

Brukerforsøk med multimodal demonstrator

Asbjørn Aas

Master i kommunikasjonsteknologi

Oppgaven levert: Juni 2006

Hovedveileder: Magne Hallstein Johnsen, IET

Biveileder(e): Knut Kvale, Telenor FoU

Narada Dilp Warakagoda, Telenor FoU

Oppgavetekst

Multimodalitet på mobilen betyr at brukeren kan kommunisere med mobilen på flere måter, f.eks. med tale og peking på trykkfølsom skjerm, også kalt "pek og preik"-funksjonalitet. En av fordelene med multimodalitet er at det lar folk velge den interaksjonsmåten de selv føler er mest naturlig for en gitt oppgave og kontekst.

Telenor R&D har de siste åra utviklet en multimodal demonstratortjeneste på PDA. Oppgaven går ut på å evaluere denne tjenesten for nye brukergrupper som kan ha spesielt nytte av valgfriheten multimodalitet gir. Vi er spesielt interesserte i å prøve ut demonstratoren på gruppene barn/ ungdom og eldre, men vi er åpne for andre forslag fra kandidaten.

Oppgaven gitt: 20. januar 2006

Hovedveileder: Magne Hallstein Johnsen, IET

FORORD

Arbeidet med masteroppgaven ble utført ved Telenor R&D, ved Telenor Fornebu, våren 2006. Den er organisatorisk under institutt for elektronikk og telekommunikasjon, ved Norges Teknisk-Naturvitenskapelige Universitet (NTNU).

Jeg ønsker spesielt å takke de som har fulgt meg faglig dette halvåret. Takk til mine veiledere Knut Kvale og Narada Dilp Warakagoda ved Telenor R&D, som har gitt meg god rettleiding og verdifull hjelp, både faglig og i forhold til rapportskrivningen. Takk går også til faglærer Magne Hallstein Johnsen, som har hatt overoppsyn med oppgaveskrivningen og vært kontakten min mot universitetet.

En stor takk går til alle dere som har vært med som testbrukere i oppgaven. Uten testbrukere hadde denne oppgaven vært svært tynn. Takk for tiden deres, og for tålmodighet når testene ikke har gått som de skulle.

Jeg vil også rette en takk til Hilde F. Berg i ”Infohjørnet”, som har hjulpet meg med det praktiske rundt brukertestene.

Til slutt vil jeg takke medstudent Olav Ådne Aas som tok seg tid til å korrekturlese masteroppgaven.

SAMMENDRAG

Oppgaven handler om brukersøk med en multimodal demonstrator, utviklet innenfor BRAGE-prosjektet. Demonstratoren er fysisk sett klient-server-løsning med en PDA (Personal Digital Assistant) som klient og en vanlig laptop personlig datamaskin som server. 'Det multimodale' ligger i at det kan både trykkes på den trykksensitive skjermen og tales til klienten, gjerne samtidig, for å interagere med systemet.

Noen eksempel blir tidlig i oppgaven trukket fram fra samfunnsutviklingen som viser at multimodale datasystemer sakte øker i aktualitet.

Systemets virkemåte blir gjennomgående forklart, og det er et system som legger vekt på enkel implementasjon.

30 studenter fra et teknisk miljø er testbrukere, og disse utfører en test hvor det forsøkes å kaste lys over bruken av multimodalitet og de andre interaksjonsformene. En analyse blir gjort under hypotesen om at brukere alltid vil forsøke å bruke så lite tid og så lite kognitiv anstrengelse som mulig for å løse oppgaven.

Noen av hovedresultatene er at tale blir mye benyttet – 48,9% av all interaksjon var tale. Penn står for 39,8%, mens simultan koordinert multimodalitet står for 11,3%. Tid ser ut til å være viktig i valget for modalitet.

I slutten av oppgaven blir det foreslått en endring av systemets taledetektor, ettersom systemet underveis i testene plukket opp svært mye bakgrunnsstøy og tolket dette som inndata. Forslaget går ut på å gjøre taledeteksjon ved hjelp av konfidenser i systemets dialogmanager.

INNHOLDSFORTEGNELSE

1	<u>INNLEDNING</u>	1
1.1	VIDT PERSPEKTIV: HVORDAN SER LANDSKAPET UT RUNDT APPLIKASJONEN?	3
1.2	LIKNENDE LØSNINGER.....	5
2	<u>TEORI</u>	6
2.1	FORKLARING AV SENTRALE BEGREP	6
2.2	HYPOTESE.....	8
2.3	EVALUERINGSMETODE	8
3	<u>BESKRIVELSE AV SYSTEMET</u>	11
3.1	SYSTEMET MODUL FOR MODUL	11
3.2	HVORDAN FUSJON ER IMPLEMENTERT	19
3.3	ULIKE DIALOGSTRUKTURER	20
4	<u>GJENNOMFØRING AV TESTEN</u>	21
4.1	VALG AV BRUKERGRUPPER	21
4.2	TESTBRUKERNE	23
4.3	TEORETISK BAKGRUNN FOR TESTEN	23
4.4	INNHold I EN BRUKERSESJON	24
5	<u>RESULTATER</u>	29
5.1	OPPTELLING AV MODALITETER.....	29
5.2	SEKVENSIELT MULTIMODALE DIALOGER.....	31
5.3	REVIDERT OPPTELLING AV MODALITETER	34
5.4	NÅR TRYKES DET I EN MULTIMODAL YTRING?.....	37
5.5	ASR- FEILRATE.....	41
5.6	RESULTATER FRA SPØRRESKJEMA.....	43
6	<u>DISKUSJON</u>	49
6.1	"FØRSTE LINJES" OBSERVASJONER AV RESULTATENE.....	49
6.2	VIDERE DISKUSJON AV UTVALGTE TEMA	52
6.3	ASR-RATE DISKUSJON	54
6.4	ASPEKTER VED Å SPESIALISERE ELLER GENERALISERE SYSTEMET.....	56
6.5	GENERELLE FORBEHOLD.....	57
6.6	SAMMENLIGNING AV RESULTATER MOT HYPOTESE.....	59

7	<u>MULIGE ENDRINGER.....</u>	<u>61</u>
7.1	KRITERIET FOR NÅR AUDIOSIGNALET ANTAS Å INNEHOLDE TALE.....	61
7.2	ANDRE MINDRE ENDRINGER.....	65
8	<u>KONKLUSJON.....</u>	<u>66</u>
9	<u>REFERANSELISTE.....</u>	<u>68</u>
10	<u>VEDLEGG.....</u>	<u>70</u>
10.1	UTFYLLENDE RESULTATTABELLER.....	70
10.2	SPØRRESKJEMAETS OPPGAVE 1.....	83
10.3	ANNEN INFORMASJON OM BRUKERNE.....	86
10.4	BRUKERTEST: CASER.DOC.....	88
10.5	ANNEN INFO VIST TIL BRUKEREN VED TEST.....	89
10.6	BRUKERTEST: SPØRRESKJEMA FOR INTERVJUET ETTERPÅ.....	90
10.7	TESTLEDERENS NOTERINGSSKJEMA I BRUKERTESTEN.....	92
10.8	TESTLEDERENS HUSKELAPP FOR INTRODUKSJONEN.....	93
10.9	GRAMATIKK.....	94

FIGURLISTE

figur 1 viser kumulativ informasjon om brukergruppen per nye testede bruker i henhold til Nielsen (1994?) ved riktig bruk av testmetoder.....	10
figur 2: Skisse av systemarkitekturen.....	11
figur 3: Hvordan informasjonsflyten er i systemet, sett uten hub.	12
figur 4: Fragment av FSM i Dialogmanager med hjelpetilstander for håndtering av multimodale ytringer.	19

TABELLISTE

Tabell 1: Konsepter, konseptverdier og konseptkonfidenserKonsepter, konseptverdier og konseptkonfidenser.....	16
Tabell 2: Oversikt over mulig resultat etter gjenkjenning.....	17
Tabell 3: Resultat etter fullført fusjon i dialogmanager.....	18
Tabell 4: Oversikt over brukere og gjennomførte oppgaver. Et kryss betyr fullført oppgave.....	30
Tabell 5: Andel turns av ulike typer, telt ved hjelp av script.....	31
tabell 6: Dialoger som inneholder sekvensiell eller simultan koordinert multimodalitet. Sortert på bruker.....	33
tabell 7: Dialoger som inneholder inneholder sekvensiell eller simultan koordinert multimodalitet, sortert på oppgave.....	33
tabell 8: Revidert andel inndaturns, sortert per bruker og modalitet.....	36
tabell 9: oppsummering, revidert andel turns, sortert på oppgave og modalitet.....	37
tabell 10: ASR-suksessrate.....	43
tabell 11: Forsøk på å gi inn all semantisk relevant informasjon i løpet av den multimodale ytringen.....	51

1 INNLEDNING

Demonstratoren som forsøkene gjøres på i denne oppgaven er utviklet i det norske BRAGE-prosjektet, som er underlagt forskningsprogrammet KUNSTI¹. BRAGEs mål er å forbedre talesentriske dialogsystemer. BRAGE er et samarbeidsprosjekt mellom Sintef IKT, Telenor R&D og Norges Teknisk-Naturvitenskapelige Universitet (NTNU). Det ble startet opp høsten 2002 og er planlagt avsluttet ved utgangen av 2006.

På tross av at datamaskiner har overtatt en rekke funksjoner i de siste to årtiene, har grensesnittet mellom bruker og system for det meste vært uforandret – stort sett begrenset til tastatur og mus, eller til spesielt utviklede knapper ved spesielle installasjoner, for eksempel på en minibank. I denne perioden har det vært en stor utvikling i økt prosessorkraft og lagringsplass, og gjort plass for teknologi som talegjenkjenning, talesyntese, dialoghåndtering m.m, slik at det har vært mulig å se på alternative interaksjonsformer mot datasystemer. Multimodale systemer (se definisjon i kapittel 2.1) har i denne perioden fått økt fokus, ettersom de grunnleggende teknologiene har kommet på plass. Fordelen med multimodale systemer er at interaksjonen i større grad enn ved et monomodalt system kan skje på brukerens preferanser. Dette åpner for muligheten for en mer effektiv omgang med systemet for vanlige brukere, men det kan også åpne for et bredere spekter av brukere, for eksempel blinde. Modaliteter som kan tas i bruk av multimodale systemer er for eksempel trykking med penn, tale, håndskriftgjenkjenning, tastaturtrykking og ulike former for kroppsbevegelser. Multimodale systemer som er i kommersiell eller forskningsmessig bruk i dag, benytter seg av blant annet tale, trykking, håndbevegelser, hode- og kroppsbevegelser².

Systemet som denne oppgaven er basert på, inneholder en liten mobil klient som brukeren kommuniserer med.

Rapporten er strukturert som følger. Kapittel 1 er et innledende kapittel. Kapittel 2 inneholder relevant teori, med fokus på systemets virkemåte. Kapittel 4 beskriver utføring av testen, mens kapittel 5 og 6 inneholder resultater og diskusjon av disse. Kapittel 7 foreslår mulige endringer, før konklusjoner kommer i kapittel 8.

Her følger en paragraf om når multimodalitet brukes, eksemplifisert ved et bimodalt system med tale og trykking. Anta at en bruker ønsker informasjon fra et multimodalt system, og at forespørselen er kompleks. Om brukeren ønsker å bruke talemodaliteten til å utføre denne forespørselen, kan dette føre til en lang ytring. Da kan det bli vanskelig for et system å trekke ut meningen av denne ytringen. Om brukeren ønsker å benytte en pek-og-klikk-tilnærming, kan dialogen blir lang og omstendelig. I et slikt tilfelle kan det være fordelaktig med bruk av multimodal interaksjon. Dette synet deles av brosjyren ”MUST –

¹ Hjemmeside for KUNSTI: <http://www.forskningsradet.no/kunsti>

² Se for eksempel Center for Human-Computer Communication. <http://www.cse.ogi.edu/CHCC>

multimodal and multilingual services for small mobile terminals” (Boves og den Os, 2002, s.10). Der står det følgende:

”Laboratorieeksperimenter med tjenester som ikke ’naturlig’ kan presenteres som en skjema utfyllingsoppgave, peker på at noen applikasjoner vil trenge hele kompleksiteten til simultan koordinert interaksjon” (sitatet er oversatt av forfatteren).

Videre gis et eksempel på applikasjonen i tjenesten som brosjyren omhandler; turistinformasjon i Paris: ”Når en bruker peker på tårnet på skjermen (...) og på samme tid sier ’hva er dette?’”. Dette spørsmålet indikerer at systemet inneholder usynlig informasjon om synlige objekter på skjermen.

Brukeren vil i følge litteraturen vurdere å bruke simultan koordinert multimodalitet om applikasjonen er slik at bruken av kun tale blir vanskelig på grunn av manglende kunnskap av navn på steder som er av interesse, men som er synlige og lett trykkbare på skjermen. Det er en motivasjon for å bruke simultan koordinert multimodalitet om bruken av kun penn eller kun tale fører til tidkrevende og omstendelig lang dialog med et høyt antall dialogturns, og dette kan reduseres ved bruk av multimodalitet.

Multimodalitet blir sett på som fordelaktig å bruke på terminaler med små skjermer fordi de er komplementære interaksjonsformer; det som er den enes svakhet er den andres styrke. Å kunne snakke til terminalen er gunstig for en terminal med skjerm med lite areal, mens det er gunstig å kunne få svar i form av tekst, som for et menneske er mye lettere å prosessere og huske enn den sekvensielle formen som tilbakemelding i audibel form. Forespørsler om for eksempel bussruter kan gjøres enklere ved bruk av multimodalitet ved at avreise/ankomststed kan angis med penn, uten å måtte vite navnet på bussholdeplassen. På samme måte kan avreise/ankomststed angis med tale uten å måtte vite beliggenheten.

Teknisk sett har det ulike fordeler å implementere et multimodalt system på små, mobile klienter. Å både kunne prate og trykke for å gi inndata er gunstig når klientens fysiske størrelse er liten. Tale lar en få uttrykke komplekse spøringer på en praktisk og tidsbesparende måte, og overfører mye av jobben med å tilpasse seg, fra bruker til systemet. Samtidig kan det trykkes når dette føles mer naturlig, eller hvis tale til systemet feiler.

I et mer samfunnsorientert perspektiv kan det også være fordeler med multimodale system. Ulike brukergrupper som ikke kan benytte seg av tjenester som krever lese- og skriveferdigheter, kan tenkes å kunne få tilgang til et en del av disse ved hjelp av bruk av multimodale system. Det multimodale systemet kan være i dette tilfellet noe lignende som det som blir presentert i denne oppgaven. Å lage apparater og tjenester slik at de passer en så stor mengde brukere som mulig, har de siste årene fått økt fokus, og er blant annet diskutert i artikkelen ”Design for all” (Mellors, 2004). Her presenteres paradigmat ”design for all”, med en argumentasjon om at et tiltagende elektronisk samfunn fortsatt

må være aksessibelt for alle dets borgere, uavhengig av funksjonsnivå. Multimodale systemer kan være et redskap for å muliggjøre dette.

Systemet som er brukt i denne oppgaven er en klient-server-løsning, hvor klienten er en håndholdt mobil terminal. Det er implementert støtte for å gi inndata ved trykking på en trykksensitiv skjerm, samt tale. Utdata presenteres som syntetisert tale, samt grafikk på skjermen. To modaliteter er altså brukt; det er mao. et bimodalt system. Systemet er bygd opp av moduler, og det er i oppbyggingen lagt vekt på å skille applikasjonen fra den underliggende multimodale plattformen. Plattformen ble utviklet i det tidligere prosjektet ”MUST”³, gjennomført av EURESCOM⁴, hvor Telenor R&D deltok.

1.1 VIDT PERSPEKTIV: HVORDAN SER LANDSKAPET UT RUNDT APPLIKASJONEN?

Dette delkapitlet viser noen samfunnsmessige aspekter rundt multimodalitet.

1.1.1 FNs “disability convention” i emning

FNs generalforsamling etablerte i resolusjon 56/168 av 19. desember 2001 en komité i den hensikt å

”to consider proposals for a comprehensive and integral international convention to promote and protect the rights and dignity of persons with disabilities, based on the holistic approach in the work done in the fields of social development, human rights and non-discrimination and taking into account the recommendations of the Commission on Human Rights and the Commission for Social Development” (United Nations Enable 2006).

Per i dag står saken slik at det skal holdes en åttende samling, som avholdes mellom 14. og 25. august 2006, som har som mål å gjøre ferdig diskusjoner rundt en revidert versjon av utkastet. Konvensjonen skal blant annet sikre funksjonshemmedes integrasjon og deltagelse i et mer og mer teknologisk samfunn. Artikkel 9, som handler om tilgjengelighet, sier blant annet:

“1. To enable persons with disabilities to live independently and participate fully in all aspects of life, States Parties shall take appropriate measures to ensure to persons with disabilities access, on an equal basis with others, (...) to transportation, to information and communications, including information and communications technologies and systems, (...) These measures, (...) shall apply to, inter alia: (...)

³ MUST – Multimodal and Multilingual Services for Small Mobile Terminals. Se referanseliste for adresse til hjemmeside

⁴ European Institute for Research and Strategic Studies in Telecommunications. <http://www.eurescom.de/>

(b) Information, communications and other services, including electronic services and emergency services. (...)”

En FN-konvensjon om funksjonshemmedes rettigheter kan bidra til økt tilrettelegging av forskjellige IKT-tjenester. Multimodale tjenester kan tenkes å spille en sentral rolle i dette.

1.1.2 Americans with Disabilities Act

Det ble i USA i 1990 vedtatt en lov mot diskriminasjon av funksjonshemmede personer i transportsektoren (National Academy Press 2001). Loven ble laget for å motvirke diskriminasjon av personer med funksjonsnedsettelse bl.a. innenfor offentlig transport. Loven, som er kalt ADA (eng. Americans with Disabilities Act), sier blant annet:

”It shall be considered discrimination (...) for a public entity which operates a fixed route system to purchase or lease a new bus, a new rapid rail vehicle, a new light rail vehicle, or any other new vehicle to be used on such system, if such bus, rail vehicle or other vehicle is not readily accessible to and usable by individuals with disabilities, including individuals who use wheelchairs”.

Det er i denne loven nedfelt at hensyn skal tas til reisende med ulike hørsel/syn/talebegrensninger, og finne måter å kommunisere med disse på på deres premisser så langt dette er mulig. Verktøy som er vurdert her inkluderer tekst-til-tale- og talegjenkjenningssystemer, trykkbare kart/skilt, og et bredt spekter andre visuelle, akustiske og multimodale hjelpemidler.

1.1.3 Multimodalitet forskningsområde for EU

EU utarbeider hvert 4. år et såkalt RTD (eng. Research, Technical development) rammeprogram som definerer forsknings- og utviklingsområder som er kvalifiserte for økonomisk støtte. I det 6. rammeprogrammet, som er aktivt i perioden 2002 til 2006, er det definert 7 hovedområder, såkalte prioriteter, innenfor rammeprogrammet. Under prioriteten IST (eng. Information Society Technologies) er ”knowledge and interface technologies” oppført, hvor multimodalitet et av forskningsområdene for perioden 2005 til 2006 (IST hjemmeside 2006).

1.1.4 Et leserinnlegg

På Landsforbundet for utviklingshemmede og pårørendes hjemmesider⁵ finnes et leserinnlegg som uttrykker frustrasjon over bl.a. IKT-tjenester i dagens samfunn. Forfatteren beskriver og kommenterer på en subjektiv måte at overgangen fra manuell til automatisk behandling av en rekke tjenester i samfunnet skaper et større gap mellom brukergruppen ”funksjonsfriske” og personer som av ulike årsaker ikke tilhører denne gruppen. Forfatteren ser ut til å ha nær kontakt med eller stiller selv i denne gruppen. Innlegget er aktuelt i diskusjonen rundt multimodalitet, ettersom forfatteren av innlegget setter ord på noe av utfordringen man håper at multimodale grensesnitt kan bidra til å løse.

⁵ <http://portal.landsforbundet-lupe.no/modules.php?name=News&file=article&sid=50>

” Ser vi langt nok bakover i tid, gjerne femti år, finn vi at samfunnet vårt har endra seg mykje, og aller mest når det gjeld teknologi, kommunikasjon og økonomi. (...)

Så innfløkt, utilgjengeleg og uoversiktleg den kvardagsleg nødvendige informasjonen er for dei fleste av oss, har dei avanserte systema gjort langt fleire menneske funksjonshindra og vanhjelpte enn vi var under enklare vilkår. (...)

Korleis går det midt i dette med læringsmålsettinga for menneske med psykisk utviklings; sjølstendigjering"?

Realismen i målsettinga minkar med utviklinga av komplisert teknologi, system og apparat, med aukande avstand til tenester og samfunnsinstitusjonar, og med fjerning av levande menneske som kan gje individuell hjelp.”

Utviklingen mot en hverdag med stadig flere automatiske systemer vil etter all sannsynlighet ikke stoppe. Utfordringen blir å utvikle og standardisere verktøy som kan gjøre kommunikasjon med systemene lettere enn før.

1.2 LIKNENDE LØSNINGER

1.2.1 Trafikanten på WAP

Applikasjonen som er installert på den multimodale plattformen som er utprøvd i denne oppgaven, er også aksessibel via WAP⁶. For funksjonsfriske mennesker er dette en god tjeneste, men er ikke en selvforklarende tjeneste. Den krever noe teknisk innsikt for bruk.

1.2.2 ScanSoft: Nuance X|mod

Firmaet Nuance er en stor aktør innen talegjenkjenningsapplikasjoner og talesyntese internasjonalt. De leverer løsningen X|Mod⁷, som er en nettverksbasert multimodal plattform. Systemet er en tynn klient-serverløsning, og muliggjør tale- og taste-innsignal samt audio og grafikk ut. Simultan tale og tasting er mulig. Systemet støtter et bredt spekter av fysiske kanaler; GPRS, CDMA og EDGE, og kan implementeres på følgende plattformer: Symbian, Nokia Series 60, UIQ, Microsoft Pocket PC, Smartphone, J2ME, BREW, Windows CE/NT/2000/XP. Gjennom N-best visning av ASR-resultater på skjermen løses svakheter mtp. talegjenkjenning. I hvor stor grad X|mod er i kommersiell bruk pr. i dag er uvisst. På produktets hjemmesider er det en oversikt over 4 demo-applikasjoner for plattformen, som inkluderer multimodalt interface for mobil (sende sms, surfe på nettet), mobil navigasjon, støtte for gule sider og støtte for nedlasting av ringetoner.

⁶ Se trafikanten.no for mer informasjon.

⁷ <http://www.nuance.com/xmode/>

1.2.3 EMMA

W3Cs Multimedia Interaction Working Group har som målsetting å utvikle spesifikasjoner for å muliggjøre aksess til web gjennom multimodal interaksjon. I dette arbeidet jobbes det for å standardisere et XML-basert språk for å kunne håndtere inndata f.eks fra penn og tastatur, tale og gester. Dette språket er kalt Extensible MultiModal Annotation markup language (EMMA). Spesifikasjonen beskriver merking for å representere ulike former for brukerinteraksjon, konfidenstall, timestamps m.m. EMMA er en del av forslaget for "the W3C Multimodal Interaction Framework"⁸. Utviklingen av EMMA er av W3C klassifisert som "work in progress" per dags dato.

1.2.4 Kirusa

Firmaet Kirusa utvikler mobile data- og taleløsninger. Deres produkter er Kirusa Voice SMS (KV.SMS), Multimodal Solution for Circuit Voice (KMS-CV) og Kirusa Multimodal Solution for IP voice (KMS-IPV). KV.SMS er et talegrensesnitt for SMS. KMS-CV og KMS-IPV er multimodale løsninger for mobile terminaler, som muliggjør simultan og sekvensiell multimodalitet jamfør klientens egenskaper, jamfør firmaets hjemmesider.⁹

1.2.5 Bozooka

Det er kjennskap om at det finnes ytterligere en multimodal tjeneste med navn Bozooka. Det har imidlertid vært vanskelig å frambringe informasjon om denne, og den blir dermed stående ubeskrevet her.

2 TEORI

2.1 FORKLARING AV SENTRALE BEGREP

Definisjonene som omhandler multimodalitet eller multimodale system på noen måte, ble tatt fra W3Cs online hjemmesider¹⁰.

2.1.1 Sekvensiell multimodalitet

En sekvensiell multimodal applikasjon er en applikasjon hvor brukeren kan interagere med kun en modalitet om gangen, svitsjende mellom modalitetene etter ønske.

2.1.2 Simultan, koordinert multimodalitet

Simultan, koordinert multimodal inndata er inndata motatt på multiple modaliteter på samme tid og behandlet som en integrert spørring ved nedstrøms prosesser. Simultan

⁸ <http://www.w3.org/TR/mmi-framework/>

⁹ <http://www.kirusa.com/>

¹⁰ <http://www.w3.org/TR/2003/NOTE-mmi-reqs-20030108/#Appendixb>

koordinert multimodalitet er det norske uttrykket som blir brukt i oppgaven for det engelske uttrykket ”composite input”.

2.1.3 Simultan multimodal inndata

Simultan inndata betegner inndata som kan komme fra forskjellige modaliteter men som ikke blir behandlet koordinert. Simultane multimodale inndata impliserer at inndata fra forskjellige modaliteter blir prosessert i den tilfeldige rekkefølgen de ble mottatt i, i stedet for å bli kombinert før prosessering.

