalitativ og kvantitativ data, kan de overnevnte effektene forbedres, men også gi muligheter for andre norske bedrifter. Vi har blant annet spekulert i bruken av talegjenkjenning for politiet, spill, språklæring eller advokattjenester. Spesifikt for de norske bedriftene er at mulighetene økes dersom modellene trenes opp på de mange dialektene vi har her i landet.

## 5.5 Implikasjoner

Vårt studie har noen implikasjoner som kan være avgjørende for kvaliteten på vår konklusjon. Det hadde blant annet vært interessant å kunne utføre et tilnærmet lik dokumentstudie av Omilons kunder i helsesektoren. Dette kunne da innebære data om spart tid for leger og sekretærer samt oversikt over forbedret ventetid og omløpshastighet. På den måten kunne vi kanskje konkludert sterkere med spekulasjonen om at kundesenteret eller undervisningssammenhengen bør bruke teknologien på samme måte som pasientjournaler.

Dokumentene vi allerede har er også påvirket av lengden på studiet. Datamaterialet som ble benyttet til dokumentstudiet inneholder ikke nyere data enn fra 2021. For bedrifter er det tidkrevende å hente ut og anonymisere datamateriale for å gi oss tilgang på det, og det ble derfor ikke tid nok for bedriften å hente ut nyere datamateriale for 2022 til oss. Det kunne vært interessant å se om bedriften har oppnådd like store ressursbesparelser i 2022, og det kunne styrket datagrunnlaget vårt ytterligere. Likevel anser vi det som svært sannsynlig at bedriften har oppnådd samme resultater, ettersom kostnadene i 2022 skal være de samme og talegjenkjenning benyttes fortsatt i alle avdelinger.

Dersom studiet hadde vært gjennomført over en lengre tidsperiode hadde det vært mulig å ha flere intervjuer med informantene våre, samt hatt intervjuer med andre relevante informanter. I de første intervjuene våre hadde vi mer generelle spørsmål, før vi konkretiserte spørsmålene våre etter hvert som vi gjorde relevante funn og tilspisset problemstillingen vår. I etterkant ser vi at det kunne vært relevant å stille flere konkrete spørsmål relatert til problemstillingen tidligere i forskningen til flere informanter, ettersom det ville styrket datagrunnlaget vårt ytterligere. Vi anser likevel funnene vi har gjort som verdifulle, og funn som går igjen mellom ulike bedrifter kan være med på å antyde at funnene kan gjelde bedrifter utover vårt utvalg.

Mange av resultatene spekulerer også i fremtiden, og slikt som litteraturen beskriver er ikke dette nødvendigvis den nærmeste fremtid basert på at det er ekstremt omfattende teknologi, noe som faktisk kanskje gjør at teknologien aldri vil bli tilstrekkelig nok. På den måten kan tidsperspektivet gjøre at mye av de beskrevne mulighetene ikke vil ha like stor effekt for de norske bedriftene som vi i utgangspunktet spekulerer i.



Det er også viktig å nevne at enkelte av våre informanter er leverandører av talegjenkjenning og selger denne teknologien til bedriftene vi omtaler i studiet vårt. Dette kan bety at noen av funnene våre er påvirket av deres utelatende eller overdrevent positive syn på teknologien. I vårt tilfelle ville det gitt større verdi dersom vi i tillegg for eksempel hadde fått en informant, en lege fra helsesektoren som selv anvender talegjenkjenning hver dag. Likevel refererte mange av informantene våre til konkrete hendelser og tall, noe som på den andre siden øker troverdigheten til tross for deres til tider inhabilitet.