2.1.4 GUI

GUI står for Graphical User Interface og betegner den delen av systemet som kan ses og kommuniseres med grafisk.

2.1.5 Kognitiv anstrengelse

Det blir i dette dokumentet snakket en del om ”den kognitive anstrengelsen”. Dette er inneholdt når det prates om ”brukerens ressursbruk”, og det er dette det er tale om når det snakkes om ”den kognitive kraft” eller ”den kognitive last”. Det brukes ulike ord på dette fenomenet ettersom det er vanskelig å definere det presist, og ettersom det ligger utenfor forfatterens fagfelt. Fenomenet har å gjøre med i hvor stor grad man må fokusere, konsentrere seg for å kunne gjennomføre en oppgave. ”Den kognitive anstrengelsen” kan bli lettere å forstå ved å tenke på at det ofte er vanskeligere å konsentrere seg om en krevende oppgave i bråkete omgivelser sammenlignet med rolige omgivelser. I rapporten blir det også brukt andre uttrykk for dette samme, som alle inneholder ordet ”kognitiv”.

2.1.6 Sesjon

I denne oppgaven er uttrykket sesjon brukt om hele tiden fra en person kommer, gjør testen, til personen går igjen.

2.1.7 W3C

”The World Wide Web Consortium” (W3C) er et internasjonalt konsortium for utvikling av standarder for internett.

2.1.8 Ytring

I rapporten er uttrykket ”ytring” brukt om en interaksjon fra brukeren som kan inkludere en eller flere modaliteter. Uttrykket er definert for å kunne håndtere tilfeller der brukeren uttaler seg på en multimodal måte, men hvor det ikke blir lagret slik i systemet.

2.1.9 Modalitet

Typen kommunikasjonskanal som blir brukt for interaksjon. Ordet dekker også måten en idé er uttrykt eller forstått, eller måten en handling blir utført på.

2.1.10 Multimodalt system

Et multimodalt system støtter kommunikasjon med brukeren gjennom forskjellige modaliteter så som tale, trykking og ulike kroppsbevegelser.

2.1.11 Turn

Sett av inndata samlet fra brukeren før oppdatering av tilbakemelding fra systemet.

2.2 HYPOTESE

Ved forespørsel til et datasystem vil (1) så kort tid og (2) så lite kognitiv anstrengelse som mulig brukes for å skaffe informasjon.

Hypotesen er relevant for multimodalitet om man ser den i lys av slutningen i artikkelen av Rugelbak og Hamnes (2003) om at multimodalitet krever mer kognitiv anstrengelse enn kun tale eller kun peking alene.

2.3 EVALUERINGSMETODE

2.3.1 Bakgrunn for valget

Ulike evalueringsmetoder ble vurdert. Her følger kort hvilke.

I utvalget av evalueringsmetoder finnes en del metoder hvor såkalte ”eksperter” er brukeren, og hvor resultatet av testen spesielt fokuserer på avdekking av styrker og svakheter med systemets virkemåte. Spesielt er disse metodene egnet på en tidlig fase av produktet. Eksempler er ”walkthrough”-metoder og forskjellige heuristiske metoder (begge: Nielsen og Mack 1994). En type metode som kan brukes i et enda tidligere stadium i produktets utvikling er en Wizard Of Oz-gjennomgang. I dette tilfellet har man ikke et fungerende produkt, men en person simulerer systemet og gir tilbakemelding til brukeren jamfør en systemspesifikasjon. I senere stadier av utviklingsperioden finner man metoder som tester systemet på ”tilfeldige” brukere. Det kan gjøres for eksempel gjennom laboratorietesting eller såkalt ”rett hjem”-testing. Begge typer kan gi resultater fra en ønsket målgruppe. En labtest tar imidlertid systemet bort fra sitt naturlige bruksmiljø, noe som vil påvirke testresultatene, men tar mindre tid og ressurser enn ”rett hjem”-testing. ”Rett hjem”-testing innebærer å utruste de aktuelle brukerne med det aktuelle testsystemet og i størst mulig grad søke å samle data om normal bruk i normal omgivelse. For en god gjennomgang av relevante evalueringsmetoder henvises ellers til Kristiansens masteroppgave (2004).

I valget av evalueringsmetode ble det lagt vekt på at demonstratoren var vurdert med ulike tester før. For eksempel finnes en ”walkthrough”-gjennomgang i Rugelbaks og Hamnes’ artikkel fra 2003, samt en ”test med ekstern ekspert” i Kristiansens masteroppgave fra 2004. Ny test med lik eller lignende metode ble antatt å ikke komme til å inneholde mye ny informasjon, selv om systemet var noe endret. Disse ble dermed ikke vurdert som et godt valg. Den optimale produkttesten i ”rett-hjem-testing” var

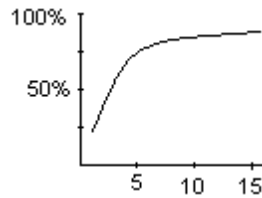
uoppnåelig på grunn av at demonstratorens klient-server-kobling var basert på WLAN/IEEE 802.11b, og at serveren benyttet seg av kabel for tilkobling til internett. I tillegg var systemet ved testtidspunkt fortsatt for ustabil for å kunne overlates uerfarne brukere. ”rett-hjem-testing” var derfor uaktuelt.

I ulike artikler om demonstratoren var det mot slutten skrevet om forslag for videre arbeid, og flere av disse pekte på brukertester (Kvale og Warakagoda, 2003), (Rugelbak og Hamnes 2003). Områder som var ønsket å få utredet var bl.a. brukertilfredsstillelse, dialogsuksessrate og dialoggjennomkjøringstid, kognitiv last, tidspunkt for trykking i en multimodal ytring, terskel for bruk av multimodale interaksjonsmetoder.

Jamfør argumentene framført i paragrafene over, ble labtest av brukere valgt. Det ble valgt å innhente informasjon om demonstratorens egenskaper gjennom en rekke tester, og det ble forsøkt å vinkle testen slik at resultatene kunne være relevante for noen av spørsmålene stilt i noen av de tidligere artiklene om demonstratoren.

2.3.2 Er det nok brukere til at dataene er til å stole på?

Resultater gode nok til å fungere som grunnlag for statistisk gyldige slutninger har behov for et større antall brukere enn det som var mulig innenfor rammene for denne oppgaven. Derfor blir ikke dataene behandlet med statistiske tester for å bekrefte eller forkaste hypotesen. Det vil bli pekt på tendenser og trender. Derimot foreligger det teori på at testinformasjon fra lite utvalg personer likevel kan gi et gyldig datagrunnlag for vurderinger om gruppen som testpersonene er tatt fra. Det hevdes at det er et spørsmål om hvilke testmetoder man bruker. Disse må være satt opp slik at man får fram mye informasjon per bruker. I dette forsøket er det brukt en laborietest og en spørreundersøkelse i intervjuform. Dette er ikke veldig ulikt kriteriene Nielsen (1994) har for at følgende forhold skal være gyldig:



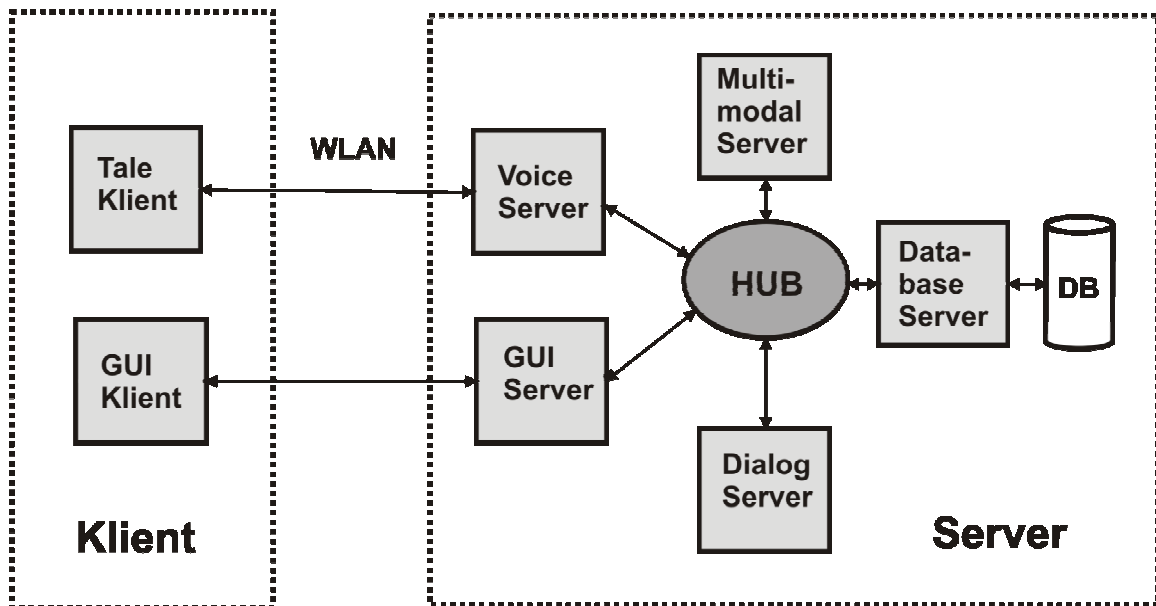
figur 1 viser kumulativ informasjon om brukergruppen per nye testede bruker i henhold til Nielsen (1994) ved riktig bruk av testmetoder.

Dette blir brukt til å forsvare antall brukere bak tallmaterialet for diskusjonen i denne oppgaven. Det blir ikke samlet så mye informasjon fra hver bruker som det potensielt kunne vært, derfor vil det reelle stigningstallet for denne oppgaven være lavere sammenlignet med stigningstallet før knekkpunktet i figur 1. Også knekkpunktet, om man følger Nielsens teori, vil komme etter et høyere antall personer. Akkurat hvor disse ligger er en unøyaktig vitenskap. Oppgaven inneholder en god del metadata om de ulike brukerne som kan finnes i kapittel 5.6, samt i vedlegg, kapittel 10.2 og 10.3. Antall brukere i testen er rundt 5 ganger større enn hvor knekkpunktet ligger i Nielsens graf. Sammen med den innsamlede tilleggsinformasjon om brukerne danner dette argumentet for hvorfor resultatene fra testene med en viss sikkerhet representerer resultatet fra den aktuelle brukergruppen.

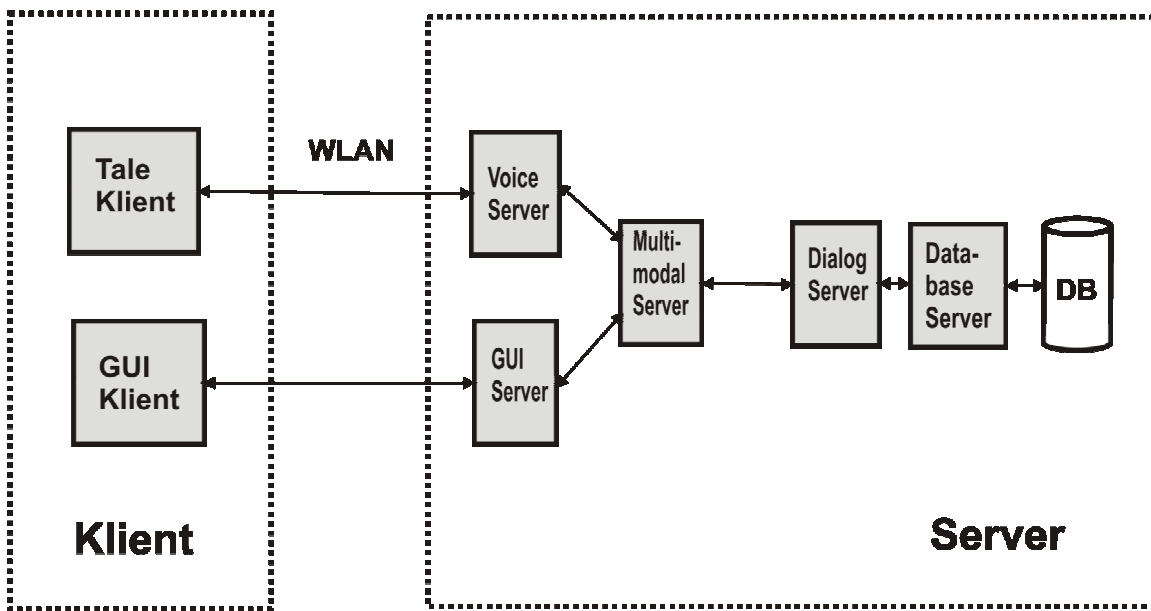
3 BESKRIVELSE AV SYSTEMET

3.1 SYSTEMET MODUL FOR MODUL

Dette kapitlet er basert på artiklene ”Speech Centric Multimodal Interfaces for Mobile Communication Systems” av Kvale et al (2003) og ”Implementation of Simultaneous Coordinated Multimodality for Mobile Terminals”, Warakagoda et al (2003).



figur 2: Skisse av systemarkitekturen



figur 3: Hvordan informasjonsflyten er i systemet, sett uten hub.

Systemet er strengt tatt *bimodalt*. Klienten har, som det kan ses på figuren over, en grafikkdel og en taledel. Systemet støtter tale og pennetrykking inn, og kan gi syntetisert tale og grafikk ut.

3.1.1 Dialogserver

Dialogserveren er hjernen i systemet. Den inneholder den sentrale prosesseringsenheten og all desisjonslogikken. Blant dens oppgaver er å prosessere brukerdataba, utføre "late fusion", holde tilstandskontroll/holde rede på konteksten, avgjøre dialogflyt og feilhåndtering. Dette oppnås gjennom at dialogserveren er implementert som fire klasser med dedikerte oppgaver: en kontekst-kontroller (klassen `dialogServer`), en bruker-inndata-prosessor (klassen `UserActObjectFactory`), en systemrespons-generator (klassen `SystemActObjectFactory`) og en XML-prosessor. Klassen `dialogServer` er hovedmodulen. Den inneholder en endelig tilstandsmaskin (eng. Finite State Machine - FSM) hvor tilstandene tilsvarer dialogtilstander. En endring til en ny tilstand skjer på bakgrunn av inneværende tilstand S_t og inndata I_t , og utløser initialisering aksjoner som databas spørringer, og hvilken systemrespons O_t som skal gis brukeren. Denne klassen initialiserer bruker-inndata-klassen og har som oppgave å finne ut hvilke av konseptene i hver turn som skal godtas som inndata til FSM. Den gjør dette ved å for hver turn fylle en tabell med to kolonner, hvor første kolonne består av konsepter og andre kolonne med konseptverdier. For hvert konsept blir det sammenlignet med hva som er i tabellen fra før, hvor kriteriet er konfidens. Når alle konsepter i ankommet turn er gjennomgått, inneholder tabellen "vinner-konseptene", som blir stående som det som ytringen inneholdt. I denne prosessen løses konflikter mellom konseptverdier tilhørende like konsepter som inndataen måtte inneholde. Integrasjon mellom penn og tale fullendes med denne prosessen, og dialogmanageren er dermed modulen som utfører "late fusion". Systemrespons-generatoren genererer som navnet røper – systemresponsen O_t . Dette gjør den på bakgrunn av I_t og S_t . O_t består av to deler: en del som sendes for konvertering til

tale, og en del som skal konverteres til grafikk. Det gjennomføres altså "early fission" her. Den grafiske delen av O_t kan inneholde tre typer elementer; forhåndslagrede tekststrenger, parametre fra I_t , og resultat fra databasespørring. XML-prosessoren er involvert i omdanningen av den grafiske delen av O_t til XML, ettersom den skal sendes videre til GUI-server som en XML-streng.

3.1.1.1 FSM i dialogserver

I dialogserveren er det implementert en endelig tilstandsmaskin. Denne har en rekke tilstander, som er organisert i to nivå: hovedtilstander og subtilstander. Hovedtilstandene er:

- Setup
- Start
- Poi. (tilstand hvor et av subkartene vises.)
- Fac
- TrafChoice
- Confirm
- Bekreft

Subtilstandene er:

- F. (Knappen "From" er trykket.)
- T
- 0-missing. (Alle konsepter påkrevd for DB-spørring er fylt.)
- 1_missing
- 2_missing
- 1_missing_pen
- 2_missing_pen_F_or_T
- Regular. (Ingen tillegg)

Gjennom denne ordningen blir tilstandene organisert på et mer oversiktlig vis. En tilstand i FSM er en kombinasjon av en hovedtilstand og en subtilstand, som for eksempel "poi-1_missing" eller "Start_regular". Det komplette tilstandsdiagrammet til FSM er komplisert, ettersom mange av subtilstandene former tilstander med mange av hovedtilstandene, og ettersom mange av tilstandene har mange grener. "Start_regular" er en spesiell tilstand, fordi dette er utgangsposisjonen som nås ved oppstart, og kan nås fra alle andre tilstander ved hjelp av knappen "home".

3.1.2 Multimodal server

Multimodal servers oppgave er å utføre "tidlig fusion", det vil si å binde inndata som ankommer innenfor et visst tidsvindu, sammen i en ytring. Fusjonen som utføres i dette systemet kan maksimalt fusjonere to pennetrykk og et tale-innsignal. Fusjonen gjennomføres ved at det ved ankomst av pennetrykk fra GUI-server eller Tale-server startes en timer, og hvor så inndata som ankommer før denne har løpt ferdig, blir inkludert i en fusjonert turn. Om til sammen to pennetrykk og et talesignal er mottatt før

timeren har løpt ut, avbrytes den umiddelbart. Det presiseres at tale markeres som ankommet når den er ASR-prosessert, pakket i galaxy-kompatibelt meldingsformat og denne meldingen er ankommet multimodal server. Det er to forskjellige timere som startes avhengig av hvilken modalitet som ankommer først. Om melding med et pennetrykk mottas, er timeren i testkonfigurasjonen satt til 3 sekunder. Om en tale-input mottas er timeren 0,85 sekunder. En ulempe med tidlig fusjon implementert på denne måten er at også knapper som "home" og "help" aktiverer disse timerne.



Figur: klienten i bruk.

3.1.3 Klienten

Klienten er implementert i en HP iPAQ Pocket PC med Windows PocketPC 2002. Systemet er satt opp slik at så mange funksjoner som mulig ikke kjører på klienten, men på serverside, for at en dårlig prosessorytelse i klienten ikke skal være avgjørende for systemets funksjon. Klienten består logisk sett av to deler; en GUI-klient og en Taleklient. Disse blir her forklart etter tur. GUI-klienten har en TCP/IP-socket-kobling til GUI-serveren, styrt av en socket-server. Det fysiske laget var i denne oppgaven WLAN/IEEE802.11b. Klienten er implementert i C++ og er basert på Pocket Internet Explorer-teknologi. Et kart med trykkbare områder på blir vist på skjermen ved at klienten mottar en URL fra GUI-server som angir en side på GUI-serveren. Ettersom GUI-klienten er basert på Internet Explorer, blir denne siden lastet ned til klienten og vist til brukeren. Kartet inneholder ActiveX-kontrollere som styrer GUI-komponenter som knapper, lister og tekstbokser og gjør siden interaktiv. Ved trykking av ikoner på skjermen, fyres det av events som av socket-serveren legges på kanalen til serveren. Taleklienten er enkelt implementert; ved detektering av et audiosignal på mikrofoninngangen, blir dette sendt gjennom en Voice Activity Detector (VAD), som er i kapittel 3.1.3.1. Om audiosignalet passerer, blir det sendt uten videre prosessering på en VoIP-kanal til tale-server. I motsatt tilfelle, når syntetisert tale sendes fra tale-server, blir det kopiert fra utgangen av VoIP-kanalen til mikrofonen.

3.1.3.1 VAD

Voice Activity Detector (VAD) er i oppgavens versjon av applikasjonen implementert i klienten. Den kunne også ha vært implementert i tale-server. Etter digitalisering i klienten passerer talesignalet gjennom en Voice Activity Detection (VAD)-enhet. I utgaven for denne oppgaven er denne implementert i klienten. VAD har til hensikt å detektere når det snakkes til systemet og skille dette fra støy, og sende talesignalet videre til talegjenkjenning. Dette gjøres ved hjelp av en enkel måling av energiinnhold på en ramme-for-ramme basis i tidsplanet. En ramme i VAD er på 128 samples, som med klientens 8 kHz samplingsrate gir en rammelengde på 16ms. En ramme blir betegnet som høyenergiramme/ramme med tale om den gjennomsnittlige energien i rammen overstiger en viss terskelverdi. Om gjennomsnittlig energi i rammen er under denne terskelen vil rammen få betegnelsen lavenergiramme/ramme med bakgrunnsstøy. Terskelen er satt til 59 dB, fordi klientens mikrofon er angitt å gi en bakgrunnsstøy på 50dB, og 9 dB over dette er satt til å være grense for menneskelig tale. Energiberegning, matematisk:

$$\bar{E} = \sum_0^{N-1} x^2(i)$$

E er energi, N er antall samples i en ramme. $N=128$ i dette tilfellet. Variabelen x er sample-verdien for sample nummer i .

3.1.4 GUI-server

GUI-serveren er en gateway som sørger for kommunikasjonen mellom GUI-klienten og resten av serversiden. Når det ankommer en event fra klienten over socket-koblingen, blir denne mottatt av socketserveren og pakket inn i en galaxy-ramme av en galaxy-interfacemodul og deretter sendt videre ut av GUI-serveren for prosessering. GUI-serverens jobb er noe større når informasjon skal fra serverside ut til klienten: Fra dialogmanageren kommer data om hvordan grafikken skal se ut, lagret på XML-format, som igjen er lagret i galaxy-format for å kunne sendes innen serveren. I GUI-server blir denne informasjonen parset fra galaxy-rammen av galaxy-interfacet, og av XML-data en HTML-fil bygd opp og lagret på en web-side tilgjengelig for nedlasting av klienten. Byggingen av en HTML-fil fra XML blir utført av en "extensible style sheet transformer" (XSLT) -modulen. Samtidig som HTML-siden blir lagt ut for nedlasting, blir en URL generert av galaxy-interfacet som peker på HTML-siden og oversendt klienten.

3.1.5 Tale-server

Tale-serveren inneholder elementene i taledelen av systemet som krever høy prosessorkraft og minne. Delene i denne serveren er talegjenkjenning (eng. Automatic Speech Recognition, ASR), en tekst-til-talem modul, en socket-server for kommunikasjon over IP-nettet mot tale-klienten, samt et galaxy-interface. Tekst-til-tale (eng. Text To Speech, TTS)-løsningen er den Telenor-utviklede Talsmann. Talsmann er utviklet i Microsoft SAPI 4.0, og støtter alle kontrolltagger definert der. Talegjenkjenneren er den proprietære Philips Speechpearl versjon 8.0., utviklet av Philips. Den benytter seg av skjulte markovmodeller i gjenkjenningen (SpeechPearl 8.0 Dokumentasjon 2002).

Speechpearl er trent på SpeechDat2, som har tale med fasttelefonkvalitet. For videre lesing om teori bak talegjenkjenning kan for eksempel teoridelen av prosjektet som ble gjort i forkant av denne masteroppgaven (Aas 2005) leses.

I bruk fungerer tale-serveren som følger: Når et audiosignal ankommer fra tale-klienten, mottas det i socketserveren og sendes videre til ASR-modulen. Her blir gjenkjenning gjort. Gramatikken tilhørende ASR-modulen holder på et sett med elleve forhåndsdefinerte klasser med ord som det søkes etter – ’konsepter’, som kan fylles med en eller flere forhåndsdefinerte ’konseptverdier’. Konseptverdiene er ordene som bærer den semantisk relevante informasjonen i brukerens ytring. Eksempler på hva som er konsepter og hva som er konseptverdier: Konseptet <buss_stasjon> blir fylt fra en setning idèt konseptverdien ”Jernbanetorget” blir funnet, eller konseptet <temporal> kan bli detektert idèt konseptverdien ”når” blir funnet. Resultatet ut fra ASR-modulen er altså en liste med en eller flere konsept-konseptverdipar, med dertil hørende konfidenser. En setning ”neste buss fra Telenor Fornebu til Jernbanetorget” vil kunne se slik ut, på utgangen av ASR-modulen:

<Temporal>	Neste buss	300
<Temporal>	Fra	700
<Buss_stasjon>	Telenor Fornebu	1000
<Buss_stasjon2>	Til Jernbanetorget	1000

Tabell 1: Konsepter, konseptverdier og konseptkonfidenser

Dette blir pakket inn i galaxy-rammer for forsendelse til multimodal server.

3.1.6 HUB

Det finnes flere alternativer på markedet for å implementere et distribuert software-system. Galaxy Communicator, som er en åpen kildekode-versjon av DARPA Communicator, er en av dem. Andre er Open Agent Architecture (OAA) fra SRI international (Open Agent Architecture, 2002), Adaptive Agent Architecture (AAA) fra Oregon Graduate Institute (OGI) (Adaptive Agent Architecture, 2002). Av disse ble Galaxy Communicator vurdert til å være den best passende da denne sentrale enheten i systemet skulle velges.

Det distribuerte systemets sentrum er den åpne kildekode-arkitekturen Galaxy Communicator. Den er spesielt designet for å kunne inkluderes i system med tale-grensesnitt, og kan håndtere dette samtidig som interfacer som inkluderer grafikk, kart, peking og kroppsbevegelser. Galaxy Communicator er basert på en distribuert infrastruktur hvor de ulike delene kommuniserer med hverandre gjennom sending av meldinger til hverandre, og hvor Galaxy Communicator er plassert i midten og forbinder alle med alle som et nav i et hjul. Galaxy Communicator ble utviklet spesielt med hensyn på implementasjon i datasystemer som styrer dialog. Galaxy Communicator HUB er implementert i C. Galaxy Communicator ble utviklet av Defense Advanced Research

Projects Agency (DARPA) under the United States Government. Utviklingen av arkitekturen er nå avsluttet. Selskapet MITRE vedlikeholder idag åpen-kildekode distribusjonen for DARPA Communicator-programmet. Første åpne kildekode-distribusjonen skjedde i juni 2000, og siste oppdatering av distribusjonen var i september 2003¹¹

Den delen av serveren som har kontakt med alle de andre serverne er en prosess som kalles hub. Rammeverket den er en del av fokuserer på modularitet, sømløshet og integrasjon. Alle servere som vil kommunisere med andre servere, sender sine meldinger i et format angitt av hub-arkitekturen. Dette formatet er dynamisk, ettersom det er laget med tanke på et dialogsystem som inkluderer et tale-interface. Formatet bygger på attributt-og-verdipar. En god forståelse for hubens virkemåte kan fås ved gjennomgang av kapittel 3.1.7.

3.1.7 Forfølgelse av en multimodal interaksjon gjennom systemet

Det blir her skissert hvordan en multimodal interaksjon vil prosesserers av systemet.

Systemet står i tilstand "poi_regular", hvor bussholdeplasser vises som ikoner på et subkart. Brukeren bestemmer seg for å spørre demonstratoren om bussruteforbindelsen mellom "Oslo gate" og "Jernbanetorget", og sier følgende ytring: "Når går neste buss herifra til Jernbanetorget". Rett før "herifra" blir det trykket på skjermen, på ikonet som markerer geografisk plassering for "Politihuset". Audiosignalet blir registrert av tale-klienten, som måler energinivået vha. VAD til å være over terskelverdien som er satt, og avgjør at dette audiosignalet er godkjent og skal sendes videre til tale-server. Audiosignalet blir dermed sendt til socketserveren i klienten som kopierer signalet over på kanalen. Trykket på skjermen avfyrrer en event som fanges i klienten, og sendes videre til socketserveren og videre ut på kanalen for mottak i GUI-server. Ved ankomst i GUI-serveren blir eventen pakket inn i en galaxy-ramme og sendt videre til huben, som videresender den til multimodal server. Audiosignalet, som ligger på inngangen til tale-server, plukkes opp fra kanalen av socket-serveren i taleserver og går til talegjenkjenning. På utgangen av talegjenkjenningsmodulen ligger så i eksempelet dette resultatet¹²:

Konsept	Konseptverdi	Konseptkonfidens
Parallell_ref	Parallell	77
Temporal	Bussrute	1000
Parallell_ref	Parallell	220
Buss_stasjon2	Jernbanetorget	1000

Tabell 2: Oversikt over mulig resultat etter gjenkjenning

¹¹ Jamfør den eksterne hjemmesiden for Galaxy: <http://communicator.sourceforge.net/>

¹² Faktisk resultat fra en multimodal turn gjort av bruker 15 i oppgave 2 av testen.

Konfidens-skalaen går fra 0 til 1000. Her er ”til Jernbanetorget” konseptverdien for konseptet ”buss_stasjon2”. ”Neste buss” er det som ble gjenkjent som ”temporal”-konsept. Den første ”Parallell_ref” refererer til ”når går”, som jamfør gramatikken skulle vært skrevet som to konsept, men er i dette tilfellet ikke med hell gjenkjent som noe som helst. ”herifra” er også gjenkjent som en veri i ”parallell_ref”-konseptet, kanskje ”herog”, men kun med 220 i konfidens. ”Herifra” er ikke angitt som egen verdi i gramatikken. Disse konseptene pakkes inn i galaxy-rammer og sendes videre til multimodal server. I multimodal server kommer penn-rammen inn noe før tale-rammen, noe som henger igjen fra det faktum at tale-rammen ikke ble sendt fra tale-server før hele signalet var ferdig mottatt. En timer starter idèt penn-rammen kommer inn, og er ikke nedtelt innen tale-rammen ankommer, med det resultat at de pakkes inn i samme galaxy-ramme og sendes videre for behandling i dialogserveren. Denne første sammenkobling av resultatene tilsvare en delvis fusjon, ettersom den er gjort kun på temporal basis og ikke tar hensyn til det semantiske innholdet. I dialogmanager blir sen fusjon gjort – det blir trukket ut av ankommet ramme ”vinnerkonseptverdiene” for hvert konsept. Etter sen fusjon ser informasjon fra bruker slik ut:

Konsept	Konseptverdi	Konseptkonfidens
Temporal	Bussrute	1000
Buss_stasjon	Politihuset	1000
Buss_stasjon2	Jernbanetorget	1000

Tabell 3: Resultat etter fullført fusjon i dialogmanager.

Legg merke til at ingen av ”Parallell_ref”-tuplene er med her; dialogserveren har en konfidensterskel på 500 for innfylling av konsepter. Legg også merke til at det nå ikke lenger er noen forskjell på konsepter med opprinnelse i gui-klient og tale-klient; fusjonen er ferdig gjennomført. Etter dette blir informasjonen matet inn på FSM som er implementert i dialogmanageren. Denne står i tilstanden ”start_regular”¹³, som vil si at ingen inndata foreløpig er registrert. Tilstandsmaskinen tar inn konseptene som inndata I_t , og finner at denne informasjonen er tilstrekkelig for en spørring mot databasen, som i dette tilfellet er trafikanten Svaret blir mottatt, og gjort klart for svar til bruker som to rammer – en for GUI-klienten og en for tale-klienten. Tale-rammen inneholder kun en tekststreng, mens gui-rammen inneholder informasjonen på XML-format. Tale-rammen sendes via hub og multimodal server til tale-server hvor den konverteres til tale ved hjelp av TTS-motoren og legges på kanalen tale-klienten som audio. Klienten plukker opp denne og legger den på utgangen for lyd. Rammen med gui-data oversendes GUI-server, hvor XML-taggene blir lest og HTML-siden med ActiveX-kontrollerte GUI-elementer blir bygd opp av XSLT-enheten. Siden blir lagt på på en webside samtidig som en URL oversendes GUI-klienten. Denne blir mottatt og innholdet lastet ned. På dette tidspunktet kan brukeren høre og lese tilbakemeldingen fra systemet: ”Buss 37 mot Nydalen T fra Politihuset kl 14.50 til Jernbanetorget T kl 14.56”.

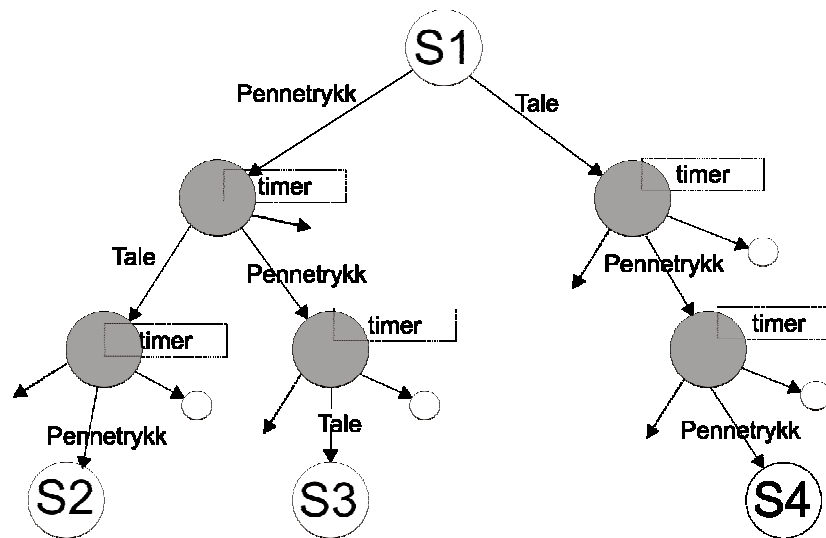
¹³ Faktisk resultat fra dialogen bruker 15, oppgave 2

3.2 HVORDAN FUSJON ER IMPLEMENTERT

Fusjon er vanligvis implementert i to klasser, *tidlig fusjon* og *sen fusjon* (Kvale et al, 2003 s 109). Denne framgangsmåten er også brukt i bussdemonstratoren. Tidlig fusjon er implementert i multimodal server og gjør temporal fusjon, andre steg er semantisk fusjon i dialogserveren. Å gjøre fusjon i dialogmanageren i stedet for i multimodal server går imot prinsippet om modularitet, som er tilstrebet i systemet. Dette ble gjort fordi dette er en enklere løsning, ettersom sen fusjon krever informasjon om nåværende tilstand, som er lagret i dialogmanager. Semantisk fusjon skjer ved at en tabell fylles med de konsepter som mottas, og hvor konflikter blir løst ved hjelp av konfidens-score. For videre informasjon om hvordan semantisk fusjon skjer henvises til gjennomgang av systemet, kapittel 3.1.7

Temporal fusjon i multimodal server er basert på timere. To timere med ulike tider kan trigges i serveren, avhengig av om en tale-inndata eller en penn-inndata ankommer først. Ved ankomst av en ramme med tale-inndata startes timeren "short_fusion" på 0,85s, mens det ved ankomst av en ramme med penn-inndata startes "long_fusion" på 3,0s. Tanken bak dette er at et eventuelt pennetrykk vil ankomme relativt kort tid etter en tale-inndata om denne ankom først, mens om pennetrykket ankommer først forventes det at talesignalet vil bruke noe mer tid før det ankommer, på grunn av at talesignalet ikke er en momentan handling samt økt tid til prosessering.

Fusjonstimerne er satt til å avslutte momentant ved deteksjon av til sammen 2 pennetrykk og ett tale-innsignal. Dette er maksimal interaksjon som er satt i systemet til å kunne mottas i løpet av en turn. Dette er implementert for å muliggjøre den mest komplekse versjonen av simultan koordinert multimodalitet – tale og to pennetrykk i samme ytring. For håndtering av disse har man valgt å utvide FSM i dialogmanageren med et antall tilstander mellom hver "normaltilstand", se figur 4.



figur 4: Fragment av FSM i Dialogmanager med hjelpetilstander for håndtering av multimodale ytringer.

Tilstandene S1, S2, S3 og S4 er normaltilstander, hvor systemet går inn i idle og venter på ny interaksjon fra brukeren. Det er 3 hopp fra en "normaltilstand" til en ny normaltilstand, som refererer til de maksimale 3 inndata som kan mottas av systemet i løpet av en turn. Introduksjonen av hjelpetilstandene er først og fremst en løsning for å implementere tidsbegrensede tilstander. Når systemet befinner seg i en slik tilstand, har det som figuren viser, 3 mulige ut-transisjoner. I traversering fra en normaltilstand til en annen normaltilstand kan systemet enten motta 3 "gyldige" inndata, for eksempel penn-tale-penn, hvor da korresponderende transisjonen i figuren vil følges. Om en ugyldig sekvens av inndata mottas, for eksempel tale-tale, vil systemet gå over i en feilhåndteringstilstand fra hjelpetilstanden øverst til høyre. Om systemet blir stående i en hjelpetilstand og timeren teller ut, vil til da innsamlet inndata bli tolket, mening forsøkt uttrukket og systemrespons bli generert på bakgrunn av dette.

Ved å implementere denne løsningen får FSM et stort antall tilstander, men dette er ikke et problem ved nåværende, begrensede system.

Det kan ses av implementasjonen av FSM i figur 4 at systemet støtter to penneklikk på samme skjerm før oppdatering. Dette er kalt simultan penn-interaksjon. Denne interaksjonsformen er bare gjengitt ettersom systemet har denne implementert. Den brukes ikke i denne oppgaven.

3.3 ULIKE DIALOGSTRUKTURER

Systemet har fått implementert en dialog med blandet initiativ. Gjennom blandet initiativ tillates så vel bruker som system å ta initiativ å gi informasjon og spørre om informasjon, og informasjon kan mottas av begge parter. Dette er en bedre og mer naturlig dialogstruktur enn systemdrevet struktur, hvor systemet styrer og brukeren uten valgfrihet må forholde seg til dette. Det er også vurdert som en bedre dialogstruktur enn det motsatte fallet – fullstendig brukerorientert struktur, som dialogstyringsmessig er meget komplisert. Barge-in er ikke tillatt gjennom talemodaliteten, men er mulig ved penneklikking. Gjennom denne spesifikasjonen er dialogstyringen implementert gjennom en endelig tilstandsmaskin i systemets dialogserver, se kapittel 3.1.1 For utdypende lesning se artikkelen "Spoken language systems - beyond prompt and response" av Wyard et al (1996).

4 GJENNOMFØRING AV TESTEN

Dette kapitlet beskriver hvordan testene ble utført, både før, under og etter. Oppgaven er også beskrevet her.

4.1 VALG AV BRUKERGRUPPER

4.1.1 Definere brukergrupper

Valget av brukergrupper henger sammen med oppgavestillinga. Oppgaveformuleringa sier blant annet:

(...)Oppgaven går ut på å evaluere denne tjenesten for nye brukergrupper som kan ha spesielt nytte av valgfriheten multimodalitet gir. Vi er spesielt interesserte i å prøve ut demonstratoren på gruppene barn/ ungdom og eldre, men vi er åpne for andre forslag fra kandidaten.

Ut fra dette ble det satt opp fire hovedgrupper brukere:

- Vanlige brukere (i denne sammenhengen)
- Eldre; personer over 60 år
- Personer med fysiske handikap
- Personer med psykiske utviklingshemninger

Dette er store grupper med store forskjeller mellom brukerne innad i gruppene. Ytterpunktene er kanskje spesielt langt fra hverandre i gruppene ”personer med fysiske handikap” og ” personer med psykiske utviklingshemninger”. På grunn av manglende kompetanse deles likevel ikke gruppene mer opp. Det ble gjort en vurdering på at om noen av disse to gruppene ble valgt, burde en pragmatisk tilnærming benyttes. Det vil si å motta alle dem som sier ja til å være med i test med åpne armer, heller enn å sette opp testperson-kriterier først og lete etter personer som passer. Dette ble vurdert som tidkrevende og antagelig vanskelig å lykkes med i praksis.

4.1.2 Kriterier i valg av brukergrupper

Telenor FoU er klare over multimodalitets mulige støtte for mennesker med ulike begrensninger, og ønsker dokumentasjon for en del av disse. Det har vært gjort noen brukertester i før, hvor fysisk handicappede deltok, se artikkelen ”Evaluation of a Mobile Multimodal Service for Disabled Users” (Kvale et al 2005) . Demonstratoren er siden den gang blitt forbedret, og en mulighet besto derfor i å teste på nytt noen av de samme gruppene.

Det økonomiske aspektet ble også skjelet til i vurderingen. En multimodal tjeneste vil sannsynligvis måtte være populær i den store gruppen ”vanlige brukere” ettersom denne gruppen ville komme til å være gruppen som bærer prosjektet økonomisk.

Om enn ikke avgjørende for beslutningen, men like fullt viktig for disponering av tid, ble det også vurdert praktiske sider ved testene; hvor mange brukergrupper vi ville kunne få til på den gitte totale tida, hvor vanskelig tilgjengelig var gruppa – dvs hvor mye måtte jobbes før man sto klar med en bruker, tidsbruk per pers osv.

4.1.3 Valg av brukergrupper

Forsøk med personer med fysiske handikap var noe dokumentert allerede (Kristiansen, 2004) (Kvale et al 2005). Disse resultatene ses det ingen grunn til å etterprøve eller komplettere.

Både de eldre og de psykisk utviklingshemmede er store grupper, med de eldre som den sannsynligvis største gruppa. Om gruppene sammenlignes, vil det sannsynligvis være lettere og raskere å gjennomføre tester blant de eldre enn blant psykisk utviklingshemmede, og de vil antagelig være lettere tilgjengelige. Gruppa med eldre var også tidlig foreslått fra Telenor R&D, ettersom dette er ei gruppe som ikke er utforsket i stor grad. Det samme gjelder imidlertid for psykisk utviklingshemmede. Begge representerer altså interessante, lite utforskede grupper i denne sammenheng.

Et kriterium som gruppene antas å være nokså forskjellige på, er hvor villige de vil være til kjøpe en eventuell kommersiell versjon. I gruppa for psykisk utviklingshemmede jobber gjerne hjem og skole/kommune sammen om tilrettelegging for den aktuelle, og i miljøet er en god del spesielle tekniske hjelpemidler i bruk i dag. Før valget av brukergrupper ble tatt, var det allerede tatt kontakt med Dalgård skole i Trondheim¹⁴, hvor IKT-ansvarlig hadde bekreftet at de ønsket alle tekniske løsninger som kunne hjelpe elevene, velkommen.

I eldregruppa forventes det ikke det samme. Mange av dagens eldre er fremmede for en stor del av IK-tjenestene som finnes i dag. Årsakene kan være forskjellige, men én av dem er antagelig at man ikke ser mange av disse tjenestene som nødvendige, og det må forventes at en multimodal bussruteopplysning heller ikke blir sett på som det. Selv om eldre-gruppa er større enn de med psykisk utviklingshemming, blir i denne oppgaven sistnevnte argument vektlagt mer.

Det var også et ønske om å se på den største gruppa brukere – de i denne sammenheng ”vanlige” brukerne. Disse kan i utgangspunktet klare seg uten multimodalitet, men kan komme til å bruke for å forenkle oppgaver som blir gjort. Om det i framtida skulle lanseres en kommersiell versjon av den multimodale bussruteopplysningen, vil denne gruppa være den som bidrar med mest penger. Følgelig vil det være en fordel å vite noe om hvordan tjenesten vil bli mottatt her.

Det er også kjent at for eksempel studenter og unge er spesielt tidlige brukere av ny teknologi. Gjennom å teste i første omgang universitetsstudenter fra NTNU, og eventuelt

¹⁴ Dalgård skole: <http://www.skole.trondheim.kommune.no/dalgard/>

deretter tilfeldig valgte ungdomsskole-elever, var håpet å treffe en del av dem som er tidligst ute med å ta ny IKT-teknologi i bruk.

En prioritering av brukergruppene blir da som følger:

- 1) Studenter, NTNU
- 2) Psykisk utviklingshemmede
- 3) Skoleelever, ungdomsskole
- 4) Eldre.

Det ble sikta mot å gjennomføre brukeforsøk i gruppene en, to og tre, men på grunn av dårlig tid ble det bare tid til å gjennomføre testen for en gruppe.

4.2 TESTBRUKERNE

I forkant av undersøkelsen ble det sendt ut invitasjon til å delta på brukertester på epost til venner av forfatteren. Det ble underveis i testperioden sendt ut ytterligere en invitasjon til samtlige studenter på forfatterens studium¹⁵. En stor del av disse fikk også personlig forespørsel om å delta i testen. Før periodens start ble det hengt opp invitasjoner på felles oppslagstavler på NTNU Gløshaugen. Underveis i perioden ble det ved ett tilfelle gjort oppsøkende virksomhet overfor tilfeldige studenter i universitetets vrimlearealer.

Kun en testbruker kom som følge av opphengte informasjonsark. En testbruker kom som følge av oppsøkende virksomhet. Resten kom som følge av forespørsel, enten på epost eller ved direkte forespørsel.

Testene med personer i et teknisk miljø skjedde på NTNU, Trondheim, i tiden 24/3 til 6/4 2006. 25 av de deltagende studerte teknologi ved NTNU, samtlige var studenter. I gruppen var 18 personer avgangselever ved fakultet for informasjonsteknologi, matematikk og elektronikk (IME), og må sies å ha god kjennskap til ulike former for datasystem og grafiske brukergrensesnitt fra studiet. Alderen på de testede studentene var mellom 20 og 28 år, med hovedvekt på gruppa 24-26 år (21 personer). 8 kvinner og 22 menn deltok. 2 tester, en fra hvert kjønn, er ikke med i sluttresultatet på grunn av dataproblemer som gjorde at testen aldri ble avholdt. Ni av 28 personer var kjent med bruk av PDA fra før, resten anga seg som ukjent for en PDA. Se vedlegg for videre informasjon om brukerne.

4.3 TEORETISK BAKGRUNN FOR TESTEN

Testen (se vedlegg 10.4) er blant annet laget med tanke på å søke å belyse spørsmålsstillinger som er blitt framsatt i tidligere artikler om emnet. I artikkelen

¹⁵ Linje for kommunikasjonsteknologi ved fakultet for informasjonsteknologi, matematikk og elektronikk (IME).

”Speech Centric Multimodal Interfaces for Mobile Communication Systems” (Kvale et al. 2003) står det i kapittel 5.3, i avsnittet ”human factors”:

It is therefore necessary to run user experiments and quantify the advantages of the multimodal system in terms of user-satisfaction, dialogue success rate and dialogue completion time. These experiments are high on our agenda.

I artikkelen “Multimodal Interaction – Will Users Tap and Speak Simultaneously?” (Rugelbak og Hamnes, 2003) står følgende i konklusjonen:

The main goal for this experiment has been to identify research issues to be studied further in a planned user experiment within the MUST project.

Konklusjonene i disse to artiklene pekte på visse områder man ønsket å vite mer om i forhold til multimodal interaksjon; noen av disse var *når* multimodalitet ble brukt, og hvordan trykket i forhold til når det ble pratet.

Casene ble utformet med ulik mengde forhåndsinformasjon for å skape varierte oppgaver, og de ble utformet for å kunne kaste lys over noen av punktene fra artiklene nevnt over.

4.4 INNHOLD I EN BRUKERSESJON

4.4.1 Avveininger ved utforming av brukersesjon

Det var avgjort at målet for testen var å teste interaksjonen med systemet fra en bruker som hadde kommet over den første introduksjonsfasen til systemet. Som beskrevet i (Rugelbak og Hamnes, 2003) synes lærekurven å være bratt. Gjennom en god introduksjon hvor alle aktuelle modaliteter ble vist, i tillegg til en fase med egen læring gjennom utprøving, ble det forsøkt å få testpersonen så langt opp på denne lærekurven før selve testens start. Dette ville belegge rundt 20 minutter hos hver testperson. Sammen med spørreskjemaet etter testen på rundt 8 minutter satte dette en nedre grense på tidsbruk på testen på rundt en halv time.

Testpersonene var i begynnelsen forespeilet en test på rundt en halv time, men dette ble korrigert opp til 45 minutter etter første dag. Ved problemer med å holde tidsbegrensingen ble det prioritert å avbryte testen for å få gjennomført spørreskjemaet, for å ikke holde testpersonen lengre enn avtalt.

4.4.2 Brukerens ressurser

Ved sesjonens start hadde brukeren enkelte ressurser for hånden:

- PDA
- Et ark med casene (vedlegg: kapittel 10.4)
- Et ark med navn på alle gyldige bussholdeplasser. (vedlegg: kapittel 10.5)
- Et ark med oversikt over unormale begrensninger brukeren bør vite om. (vedlegg: kapittel 10.5) Dette ble fjernet etter bruker 7 ettersom arket ble vurdert å bidra mer til å gi et rotete inntrykk av brukers sitteplass enn det ga hjelp. Informasjonen ble heretter gitt muntlig i slutten av ”kameratintro”.

- Et standard turistkart over Oslo. Kartet er et standard gratis turistkart som ble skaffet fra Oslo turistinformasjon.

4.4.3 Kameratintroduksjon

4.4.3.1 Begrunnelse for denne typen introduksjon

Av artikkelen ”Multimodal Interaction – Will Users Tap and Speak Simultaneously?” av J. Rugelbak og K. Hamnes (2003) framgår det av konklusjonen at multimodalitet er en ny form for interaksjon og at brukere vil trenge en introduksjon for å bli klar over og forstå denne nye funksjonaliteten.

For å introdusere en bruker til systemet eksisterte i dette tilfellet minst tre muligheter: Skreven introduksjon, introduksjon på video, samt kameratintroduksjon. Av disse ble kameratintroduksjon ble valgt. Denne formen for introduksjon gjøres ved at brukeren vises bruksmåten til systemet fra en allerede erfaren person, i en situasjon som forsøker å etterligne en situasjon hvor en person lærer opp en kamerat. Denne formen for introduksjon ble valgt jamfør konklusjonen i artikkelen ”How do non-expert users exploit simultaneous inputs in multimodal interaction?” (Kvale et al 2003). Denne viser at formen introduksjonen blir gitt på, har en notèrbar effekt på brukerens bruk av systemet. Det var et ønske at brukeren skulle kunne benytte seg av alle former for interaksjon systemet tilbød, i testen, derfor ble kameratintro valgt. En ulempe med å benytte kameratintroduksjon er at ingen person ville få den eksakt samme introduksjonen, og dermed få noe ulik forkunnskap for testen. Dette ble forebygget så godt som mulig ved å følge en detaljert mal.

4.4.3.2 Gjennomføring

Ved start av introduksjonen ble mikrofonen først skrudd av ved hjelp av mute-funksjonen. Deretter ble det opplyst om at det man hadde foran seg, var et program for bussinformasjon i Oslo, og at en av systemets styrker var muligheten for tale og trykking – samtidig. Det ble også opplyst om at taledelen kunne være noe streng, men at det ved normal tale som regel ville gå bra. Etter dette ble det opplyst om at de ulike interaksjonsformene ville bli presentert brukeren etter tur, og at det ville bli startet med kun penn.