## 6.0 Konklusjon

Vår utredning bidrar til den nåværende, snevre litteraturen om norsk talegjenkjenningsteknologi og hvilke effekter og muligheter det kan gi norske bedrifter. Vi startet studiet ved å motiveres av hvilke effekter teknologien faktisk gir, men ikke minst hvilke fremtidige muligheter vi kan identifisere. Vårt hovedfunn rundt effekter er at talegjenkjenning allerede kan være svært lønnsomt for norske bedrifter. Dette viser vårt dokumentstudie av kundesenteret helt konkret med en avkastning på investeringen på 60.2%. I tillegg har vi identifisert lønnsomhet og effektivitet i helsesektoren der blant annet omløpshastigheten er betraktelig forbedret. Vårt neste funn rundt effekter er de ikke-kostnadsrelaterte fordelene, der blant annet kvalitet, informasjonsflyt og ergonomiske forhold er forbedret ved bruk av talegjenkjenning.

En avgjørende grunn for hvorfor talegjenkjenning gir en såpass god effekt for noen bedrifter er at teknologien klarer å tjene dens hovedsakelige formål. Studiet vårt viser at dette avhenger blant annet av hvor godt spesifikt trente modellene for talegjenkjenning er og i hvilken grad teknologien brukes optimalt innad i bedriftene. Det er disse forholdene som kan være grunnen til at andre norske bedrifter ikke opplever like gode effekter. Vi har derfor spekulert i om bedriftene bør bruke talegjenkjenning på en annen måte, og spekulerer blant annet i å bruke teknologien i etterkant for å transkribere notater eller sammendrag, på lik linje som pasientjournaler. I tillegg nevnte vi i introduksjonen hvor stor investering en implementering av talegjenkjenning kan være for bedrifter, men vi kan konkludere med at dersom teknologien kan kjøpes per bruker trenger hverken driftskostnader eller enhetskostnader å være av betydelige summer. På den måten trenger også teknologien å kun gjøres tilgjengelig for de enhetene i bedriften som faktisk gir stor nok effekt.

En av de største mulighetene for norske bedrifter ligger i den teknologiske utviklingen. Vi ønsker derfor at vårt studie kan bidra til å belyse viktigheten av å dele data, koder og innsikt for at modellene blir trent godt nok til å kunne tjene dens formål; Om det er i de bedriftene som allerede bruker teknologien eller forhåpentligvis nye bedrifter som tar det i bruk. Mer konkret ser vi viktigheten av å trene modeller og utvikle teknologien med det norske språket og dens mangfoldige innhold av dialekter og skriftspråk. Studiet vårt viser også verdien av å kalkulere lønnsomheten og hvordan dette kan gjøres i noen av bedriftene. Alt dette er med på å forbedre teknologiens nøyaktighet og derav gi flere muligheter, sånn som mer inkludering

på arbeidsplasser og åpner også opp for andre norske bransjer. På den måten kan nåværende utfordringer omdannes til fremtidige muligheter for bruk av talegjenkjenning.

Avslutningsvis er det essensielt å påpeke at forskningsspørsmålet vårt er relativt stort og inkluderer norske bedrifter i en generell forstand. Siden utredningen er en avsluttende del av masterstudiet og har vært gjennomført over en tidsperiode på ett semester, måtte vi åpenbart begrense oss til et utvalg av norske bedrifter og snevre ned vårt forskningsspørsmål til dette. Vi har naturligvis derfor ikke informanter fra de fleste bedriftene, noe som følgelig vis begrenser datagrunnlaget vårt. Vi kan dermed ikke generalisere og konkludere med at våre funn gjelder for alle bedrifter i hele Norge. Likevel har vi under utredningen diskutert hvordan mange av våre funn kan anvendes til flere norske bedrifter, og mener også at det gir verdi for flere enn de som er nevnt. På bakgrunn av dette mener vi at videre forskning innenfor dette temaet bør strekke seg over en lenger periode enn dette studiet. På den måten vil man få mulighet til å inkludere mer tallgrunnlag fra flere bedrifter og at kvantitativ data vil implisere effektene av talegjenkjenning enda mer nøyaktig. I tillegg er flere av mulighetene vi beskriver basert på fremtidig utvikling og det er derfor interessant for fremtidig forskning å se om det vi spekulerer i faktisk er virkeligheten og gir de effektene som vi diskuterer at det kan gjøre.