Fortsatt med mikrofonen avslått, ble trykket på øvre høyre subkart. Det ble trykket på til-knappen, ventet på skjermoppdateringen, og deretter en vilkårlig holdeplass. Etter å ha fått skjermoppdateringen ble det så trykket på en ny holdeplass i samme subkart, mens det ble forklart at dette ville bli forstått som fra-holdeplass ettersom til-holdeplass allerede var valgt. Da ble det informert om muligheten til å velge holdeplasser i ulike subkart ved hjelp av zoom ut-knappen. ”Hjem”-knappen ble så trykket.

Interaksjonsform to var tale, som ble demonstrert ved at testleder aktiverte mikrofonen og ba om ”neste buss fra Telenor Fornebu til Majorstua”. Majorstua ble uttalt med a-ending. Ved gjentatt forespørsel fra systemet om ankomstholdeplass, ble dette i noen tilfeller

forandret til en-ending. Etter ytringen ble mikrofonen slått av og ”hjem”-knappen trykket. Etter ytringen ble det repetert den informasjon som systemet måtte etterspørre.

Tredje interaksjonsform var ”peik og preik”. Dette ble demonstrert ved at testleder trykket på øvre høyre subkart, aktiverte mikrofonen og sa ”neste buss til Majorstuen herifra”. Idet ”herifra” ble sagt, ble et vilkårlig bussholdeplass-ikon på subkartet trykket. Etter ytringen ble det repetert den informasjon som systemet i enkelte tilfeller måtte etterspørre.

4.4.3.3 Endringer

Etter forsøksperson 7 ble det på dette tidspunktet gitt noen ekstra opplysninger. Det ble gjort oppmerksom på at testbrukeren i testen skulle simulere å ha leilighet i nærheten av busstoppet ”Harald Hardrådes plass”. Brukeren ble opplyst om det begrensede utvalget bussholdeplasser i ordboka, og at verken andre bussholdeplasser eller andre stedsnavn ville bli gjenkjent. Fra forsøksperson 12 ble det gjort oppmerksom på at brukeren kunne studere navn på bussholdeplasser uten å starte en spørring. Før dette ble denne informasjonen gjort kjent kun på spørsmål. Det ble til en del, men ikke til alle, forklart funksjonen til HOME-knappen.

4.4.4 Egen tid

Etter introduksjonen ble det gitt noe tid for brukeren til å bli kjent med systemet på egen hånd. Det ble gitt frihet til å holde PDAen eller ha denne i dokken. Brukeren ble informert om at tilgjengelig tid var mellom fem og ti minutter, og at brukeren selv kunne gi tilbakemelding om når han/hun var klar.

4.4.5 Test

Ved start av testen ble tidspunkt notert. Testleder satt i testen overfor brukeren, slik at kun brukeren så skjermen på PDAen. Testleder hadde serveren å se på. Under testen ble brukerens tilnærming til systemet observert. Observasjonsskjemaet er gjengitt i vedlegg 10.7.

4.4.5.1 OPPGAVE 1

I oppgave 1 er brukerens eneste spørsmål *når*. Fra-stasjon og til-stasjon er kjent, og brukeren henvender seg til systemet for å få rede på tidspunkt for neste buss.

Ved start av testen ble første del av oppgave 1 lest opp for brukeren. Det ble ventet til brukeren hadde gjennomført oppgaven før man fortsatte med oppløsing av andre del.

Færrest mulig dialogturns/kortest mulig tid oppnås for denne oppgaven idet man sier inn hele ytringen med monomodal tale i én turn.

4.4.5.2 OPPGAVE 2

I oppgave 2 vet brukeren en av to stasjoner (Jernbanetorget), og blir bedt om å finne to bussløsninger nær sitt hjemsted. Brukeren trenger altså informasjon om *når* og *hvor*.

Etter at oppgave 1 var ferdig løst, ble tiden notert og oppgave 2 opplest for brukeren. Det ble for noen brukere spesifisert på spørsmål at ”andre stopp i din bydel” pekte på bussholdeplasser rundt den tenkte boligen, men unntatt busstoppet ”Harald Hardrådes Plass”.

Nest etter forrige case, antas dette å være den hyppigst forekommende typen forespørsel for denne type applikasjon; Brukeren kjenner navnet (eller eksakt posisjon i kart) for en av holdeplassene, og ønsker informasjon om hvordan man kommer seg til en holdeplass nr. 2, samt tid for dette.

Færrest mulig dialogturns/kortest mulig tid oppnås for denne oppgaven idet man zoomer inn i subkartet som inneholder den ukjente bussholdeplassen og bruker en simultan koordinert multimodal ytring til å angi denne med penn, mens man gir konseptet ”temporal” og den andre bussholdeplassen gjennom tale.

4.4.5.3 OPPGAVE 3

I oppgave 3 er brukerens mål å vite *når*, *hvorfra* og *hvortil*. Dette representerer en helt ny situasjon, hvor brukeren ikke kan støtte seg på tidligere erfaringer.

Etter at oppgave 2 var ferdig løst, ble tiden for dette notert og oppgave 3 lest opp. Det ble henvist til kartet som var tilgjengelig for brukeren. Fra og med bruker 4 ble det notert brukeren at navnene ”Aker Brygge” og ”Oslo Spektrum” ikke var bussholdeplassnavn. Det ble lagt inn en kommentar fra og med bruker 13 at testens formål ikke var å teste brukerens kjennskap til byen, derfor ville det ved behov bli gitt hjelp til å få øye på de ulike stedene som ble nevnt. Det ble også nevnt at det ikke fra testens side var viktig om brukeren fant fram til den bussholdeplassen som lå nærmest det ønskede stedet, om det bare var i ”rimelig gangavstand”.

Oppgave 3 inneholder en komponent av leting i kart, på PDAen og/eller på et turistkart. Kartene har ulik målestokk. De digitale kartene har begrenset visuell og funksjonell kvalitet. Dette, samt at oppgaven krever noe sammenligning av kartene, som gjør oppgaven noe tidkrevende.

Denne oppgaven er bare mulig å løse ved aktiv pennebruk; oppgaven er en typisk ”penneoppgave”. Man må zoome inn på et subkart først, og trykke på en stasjon. Deretter må man zoome ut og inn i et nytt og trykke på nytt på en av holdeplassene for å merke holdeplass nummer to. Et alternativ som går omtrent like fort, er å finne holdeplassnavn på en av holdeplassene, zoome inn i det andre og gjøre en simultan koordinert ytring. I dette tilfellet trenger man bare å huske på et av holdeplassnavnene, men ”utsetter seg for” faren for mislykket gjenkjenning (på om lag 43%, skal det vise seg).

4.4.5.4 Momenter i oppgave 3

Oppgave 3 etterspør funksjoner i systemet som ved testens tidspunkt ikke fantes i demonstratoren. Dette er: Gateadresser og navn på hus og institusjoner i byen lagt inn i systemets bibliotek, en bedre zoom-funksjon; å kunne presentere kartet i mer enn to målestokker, og panering; å kunne dra kartet til siden. At brukeren gjennom leting på kartet selv må finne bussholdeplassen fordi ønsket hus eller institusjon ikke er lagret i biblioteket, kan til noen grad forsvares ettersom systemet kun vil holde på en viss mengde av dette, og alltid vil være mangelfull. Dette gjør at oppgaven bare i begrenset grad kan brukes til evaluering av interaksjonsformen. Oppgavens resultater kan kun avsløre tendenser for hvordan en slik case vil løses.

En slik case kan tenkes å være relevant når brukeren er i en situasjon med god tid og uten ytre faktorer som påvirker brukerens interaksjon med systemet. Brukeren kan kanskje tenkes å i denne typen forespørsler være klar for å stille til rådighet en stor mengde kognitiv anstrengelse for å få svar, jamfør hvor viktig svaret er for brukeren.

4.4.6 Spørreskjema

Etter testslutt ble et spørreskjema gjennomgått for å samle informasjon som ikke lot seg måle i testen. Skjemaet som ble brukt til dette finnes som i vedlegg 10.6. Spørsmålene ble stilt av testleder og denne noterte også svarene.

5 RESULTATER

Dette kapitlet presenterer resultatene som ble funnet gjennom brukertestene, hvordan det ble telt, samt begrunnelse der hvor det er naturlig.

5.1 OPPTELLING AV MODALITETER

5.1.1 Slik ble opptellingen gjennomført

I opptellingen av turns med inndata ble et perl-script, "anal_log_2.pl", og deretter et nytt perl-script, "hoyNivaaPerlscript.pl" benyttet. "anal_log_2.pl" ble brukt til å legge inn relevant informasjon om samtlige dialog-logger, inn i en tekstfil, og luke bort resten. "hoyNivaaPerlscript.pl" ble brukt til å traversere denne tekstfilen og foreta opptellingen av de ulike typene modaliteter.

5.1.2 Testgjennomføringsmessige merknader

Bruker	oppg. 1	oppg. 2	oppg. 3
1	x	x	x
2			
3	x	x	x
4	x	x	x
5	x	x	x
6	x	x	x
7	x		
8	x	x	x
9	x	x	x
10	x	x	x
11	x	x	x
12	x	x	x
13	x	x	x
14	x	x	x
15	x	x	x
16	x		
17	x	x	x
18			
19	x	x	x
20	x		
21	x	x	x
22	x	x	x
23	x	x	x
24	x	x	
25	x	x	x
26	x	x	x
27	x	x	x

28	x	x	x
29	x	x	x
30	x	x	x

Tabell 4: Oversikt over brukere og gjennomførte oppgaver. Et kryss betyr fullført oppgave.

Bruker 2: Ved teststart for brukeren ble det erfart mye systemproblemer som vedvarte. Ettersom dette var tidlig i testfasen tok det lenger tid å håndtere dette, og testen viste seg å måtte utgå i sin helhet.

Bruker 5 ble enig med testleder etter oppgave 1 å gjøre resten av testen med penn. Det oppleves store problemer med å få systemets talemønstre til å virke som det skal.

Bruker 7: Det oppleves problemer med TaleServer. Ved enkelte ganger tar det unormalt lang tid for systemet å gi tilbakemelding etter at bruker har endt sin stemme-interaksjon. Systemet slutter å prosessere inndata og må restarteres 9 ganger i løpet av testen. Dette tar mye av tida, og påvirker brukers motivasjon til å fullføre testen. Testen avsluttes etter 20 minutter uten å ha fullført oppgave 1.

Bruker 16: Det oppleves treg respons fra systemet, og en er plaget med ofte krasj. Testen avbrytes etter gjennomføring av oppgave 1 på grunn av dette.

Bruker 18: Det lykkes aldri å få systemet i en stabil tilstand over noe tid. Testen blir derfor ikke gjennomført.

Bruker 20: Brukeren opplever dårlig talegjenkjenning, samt at brukeren ikke venner seg til systemets responstid og får mye feilmeldinger. Det blir gjort noen restarter før og under test. Tidsperspektivet gjør at en for tidlig avslutning av testen velges.

Bruker 24: Brukerens tale gjenkjennes dårlig, og brukeren blir ikke vant med systemets langsomme responstid. Dette gir en del ASR-relaterte problemer, og gjør at testen tar lang tid samt at brukeren blir frustrert. Testen gjøres på en dag da til sammen 5 tester er avtalt, noe som for denne testen gir et trangt tidsskjema. Testen avsluttes i samråd med testpersonen etter oppgave 2 etter ti minutter.

5.1.3 Andel turns av ulike typer

OPPGAVE 1:		Antall	Prosent
Antall Penn-turns		78	34,21 %
Antall Tale-turns		148	64,91 %
Antall MM turns		2	0,88 %
Totalt antall turns		228	100,00 %
OPPGAVE 2:			
		Antall	Prosent
Antall Penn-turns		86	40,00 %
Antall Tale-turns		122	56,74 %
Antall MM turns		7	3,26 %
Totalt antall turns		215	100,00 %
OPPGAVE 3:			
		Antall	Prosent
Antall Penn-turns		64	32,99 %
Antall Tale-turns		124	63,92 %
Antall MM turns		6	3,09 %
Totalt antall turns		194	100,00 %
I ALT			
		Antall	Prosent
Antall Penn-turns		228	35,79 %
Antall Tale-turns		394	61,85 %
Antall MM turns		15	2,35 %
Totalt antall turns		637	100,00 %
Gjennomsnittlig antall turns			
oppgave 1	173/28	6,18	28 brukere gjennomførte oppgave 1.
oppgave 2	158/26	6,08	26 brukere gjennomførte oppgave 2.
oppgave 3	132/23	5,74	23 brukere gjennomførte oppgave 3.
hele testen		17,99	

Tabell 5: Andel turns av ulike typer, telt ved hjelp av script

5.2 SEKVENSIELT MULTIMODALE DIALOGER

5.2.1 Hvordan ble sekvensiell MM telt

I opptellinga ble det valgt å følge følgende definisjon på sekvensiell multimodalitet: Hvis brukeren i løpet av en dialog, slik "dialog" forstås i systemet, fyller konsepter ved hjelp av begge modaliteter, men i ulike turns, er dette en sekvensiell multimodal turn. Ettersom definisjonen for hva som er simultant koordinerte multimodale turns i denne oppgaven baserer seg på brukerens handling og ikke dialogmanagerens logg, oppstår et nøyaktighetsproblem ved tallfestelse av sekvensiell multimodal handling. Resultatene om

simultan koordinert multimodalitet angir minimumstall, som følge av at metodene for opptelling for en stor del baserte seg på auditiv analyse av de lagrede ytringene.

Det ble valgt å angi sekvensielle og simultant koordinert multimodalitet i sammen, ettersom det lettet opptellingen vesentlig, og ettersom det fortsatt gir verdifulle resultater.

En dialog kan være hel eller delvis. Brukeren hadde mulighet til å avbryte dialogen før den er ferdig ved å trykke på "home"-knappen. Dialoger fra ikke fullførte tester er regnet med.

Tallmaterialet som beregningen bygger på, er samme tallmateriale som står bak tabell Tabell 5. på grunn av egnethet for script-opptelling. Unøyaktigheten som introduseres på grunn av dette er ukjent, men antas å være begrenset på grunn av at

- Det er irrelevant om det er 10 eller 2 talesignal i en dialog.
- Telling av sim. koord. og sekvensiell multimodalitet i sammen gjør at representasjonen av en multimodal ytring i loggen som én eller to turns er irrelevant.

5.2.2 Dialoger med sekvensiell eller simultan koordinert multimodal interaksjon

Bruker nr	Antall MM dialoger	Antall dialoger totalt
1	5	7
2	-	-
3	11	12
4	7	9
5	1	9
6	4	11
7	4	7
8	5	8
9	1	8
10	6	16
11	1	8
12	4	9
13	8	14
14	2	10
15	5	10
16	2	2
17	6	9
18	-	-
19	9	16
20	1	7
21	3	6
22	0	6
23	3	9
24	1	2

25	3	6
26	2	5
27	3	6
28	7	12
29	4	5
30	6	8

tabell 6: Dialoger som inneholder sekvensiell eller simultan koordinert multimodalitet. Sortert på bruker

OPPGAVE 1		
MM dialoger:	35	50,0 %
Antall dialoger totalt:	70	
OPPGAVE 2		
MM dialoger:	37	46,3%
Antall dialoger totalt:	80	
OPPGAVE 3		
MM dialoger:	42	48,3 %
Antall dialoger totalt:	87	
I ALT		
MM dialoger:	114	48,1 %
Antall dialoger totalt:	237	

tabell 7: Dialoger som inneholder inneholder sekvensiell eller simultan koordinert multimodalitet, sortert på oppgave

5.2.3 Tidsbruk

Dessverre ble nedskrivning av tid for de ulike brukerne ikke gjort med den nødvendige presisjon. Noteringen ble heller ikke gjennomført konsekvent. Det ble forsøkt å gjenfinne dialoggjennomkjøringstid ved inspeksjon av dialogmanagerens logg, men dette ble for usikkert. Resultater om tidsbruk er derfor utelatt.

5.2.4 Hvordan ble tellingen gjort

Det viste seg vanskelig å finne presise tall for dialoggjennomkjøringstid ved å gå gjennom loggene i etterkant av testen. Tallene er unøyaktige fordi det i noen tilfeller medregnes tid hvor brukeren stanser for å stille spørsmål, og hvor testleder interagerer. Dette er ikke mulig å kontrollere gjennom traversering av loggen. Noen av tallene inneholder kun tiden fra brukeren starter på oppgaven til siste svar i oppgaven er mottatt, mens andre også inneholder hele eller deler av testleders opplesing av oppgaven. Dette er bieffekter av unøyaktig tidtaking i testsituasjonen.

5.3 REVIDERT OPPTELLING AV MODALITETER

5.3.1 *Hvorfor revidert opptelling av turns som inneholder tale*

Ved kontroll av sporadiske tale-filer nevnt i dialogmanager-loggene, kom det fram at av de kontrollerte filene var en signifikant andel bestående av ikke-ord - mumling og fylte pauser, bakgrunnsus, dunking. Det var også en del lav prat som åpenbart ikke var tenkt som inndata til systemet, eller på annen måte auditiv informasjon ikke var ytret av bruker i den hensikt å kommunisere med systemet.

Dette var innsignal som var uønsket, og som var med på å gi resultatet et bias. Denne inndataen var av scriptet telt av to grunner:

1. Uppreis angivelse i opptellingsskript om start- og slutt-tidspunkt for oppgavene, som førte til telling av inndata som ikke hørte med i oppgavene
2. Ytringer eller lyder i løpet av testen som ikke var en spørring til systemet kom som følge av at systemet mangler en god teknikk for å skille reell inndata fra bakgrunnsstøy.

Testmessig kunne en del av disse ”søppelytringene” vært skrelt bort ved hjelp av en nøyaktigere angivelse av hver oppgaves start og slutt, og bare la skriptet telle inndata innenfor. Andelen turns med tale i tabell Tabell 5 ble på bakgrunn av disse manuelle, sporadiske sjekkene vurdert som høyst sannsynlig mer enn at det kunne beskrives som ”ubetydelig”. Det ble derfor valgt å gjøre en ny, manuell inspeksjon av talefilene fra testene, hvor nå kriteriet for å bli telt at brukerens *intensjon* i lydfilen skulle være å kommunisere verbalt med systemet.

Noen av talesignalene som i den reviderte opptellingen ble forkastet, kommer fra periodene mellom oppgavene, hvor nye oppgaver har blitt opplest for brukeren eller hvor det på annen måte har vært kommunikasjon mellom testleder og bruker. Noen av de forkastede audiofilene er opptak av støy uten noen form for muntlige lyder.

5.3.2 *Hvorfor revidert opptelling av simultan koordinerte multimodale turns*

Ved inspeksjon av dialogmanagerens loggfiler og lytting til lydfilene generert av brukerne ble det oppdaget at penndata og taledata i noen tilfeller henter sine timestamps fra ulike klokker. I loggfil bruker3\2006-03-27_11_46_19\main_log_file.xml er turn sju og åtte egentlig en multimodal turn. Denne blir ikke notert som en multimodal turn i loggen, selv om brukeren i lydfilen sier "fra Telenor Fornebu til ", og hvor det tydelig høres et dunk i slutten av fila, som høyst sannsynlig er lyden av pennen som dunker mot den trykksensitive skjermen. Bruker 3 forsøker flere ganger på samme multimodale framgangsmåte, som gjentatte ganger blir logget som to turns, hvor en penn-turn kommer først, etterfulgt av en turn med tale. Denne svakheten fører til en skeivhet i dataene mhp. opptelling av antall turns av de forskjellige modalitetene.

Denne nye opptellingen av simultan koordinerte multimodale turns skjer altså med en ny definisjon av ”simultan koordinert multimodalitet”. *Bruker anvender tale og penn i løpet av én og samme ytring* er definisjonen det nå blir arbeidet med. Merk at definisjonen til forskjell fra definisjonen implementert i systemet, ikke inneholder tidsinformasjon.

5.3.3 Hvordan ble manuell telling gjort

Ovenstående problem med loggene oppfordrer til telling av multimodale turns manuelt. Det ble gjort en manuell traversering av dialog-serverens loggfiler, hvor samtidig lagrede taledata ble kontrollert. To typer ytringer ble vurdert:

- De 15 turnene som systemet allerede kategoriserte multimodale.

For å finne andre multimodale turns ble samtlige taledata deretter målt opp mot disse kriteriene, og vurdert individuelt:

- Når dialogserver-loggen inneholdt en penn-turn umiddelbart etterfulgt av en tale-turn med et start-timestamp fra *før* penn-turnen. Dette indikerer et talesignal med start før pennetrykket og slutt omtrent samtidig med eller etter pennetrykket.
- Tilstedeværelse av et deiktisk ord i setningen
- Når det ikke sies noe deiktisk ord, men det ”mangler” et av konseptene ”buss_stasjon” eller ”buss_stasjon2”, og det i stedet for dette kan høres en karakteristisk kneppe lyd (noe som er tilstede på de fleste opptak av multimodale turns), antas dette å være pennens trykk mot klientens skjerm.

Antallet penn-turns fra skriptets opptelling fikk naturligvis subtrahert gruppen av nye multimodale turns som i loggen ble ført opp som to adskilte monomodale turns.

5.3.4 Revidert resultat av bruk av ulike modaliteter

Bruker	Penn	Tale	MM
1	ble ikke fullført		
2	ble ikke fullført		
3	3	1	12
4	11	6	4
5	10	4	1
6	12	7	0
7	ble ikke fullført		
8	11	8	0
9	11	2	0
10	2	19	7
11	7	8	0
12	2	12	2
13	8	12	0
14	13	8	0
15	11	7	4
16	ble ikke fullført		

17	10	13	0
18	ble ikke fullført		
19	15	18	0
20	ble ikke fullført		
21	7	6	0
22	0	11	0
23	1	7	2
24	ble ikke fullført		
25	2	6	4
26	4	9	0
27	2	2	6
28	10	10	3
29	5	6	3
30	7	14	0

tabell 8: Revidert andel inndaturns, sortert per bruker og modalitet

OPPGAVE 1:		Antall	Prosent
Antall Penn-turns		60	34,7 %
Antall Tale-turns		98	56,6 %
Antall MM turns		15	8,7 %
Totalt antall turns		173	100,0 %
OPPGAVE 2:		Antall	Prosent
Antall Penn-turns		69	43,7 %
Antall Tale-turns		67	42,4 %
Antall MM turns		22	13,9 %
Totalt antall turns		158	100,0 %
OPPGAVE 3:		Antall	Prosent
Antall Penn-turns		55	42,0 %
Antall Tale-turns		61	46,6 %
Antall MM turns		15	11,5 %
Totalt antall turns		131	100,00 %
I ALT		Antall	Prosent
Antall Penn-turns		184	39,8 %
Antall Tale-turns		226	48,9 %
Antall MM turns		52	11,3 %
Totalt antall turns		462	100,0 %

Gjennomsnittlig antall turns		
oppgave 1	173/28	6,18
oppgave 2	158/26	6,08
oppgave 3	132/23	5,74

28 brukere gjennomførte oppgave 1.
26 brukere gjennomførte oppgave 2.
23 brukere gjennomførte oppgave 3.

hele testen		17,99
-------------	--	-------

tabell 9: oppsummering, revidert andel turns, sortert på oppgave og modalitet

Interaksjonsmønster for de 12 brukerne med 0 multimodale turns: Penn 47,51%, tale 52,49%.

5.4 NÅR TRYKES DET I EN MULTIMODAL YTRING?

5.4.1 Liste over registrerte ytringer i multimodale turns

På grunn av at timestampet for taleinteraksjonen i loggen ikke var presist, kunne ikke dette brukes til å stadfeste eksakt når pennetrykket kom i forhold til talesignalet start. Eksempel: I loggen står turn seks oppført med timestamp 2006_03_27_11_47_38_951. I turn sju står penn-turnen oppført med timestamp 2006-03-27_11_47_46_492, mens turn åtte står oppført med tale som interaksjon og med timestamp 2006-03-27_11:47:05.162. Merk at formatet på siste turn er forskjellig fra resten. Disse tidspunktene er omkring 41 sekunder fjernt fra hverandre, mens audiosignalet bare er en brøkdel av denne lengden. Kanskje kan dette tilbakeføres til forskjellige referanseklokker i forskjellige deler av systemet.

I oppstillingen av fra brukers side multimodale ytringer er det forsøkt å plassere piler så presist som mulig hvor trykket forekom. Presisjonen er forsøkt gjort så nøye at om pilen står over starten på et ord, forekom trykket idet brukeren startet på dette ordet. Om brukeren trykket i slutten av sist uttalte ord eller ett sekund etter, er dette forsøkt angitt med pilen hhv. på og med noe avstand fra siste ord. Pauser i talen er forsøkt antydnet med åpenrom i setningen. Setningene følger den foreslåtte standarden for transkribering av spontan tale foreslått av BRAGE-prosjektet (Markhus et al, 2004). Setningene står oppført med ytringene fra første bruker først, og ender med ytringene fra siste bruker.

Oppgave 1

↓
Fra til Telenor Fornebu
↓
Fra til Telenor Fornebu
↓
Fra til Telenor Fornebu
↓
Fra Telenor Fornebu til
↓
Prøver vi en gang til da ... Fra Telenor Fornebu til
↓
Fra Telenor Fornebu til
↓
Fra Telenor Fornebu til
↓
Fra Telenor Fornebu til
↓
Fra Telenor Fornebu
↓
Neste buss herfra til Telenor Fornebu
Neste buss herfra til Fornebu
Neste buss fra her til Fornebu
↓
Til Fornebu fra
↓
Når går neste buss fra til Telenor Fornebu
↓
Når går neste buss fra Telenor Fornebu til Harald Hardrådes –

Oppgave 2

Fra til Jernbanetorget

↓

Fra til Jernbanetorget

↓

Fra til Jernbanetorget

Neste buss herfra til Jernbanetorget

Neste buss [spk-f] herfra [spk-f] til Jernbanetorget

Buss til Jernbanetorget fra denne stasjon

Buss fra [spk-e]

↓

Til Jernbanetorget fra denne stasjonen

Til Jernbanetorget fra [spk-e]

↓

Fra her til Jernbanetorget

↓

Fra [spk-e] her til Jernbanetorget

Når går neste buss herifra til [spk-e] Jernbanetorget

↓

Når går neste buss herifra til Jernbanetorget

↓

-este buss fra til Jernbanetorget

↓

Neste buss fra til Jernbanetorget

↓

-ste buss fra til Jernbanetorget

↓

Neste buss fra til Jernbanetorget

↓

-r går neste buss fra til Jernbanetorget

↓

Når går neste buss herifra til Jernbanetorget

↓

Når går neste buss herifra til Jernbanetorget

↓

Når går neste buss herifra til Jernbanetorget

Neste buss fra her

Oppgave 3

Fra ↓
Fra ↓ til Jernbanetorg-
Fra ↓
Til ↓
Fra Stortorget til
Fra Stortorget til
Når går neste buss herifra til Jernbanetorget
Når går neste buss herifra til Jernbanetorget
Neste buss fra ↓ til Jernbanetorget
Når går neste buss fra Jernbanetorget til ↓
Til ↓ Observatoriegata
Neste buss fra ↓ til Observatoriegata
Neste buss fra ↓ til Observatoriegata
Neste buss fra Oslo Domkirke til hit
Neste buss fra Jernbanetorget til hi-

5.4.2 Om konsepter i en multimodal ytring

Antall konsepter: 183

Antall ytringer: 55

Gjennomsnittlig antall konsepter per ytring i en fra brukers ståsted simultan koordinert multimodal turn: 3,33

Konseptene som ble telt var alle konseptverdier tilstede i gramatikken, i tillegg til penneklikk. En konseptverdi blir telt to ganger om den blir funnet i penn- og tale-modaliteten. Dette ettersom det søkte tallet på gjennomsnittlig antall konsepter per ytring forsøker å tallfeste kompleksiteten i en multimodal ytring.

5.5 ASR- FEILRATE

5.5.1 Hvordan ble tellingen gjort

I fastsettelsen av feilraten for ASR-modulen ble de originale lydfilene gjennomlyttet og sammenlignet med resultatet på utgangen av modulen. Det ble avgjort å måle antall innsnakkede konsepter og måle feilraten ved å se på hvor mange som ikke ble riktig oppfattet. Det ble avgjort å ikke spesifisere dette nærmere i *hvilke* konsepter dette gjelder, selv om det kanskje kan gi ekstra informasjon. Grunnen er den økte arbeidsmengden dette vil medføre. Alle konsepter listet i gramatikken ble dermed valgt å bli telt opp. Om brukeren i ytringen skulle komme til å repetere noen av konseptene flere ganger, telles også repeteringene. I testen var oppsettet slik at en konfidens på 500 eller mer var nødvendig for konseptet for å bli akseptert.

Om systemet skulle komme til å prøve å dekode tale som ikke var ment som inndata for systemet for eksempel når bruker snakker til testleder, blir dette ikke regnet med i gjenkjenningssraten. Dette fordi det er gjenkjenningssrate på inndata-ytringer som er i fokus i denne tellingen. Det er uinteressant i akkurat denne tellingen hvor god taledetektoren er, og hvordan den eventuelt påvirker resultatet. At brukeren ikke forholder seg nøyaktig til konseptverdiene i gramatikken, men gir som inndata deler av en konseptverdi, som for eksempel "Fornebu" i stedet for "Telenor Fornebu", blir regnet som en gyldig ytring, hvis ikke ytringen er så ufullstendig (kuttet av i en ende eller lignende) at det vil være vanskelig for både menneske og maskin å forstå meningen. Hvis en verdi i buss_stasjon2 blir gjenkjent som en verdi i buss_stasjon, blir det vurdert som feil, selv om dialogmanageren i visse tilfeller kan løse slike feil. Konseptet "garbage" ble også telt. Ugyldige konseptverdier som for eksempel "til Sentralstasjonen" ble ikke telt som et konsept, ettersom sjansen for å dekode dette riktig er lik null.

Det må medregnes i tallene noe usikkerhet, ettersom arbeidet ble manuelt gjort, og det var relativt store mengder informasjon.

Det ble telt på følgende måte: Setningen ”Når går neste buss fra Harald Hardrådes Plass til Telenor Fornebu” vil gi 6 konsepter: Når (temporal), går (garbage), neste buss (temporal), fra (temporal), Harald Hardrådes Plass (buss_stasjon), til Telenor Fornebu (buss_stasjon2). Opptellingen skjedde ut fra fila vaskaLogg.txt, ei fil generert av perl-scriptet ”anal_log_2.pl”. Konseptene listet opp i fila er tatt fra loggens ”ConceptChoice”-tagg, slik at det ennå ikke er blitt gjort noe valg på hvilket konsept som skal foretrekkes ved mer enn ett tilstede. Vet man i tillegg konfidensterskelen som er satt for konseptene, kan man sammenligne med audio-filene som er lagret fra TaleServer og høre hvilke konsepter som virkelig var inndata, og utarbeide en feilrate på bakgrunn av dette. Slik ble ASR-feilraten utarbeidet. ASR-feilraten sier altså hvor god talemengdegjennakeren er ved bruk av en ”perfekt taledetektor”.

5.5.2 ASR-rate resultater

Gjenkjenningssrate med konseptkonfidens = 500 som grense

Bruker	Antall konsepter i ytring	Riktig gjenkjent	Suksessrate	Feilrate
1	7	6	85,71 %	14,29 %
2	-	-	-	-
3	21	12	57,14 %	42,86 %
4	27	21	77,78 %	22,22 %
5	11	2	18,18 %	81,82 %
6	12	10	83,33 %	16,67 %
7	23	14	60,87 %	39,13 %
8	17	6	35,29 %	64,71 %
9	6	5	83,33 %	16,67 %
10	54	28	51,85 %	48,15 %
11	21	12	57,14 %	42,86 %
12	23	17	73,91 %	26,09 %
13	25	8	32,00 %	68,00 %
14	29	19	65,52 %	34,48 %
15	38	11	28,95 %	71,05 %
16	0	0	0,00 %	100,00 %
17	12	6	50,00 %	50,00 %
18	-	-	-	-
19	75	38	50,67 %	49,33 %
20	15	7	46,67 %	53,33 %
21	10	7	70,00 %	30,00 %
22	41	22	53,66 %	46,34 %
23	29	23	79,31 %	20,69 %
24	9	3	33,33 %	66,67 %

25	29	17	58,62 %	41,38 %
26	12	11	91,67 %	8,33 %
27	35	14	40,00 %	60,00 %
28	33	26	78,79 %	21,21 %
29	24	19	79,17 %	20,83 %
30	28	17	60,71 %	39,29 %
Sum	666	381	57,21 %	42,79 %

tabell 10: ASR-suksessrate

5.6 RESULTATER FRA SPØRRESKJEMA

5.6.1 spørsmål 1

”Dette er det jeg tenker på nå/umiddelbar respons:”

I dette spørsmålet kom det en mengde tilbakemeldinger. På grunn av spørsmålets åpne form er det gjort en samling av noen tema som gikk igjen, som gjengis her. Se ellers vedlegg i kapittel 10.2 for komplett liste.

Noen av punktene som ble nevnt gikk på (uten angivelse av antall svar bak hvert punkt):

- Positivt med valgfrihet i spørreform – greit å kunne skifte avh. av oppgavens art
- Lang responstid
- Ustabil system - omstarter
- Talegjenkjenning – noen fornøyd, andre misfornøyd
- GUI. Ikke gode kart, mangler vesentlige kartverktøy, for eksempel panoreringsmulighet. Savn av indikatorer for om F- eller T-knapp er trykket. Savn av prosesserings-indikator. Omstendelig å bruke penn-løsning.
- Selve idéen med bussinfo
- Merknader om talesyntesen

5.6.2 Spørsmål 2

”Kan du tenke på andre situasjoner enn bussruteinformasjon hvor multimodalitet kan være nyttig?”

- Nei (5)
- Bestille noe. Tale vs 25-digit refnr
- Andre transportsystem (3)
- Nettsøk
- Turistinfo/turistattraksjoner (4)
- Åpningstider, butikker
- Bil: navigasjonssystem

- System til å stå på bussholdeplasser. -> mer enn bare bussruter
- Geografisk/kultur/butikk-orienteringssystem (7)
- Styre ting hjemme
- Billettbestilling (2)
- Eldre, hjemmetjenester
- Nummeropplysning (2)
- PC-bruk
- Leger. Folk som trenger hendene frie.
- Finne nærmeste bar. Hjelp til pub-til-pub-runden innenfor skjenkeringen (stiller krav til ASR-delen)

5.6.3 Spørsmål 3

”Hva var mest naturlig for deg å bruke – tale? Peking? Peik&Preik? Evt situasjonsbetinget?”

”Tale”: 4

- ”På grunn av nysgjerrighet”
- ”Når den fungerer. Hvis den blir mer robust”
- ”Fortest med tale. Er usikker på kartet”
- Uten kommentar: 1

”Peking”: 9

- ”I utgangspunktet peking, men ville få testa ASR her.”
- ”med dette systemets forutsetninger. Kanskje noe annet om 3 år”
- ”Lyst til å bare trykke. Men enklere å prate”
- ”Pek. Men kjører heller tale enn å gå ut av subkartet”
- ”Hvis man er ute”
- Uten kommentar: 4

”P&P”: 5

- ”pga treg gui (zooming)”
- ”Peik&Preik eller tale. På grunn av at man i forkant ikke alltid vet hvor holdeplassen ligger”
- Peker mens jeg prater, men ikke samtidig
- ”Peik&Preik best. Hvis du vet hvor du skal, men ikke veit hvor start er. Tale forutsetter kunnskap om navn.”
- Uten kommentar: 1

”Situasjonsbetinget”: 10

- ”Avh. av om kjent stasjonsnavn. Navn -> tale. Generelt kortest tid.”
- ”Kommer an på situasjonen. Kommer an på om info om [holdeplass] A og B er tilstede”
- ”Penn. Men å venne seg til talen fører til kortere tid. Penn er dårlig implementert”
- ”Situasjonsbetinget, men tale er førstevalg alltid”
- ”Lett å uttale. Ved dårlig gjenkjenningssrate velges trykking”

- ”Er man kjent i Oslo er det lettere å prate.
- ”Ved mye trafikk: peike. Nå: preike. På bussen: peike. Peike mer knotete enn preike”
- ”Hvis ukjente holdeplasser: peking. Hvis stasjonene er kjent: Trykke hit og si holdeplass 2. Så lite zoom som mulig. Det er tungvint.”
- Uten kommentar: 2

Svarene på dette spørsmålet var i stor grad betingede svar, selv fra mange av dem som ikke svarte ”situasjonsbetinget”. Det ble i de fleste svar sagt ønsket modalitet først, etterfulgt av situasjoner hvor brukeren enten ville valgt annerledes eller etterfulgt av en grunn for hvorfor han eller hun valgte som det ble gjort, basert i det utprøvde systemet.

28 personer besvarte dette spørsmålet.

5.6.4 Spørsmål 4

”Er du i noen grad villig til å snakke normert for å øke suksessraten til stemmegjenkjenneren når denne ønskes brukt? I hvor stor grad?”

Ja: 27

- ”ikke uvillig. Gaten i stedet for gata.”
- ”gjorde nok det allerede” (2 pers)
- ”absolutt”
- ”Jo; tydeligere. Langsommere”
- ”Villig, men det er et irritasjonsmoment. Særlig når det er en maskin”
- ”Ja. Kan godt ofre. Men naturlig tale er å foretrekke. Slitsomt å tilpasse seg”
- ”Må jo, liker det ikke. En av de bedre stemmegjenkjennerne jeg har prøvd”
- ”nokså stor”
- ”villig”
- ”Prøvde også dette.”
- ”Gjør det automatisk”
- ”Fungerte allerede i veldig stor grad for min stemme”
- ”men ikke i stor grad”
- ”Ja, klart.”
- ”Ja, mye.”
- ”Middels på skalaen liten-middels-stor”

Nei: 0

Ikke valgt: 1

28 personer besvarte dette spørsmålet.

5.6.5 Spørsmål 5

”For talemодuset: Hvordan vurderer du å bruke noen minutter til høytlesing for å opprette en personlig profil, om du vet at gevinsten er økning i gjenkjenningsrate?”

Ja: 24 (88,89%)

- ”5 minutter?”
- ”3-5 minutter”
- ”5 minutter. Ikke en halv time.”
- ”Aktuelt.”
- ”Ja. En halv time”
- ”Et halvt minutt”
- ”tre minutter.”
- ”Gjerne et time”
- ”Ja, noen minutter. Innen 15 minutter.”
- ”Ja. Så lite som mulig. Kanskje ti minutter.”
- ”Til det fungerer bra. Halvtime? Mye tid.”
- ”Inntil 5 minutter”
- ”Sikkert med på det. Ett minutt burde være nok. Par minutt maks.”
- ”Ja, villig. Ett til to minutter. Ikke særlig lengre.”

Nei: 2 (7,41%)

- ”Ikke stor interesse”
- ”liten. Vanskelig å si.”

Ikke svart: 1 (3,70%)

Tot: 27 svar

5.6.6 Spørsmål 6

”Liker du å bruke talegrensesnittet? Hvis ja: Fortsetter du å bruke det om antall korrekte svar er...

- 9 av 10
- 4 av 5
- 2 av 3
- 1 av 2
- Annet:”

Det ble i løpet av testen flere ganger bemerket at svaralternativene var vanskelige, og de ble derfor forandret i flere omganger. Dette gjør det upassende å sette sammen tallene i et felles tallmateriale.

Nei:

3 pers

Ja:

Alternativer gitt:

9 av 10, 4 av 5, 2 av 3, 1 av 2

Stopper å bruke tale på:

4 av 5: 1 pers

2 av 3: 3 pers

Nye alternativer gitt: 9 av 10, 4 av 5, 3 av 4, 2 av 3, 1 av 2, annet

Stopper å bruke tale på:

4 av 5: 1 pers

3 av 4: 6 pers

2 av 3: 1 pers

Nye alternativer gitt: 9 av 10, 8 av 10, 7 av 10, 6 av 10, 5 av 10, 4 av 10, annet

Stopper å bruke tale på:

9 av 10: 1 pers

7 av 10: 3 pers

6 av 10: 6 pers

5 av 10: 2 pers

4 av 10: 1 pers

Tot: 28 pers

5.6.7 Spørsmål 7

”En ting du likte godt:”

- Peik og Preik (4)
- Når den fungerte, fungerte den fint. (2)
- Mulighet for tale (2)
- Hurtigere mtp navigering med Peik&Preik
- Mute-knappen
- Konseptet med bussinformasjon
- Kartinterfacet
- Å kunne lese bussholdeplassnavn uten å måtte sette i gang en spørring
- ”Snakkende elektronisk rutetabell”
- PDAens indiske aksent
- Mulighet for å kunne velge [interaksjonsform. forfatters anmerkning]

- Mulighet for å kunne angi punkt A og B i to omganger
- Kult med ny datating
- Oversiktskartet
- Idéen
- Gjenkjenningssrate

Tot: 21 pers

5.6.8 Spørsmål 8

”Det du var minst fornøyd med: ”

- Dårlige kart/GUI (5)
- Responstid (4)
- Systemets generelt uferdige tilstand (4)
- Gjenkjenningssraten (2)
- Ingenting (2)
- ”Den ble ført litt utålmodig”
- ”Vippetangen” var ikke med
- Ikke automatisk å prate

Et nedskrevet svar lot seg ikke forstå ved analyse av svarene, på grunn av hastig notering.

Tot: 21 pers

6 DISKUSJON

6.1 "FØRSTE LINJES" OBSERVASJONER AV RESULTATENE

6.1.1 "Første linjes" observasjoner av resultater i kapittel 5.1 og 5.3

Både fra telling med script og fra revidert telling er det en høy andel tale. I opptelling med script er andelen tale så høy som 61,85%, i revidert opptelling 48,9%. Penn-interaksjon er i begge tilfeller noe under, med 35,79% i script-opptelling, 39,8% i revidert opptelling. Andel multimodal interaksjon stiger fra en lav prosentandel ved script-opptelling på 2,35% til 11,3% i den reviderte versjonen.

Hvis man tar for seg bruken av modalitetene i hver oppgavene, ser man følgende. I første telling viser oppgave 1 omkring to tredjedelers bruk av tale og en tredjedels bruk av penn. Oppgave 2 viser en oppgang for penn til 40,00% og en nedgang for tale, mens det blir telt litt over 3% multimodale turns. Dette tallet holder seg for oppgave 3, men penn synker til 32,99% og tale stiger til 63,92%. I andre, reviderte opptelling er andelen tale generelt redusert, til fordel for høyere tall for penn og multimodale turns. Oppgave 1 skiller seg noe ut fra andre to, med en andel tale på 57%. Multimodalitet ble brukt minst her – 8,7%. Oppgave 2 og 3 ligger ikke så veldig langt fra hverandre mtp. andelen turns av de ulike sortene – penn viser hhv. 43,7 og 42%, tale ligger mellom 42 og 47 prosent, multimodale turns 13,9 og 11,5 prosent – altså en liten topp på oppgave 2.

Antall totale turns brukt i de forskjellige oppgavene er nokså lik, og ligger på ganske nøyaktig 6 turns, med en varians på xx for for de ulike oppgavene. [*kunne* ha angitt varians for antall turns på de ulike oppg] Det finnes en *liten* reduksjon i turngjennomsnittet fra oppgave 1 til oppgave 3, men denne er såpass liten at det med det lave antallet brukere ikke kan sies om er en trend eller tilfeldig.

Når det gjelder multimodale turns, var det bruker 3 som benyttet seg oftest av dette. Brukeren hadde til sammen 4 monomodale turn, mens hele 12 var multimodale. Brukeren hadde ikke spesielt lang sesjon. Bruker 10 var nest mest aktiv til å bruke multimodale turns med 7 stykk. Bruker 10 står oppført med den lengste sesjonen av brukerne. Til sammen står disse to brukerne for 36,5% av registrerte multimodale turns. 12 brukere ligger på samlet sisteplass i bruk av multimodalitet – til sammen 0 turns er kreditert denne gruppa. Disse fikk som alle andre demonstrert multimodalitet i introduksjonen. Om brukerne testet det i løpet av egen tid før testen er ikke kjent. I denne gruppa fordeler ellers interaktiviteten seg på penn med 47,51% og tale med 52,49%.

Vi ser at totalt antall taleturns går ned fra 394 til 226 turns, en nedgang på 42,64%. Blant brukerne varierer det sterkt hvor mange turns som blir tatt bort. Den store forskjellen kan for en stor del føres tilbake til om og hvor aktivt brukerne brukte MUTE-funksjonen.

6.1.2 "Første linjes" observasjoner av resultater i kapittel 5.2

Det kan ses i resultatene for opptellingen av dialogene at en stor del er sekvensielt multimodale. Nesten halvparten – 48,1% - av dem inneholder både tale og penn. På tross av at brukeren i oppgavene har forskjellig forhåndskunnskap, er andelen tilnærmet uforandret, med laveste andel i oppgave 2: 46,3 %, høyeste andel oppgave 1; 50,0 %. Dette er et atskillig høyere tall enn tallet for simultan koordinerte turns som er registrert. Blant brukerne er det stor forskjell på hvor mye sekvensiell multimodalitet er brukt. Noen har ingen eller svært få multimodale turns i løpet av testen, andre har de aller fleste eller samtlige, og andre igjen ligger på gjennomsnittet.

6.1.3 "Første linjes" observasjoner av resultater i kapittel 5.4

Det kan ses fra oversikten over multimodale setninger at det er noen flere registrert for oppgave 2 (22 stk) enn i de andre to (hver 15 stk). Det er totalt 52 multimodale turns. 36 av disse har blitt merket med tidspunkt for pennetrykk. Etersom setningene står gruppert bruker for bruker nedover, kan det ses at noen brukere har repetert en ytring flere ganger. (dette burde absolutt vært nummerert!). Her følger 4 ulike måter å systematisere observasjonene på:

1. Prates det idet det trykkes eller tier brukeren?
Brukeren tier: 25 (69,4%)(herav trykk før første ord: 3, trykk etter siste ord: 8, trykk mellom to ord: 14)
Brukeren prater: 11 (30,5%)
Kun setninger med piler ble vurdert. Mange av ytringene hvor det prates samtidig som det trykkes, er ytringene som inneholder deiktiske ord, og trykket kommer ofte på disse ordene. Et tydelig system i setningene hvor brukeren er stille akkurat mens det trykkes, er at trykket kommer på stedet i setningen hvor det ville vært naturlig å si et deiktisk ord.
2. Antall setninger med/uten deiktiske ord: 19/33 (36,5% / 63,5%). Alle setninger var med i vurderingen.
3. Ytringer med samme informasjon i to kanaler: 2 ("Når går neste buss fra [trykk] Telenor Fornebu til Harald Hardrådes –", og "Til [Trykk] Observatoriegata")
4. "alt-på-en-gang": I hvor mange av de multimodale ytringene prøver brukeren å fylle alle de semantisk relevante konseptene buss_stasjon, buss_stasjon2 og temporal i løpet av en enkelt turn:

Oppgave 1	samtligte ytringer	100%
Oppgave 2	21 av 22 ytringer	95,45%
Oppgave 3	elleve av 15 ytringer	73,33%

tabell 11: Forsøk på å gi inn all semantisk relevant informasjon i løpet av den multimodale ytringen¹⁶.

Det observeres at blant setningene er det nokså stor variasjon i antall ord ytret og hvor i setningen pennetrykket er. Noen av setningene er hele setninger, mens andre er kun et par ord. Pennetrykket kommer av og til som et tillegg til ytringen, av og til ser det ut som pennetrykket erstatter et (deiktisk) ord.

Det ble observert av testleder under testene at flere av de multimodale ytringene som har trykk *på* et ord, er repetisjoner av en tidligere ytring som ikke lyktes. Det ble sett flere eksempler på at en bruker ytret en setning hvor han tiet i det øyeblikket det ble trykket, og når denne ytringen feilet, ble den gjentatt, men denne gangen med trykket *på* et av ordene.

I oppgave 3 ser det ut til at en del av de multimodale turnene har kommet etter at brukeren i en fase har lett etter bussholdeplassnavn *uten* å markere fra- eller tilholdeplass, og så etterpå gått over til å gjøre forespørselen. Oppgavens art krever leting etter bussholdeplasser på to ulike subkart. En sekvensiell framgangsmåte vil være fordelaktig, og det er derfor overraskende med andelen simultane koordinerte turns. De simultant koordinerte ytringene i oppgave 3 er likevel brukt i større grad i denne oppgaven (fire av 15 multimodale ytringer inneholder ikke alle tre nødvendige konseptene mot 1 og ingen i de to første oppgavene) som ledd i en sekvensiell framgangsmåte for å gi inn nødvendige konseptene.

For dataene er det et stort forbehold: Disse ytringene er ikke representative ettersom flere av setningene er ytret av samme bruker etter hverandre, og mange brukere ikke er representert.

6.1.4 "Første linjes" observasjoner av resultater i kapittel 5.5

Gjennomsnittlig ASR- konseptgjenkjenningssrate fra 666 inntalte konsept, med 28 forskjellige brukere, ligger på 57,21% når kravet for gjenkjenning av et konsept ligger på en konseptkonfidens på 500. De tre brukerne med høyest antall inntalte konsepter hadde en gjenkjenningssrate på hhv. 50,67% (bruker 19, 75 konsepter), 51,85% (bruker 10, 54 konsepter) og 53,66% (bruker 22, 41 konsepter). Det finnes gjenkjenningssrater med både høye og lave verdier: bruker 26 – 91,67% gjenkjent, 12 konsepter, bruker 5 – 18,18% gjenkjent over 11 ytrede konsepter. Disse har imidlertid relativt lite grunnlag for beregning av rate, og tallene må betraktes med forbehold.

¹⁶ Det ble i denne tellingen telt med *alle* opplistede turns. Det betyr at det for setningene hvor ingen peking er notert, men hvor setningen tyder på at dette ble gjort, er pennetrykket telt med.

6.2 VIDERE DISKUSJON AV UTVALGTE TEMA

6.2.1 *Definisjonen av en simultan koordinert multimodal turn*

Testresultater for simultan, koordinert multimodalitet fra opptelling med script av dialogserverloggene viser lave tall for samtlige oppgaver, med 0,88% i første oppgave og hhv 3,26% og 3,09% for oppgave 2 og 3. Gir opplysningene som systemet holder på samt valgmulighetene som systemet tilbyr i sin nåværende form, for liten ”gevinst” for at den jevne bruker skal finne simultan koordinert multimodalitet verdt å bruke? Er applikasjonen som testen er blitt utført på, for enkel for den noe krevende interaksjonsmetoden som simultan koordinert multimodalitet? Dette er et spørsmål som springer fram når man ser på argumentasjonen for multimodalitet gitt i brosjyren ”MUST – multimodal and multilingual services for small mobile terminals” (Boves og den Os, 2002, s.10) som ble referert i innledningskapitlet.