## Litteraturliste

- Al.Aiad, A., Momani, A.K., Alnsour, Y., Alsharo, M. (2020). *The Impact of Speech Recognition Systems on The Productivity and The Workflow in Radiology Departments: A Systematic Review*. AIS Electronic Library
- Auforas, T., Chung, J., Senior, A., Vinyals, O. & Zisserman, A. (2018). *Deep Audio-visual Speech Recognition*. United States: IEEE
- Callaway, E.D., Sweet, C.F., Siegel E., Reiser, J.M & Beall, D.P. (2002). *Speech Recognition Interface to a Hospital Information System Using a Self-Designed Visual Basic Program: Initial Experience*. Journal of Digital Imaging
- Durling, S. & Lumsdel, J. (2008) *Speech recognition use in healthcare applications*. Association for Computing Machinery, NY.
- Engan, O. & Tørdal, R.M. (2018). *Språkets byggeklosser*. NDLA  
<https://ndla.no/nb/subject>
- Hagiwara, M. (2021). *Real-world natural language processing: practical applications with deep learning*. Manning Publications Co. LLC.
- Huang, X., Baker, J. & Reddy, R. (2014, januar.). *A Historical Perspective of Speech Recognition*.  
<https://cacm.acm.org/magazines/2014/1/170863-a-historical-perspective-of-speech-recognition/fulltext#R5>
- Huang, Y., Liu, C., Shadiev, R., Shent, M. & Huang W. (2015). *Investigating an application of speech-to-text recognition: a study on visual attention and learning behaviour: Application of speech-to-text recognition technology*. Journal of computer assisted learning
- Johannessen, A., Christoffersen, L. & Tufte, P. A. (2011). *Forskningsmetode for økonomisk-administrative fag* (3.utgave). Abstrakt Forlag.
- Johannessen, A., Tufte, P. A. & Christoffersen, L. (2016). *Introduksjon til samfunnsvitenskapelig metode*. Abstrakt Forlag.
- Juang, B.H & Rabiner, L.R. (2004). *Automatic Speech Recognition – A Brief History of the Technology Development*. Georgia Institute of Technology.
- Kidawara, Y. Sumita, E. & Kawai, H. (2019). *Speech-To-Speech Translation*. Springer Singapore Pte. Limited.

- Koivikko, M.P., Kauppinen, T. & Ahovuo, J. (2008). *Improvement of Report Workflow and Productivity Using Speech Recognition—A Follow-up Study*. Journal of Digital Imaging
- Kunnskapsbanken (2020, 6.mai). *Norsk talegjenkjenning*  
<https://www.kunnskapsbanken.net/norsk-talegjenkjenning/>
- Lincoln, Y. S. & Guba, E. G. (1985) *Naturalistic inquiry*. SAGE Publications.
- Saunders, M., Lewis, P. & Thornhill, A. (2016). *Research Methods for Business Students* (7.utgave). Pearson Education Limited.
- Marr, B. (2021, 27.september). *The 5 Biggest Technology Trends in 2022*. Forbes.  
<https://www.forbes.com/sites/bernardmarr/2021/09/27/the-5-biggest-technology-trends-in-2022/?sh=d1e891624147>
- Mustaquim, M.M. (2011). *Automatic speech recognition- an approach for designing inclusive games*. Springer US.
- Nasjonalbiblioteket. (2019, 18.september). *Talegjenkjenning på norsk*.  
<https://www.nb.no/hva-skjer/talegjenkjenning-pa-norsk-innspillseminar-om-talegjenkjenning-og-maskinlaering-pa-norsk/>
- Norris & McQueen. (2008). *Shortlist B: A Bayesian model of continuous speech recognition*.
- Nuance. (U.å). *Medical Speech recognition solutions*.  
<https://www.nuance.com/healthcare/provider-solutions/speech-recognition.html>
- SCRIBE. (2021). *Research. Novelty and ambition*.  
<https://scribe-project.github.io/research/>
- Shadiev, R., Hwang, W., Chen, N. & Huang Y. (2014). *Review of Speech-to-Text Recognition Technology for Enhancing Learning*. Department of Engineering Science
- Solberg, P.E. & Ortiz, P. (2022, 26. januar). *The Norwegian Parliamentary Speech Corpus*. National Library of Norway, Telenor Research.
- O'Shaughnessy, D. (2008). *Invited paper: Automatic speech recognition: History, methods and challenges*. Elsevier Ltd.
- SSB. (2021, 28. februar). *Her kan du sjekke lønna i ditt yrke*.  
<https://www.ssb.no/arbeid-og-lonn/lonn-og-arbeidskraftkostnader/artikler/her-kan-du-sjekke-lonna-i-ditt-yrke>
- Store Norske leksikon. (2019, 31. oktober). *Fonem*.  
<https://snl.no/fonem>
- Store Norske leksikon. (2020, 20. januar). *Kunstig intelligens*.  
[https://snl.no/kunstig\\_intelligens](https://snl.no/kunstig_intelligens)