Som det ble referert til i argumentasjonen for en revidert, manuell opptelling i kapittel 6.3, eksisterer det en rekke taleturns hvor det i løpet av ytringen også har blitt benyttet penn, som ikke er blitt registrert som simultant multimodale. En transkribering av alle simultan koordinerte turns i testen basert på lytting (kapittel 5.4) viser at det ikke er noen åpenbar forskjell på registrerte og uregistrerte simultan koordinerte turns. Dette taler for at kriteriet for merking av simultan koordinerte turns, dvs. det implementerte fusjonskriteriet¹⁷, er noe unøyaktig. Det er i den reviderte opptellingen i oppgaven (kapittel 5.3) og utover tatt utgangspunkt i brukerens ståsted når det telles simultan koordinert multimodalitet.

Fusjonsteknikken som er implementert i nåværende demonstrator er en teknikk som fungerer greit for korte talesignal. Når talesignalet øker i lengde, er følgen at kun ytringer med pennetrykket mot slutten av setningen har sjansen til å bli merket multimodale. Om trykket i en lang ytring kommer tidlig, ankommer meldingen med pennetrykket multimodal server tidlig før taleytringen er fullført (og prosesseringstid mellom klient og MM-server kommer i tillegg), og umuliggjør fusjon. Den multimodale plattformen ble opprinnelig utviklet for en turist-opplysningsapplikasjon, ”MUST” (Boves og den Os, 2002). Kanskje var ytringene for denne applikasjonen noe kortere, slik at implementasjonen av den gjeldende fusjonsmekanikken fungerte etter hensikten. Men med bakgrunn i det store antallet ytringer som er multimodale fra brukers side testen, og som ikke blir merket som dette (37 av 52 – 71,15%), pekes det på at en utvikling av fusjonsteknikken kan være formålstjenlig. En ny implementasjon som gjør plattformen i stand til å fusjonere tale av ulike ulik lengde og trykking som varierer fra tidlig til sent i ytringen, antas å kunne gjøre fusjonen mer robust og applikasjonsuavhengig.

6.2.2 *Diskusjon av andelen sekvensiell multimodalitet*

At sekvensiell multimodalitet ligger på rundt 50% for alle tre oppgavene er overraskende. Det ble ventet en noe lavere verdi for oppgave 1, ettersom dette er en oppgave som raskest løses med kun tale. At verdien er såpass jevn for alle tre oppgavene gjenspeiler

¹⁷ I dette avsnittet menes det den temporale fusjonen når det snakkes om fusjon.

kanskje det faktum at gjenkjenningsraten gjør at brukeren går over til penn etter noen gangers feiling, eller at mange av brukerne brukte fordelen med multimodalitet i flere av oppgavene til å kun finne navnet på én av busstasjonene, og trykke på ikonet på den andre. Noen brukere kan ha foretrukket å vente på resultatet fra en turn, for så å gi inn et til – altså en bevisst sekvensiell tilnærming for å øke graden av kontroll over en ny, ukjent applikasjon.

6.2.3 Antydninger mhp kognitiv anstrengelse

Det var i alt 69,4% (25 av 36) av turnene hvor pennetrykket var merket av, hvor brukeren tiet ved trykking. Noen av ytringene hvor det er trykket samtidig som det er snakket, er også ytringer som er identiske repetisjoner av tidligere ytringer. Sammen kan dette peke på at det skjer en omfordeling av fokus fra prat til håndtering av pennen, og dette ser ut til å få brukeren til å stanse idet pennen trykkes. Det kan kanskje se ut som repetisjoner av allerede tidligere multimodale ytringer krever mindre kognitiv anstrengelse.

6.2.4 Mye bruk av tale

Det er i revidert opptelling av modalitetene en andel på 48,9% tale. Dette må betraktes som et høyt tall, tatt i betraktning at flesteparten av brukerne var fra før vant til kommunikasjon med datamaskiner ved hjelp av pek og klikk, med mus og tastatur, som normalt ville gjøre pennen til ønsket interaksjonsredskap. Et sitat fra en av brukerne fra en situasjon hvor personen måtte gå over til bruk av penn, forteller noe om hvordan denne personen tenkte på forholdet mellom bruken av tale og penn:

”Nei, det var jo greit å gjøre det med stemmen, men da får jeg vel gjøre det manuelt.” (bruker 5)

At brukeren ser på trykking som manuelt, indikerer at brukeren har en forståelse av bruk av tale som mer *automatisk*.

En opplagt fordel med tale som modalitet som forklarer noe av denne høye raten er sjansen til å få gjenkjent alle konseptene i en og samme turn og få tilbakemelding med en gang. Den høye raten, om man ser på ASR-raten samtidig, betyr at det sannsynligvis har vært en del repetering av ytringer. Dette ble observert av testleder under testene. En annen mulig forklaring er at det ikke har vært mange muligheter tidligere for mange av deltakerne til å prøve automatisk talegjenkjenning, og at mange av dem så på dette som en sjanse til å prøve noe nytt. Det er altså ikke sikkert at dette er representative verdier for en tenkt daglig bruk.

6.2.5 Simultan koordinert multimodalitet – en ”alt på en gang”-løsning

Om en av testpersonene brukte simultan koordinert multimodalitet, var det for å kunne gjøre hele dialogen på én enkelt turn. Simultan koordinert multimodalitet ser ut som en høyst ”alt-på-en-gang”-framgangsmåte. tabell 11 viser at dette var tilfellet for alle ytringene i oppgave en og 21 av de 22 ytringene i oppgave to, og 11 av 15 ytringer i oppgave tre. Brukeren sier hele forespørselen sin i 90,38% av forekomstene av simultan koordinert multimodalitet. Selv om brukeren i oppgave to og tre i starten av oppgaven mangler informasjon om ett eller flere konsepter, holder dette tallet seg meget høyt.

Brukerens framgangsmåte har i disse tilfellene altså vært å først lete seg fram til de manglende konseptene, og deretter spørre systemet gjennom en multimodal turn. Ved dette resultatet må det understrekes at i denne applikasjonen er ikke veien lang for å gjøre en ”full” ytring (dialogmanageren kan sette i gang databaseoppslag) når en multimodal ytring gjøres – en multimodal turn inneholder minst to konsepter, og minimum for en full ytring er tre konsepter.

6.2.6 Diskusjon av introduksjonen

Presentasjonen av interaksjonsformen *tale* blir i etterkant vurdert som godt presentert, ettersom det i løpet av testen viste seg at det ikke var nødvendig å endre noe på den. Den vurderes også som å ha oppnådd å vise det essensielle for å kunne bruke talemødaliteten. Presentasjonen av interaksjonsformen *trykking* viste seg å ha en svakhet fordi den ikke demonstrerte bruk av zoom-funksjonen, som muliggjør avreise og ankomst i forskjellige subkart. Dette skapte problemer for flere brukere. Zoom-funksjonen viste seg å heller ikke være selvforklarende for flere brukere. Multimodaliteten som ble vist i introduksjonen er synkron, men sekvensiell multimodalitet ble ikke vist. Dette kunne også vært brukt for å få frem dette. Det hadde ikke vært nødvendig for testlederen å si ”herifra” ved trykking på skjermen for å angi avreiseholdeplass, og dette kunne ha blitt forklart.

6.3 ASR-RATE DISKUSJON

Generelt gikk det fram av manuell inspeksjon av dialogserverloggene/taleserverloggene at lengre ytringer hadde lavere gjenkjenningsprosent enn kortere. En veldig ofte forekommende feil var ”fra” gjenkjent som ”ja”. ”Filler”-konseptverdien ”går” ble sjelden riktig gjenkjent, ettersom brukerne hadde en tendens til å uttale utydelig. En ofte forekommende konstellasjon var ”når går neste buss...”, hvor mange brukere uttalte ”når” og ”neste buss” med trykk, mens ”går” ble trykklett uttalt og i enkelte tilfeller nesten slukt.

Det er påfallende mange ganger at Voice Activity Detector har blitt aktivert på grunn av tale i bakgrunnen, fra samtale mellom bruker og testleder. Det er enda mange flere audiofiler som systemet korrekt har forkastet som søppelfiler, etter at lyder aktiverte klientens mikrofon. Det ble ikke gjennomgått systematisk alle søppelfilene, men tatt sporadiske tester på hva innholdet var. Det ble ikke i noen funnet tale ment for dekodning. Dette peker på at terskelen for aktivisering av dekoderen kan forskyves noe oppover i decibel.

Det er ved et veldig stort flertall av brukerne som ytrer lyder som blir forvrengt i audioopptaket. Dette gjelder ustemte lyder som ’s’-lyd, ’f’-lyd, ’kj’-lyd og ’skj’-lyd, og det gjelder plosiver. Dette påvirker gjenkjenningsraten. Avstand fra mikrofonen, mikrofonens utforming, brukerens taleretning og volum og mikrofonens sensibilitet er blant de faktorene som bidrar til dette. Det er sannsynlig at gjenkjenningsraten påvirkes av dette i merkbart grad.

6.3.1 Andel taleturns og brukertilbakemeldinger

Det er interessant å se at andelen taleturns i snitt ligger på 48,9% når flesteparten av brukerne angir en nedre grense for bruk av talemodalitet (se kapittel 5.6.6) over den funnede gjenkjenningssraten for systemet – 57,21%. Dette misforholdet kan komme av enten at brukerne av en eller annen grunn ikke handler i testen jamfør det de ville gjort i virkeligheten, eller at brukerne har en misforestilling om egne grenser i forhold til feil i talegjenkjenning. Den første antagelsen er ikke usannsynlig, men det er heller ikke den andre; ettersom talegjenkjenning ikke er særlig utbredt enda.

6.3.2 Høy gjenkjenningsrate for bruker 23

Bruker 23 prater tydelig og uttaler hvert ord fullstendig, og brukeren snakker relativt rolig. Dette medfører en høy ASR-rate for denne personen. Brukerens dialekt er fra området rundt Bodø/Rognan.

6.3.3 Årsaker til lav gjenkjenningsrate

Bruker 24 snakker lavt og har en lys stemme. Det er en kvinne. Eksempel: filene med timestamp 2006-04-04_15:16:36.906 og 2006-04-04_15:17:17.234 har like ytringer. De er begge kuttet i enden, i midten av siste konseptverdi ”jernbanetorget”, slik at kun ”jernbanetor” er med i lydfilen. Opptaket som er lagret på siste fil har i tillegg ikke registrert starten av ytringen, slik at konseptverdien ”neste buss” er redusert til ”te buss”. Konseptet ”temporal” blir likevel fylt ettersom en konseptverdi med konfidens på 892 registreres. Dette kommer sannsynligvis ikke fra verdien ”neste buss”, men fra verdien ”buss”. De to filene har følgende resultat:

```
There are 2 concepts from voice here. file: 25.wav
Voice. Concept: Temporal      ConceptConfidence: 340 value: Bussrute
Voice. Concept: Buss_stasjon2 ConceptConfidence: 220 value: Klingenberg
Timestamp: 2006-04-04_15:16:36.906
```

```
There are 2 concepts from voice here. file: 26.wav
Voice. Concept: Temporal      ConceptConfidence: 892 value: Bussrute
Voice. Concept: Buss_stasjon2 ConceptConfidence: 134 value: Jernbanetorget
Timestamp: 2006-04-04_15:17:17.234
```

Brukeren opplever til dels store problemer med å få høy nok konfidens på grunn av en talemåte med mye ”s”-lyd i talen, som skaper problemer for gjenkjenneren.

En annen persons tale viser en annen årsak til lav gjenkjenningsrate: ”Jernbanetorget” ble i to etterfølgende ytringer sagt to ganger. Første gang ble det tatt opp ”ernbanetorget”, andre gang ”Jernbanetorget”. Andre gang ble navnet sagt mer med trykk på alle stavelser. Personen var en kvinne, som snakket normalt høyt og hadde en alt-stemme.

```
***Turn 51:
There are 2 concepts from voice here. file: 19.wav
Voice. Concept: Buss_stasjon ConceptConfidence: 172 value: Toyen
Voice. Concept: Buss_stasjon ConceptConfidence: 53 value: Ankertorget
Timestamp: 2006-04-06_10:53:19.826
***Turn 52:
There are 1 concepts from voice here. file: 20.wav
Voice. Concept: Buss_stasjon ConceptConfidence: 1000 value: Jernbanetorget
Timestamp: 2006-04-06_10:53:27.998
```

Det kan ses i turn 51 at kuttingen i starten på ytringen, forårsaket av VAD, førte til at konseptet ikke ble gjenkjent. I turn 52 kan det ses at konseptet er blitt gjenkjent med full score. Starten av ordet har her altså hatt en elementær plass i ASR-gjenkjenningen. Ettersom den benyttede ASR-løsningen er lukket, er det uvisst hvordan beregning av konseptkonfidenser gjøres. Det viste eksempelet viser at det ligger en potensiell gevinst i gjenkjenningssrate i gjensidig tilpasning av ASR-løsningen og utstyret som tar opp ytringen.

Utbredelsen for denne typen kutting av ytringene er ukjent. Det kunne vært gjort en kontroll av dette ved å manuelt høre på talefilene som er tilgjengelige for å få en oversikt, men dette ble lagt til side av tidsmessige grunner.

6.4 ASPEKTER VED Å SPESIALISERE ELLER GENERALISERE SYSTEMET

I det følgende blir aspekter vedrørende å optimalisere systemets deler opp mot en enkelt applikasjon versus å holde det så generelt som mulig forsøkt belyst.

I utviklingen av systemet er det mulig å sikte på ulike mål. Her vil det bli diskutert hvordan situasjonen vil bli om systemet utvikles i den ene eller andre retningen – mot å være en generell multimodal plattform eller bli optimalisert for én type applikasjon. I dag er systemet utviklet slik at det er en så klar skillelinje som mulig mellom system / plattform og applikasjon. Fordelen med dette er at dette gjør det mulig å tilby en multimodal plattform til nye applikasjoner på en så enkel måte som mulig. Systemet vil da tilby et felles grensesnitt som en tredjepart forholder seg til for å lage programvare. Å optimalisere ulike deler av systemet for én enkelt applikasjon har fordelene bedre spesiell applikasjonsytelse på områder som tid, krefter brukt for å få informasjon, og kanskje en generell følelse av ”skreddersøm”. Prisen for dette vil være større utviklingskostnader for applikasjonen.

De fleste av systemendringene som er foreslått i denne oppgaven er med på å trekke systemet over til å bli mer applikasjonsspesifikt om de gjennomføres. Konfidensterskelen er foreslått senket på grunnlag av den treffsikkerhet som TaleServer hadde med den grammatikk som gjaldt ved testøyeblikket, dialogmanageren har blitt foreslått utvidet med et begrenset antall inndata som starter tilstandsmaskinen – hvor antall inndata og deres innhold er tilpasset en bussruteopplysningstjeneste. (Grammatikken er blitt foreslått utvidet med et deiktisk ord ”herifra”, men dette tjener en multimodal plattform generelt.)

Hvor generell er det mulig å gjøre en multimodal plattform? Hvor mye tilpasning vil måtte trenges til en ny applikasjon? I applikasjonen som oppgaven er gjort på, har man en dialogmanager i form av en utvidet FSM, som er høyst applikasjonsspesifikk. Det er usikkert hvor generell denne delen av systemet kan lages, men ut fra modulens viktighet for hele systemet er det sannsynlig at en dialogmanager må skreddersys for hver ny applikasjon. Kanskje kan arbeidet gjøres lettere ved systemutviklingsteknikker som

opprettelse av biblioteker med generell kode for ulike oppgaver, slik at noe kodegjennbruk kan finne sted. Gramatikkens innhold er tett forbundet med dialogmanageren, ettersom den definerer hvilke konsept som skal være mulige som inndata på tilstandsmaskinen. Gramatikken inneholder i busdemonstratoren settet av gyldige bussholdeplasser og gyldige verdier for konseptet temporal. Dette er informasjon som er applikasjonsavhengig. Gramatikkens oppgave er i et talegjennkjenningsystem å forbedre suksessraten ved å innskrenke antall valg for brukeren. Disse valgene må være tilpasset, slik at de slipper gjennom ytringer som kommer på riktig form, og blokkerer for ytringer som kommer på feil form. For at gramatikken skal være hensiktsmessig i bruk, må den tilpasses hver enkelt applikasjon.

Talegjennkjennning ved bruk av skjulte markovmodeller er per i dag en unøyaktig teknologi, som har en gjennkjenningsrate under det som er ønskelig, og spesielt når det gjelder bruk i en omgivelse med spontan tale, gjerne uten brukertilpasning. Det er sannsynlig at en hvilken som helst applikasjon som ønskes kombinert med en multimodal plattform som i denne oppgaven, vil kreve en god del tilpasning av plattformen. Om dette vil forandre seg i fremtiden er usikkert, men i dagens situasjon, med begrenset kunnskap om multimodale system og et begrenset antall av dem i bruk, kan det se ut som om det er en fordel å gjøre flest mulig valg til fordel for en høy dialogsuksessrate. Dette krever en applikasjonstilpasning, i form av gramatikk, tilstandsmaskin, konfidensterskler, støynivåterskler m.v. Idealet er en god multimodal plattform med et felles grensesnitt som kan brukes "out of the box" i så høy grad som mulig, men et enda høyere krav er en tjeneste som gir rett svar så ofte at den i det hele tatt oppleves nyttig. I dagens situasjon ser det ut som dette krever en høy grad av tilpasning av systemet til den aktuelle tjenesten som gjør systemet mindre generelt, men gir bedre suksessrate.

6.5 GENERELLE FORBEHOLD

6.5.1 Måten testbrukere ble funnet på

Tallene som er funnet i testen, har blitt funnet ved testing av personer i et teknisk miljø, og hvor de fleste er i slutten av deres studier. Dette gjør at de har en høy kompetanse innen IKT. Innvirkningen dette er ventet å ha på resultatet er en høyere enn normalt tilvenningskurve, og evne til å "gjennomskue" og arbeide "på lag med" systemets styrker og svakheter.

Resultatene må ventes å ha blitt påvirket av at de fleste av brukerne var personlig bekjente av testleder, gjennom en noe mer spent atferd hos bruker på grunn av "frykten for ikke å forstå". Det kan ytterligere føre til et bias i resultatene at testene ikke var fullstendig objektive på grunn av relasjonen mellom testleder

6.5.2 Endringer i løpet av testen

Det ble gjort visse endringer i løpet av testen. MUTE-knappen ble bedt brukt etter få brukere når de ikke direkte var i en oppgave. Andre halvpart av brukerne kom til en test som var mer strømlinjeformet og lik enn den første halvparten, hvor hva testleder sa,

tidsbruk osv. var mer skiftende. Det ble gjort en del erfaring i å holde tester i løpet av perioden, som gjorde de testene som kom på slutten bedre enn de i starten. Testleder ble bedre til å observere mot slutten enn i starten. Antall systemkrasj, dalte fra starten til slutten ettersom systemet ble mer og mer kjent for testleder. Systemkrasj hadde mye å si for testleders evne til å beholde fokus. Hyppige systemkrasj førte oftere til formelle feil som for eksempel å glemme å informere brukeren informasjon som skal gis. Dette har påvirket resultatene, og gjort dem mindre sammenlignbare.

6.5.3 Realistiske bruksomgivelser?

Testen ble holdt i et avskjermet rom uten ytre påvirkninger. Applikasjonens naturlige omgivelse vil i mange tilfeller være travelt utemiljø. Denne oppgaven kan bare gi antydninger til hvordan applikasjonen vil bli brukt i et slikt scenario. De innhentede resultatene betinger stille omgivelser. Ikke-stasjonær bakgrunnsstøy representerer en stor utfordring for automatisk talegjenkjenning i dag.

6.5.4 Effekter av begrensninger i modellen

Modellen hadde totalt 63 bussholdeplasser implementert, spesielt utvalgt etter unikhet i uttale, for å øke gjenkjenningsraten av konsepter. Med denne begrensningen fjernet vil feilraten gå opp vesentlig, samtidig som gjenkjenningsraten vil synke. Prosesseringstiden vil gå opp noe.

I modellen er det kun mulig å spørre etter ”neste buss”. I en fullskalamodell vil dette komme til å utvides. Dette vil ha innvirkning på gjenkjenningsraten gjennom mer komplekse spørringer til systemet.

Det er som tidligere sagt, ikke lagt skjul på at GUI ikke er optimal, og dette er bemerket av flere av brukerne at gir en bias mot bruk av tale. En god GUI vil antagelig heve andelen av pennebruk noe.

6.5.5 Effekter fra suboptimal implementasjon

Flere brukere gjorde det kjent at det ble interagert og svart som det ble gjort på grunn av løsninger i implementasjonen som var mindre gode enn det de var vant med i hverdagen. Begrenset GUI fikk en rekke brukere til å velge oftere tale; treg systemrespons som følge av VAD og fusjonstimere oppfordret til interaksjon som krevde få turns i dialogen – multimodalitet og tale, og absolutt ikke zoom ut og inn, som ble sett på som veldig tregt.

6.5.6 Eksempelets makt: introduksjonen

Introduksjonen av demonstratorens virkemåte i begynnelsen av brukersesjonen har uten tvil lagt en sterk føring på hvordan interaksjon med systemet blir for brukeren, og antagelig spesielt multimodal interaksjon, ettersom dette er en ny, ukjent teknikk. En annen introduksjon ville skapt andre resultater.

6.6 SAMMENLIGNING AV RESULTATER MOT HYPOTESE

Hypotesen, repetert slik den står i kapittel 2.2:

Ved forespørsel til et datasystem vil (1) så kort tid og (2) så lite kognitiv anstrengelse som mulig brukes for å skaffe informasjon.

Hypotesen er relevant for multimodalitet om man ser den i lys av slutningen i artikkelen av Rugelbak og Hamnes (2003) om at multimodalitet krever mer kognitiv anstrengelse enn kun tale eller kun peking alene.

6.6.1 Hoveddiskusjon av hypotese mot data. Når brukes multimodalitet?

En gjennomgang av applikasjonen viser at bruk av simultan koordinert multimodalitet kun gir en direkte tidsgevinst sammenlignet med monomodal penn og monomodal tale, i situasjoner der brukeren vet navnet på en av avreise/ankomstholdeplassene. Ved andre tilfeller er det ingen tidsgevinst; Om begge navn er kjent, kan monomodal taleinteraksjon gjøre samme jobb, og om begge navn er ukjent må det brukes penn for begge punktene. I testen er det noen flere multimodale ytringer i oppgave 2, som nettopp har utgangspunktet med et kjent og et ukjent navn. Dette kan tyde på at første del av hypotesen er korrekt; at det søkes etter metoder for å gjøre oppgaven på så lite tid som mulig.

I oppgave 1 var begge bussholdeplassnavn gitt, og dette er en oppgave som raskest, og med minst kognitiv anstrengelse, løses ved bruk av monomodal taleinteraksjon. Oppgave 2 er som beskrevet over en oppgave som løses raskest med en multimodal turn. Det kan føyes til at den vil være raskest løst med simultan koordinert interaksjon, men sekvensiell multimodalitet vil kun bruke en turn ekstra for det samme. Oppgave 3 oppfordrer til bruk av penn; den gir ingen bussholdeplassnavn, og brukeren er tvunget til å zoome inn etter tur de forskjellige bussholdeplassene. I lys av den ekstra ytelsen man må framvise for å bruke simultan koordinert multimodalitet, forventes det i følge hypotesen for oppgave 1 og 3 en lav andel simultane koordinerte turns. Dette bekreftes ikke av testdataene. For oppgave 1 lyder andelen 8,7%, oppgave 2 13,9%, oppgave 3 11,5%. Det er riktignok noen flere simultane koordinerte turns i oppgave 2 enn i de andre, spesielt oppgave 1, men oppgave 3 har et prosenttall som ligger tett opptil oppgave 2, på tross av at det ikke gir noen gevinst objektivt sett å bruke simultan koordinerte turns her. Om man ser på hvordan disse ytringene ser ut (kap 5.4) ser man at de er, som i oppgave 2 og 1, stort sett "fulle" ytringer - hvor alle nødvendige konsepter er med for å få tilbakemelding med en gang. Dette kan peke på at noen brukere har foretrukket å dele opp oppgaven i to faser – en hvor man samler inn data, og en hvor man gir inn denne informasjonen til systemet.

Og da har man kanskje nøydt seg med å finne navn på en av holdeplassene, samt lokalisert nummer to på kartet.

Ettersom multimodalitet er en ny interaksjonsform, kan det tenkes at utforskningslysten hos brukerne har påvirket resultatet. Brukerne er i et miljø hvor nye IKT-løsninger tidlig tas i bruk. Dette kan ha økt bruken av simultan koordinert multimodalitet til en urealistisk høy verdi.

Andelen sekvensiell multimodalitet er tilnærmet lik, og ganske høy, i alle oppgaver (50,3%, 46,3%, 48,3%). At tallene ligger rundt 50 prosent peker på at det ikke ser ut som det representerer noen særlig terskel mtp. kognitiv anstrengelse å bytte modalitet mellom hver turn. At tallene er høye for alle tre oppgavene peker på at brukerne setter pris på valgfriheten som tilbys. Men det kan også bety at en skifte av multimodalitet i en del dialoger er en utvei etter (gjentatt) feiling med den først valgte modaliteten. Dette kan også være en av grunnene til at det finnes lik mengde sekvensiell multimodalitet i oppgave 1 som i oppgave 3. Ettersom sekvensiell multimodalitet kun i oppgave 2 objektivt sett sparer brukeren for tid, og resultatene for oppgavene varierer med under 10% i forhold til hverandre, kan det ikke sluttet ut fra resultatene i denne oppgaven at brukerne bruker sekvensiell multimodalitet i den hensikt å spare tid. Det *kan* sluttet at sekvensiell multimodalitet er hyppig brukt, men om dette er fordi det reduserer den kognitive anstrengelsen for brukeren, eller om det skjer av andre grunner, kan ikke sies.

Det blir ikke benyttet statistiske metoder for å akseptere eller forkaste hypotesen, ettersom det ikke finnes godt nok statistisk grunnlag for dette. Tallet på deltagere er lavt, noe som gjør at resultatene som baserer seg på kun et utvalg av testgruppen, i virkeligheten støtter seg på kun et fåtall personer. Dette er tilfellet for simultan koordinert multimodalitet, hvor 12 av 28 personer aldri benyttet seg av dette, og hvor tre av deltagerne til sammen sto for 36,5% av resultatet.

Om resultatene fra bruk av simultan koordinert multimodalitet skulle representere brukergruppen på en korrekt måte og at forbeholdene som er tatt, ikke gjør de funnede resultatene fullstendig ugyldige, kan det bety at den kognitive anstrengelsen forbundet med å tale og trykke i samme ytring, for disse brukerne ikke er så høy at det ikke blir brukt. I lys av andelen simultan koordinert multimodalitet i oppgave 3, kan det til og med se ut som at dette i enkelte tilfeller er en foretrukket framgangsmåte selv om det tidsmessig ikke er noe å tjene på dette.

7 MULIGE ENDRINGER

7.1 KRITERIET FOR NÅR AUDIOSIGNALET ANTAS Å INNEHOLDE TALE

I denne delen foreslås endringer i kriteriet for når audiosignalet mottatt i klientens mikrofon avgjøres å inneholde tale eller støy. En annen måte å beskrive det på er: hvordan forholdet mellom type-1-feil og type-2-feil i den gjennomførte testen kan endres. Nullhypotesen er i dette tilfellet ”Det ligger en ytring på inngangen av talekanalen”. Type-1-feil er i denne sammenhengen ”Ytring, men feilaktig forkastet som støy”, mens type-2-feil i denne sammenhengen er ”Støy, men feilaktig akseptert som ytring”. I systemet i sin nåværende form blir avgjørelsen tatt i VAD. Samtlige audiosignal som slipper forbi her, blir videre behandlet som tale, og blir gitt respons på. Det presiseres at det ikke er gjenkjenningssats som det er tale om her. En ytring er en ytring når den blir svart på av systemet. Systemet svarer på ytringer selv om det ikke har blitt gjenkjent noe i dem. Årsaken til punktet tas opp til diskusjon er den store overvekten av type 2-feil i forhold til type 1-feil tilstede i de utførte testene.

I demonstratoren brukt i denne oppgaven er spørsmålet ”når er det tale å gjenkjenne” løst ved bruk av en taledetektor. Kriteriet her var at signalet over en viss tid overskrider et visst terskelnivå. Som tidligere diskutert var en relativt høy andel av antall Tale-turns i første opptelling (Tabell 5) generert type 2-feil. For enkelte av brukerne med høy andel slike ikke tiltenkte inndata, for eksempel bruker 13, gikk antall tale-turns ned fra 28 til 11 ytringer, tilsvarende 61%. Totalt ga systemet tilsammen feilaktig tilbakemelding på 168 av 394 (42,64%) audiosignal på grunn av type 2-feil.

I kapittel 6.5.2 kan også leses hvordan brukerne etter gjennomgang av få tester ble bedt om å bruke MUTE-knappen aktivt. Dette ble gjort fordi det ble opplevd at klienten ”lyttet til” og ”forsøkte å delta” i samtalen mellom testbruker og testleder, og at den oppfattet brukerens lyder som ”hmm”, ”tja”, kremting og ulike andre former for språklige artifakter som opptrer i en spontan tale-situasjon. Dette peker på at den nåværende taledetektor-mekanismen ikke fungerer etter hensikten. Foruten generell systemustabilitet og lang responstid, som er utenfor denne oppgavens formål å løse, ble det vurdert dithen at dette var den mest åpenbare svakheten med systemet. Begrunnelsen ligger i hvor hyppig problemet forekommer kombinert med den relativt lange tiden det tar for en dialogturn å fullføre, når det ser ut som tid er en viktig faktor for brukeren.

Det blir i det følgende skissert to løsninger på dette problemet og korte diskusjoner av disse. En implementasjon og test av disse ville vært ønskelig i oppgaven, men på grunn av mangel på tid ble dette ikke gjennomført.

7.1.1 Dedikering av en hardware-knapp til å markere tale

En nærliggende løsning på problemet er å tilordne en av klientens hardware-knapper MUTE-funksjonen som allerede ble hyppig brukt i testen. Denne kan grovt sett brukes på to måter:

7.1.1.1 En push-to-talk-knapp

En push-to-talk-knapp aktiviserer mikrofonen ved nedtrykking av knappen, og deaktivering når knappen blir sluppet igjen. Den andre ville være å aktivere og deaktivere mikrofonen med to separate trykk. Fordelen med en mute-knapp implementert som en push-to-talk-knapp over den andre løsningen vil sannsynligvis være en noe hurtigere betjening. Bakelen er fare for dårlig gjenkjenning hvis brukeren ikke holder jevnt press, men slipper av og til. Bruk av simultan koordinert multimodalitet i dette tilfellet vil kreve asynkron bruk av begge hender, i tillegg til tale, noe som vurderes som usannsynlig at kommer til å bli gjort i og med at man ved bruk av allerede én hånd og tale er usikker på om den kognitive lasten er for stor. Bruk av en push-to-talk-knapp kunne kanskje forsvares ut fra den høye andelen monomodal tale – 48,9%, men dette ville da kun være en tilleggsegenskap, etter som den jfr. ovenstående diskusjon ikke egner seg som taledetektor-mekanisme for et multimodalt system.

7.1.1.2 En annen knapp

Alternativ nummer to med aktivering og deaktivering av mikrofonen med to separate trykk krever ikke bruk av begge hender. Derimot krever den noe bruk av finmotorikk av brukeren, ettersom brukeren antas å ville trykke på knappen med en av fingrene på hånden han også holder pennen i. Dette vil være situasjonen for alle i testen som holdt klienten i hånden. Dette er ikke en god funksjon med tanke på brukergrupper som har motoriske problemer, og vil for alle brukere sannsynligvis kreve noe mer kognitiv anstrengelse. Hvorvidt dette er en løsning som vil bli brukt må det nye brukertester til for å bekrefte eller avkrefte.

7.1.1.3 Andre knappeløsninger

Mellomløsninger som trykk for aktivering, mens deaktivering skjer basert på nåværende VAD-løsning, blir ikke tatt med til diskusjon. Testresultater for de foreslåtte endringene antas først å måtte kjennes før en diskusjon av en slik videreutvikling kan gi nyttig informasjon.

7.1.1.4 Implementasjon, knapp

En enkel skisse til implementasjon av en push-to-talk-knapp vil være som følger: Ved nedtrykking av knappen fyres det av en event, som fanges og sendes videre til metoden for aktivering/deaktivering av MUTE-funksjonen. Her blir MUTE slått på, inntil en ny action med informasjon om at knappen er blitt sluppet, ankommer, og skrur mikrofonen av. For en knapp hvor aktivering og deaktivering skjer i to trykk, brukes bare

informasjonen om enten nedtrykk eller slipp av knappen til å bytte status av MUTE-funksjonen.

7.1.2 Konklusjon

Trykking på en fysisk knapp for å markere når tale er å dekode, er en sikker måte å utelukke feil av type 2 på. Kostnaden med denne løsningen er mindre brukervennlighet. Det forventes at introduksjonen av en slik knapp høyner terskelen for å ta i bruk talemodaliteten. En knapp implementert etter push-to-talk-metoden kan være god i for et monomodalt talegrensesnitt, men skaper en praktisk utfordrende i en multimodal tjeneste ved bruk av simultan koordinert multimodalitet, og må derfor anses som uaktuell. En knapp som ved ett trykk aktiverer mikrofonen og ved neste deaktiverer den, hindrer ikke multimodal bruk, men vil sannsynligvis kreve noe mer finmotorikk av brukeren enn uten. Dette er en utvikling i gal retning sett i lys av ”design for all”. Om en fysisk knapp er en varig løsning vil nye brukertester kanskje gi svar på, men denne korte vurderingen antyder et negativt svar.

7.1.3 Konfidensbasert avgjørelse

Et alternativ for å løse spørsmålet ”Når er det tale å gjenkjenne?” er å gjøre dette i dialogmanageren. Det bli her kort vurdert to ulike strategier for hvordan dette kan implementeres. Årsaken til at det vurderes gjort med konseptkonfidens som kriterium er at en stor mengde av audiofilene med lave konseptkonfidenser laget under testen, virkelig viste seg å inneholde støy, og resulterte i en mengde gjenkjente filer *med lav konfidens*.

7.1.3.1 Første strategi

- I dialogtilstanden ”start_regular”, som betyr at systemet er i utgangsposisjon, implementeres følgende sjekk før videre sjekker foretas: Et av få nøkkelord må være tilstede blant de suksessfullt gjenkjente konseptverdiene. Forslag til nøkkelord er ”neste buss”, ”buss”, og ”når går”, og ”hallo”. ”Når” og ”går” gjøres her om fra hhv. temporal- og filler-verdier til en egen konseptverdi. ”Hallo” legges til som ny ”filler”-konseptverdi. I tilfellene hvor konfidensen ikke er høy nok / ingen av nøkkelordene er tilstede, blir ytringen ikke prosessert og registrert som en turn, men forkastet av den nye taledetektoren under antagelsen at innholdet er støy.

Det ble valgt å planlegge denne endringen utført på tilstanden ’start_regular’, etter som denne tilstanden er den som oftest forekommer i en sesjon, og tenkes å være den som brukeren lær systemet stå i når han er inaktiv for en periode. Resultatet av den nye taledetektoren vil bli en klient som i starttilstanden viser en mer passiv oppførsel enn inne i selve dialogen, hvor demonstratorens oppførsel forblir uforandret. I start_regular-tilstanden vil den fortsatt markere med et ”bip” når den har gjort et opptak fra mikrofon, men noe tilbakemelding fra systemet vil ikke komme før et nøkkelord er detektert med en konfidens over den gitte terskelen. Det vil være mulig å si nøkkelordene på vilkårlig plass i ytringen. Fordelen med denne endringen er et system som med FSM i utgangstilstand kan håndtere type 2-feil på en mye bedre måte i tilstand ”start_regular”. I denne

tilstanden kan den sannsynligvis stå i et miljø med relativt høy bakgrunnsstøy uten å reagere på dette. På den andre siden vil systemet i større grad kunne oppleves uforutsigbart for brukeren, ettersom det vil merkes at systemet reagerer forskjellig på støy i ulike tilstander. Dette problemet må forventes at øker med økende grad av bakgrunnsstøy. Det vil forekomme flere feil enn før av type 1 i tilstand "start_regular", hvor ytringer feilaktig blir forkastet som støy.

7.1.3.2 Andre strategi

- Alle konsepter som ankommer i alle tilstander i FSM fra tale-inndata får konfidensen kontrollert opp mot en terskel, for eksempel terskelverdien for fylling av konsepter, før videre sjekker foretas. Har ingen konsept i den mottatte meldingen høyere konfidens enn denne terskelen, blir ikke dette registrert som en turn, men alt kastet under antagelsen om at innholdet er støy. Ingen turn blir skrevet til logg, og *ingen tilbakemelding gitt til brukeren.*

Resultatet av denne kontrollen vil bli en klient som viser en mer passiv oppførsel i samtlige deler av dialogen, ved at den ikke gir tilbakemelding til brukeren før suksess i fylling av konsepter oppnås. Den vil fortsatt markere med et "bip" når den har gjort et opptak fra mikrofon, og slik vise at den har detektert og tatt opp lyd. Fordelen med denne endringen er et system som til forskjell fra endring i kapittel 7.1.3.1 ikke begrenser brukeren i start-tilstanden, og at alle tilstander er underlagt de samme tersklene. I støyende omgivelser vil systemet vise like god eller dårlig resistens mot type 1-feil og type 2-feil uansett hvor i dialogen en er. Det vil gjennom hele dialogen forekomme flere tilfeller enn før av type 1-feil - ytringer feilaktig forkastet som støy. Det vil bli mindre tilbakemelding til brukeren ved disse feilene. Systemet vil ikke ha like god immunitet mot støy i start-tilstanden som i strategi 1.

7.1.4 Konklusjon

Hvilken av de to strategiene skissert som velges, burde optimalt velges ut etter empiriske undersøkelser. Begge strategier vil føre til et mye lavere antall type 2-feil, men introdusere noen flere type 1-feil. Ut fra brukerens standpunkt kan det kanskje argumenteres for at strategi nummer to vil gi det mest forutsigbare systemet. Hvorvidt endret adferd i forhold til støy i løpet av dialogen påvirker brukeropplevelsen er umulig å anslå uten ytterligere tester. Dagens system har en klar overvekt av type 2-feil; det eksakte tallet på type 1-feil er ukjent, men det er lavt. Det er uvisst hvordan det vil påvirke brukeren å ikke få tekstlig/audibel tilbakemelding når en ytring ikke blir gjenkjent. Det er å bevege systemet mot en bedre håndtering av audiosignal å sette en tale/støy-terskel på konseptkonfidenser. Dermed kan en stor andel audiosignal med lave konseptkonfidenser korrekt forkastes som støy, med den kostnaden at noen ytringer feilaktig forkastes som støy.

I og med at terskelen for tale/støy alltid velges lavere eller lik terskelen for fylling av konsepter – ellers gir det ingen mening – er det bare tuning av grensen for når systemet skal svare "Jeg forstår ikke din forespørsel"/ingen tilbakemelding (forkaste ytringen) som

er diskutert i dette kapitlet (7.1). Det er et spørsmål om god og korrekt tilbakemelding, og har ingen direkte innvirkning på gjenkjenningsrate.

7.2 ANDRE MINDRE ENDRINGER

Her følger noen endringer som kunne vært gjennomført, basert på brukernes tilbakemeldinger.

7.2.1 GUI

Det ble i intervjuet etter endt test gitt en god del implementasjonsmessige tilbakemeldinger, som går på GUI, kart, antall valg som gis brukeren i forskjellige situasjoner, med mer.

- Linjer på kartet som viser veien bussen tar fra A-B ved visning av svar på henvendelse.
- Navn på bussholdeplassene på kartet vist hele tida som default
- Bedre grafisk framstilling av resultat. Vise begge busstopp i samme vindu, kanskje med strek i mellom. Ikke som nå, hvor (ved stemme på begge) kun bilde av avgangsposisjon er vist.
- Er F eller T-knappen trykket? Indikatorlist ble ikke sett. (bruker 14 med flere).

7.2.2 Dialogmanager

Følgende endringer foreslås utført for dialogmanager.

- Utvide start-tilstanden av FSM slik at det kan trykkes på bussholdeplasser for å få tak i info om holdeplassen uten å starte en forespørsel. Dette var en framgangsmåte som ble omfattende benyttet i oppgave 3. Info om holdeplassen kunne vært utvidet til å holde på navn, hvilke busser som passerer og navn på bydel.
- Implementasjon av bedre feilhåndtering: ”Jeg forsto ikke din forespørsel”. Hva var det den ikke forsto? Presisere tilbakemelding til bruker der hvor det er mulig. (bruker 11)

7.2.3 Gramatikk

- Legge til det deiktiske ordet ”herifra”. Gjennom hele kameratintroduksjonen brukte testleder det deiktiske ordet ”herifra”. I en ytring sier bruker fire ”herifra” idèt en bussholdeplass blir trykket. Dette deiktiske ordet er ikke med i gramatikken i nåværende tilstand, men vil være naturlig å tilføye. ”Herfra” ble også sagt i introduksjonen.

8 KONKLUSJON

Multimodalitet ble i høy grad utnyttet av personer i brukergruppen ”studenter i et teknisk miljø”, som de 30 testpersonene testet i denne oppgaven, kom fra. I 48,1% av dialogene ble det brukt enten simultan koordinert eller sekvensiell multimodalitet. Observasjonen bekrefter det som ble observert i artikkelen av Kvale, Rugelbak og Amdal (2003) om at brukere setter pris på å kunne velge modalitet på nytt for hvert turn i dialogen.

Angående tallfesting av simultan koordinert multimodalitet, ble først systemets fusjonsmekanisme brukt som definisjon. Dette ga 2,35% simultan koordinerte multimodale turns til sammen i testene. Dette må sies å være en lav verdi. Det ble så telt simultant koordinerte turns etter en ny definisjon - ”turns som var simultan koordinert sett fra brukerens side”. Med denne definisjonen lyder andelen simultan koordinerte turns 11,3%. Det ble i oppgaven sett på dette sistnevnte tallet, ettersom å følge systemets fusjonsmekanisme utelukket en rekke turns som var simultant multimodale sett fra brukerens side.

I gjennomsnitt inneholdt 48,9% av turnene monomodal tale. Turns med inndata fra penn stod i gjennomsnitt for 39,8% av turnene med inndata. Forklaringer til dette kan være en bedre implementert talemodalitet enn GUI, nysgjerrighet på talegjenkjenning blant brukerne, men også at pennebruk kan ha blitt sett på som en *mindre automatisk* framgangsmåte enn pennetrykking, noe som ble kommentert av en av brukerne.

ASR-gjenkjenningsraten ligger for denne testen i gjennomsnitt på 57,21%. Det kan være et tegn på at dette er et nivå som aksepteres av brukerne når man samtidig ser på at bruken av tale er såpass høy som den er. Det er imidlertid verdt å merke seg at de fleste personene oppgir i intervjuet en nedre gjenkjenningsrate for å bruke talegjenkjenning som er høyere enn dette.

Sammenligning av resultatene mot første del av hypotesen tyder på at den delen av hypotesen kan være korrekt. Første del av hypotesen sa at så kort tid som mulig vil brukes for å skaffe informasjon. Et argument for å bruke simultan koordinert multimodalitet i denne applikasjonen er at man kan spare tid ved å utføre en spørring uten å vite navnet på begge bussholdeplassene. Dette oppnås om man zoomer inn på og trykker på den holdeplassen som er ukjent, samtidig som man sier navnet på den andre. Oppgave to ville bli løst raskest med denne framgangsmåten, og resultatene viser en økt andel simultan koordinert multimodalitet for denne oppgaven sammenlignet med de to andre.

Sammenligning av resultatene opp mot andre del av hypotesen kan ikke ses å gi noen klare svar. Oppgave 1 (begge bussholdeplassnavn kjent fra oppgaveteksten) inneholder 8,7% simultant koordinerte turns, selv om det ikke logisk sett gir noen fordel som helst. Å gi inndata ved hjelp av tale gir brukeren mindre kognitiv last enn ved hjelp av simultan koordinert multimodalitet, og ifølge hypotesen burde dermed all interaksjon være tale.

Oppgave 3's andel av simultant koordinerte turns i representerer også et problem for den andre delen av hypotesen. Andelen ligger på 11,5%, selv om det kognitivt sett kun er tap forbundet med å løse oppgaven ved hjelp av multimodalitet. Det må oppsummeres med at "kognitiv anstrengelse" er en komplisert variabel å teste, og i denne oppgaven har det ikke lyktes å si noe om den ved hjelp av hypotesen.

I 69,4% av de 52 simultant koordinerte turnene i testen var brukeren stille idè det ble trykket. Dette kan tyde på at å tale *eksakt samtidig som* man trykker er kognitivt krevende, og at å gjøre èn ting om gangen, sett i liten tidsskala, er å foretrekke.

I de simultant koordinerte turnene i testen som hadde et deiktisk ord, var det ofte trykket på eller rundt dette ordet. Et flertall - 63,5% av de simultant koordinerte turnene - hadde derimot ikke et deiktisk ord. I mange av disse ytringene gjorde brukeren en liten pause i pratingen mens det ble trykket. På denne plassen i setningen kunne det semantisk sett ha passet med et deiktisk ord, men dette ble altså ikke sagt. I introen til oppgaven ble alltid et deiktisk ord sagt sammen med trykking.

90,4% av de simultan koordinerte multimodale ytringene i testen inneholdt alle konsepter for å kunne få direkte tilbakemelding fra trafikanten. Det ser ut til at simultan koordinert multimodalitet brukes når man kan er klar til å gjennomføre hele dialogen på èn turn.

Demonstratoren hadde en taledetektor som ikke fungerte tilfredsstillende. Den ble aktivisert for ofte av bakgrunnsstøy, og førte til stor irritasjon og hyppig bruk av MUTE-knappen hos brukeren. Det ble lagt fram to ulike forslag for å forbedre dette. Et alternativ med konfidensbasert taledetektor i dialogmanager, fortoner seg som den beste løsningen.

9 REFERANSELISTE

Johnsen, M. H., Kvale, K. *Improving speech centric dialogue systems – The BRAGE project. Proceedings NORSIG2005*, Stavanger, 22.9 - 24.9.2005.

Rugelbak, J., Hamnes, K. : *Multimodal Interaction - Will Users Tap and Speak Simultaneously?*. *Teletronikk* nr. 2, 2003, pp. 118-124 (Temanummer : Spoken Language Technology)

Boves, L., den Os, Els. *MUST – Multimodal and Multilingual Services for Small Mobile Terminals*. Heidelberg, EURESCOM Brochure series, 05.2002. (<http://www.eurescom.de/public/projects/p1100-series/p1104/default.asp>)

Markhus, V., Gajic, B., Svarverud, J., Solbraa, L.E. og Johnsen, M.J. *Annotation an Automatic Recognition of Spontaneously Dictated Medical Records for Norwegian*. Paper presentert på NORSIG, Finland 2004.

Open Agent Architecture. 2002. [online] – URL: <http://www.ai.sri.com/~oaa>

Adaptive Agent Architecture. 2002. [online] – URL: <http://chef.se.ogi.edu/AAA/>

Kvale, K., Warakagoda, N.D., Knudsen, J.E. *Speech centric multimodal interfaces for mobile communication systems*. *Teletronikk* nr. 2, 2003, pp. 104-117 (Temanummer : Spoken Language Technology)

Warakagoda, N.D. Knudsen, J.E., Lium, A. *Implementation of Simultaneous Coordinated Multimodality for Mobile Terminals*. 1st Nordic Symposium on Multimodal Communication (MUMIN), København, 23.9-24.9.2003

SpeechPearl 8.0 Dokumentasjon. Automatic Speech Recognition – the Technology behind it, Aachen 2002

Talsmann® - Norwegian TTS for Windows. [online]: http://www.telenor.no/fou/prosjekter/taletek/talsmann/tts_more.htm

(ukjent) United Nations enable. *Ad Hoc Committee on a Comprehensive and Integral International Convention on the Protection and Promotion of the Rights and Dignity of Persons with Disabilities*. [online]: <http://www.un.org/esa/socdev/enable/rights/adhoccom.htm>. 2006.

Transportation Research Board- National Research Council. *TCRP SYNTHESIS 37 Communicating with Persons with Disabilities in a Multimodal Transit Environment – A Synthesis of Transit Practice*. Washington D.C.. National Academy Press; 2001.

IST Information Desk. Information Society Technologies hjemmeside. [online]: <http://cordis.europa.eu/ist/> 19.6.2006

Mellors, W. J. *Design for all*. Teletronikk nr. 1, 2004, pp. 14-19 (Temanummer : Spoken Language Technology).

Nielsen, J., Mack R. *Usability Inspection Methods*. New York, John Wiley and Sons. 1994.

Wyard, P. J., Simons A. D., Appleby S., Kaneen E., Williams, S. H., Preston, K. R. *Spoken language systems – beyond prompt and response*. BT Technol J Vol 14. No 1. 1996.

Aas, A. *Gjenkjenning av medisinske epikriser*. Trondheim, Norge: Tapir 2005

Knut Kvale, Narada Warakagoda, Marthin Kristiansen: *Evaluation of a Mobile Multimodal Service for Disabled Users*. Proceedings MMUI, Gothenburg, Sweden, 7.4 - 8.4.2005.

Kristiansen, M. *Evaluering og tilpasning av et multimodalt system på en mobil enhet*. Trondheim, Norge: Tapir 2004

K. Kvale, J. Rugelbak, I. Amdal. *How do non-expert users exploit simultaneous inputs in multimodal interaction?*, Proceedings of the 19th International Symposium on Human Factors in Telecommunication, pp.169-176, Berlin, Germany, 1-4. December 2003.