Telenor. (2021, juni). *Telenor explores automatic voice recognition for live readings of customer sentiment in Norway.*

<https://www.telenor.com/telenor-explores-automatic-voice-recognition-for-live-readings-of-customer-sentiment-in-norway/>

Visma. (U.å) *Avkastning. Hva er avkastning?*

<https://www.visma.no/eaccounting/regnskapsordbok/a/avkastning/>

Yin, R.K. (2003). *Case Study Research: Design and Methods* (3.utgave). SAGE Publications.

Zajechowski, M. (U.å). *Automatic Speech Recognition (ASR) Software – An Introduction.*

## **Vedlegg**

### **A1 Intervjuguide**

#### **Intervjuguide: Informant 1 - Prosjektrådgiver for implementering av talegjenkjenning i Telenor**

##### **Spørsmål for introduksjon:**

1. Hva er din rolle i SCRIBE og S2T?

##### **Spørsmål om SCRIBE og S2T:**

1. Hva er forskjellen på SCRIBE og S2T?
2. Hva er/var motivasjonen deres for å sette i gang med prosjektet?
3. Hvor stort er S2T i Telenor per dags dato?
4. Hva slags forventninger har dere til SCRIBE og S2T?
5. Hvordan fungerer samarbeidet deres med bla Forskningsrådet, Nasjonalbiblioteket, NRK og Politiet?
6. Har/er det noen utfordringer ved SCRIBE og S2T som prosjekter?

##### **Spørsmål om implementering:**

7. Er S2T implementert i Norge? Er det implementert i andre avdelinger/land?
8. Er det planer om å implementere det i andre avdelinger?
9. Hvordan foregår implementeringen?
10. Er det gjort noen analyser etter en eventuelt implementering?
11. Utover produktutvikling, opplæring og kundeforståelse- er det noen andre grunner for at S2T blir implementert i avdelinger?
12. Har det vært noen utfordringer ved implementering?

##### **(Hvis implementert) Spørsmål om kommunikasjon:**

13. Har S2T forbedret kundeservice? Og på hvilken måte?
14. Har det vært noen utfordringer med S2T i kundeservice?
15. Har det vært noen reaksjoner fra ansatte eller kunder i forhold til S2T?

**Spørsmål om markedet:**

16. Speech to text finnes fra før av i forskjellige språk og former, hva skiller Telenor sin teknologi fra dette?
17. Hva slags muligheter har dere og ser dere for dere i markedet?

**Spørsmål om business:**

18. Hvordan ser dere at prosjektet er lønnsomt?
19. Ser dere for dere at prosjektet gir besparte kostnader i Telenor, for eksempel innen kundeservice?
20. Er dette teknologi Telenor selger videre og får inntekter av?

**Intervjuguide: Informant 2 - Prosjektleder for implementering av talegjenkjenning på kundesenter****Spørsmål om prosjektet generelt:**

1. Hva var motivasjonen og grunnen til at dere begynte å jobbe med talegjenkjenning?
2. Hva er bakgrunnen for at dere begynte med implementering av talegjenkjenning på kundeservice?
3. Hvor lenge har det vært implementert i kundeservice?
4. Hva har vært deres mål med prosjektet?
5. Hvordan arbeidet kundeservice før implementering?

**Spørsmål om fordeler og utfordringer ved bruk av talegjenkjenning:**

6. Hva vil du trekke frem som de viktigste fordelene ved bruk av talegjenkjenning?
7. Hva er mulighetene med talegjenkjenning?
8. Kan du nevne noen utfordringer/ulempes ved bruk av talegjenkjenning?

**Spørsmål om resultater etter implementering:**

9. Har det vært noen reaksjoner fra de ansatte?
10. Hvilke resultater har dere sett etter implementeringen?
11. Har du noen konkrete tall å vise til?
12. Hvordan er de totale kostnadene, sammenlignet med effekten dere har hatt av prosjektet?



### **Fremtiden for talegjekjenning hos selskapet:**

13. Hva vil du trekke frem som noe dere ser for dere i fremtiden med talegjekjenning hos selskapet?
14. Hvordan jobber dere med prosjektet fremover?

### **Intervjuguide: Informant 3 - Produktspesialist Omilon**

#### **Spørsmål om Omilon generelt:**

1. Vil du fortelle litt om Omilons historie?
2. Hva er motivasjonen for å satse på talegjekjenning, og har dere noen mål?
3. Hvem er deres kunder?

#### **Spørsmål om fordeler og utfordringer ved bruk av talegjekjenning:**

4. Hva er salgspitchen dere bruker som leverandør av talegjekjenning?
5. Er det noen andre fordeler dere generelt tenker rundt talegjekjenning?
6. Er det noen utfordringer dere har identifisert ved bruk av talegjekjenning?

#### **Spørsmål om implementering av talegjekjenning:**

7. Hvordan foregår implementeringen?
8. Hvordan vil du beskrive opplæringen ved bruken av talegjekjenning?
9. Utover produktutvikling, opplæring og kundeforståelse- er det noen andre grunner for at S2T blir implementert i avdelinger?
10. Har det vært noen utfordringer ved implementering?

#### **Spørsmål om resultater etter implementering:**

11. Hvordan har dere målt effekten?
12. Har dere noen konkrete tall eller resultater å vise til?
13. Har det vært noen reaksjoner fra kundene som har tatt i bruk talegjekjenning?
14. Hvordan jobber dere videre ut i fra resultater og tilbakemeldinger?

#### **Spørsmål om fremtiden for Omilon:**

15. Er det andre markeder dere ser for dere å gå inn i?
16. Er det noen ny/potensiell ny teknologi rundt talegjekjenning dere ser for dere å ta i bruk?

## **Intervjuguide: Informant 4 - Forsker innen talegjenkjenning i Telenor**

### **Spørsmål om SCRIBE generelt:**

1. Hva er din rolle i SCRIBE?
2. Hva skiller SCRIBE mot implementeringen i Telenor?
3. Hva slags mål jobber SCRIBE mot?
4. Hva er målet med samarbeidspartnerne deres i SCRIBE?

### **Spørsmål om talegjenkjenning mer teknisk**

5. Vil du fortelle litt om dataen dere bruker til å trene modellene for talegjenkjenning?
6. Har dere mangel på mer daglig språk, siden Språkbanken inneholder mye formell tale?
7. Hvordan måler man resultatene, slikt som nøyaktighet under testing av modellene?
8. For tekstanalyse i ettertid, hvordan klarer modellene å identifisere temaet for samtalene?
9. Er det noen klare tekniske utfordringer ved talegjenkjenning som dere i SCRIBE er bevisste over?
10. Er det nye markeder man ser for seg at talegjenkjenning kan brukes til?

## **Intervjuguide: Informant 5 - Foreleser ved universitet**

### **Spørsmål generelt om bruken av talegjenkjenning:**

1. Hvordan har du tatt i bruk talegjenkjenning på universitetet?
2. Hva var formålet med å ta i bruk talegjenkjenning på universitetet?
3. Har det vært etterspurt fra studenter eller andre?

### **Spørsmål om utfordringer og fordeler med talegjenkjenning:**

4. Hva vil du trekke frem som de største utfordringene ved bruk av talegjenkjenning?
5. Hva vil du trekke frem som de største fordelene ved bruk av talegjenkjenning?
6. Har du fått noen tilbakemeldinger fra studenter?
7. Har du fått noen tilbakemeldinger fra andre forelesere?
8. Hvordan vil du beskrive effekten av bruk av talegjenkjenning i undervisningssammenheng?

### **Spørsmål om fremtid:**

9. Har du interesse av å fortsette å bruke talegjenkjenning i undervisning?

### **Intervjuguide: Informant 6 - Rekrutterer for Omilon**

#### **Spørsmål om rollen som rekrutterer for Omilon:**

1. Hva er oppgaven din rundt ansettelsesprosessene?
2. Hva opplever du som minstekrav for kandidater som skal jobbe med talegjenkjenning?
3. Hva er optimale egenskaper for kandidatene?
4. Er det andre attraktive egenskaper man ser etter hos kandidatene?

#### **Spørsmål om utfordringer rundt Omilon og talegjenkjenning**

5. Har du observert noen utfordringer med talegjenkjenning som system?
6. Opplever du selv noen potensielle utfordringer?

#### **Spørsmål om fremtiden til Omilon og talegjenkjenning:**

7. Opplever du selv noen potensielle muligheter for Omilon og talegjenkjenning?
8. Er det noen andre markeder dere, eller Omilon ønsker å rekruttere mot innenfor talegjenkjenning?

### **Intervjuguide: Informant 7 - Kunderådgiver på kundesenter som har implementert talegjenkjenning**

#### **Spørsmål om rollen som kunderådgiver:**

1. Hvordan er en normal arbeidsdag for deg?
2. Hvordan måles resultater for en kunderådgiver?
3. Hvordan var arbeidsdagen før implementering av talegjenkjenning?
4. Hvor mange samtaler hadde du normalt i løpet av en dag før implementering?

#### **Spørsmål om erfaringer etter implementering av talegjenkjenning:**

5. Hva vil du trekke frem som de største endringene for din del?
6. Hva vil du beskrive som fordelene med implementeringen av talegjenkjenning på kundesenteret?
7. Har du noen ulemper med implementeringen av talegjenkjenning på kundesenteret?
8. Hvordan tror du effekten av talegjenkjenning har påvirket kundesenteret?
9. Hvor mange samtaler har du normalt i løpet av en dag etter implementering?

**Spørsmål om fremtiden til talegjenkjenning:**

10. Hva tror du om fremtiden til talegjenkjenning?
11. Har du noen formening om talegjenkjenning etter hvert vil kunne erstatte mennesket?

**Intervjuguide: Informant 8 - Ledende rolle innenfor prosesser med arbeid rundt taleteknologi i NRK****Spørsmål om NRK generelt:**

1. Hvordan jobber dere med talegjenkjenning?
2. Hva er motivasjonen for å satse på talegjenkjenning, og har dere noen mål?
3. Hvordan fungerer samarbeidet deres i SCRIBE?
4. I intervju med Telenor nevnte de at dere skulle transkribere radioarkivet deres, hva er årsaken til det?

**Spørsmål om fordeler og utfordringer ved bruk av talegjenkjenning:**

5. Hva er fordelene ved bruk av talegjenkjenning?
6. Hva er utfordringer dere har identifisert ved bruk av talegjenkjenning?
7. Hva er begrensningene til talegjenkjenning?

**Spørsmål om fremtiden for NRK:**

8. Er det andre markeder dere ser for dere å gå inn i?
9. Er det noen ny/potensiell ny teknologi rundt talegjenkjenning dere ser for dere å ta i bruk?
10. Hva er mulighetene ved bruk av talegjenkjenning?