10 VEDLEGG

10.1 UTFYLLENDE RESULTATTABELLER

10.1.1 *Andel turns av ulike typer, full tabell.*

Bruker	Oppgave	Penn	Tale	MM	Totalt (full test)
1	oppg1	5	4	1	
	oppg2	3	4	0	
	oppg3	-	-	-	
	sum	ble ikke fullført			
2	oppg1	-	-	-	
	oppg2	-	-	-	
	oppg3	-	-	-	
	sum	ble ikke fullført			
3	oppg1	6	6	0	
	oppg2	2	2	0	
	oppg3	6	4	1	
	sum	14	12	1	27
4	oppg1	4	5	0	
	oppg2	7	7	1	
	oppg3	4	7	0	
	sum	15	19	1	35
5	oppg1	3	8	0	
	oppg2	4	0	0	
	oppg3	4	0	0	
	sum	11	8	0	19
6	oppg1	6	6	0	
	oppg2	3	6	0	
	oppg3	3	2	0	
	sum	12	14	0	26
7	oppg1	4	8	0	
	oppg2	6	4	0	
	oppg3	-	-	-	
	sum	ble ikke fullført			

8	oppg1	5	8	0	
	oppg2	0	4	0	
	oppg3	6	4	0	
	sum	11	16	0	27
9	oppg1	1	3	0	
	oppg2	8	1	0	
	oppg3	2	0	0	
	sum	11	4	0	15
10	oppg1	1	13	1	
	oppg2	0	12	3	
	oppg3	1	15	2	
	sum	2	40	6	48
11	oppg1	1	6	0	
	oppg2	4	0	0	
	oppg3	2	5	0	
	sum	7	11	0	18
12	oppg1	1	5	0	
	oppg2	1	5	1	
	oppg3	1	5	0	
	sum	3	15	1	19
13	oppg1	1	8	0	
	oppg2	3	17	0	
	oppg3	4	3	0	
	sum	8	28	0	36
14	oppg1	1	4	0	
	oppg2	11	4	0	
	oppg3	1	5	0	
	sum	13	13	0	26
15	oppg1	4	3	0	
	oppg2	5	3	1	
	oppg3	4	5	1	
	sum	13	11	2	
16	oppg1	3	9	0	
	oppg2	-	-	-	
	oppg3	-	-	-	
	sum	ble ikke fullført			

17	oppg1	4	1	0	
	oppg2	2	2	0	
	oppg3	4	14	0	
	sum	10	17	0	27
18	oppg1	-	-	-	
	oppg2	-	-	-	
	oppg3	-	-	-	
	sum	ble ikke fullført			
19	oppg1	5	4	0	
	oppg2	5	8	0	
	oppg3	5	11	0	
	sum	15	23	0	38
20	oppg1	6	10	0	
	oppg2	-	-	-	
	oppg3	-	-	-	
	sum	ble ikke fullført			
21	oppg1	0	2	0	
	oppg2	5	6	0	
	oppg3	2	2	0	
	sum	7	10	0	17
22	oppg1	0	1	0	
	oppg2	0	1	0	
	oppg3	0	11	0	
	sum	0	13	0	13
23	oppg1	3	3	0	
	oppg2	1	6	0	
	oppg3	0	4	0	
	sum	4	13	0	17
24	oppg1	1	2	0	
	oppg2	1	2	0	
	oppg3	-	-	-	
	sum	ble ikke fullført			
25	oppg1	0	6	0	
	oppg2	3	3	0	
	oppg3	3	1	0	
	sum	6	10	0	16

26	oppg1	0	3	0	
	oppg2	1	5	0	
	oppg3	3	1	0	
	sum	4	9	0	13
27	oppg1	1	1	0	
	oppg2	5	7	0	
	oppg3	2	1	0	
	sum	8	9	0	17
28	oppg1	7	8	0	
	oppg2	1	3	0	
	oppg3	4	12	1	
	sum	12	23	1	36
29	oppg1	3	4	0	
	oppg2	1	3	1	
	oppg3	2	4	1	
	sum	6	11	2	19
30	oppg1	2	7	0	
	oppg2	4	7	0	
	oppg3	1	8	0	
	sum	7	22	0	29

OPPGAVE 1:			Antall	Prosent
Antall Penn-turns			78	34,21 %
Antall Tale-turns			148	64,91 %
Antall MM turns			2	0,88 %
Totalt antall turns			228	100,00 %
OPPGAVE 2:			Antall	Prosent
Antall Penn-turns			86	40,00 %
Antall Tale-turns			122	56,74 %
Antall MM turns			7	3,26 %
Totalt antall turns			215	100,00 %
OPPGAVE 3:			Antall	Prosent

Antall Penn-turns			64	32,99 %
Antall Tale-turns			124	63,92 %
Antall MM turns			6	3,09 %
Totalt antall turns			194	100,00 %
I ALT			Antall	Prosent
Antall Penn-turns			228	35,79 %
Antall Tale-turns			394	61,85 %
Antall MM turns			15	2,35 %
Totalt antall turns			637	100,00 %
Gjennomsnittlig antall turns				
oppgave 1	173/28		6,18	28 brukere gjennomførte oppgave 1.
oppgave 2	158/26		6,08	26 brukere gjennomførte oppgave 2.
oppgave 3	132/24		5,50	24 brukere gjennomførte oppgave 3.
hele testen			17,76	

10.1.2 Revidert resultat av bruk av ulike modaliteter, full tabell

Bruker	Oppgave	Penn	Tale	MM	Totalt (full test)
1	oppg1	5	0	0	
	oppg2	3	3	0	
	oppg3	-	-	-	
	sum	ble ikke fullført			
2	oppg1	-	-	-	
	oppg2	-	-	-	
	oppg3	-	-	-	
	sum	ble ikke fullført			
3	oppg1	0	0	6	
	oppg2	0	0	2	
	oppg3	3	1	4	
	sum	3	1	12	16
4	oppg1	1	2	3	
	oppg2	6	2	1	
	oppg3	4	2	0	
	sum	11	6	4	21
5	oppg1	2	4	1	
	oppg2	4	0	0	
	oppg3	4	0	0	

	sum	10	4	1	15
6	oppg1	6	3	0	
	oppg2	3	4	0	
	oppg3	3	0	0	
	sum	12	7	0	19
7	oppg1	2	3	2	
	oppg2	4	1	2	
	oppg3	-	-	-	
	sum	ble ikke fullført			
8	oppg1	5	5	0	
	oppg2	0	3	0	
	oppg3	6	0	0	
	sum	11	8	0	19
9	oppg1	1	2	0	
	oppg2	8	0	0	
	oppg3	2	0	0	
	sum	11	2	0	13
10	oppg1	1	7	1	
	oppg2	0	4	4	
	oppg3	1	8	2	
	sum	2	19	7	28
11	oppg1	1	5	0	
	oppg2	4	3	0	
	oppg3	2	0	0	
	sum	7	8	0	15
12	oppg1	1	5	0	
	oppg2	0	2	2	
	oppg3	1	5	0	
	sum	2	12	2	16
13	oppg1	1	5	0	
	oppg2	3	7	0	
	oppg3	4	0	0	
	sum	8	12	0	20
14	oppg1	1	3	0	
	oppg2	11	2	0	
	oppg3	1	3	0	

	sum	13	8	0	21
15	oppg1	4	3	0	
	oppg2	4	1	2	
	oppg3	3	3	2	
	sum	11	7	4	
16	oppg1	0	9	0	
	oppg2	-	-	-	
	oppg3	-	-	-	
	sum	ble ikke fullført			
17	oppg1	4	3	0	
	oppg2	2	2	0	
	oppg3	4	8	0	
	sum	10	13	0	23
18	oppg1	-	-	-	
	oppg2	-	-	-	
	oppg3	-	-	-	
	sum	ble ikke fullført			
19	oppg1	5	4	0	
	oppg2	5	7	0	
	oppg3	5	7	0	
	sum	15	18	0	33
20	oppg1	6	10	0	
	oppg2	-	-	-	
	oppg3	-	-	-	
	sum	ble ikke fullført			
21	oppg1	0	0	0	
	oppg2	5	4	0	
	oppg3	2	2	0	
	sum	7	6	0	13
22	oppg1	0	1	0	
	oppg2	0	1	0	
	oppg3	0	9	0	
	sum	0	11	0	11
23	oppg1	1	0	2	
	oppg2	0	4	0	
	oppg3	0	3	0	

Antall Tale-turns			98	56,6 %
Antall MM turns			15	8,7 %
Totalt antall turns			173	100,0 %
OPPGAVE 2:			Antall	Prosent
Antall Penn-turns			69	43,7 %
Antall Tale-turns			67	42,4 %
Antall MM turns			22	13,9 %
Totalt antall turns			158	100,0 %
OPPGAVE 3:			Antall	Prosent
Antall Penn-turns			55	42,0 %
Antall Tale-turns			61	46,6 %
Antall MM turns			15	11,5 %
Totalt antall turns			131	100,00 %
I ALT			Antall	Prosent
Antall Penn-turns			184	39,8 %
Antall Tale-turns			226	48,9 %
Antall MM turns			52	11,3 %
Totalt antall turns			462	100,0 %
Gjennomsnittlig antall turns				
oppgave 1	173/28	6,18	28 brukere gjennomførte oppgave 1.	
oppgave 2	158/26	6,08	26 brukere gjennomførte oppgave 2.	
oppgave 3	132/23	5,74	23 brukere gjennomførte oppgave 3.	
hele testen		17,99		

10.1.3 Dialoger med sekvensiell eller simultan koordinert multimodal interaksjon, full tabell

Bruker		MM dialog	Dialoger tot
1	oppg1		2
	oppg2		3
	oppg3	-	-
	sum		5
	sum, pst		71,43 %
2	oppg1	-	-
	oppg2	-	-
	oppg3	-	-
	sum		0

3	oppg1	5	5
	oppg2	2	3
	oppg3	4	4
	sum	11	12
	sum, pst	91,67 %	
4	oppg1	2	3
	oppg2	3	4
	oppg3	2	2
	sum	7	9
	sum, pst	77,78 %	
5	oppg1	1	5
	oppg2	0	1
	oppg3	0	3
	sum	1	9
	sum, pst	11,11 %	
6	oppg1	1	3
	oppg2	2	4
	oppg3	1	4
	sum	4	11
	sum, pst	36,36 %	
7	oppg1	4	6
	oppg2	0	1
	oppg3	-	-
	sum	4	11
	sum, pst		
8	oppg1	2	2
	oppg2	1	1
	oppg3	2	5
	sum	5	8
	sum, pst	62,50 %	
9	oppg1	0	2
	oppg2	1	5
	oppg3	0	1
	sum	1	8
	sum, pst	12,50 %	
10	oppg1	2	4
	oppg2	1	4
	oppg3	3	8

	sum	6	16
	sum, pst	37,50 %	
11	oppg1	1	1
	oppg2	0	2
	oppg3	0	5
	sum	1	8
	sum, pst	12,50 %	
12	oppg1	0	1
	oppg2	2	3
	oppg3	2	5
	sum	4	9
	sum, pst	44,44 %	
13	oppg1	1	2
	oppg2	2	4
	oppg3	5	8
	sum	8	14
	sum, pst	57,14 %	
14	oppg1	1	1
	oppg2	0	6
	oppg3	1	3
	sum	2	10
	sum, pst	20,00 %	
15	oppg1	1	3
	oppg2	2	4
	oppg3	2	3
	sum	5	10
	sum, pst	50,00 %	
16	oppg1	2	2
	oppg2	0	0
	oppg3	0	0
	sum	2	2
	sum, pst	100,00 %	
17	oppg1	1	1
	oppg2	2	3
	oppg3	3	5
	sum	6	9
	sum, pst	66,67 %	

18	oppg1	-	-
	oppg2	-	-
	oppg3	-	-
	sum	0	0
19	oppg1	2	4
	oppg2	3	7
	oppg3	4	5
	sum	9	16
	sum, pst	56,25 %	
20	oppg1	1	7
	oppg2	0	0
	oppg3	0	0
	sum	1	7
	sum, pst	14,29 %	
21	oppg1	0	1
	oppg2	2	4
	oppg3	1	1
	sum	3	6
	sum, pst	50,00 %	
22	oppg1	0	0
	oppg2	0	1
	oppg3	0	5
	sum	0	6
	sum, pst	0,00 %	
23	oppg1	1	2
	oppg2	2	4
	oppg3	0	3
	sum	3	9
	sum, pst	33,33 %	
24	oppg1	1	1
	oppg2	0	1
	oppg3	0	0
	sum	1	2
	sum, pst	50,00 %	
25	oppg1	0	1
	oppg2	1	2
	oppg3	2	3

	sum	3	6
	sum, pst	50,00 %	
26	oppg1	0	0
	oppg2	1	2
	oppg3	1	3
	sum	2	5
	sum, pst	40,00 %	
27	oppg1	0	1
	oppg2	3	4
	oppg3	0	1
	sum	3	6
	sum, pst	50,00 %	
28	oppg1	2	7
	oppg2	0	0
	oppg3	5	5
	sum	7	12
	sum, pst	58,33 %	
29	oppg1	0	0
	oppg2	2	2
	oppg3	2	3
	sum	4	5
	sum, pst	80,00 %	
30	oppg1	2	2
	oppg2	2	4
	oppg3	2	2
	sum	6	8
	sum, pst	75,00 %	
	Total:	110	230
	Total, prosent	47,83 %	

OPPGAVE 1		
MM dialoger:	35	50,0 %
Antall dialoger totalt:	70	
OPPGAVE 1		
MM dialoger:	37	
Antall dialoger totalt:	80	46,3 %

OPPGAVE 1		
MM dialoger:	42	
Antall dialoger totalt:	87	48,3 %

10.2 SPØRRESKJEMAETS OPPGAVE 1

” Dette er det jeg tenker på nå/umiddelbar respons:”

Bruker: 1

God trykking. Hadde vært en bra applikasjon med kun trykking. Det blir bare masse rot med tale. Vanskelig å vite om den prosesserer eller om den ikke har hørt ytringen. En ”prosessindikator” ville vært bra om man ikke får umiddelbart svar.

Bruker: 2

Fikk aldri gjennomført testen

Bruker: 3

Stemmegjenkjenninga var ikke helt bra. Man må gjenta seg selv. Å skifte vindu går tregt. Tung navigering i kartet! Ønskelig å kunne nulle ut spørringen så langt, og forbli i samme subkart.

Bruker: 4

”Fæncy gadget”, men treig. Systemet er treigt og trenger videre utvikling. Man blir utålmodig ved mangelfull tilbakemelding. Jobber systemet? ”Zoom ut” måtte gjøres for ofte.

Bruker: 5

Stort potensiale. Det er fortsatt tungvint å bruke om stasjonsnavn er ukjent og man ikke er god på kart. Hva med en jobbe-indikator? Savner å panorere i kartet, samt navn på stasjonene uten å trykke. Store ting (for eksempel Oslo Spektrum) i nærheten av stopp burde være kjent for systemet.

Bruker: 6

Utrolig kjekt med til- og fra-funksjon. Hadde problemer med å vite navn på holdeplassene ved oppgave 3. Ikke god tilbakemelding om jeg hadde trykt F og T. Det kom mange feilmeldinger når de først kom. Visste ikke at dette var mulig, jeg vil gjerne ha noe sånt. Oslo i firkanter var et problem. Zoom tok lang tid.

Bruker: 7

Den hang veldig ofte! Det tok lang tid før tilbakemelding kom.

Bruker: 8

Dårlig talegjenkjenning. Litt tungvint. Fungerte bra å trykke på bussholdeplassene i stedet for å finne disse på ei liste. ASR er genialt når det funker, men ikke ellers.

Bruker: 9

Foretrekker tastefunksjon. Tasting er forholdsvis enkelt for en med teknisk bakgrunn. Systemet henger av og til. Dette er frustrerende for en uten teknisk bakgrunn. Brukervennlighet på en skala fra 1 til 10: 6

Bruker: 10

- Vanskelig å finne navn å busstasjonsnavn. Man kan navnene på steder, ikke holdeplasser.
- Kronglete – ”Fra” eller ”til” kan ikke sjongleres med i tid (Det første som gis inn blir nesten alltid forstått som ”fra-stasjon”).
- Peik&Preik: Tar tasting viktigere enn tale.
- Gjorde *noen* av tingene jeg ville.

Bruker: 11

- Savnet funksjon for gjentakelse av siste melding
- Utydelig å lese tekst
- Ønsker bedre framstilling av info som gis. Presisering av busskifte
- Lettere å bruke kartet når jeg lette etter alternative stasjoner (i forhold til å bruke stemmen)
- Lettere å slå opp i busstabell enn å starte alt dette

Bruker: 12

Artig. Funker relativt bra, helt til ”Slottsparken”.

Savner ”vis meg det-og-det-området/ bussholdeplasser i nærheten av steder. Å vise bussholdeplassnavn fungerte dårlig (Bussholdeplassnavn kommer under høyre hånd som holder penna...). Store deler av nordmenn sier TE, ikke TIL. Virker som første navn sagt blir antatt å være ”fra-stasjon”.

Bruker: 13

Veldig bra, nyttig verktøy. Finpuss trengs: Når jobber den? Den avbryter seg selv. GUI. Ble imponert over talegjenkjenningen. Fikk en del ”Forstår ikke din forespørsel”.

Bruker: 14

Gikk fint å prate. Tungvint trykking. Det er dumt hvis du ikke vet hvor du skal. Kanskje lik farge på F-knapp og buss_stasjon, og til-knapp og buss_stasjon_2?

Bruker: 15

ASR-modul litt følsom på dialekt. Dårlig trykksensitiv skjerm. Ikke bra oppdateringshastighet. Liker ikke å måtte vente på ny kartoppdatering ved trykking av F- og T-knappen. Syntesen ikke super. Kunne vært bedre nøyaktighet på peking.

Bruker: 16

Plaget med ofte kræsje. Irritert over treg respons. Foretrekker uten stemme-output. Liker å kunne si holdeplasser direkte.

Bruker: 17

Tryke og prate fungerer bra med litt trening. Treg oppdatering på å få opp holdeplass 1 og holdeplass 2. Generelt ganske greit.

Bruker 18 kom aldri i gang med testen på grunn av omfattende systemproblemer.

Bruker: 19

Generelt veldig bra – noe problemer med å forstå dialekt. Bra av systemet å forstå en av to, og så spørre etter nummer to. Interessert i å få se alle bussnr som passerer en holdeplass på samme måte som man kan finne navnet til holdeplassen

Bruker: 20

Systemet gir brukeren begrensa valg. Kan ikke se på flere valg – Brukeren må ut til oversiktskart på nytt for hver spørring. Tungvint – dårlig gjenkjenning. Savner angreknapp på angivelse av holdeplass. Treig respons.

Bruker: 21

- Litt fortvilende når det tar lang tid
- Hvorfor ikke få repetert talebeskjed?
- Jobber systemet/henger det i oppgave 3? Hva med ”browserikon” som surrer når systemet jobber?
- Veldig greit med tale og trykking – vet ikke alltid hvor ting er på kartet.

Bruker: 22

Fungerte som ønsket (Men kjenner bare begrenset med steder). ASR fungerte bra. Kanskje på grunn av lite spennende dialekt?

Bruker: 23

- Veldig greit å finne fram om litt kunnskap om hvor du vil.
- Zoom ut ikke alltid åpenbar (ikke vist i demo, ikke selvforklarende)
- Veldig greit med snakk og trykk

Bruker: 24

- God idé.
- Dårlig gjenkjenning.
- Områder, ikke bare holdeplasser, burde kjennes til
- Få opp tabeller, ikke bare neste buss
- Bedre kart?

Bruker: 25

Hang seg litt ofte. Vanskelig å orientere seg på kartet. Klienten oppfatta ikke alltid trykkinga. Responstid.

Bruker: 26

Bra, skulle gjerne hatt noe slikt. Systemet gjør nytte for seg. Systemet hjelper en godt på nye plasser.

Bruker: 27

Tøft, men tungvint at det ikke fungerte hele tida. I siste oppgave har ikke snakkefunksjonen mye for seg. Burde finnes en knapp for å skru av stemme.

Bruker: 28

- Flere stasjoner!
- Oppdatering. Oppdateres den eller venter den på inndata?
- Rimelig selvforklarende.
- Med litt opplæring gikk det bedre
- Bredere valgmuligheter – etter klokka 15?
- Irriterende zoom ut. Gli på subkart med penn i kanten for å panere er ønskelig!

Bruker: 29

- GUI går seint
- Veldig greit å finne fram for folk som ikke er kjent i byen.
- Greit å trykke og prate. Slipper å zoome ut.
- Irriterende å bruke zoom ut / inn når man har en rutine-forespørsel
- Favoritter/hurtigmeny?

Bruker: 30

- Lang responstid. Trykking på GUI, få svar...
- Irriterende stemme
- Mangler stedsnavn som ikke er stasjoner
- Savner mest: raskere responstid. Det ville hjulpet med en indikator på hva som skjer.

10.3 ANNEN INFORMASJON OM BRUKERNE

Testbrukerne kommer fra:

	Alder	Antall
Arendal (2)	20	1
Birkeland (2)	21	1
Brumunddalen	22	3
Eidsvoll	23	1
Fredrikstad	24	13
Hamar	25	4
Hønefoss	26	4
Kristiansand	27	0
Larvik	28	1
Lørenskog	SUM	28

Oslo (4)

Rognan

Røros

Rørvik

Salangen

Ski

Er du vant med PDA?

Nei: 19

Ja: 9

Hvilken mobiltelefon er du vant til?

Stavanger
Tau
Trondheim (4)
Ålesund

SUM 28

Kjønnsmessig fordeling

Kvinner: 7(8)

Menn: 21 (22)

Hos to brukere ble det ikke foretatt noen test.

Siemens M65
Nokia 3310
Siemens S55
Ericsson R320S (2000-mod)
Nokia 55XX
SonyEricsson T630
Nokia 6680
"Gammel Nokia" (4)
SonyEricsson K700i
"Middels SonyEricsson" (2)
"Ny, billig"
"Nokia, 2 år gammel" (2)
(ingen angivelse)
"Ny" (3)
"Middels" (2)
"Ny Nokia" (2)
"Ny SonyEricsson" (2)

10.4 BRUKERTEST: CASER.DOC

BRUKERFORSØK MULTIMODAL DEMONSTRATOR

Oppgave 1

a) Det er morgen og du skal på jobb på Telenor Fornebu, og lurere på når neste buss går.
Prøv å finne neste bussforbindelse til Telenor Fornebu.

b) Du lurere på når neste buss går hjem igjen.
Prøv å finne neste bussforbindelse hjem fra Telenor Fornebu.

Oppgave 2

a) Du sitter hjemme en kveld og ønsker å bli mer kjent med noen av bussene som stopper i rimelig gangavstand fra leiligheten din. Hvordan er bussløsningen er mellom disse og busstoppet ”Jernbanetorget”?

Prøv å finne 2 bussløsninger mellom andre stopp i din bydel og Jernbanetorget.

Oppgave 3

I morgen er det en spesiell dag.

Du ønsker å ta en lørdag i Oslo for å besøke noen nye steder. Du har et kart du har henta på turistinformasjonen, og ønsker nå å sjekke en del bussforbindelser.

Prøv å finne en bussforbindelse mellom Aker Brygge og Oslo Spektrum.

Prøv å finne en bussforbindelse mellom Oslo Domkirke og Nobelinstituttet (sørvestre hjørne av slottsparken).

10.5 ANNEN INFO VIST TIL BRUKEREN VED TEST

Innlagte bussholdeplasser:

Adamstuen	Kampen	Skoyen stasjon
Agathe Grondhahls gate	Klingenberg	Slottsparken
Aleksander Kiellands plass	Kobenhavngata	Sofienberg
Ankertorget	Kongens gate	Sofies plass
Anton Schjots gate	Lakkegata skole	Solli
Apotekergata	Lapsetorvet	St.Hanshaugen
Arkitekt Rivertz plass	Lilleoyveien	Stensberggata
Bankplassen	Lovisenberg	Stortorvet
Brugata	Lysaker stasjon	Telenor Fornebu
Bryggetorget	Majorstuen	Torshov
Bussterminalen Gronland	Marienlyst	Torshovsparken
Carl Berners plass	Maritim	Toyen
Evald Ryghs gate	Mollerveien	Toyen kirke
Fagerheimsgata	Munchmuseet	Toyen stasjon
Falck Ytters plass	Munkegata	Toyengata
Fornebu parken	Nationalteateret	Ullevaal sykehus
Galgeberg	Observatoriegata	Uranienborgveien
Harald Hardraades plass	Oslo gate	Vakero
Homans byen	Politihuset	Vestre Aker kirke
Huitfeldts gate	Sinsenterrassen	Vika Atrium
Jernbanetorget	Sjolyst	Voyenbrua
		Aasengata

Andre Unormale begrensninger:

- KUN bussholdeplassnavn gjenkjennes. ”Oslo S”, gateadr. osv. gjenkjennes ikke.
- I forsøket simuleres at brukeren har bopel i Gamlebyen, nær bussholdeplassen ”Harald Hardrådes plass”. Denne holdeplassen er i subkartet nederst til høyre, midt i vinduet.
- Kodeord for få bussinfo: fra... til..., eventuelt ”neste buss”...

10.6 BRUKERTEST: SPØRRESKJEMA FOR INTERVJUET ETTERPÅ

SPØRRESKJEMA ETTERPÅ

Dette er det jeg tenker på nå/umiddelbar respons:

Kan du tenke på andre situasjoner enn bussruteinformasjon hvor multimodalitet kan være nyttig?

(Hvilken interaksjonsform (kun tale, kun taster, peik&preik) synes du var greiest å bruke?)

Hva var mest naturlig for deg å bruke – tale? Peking? Peik&Preik? Evt situasjonsbetinget?

Er du i noen grad villig til å snakke normert for å øke suksessraten til stemmegjenkjenneren når denne ønskes brukt? I hvor stor grad?

For talemødet: Hvordan vurderer du å bruke z minutter til høytlesing for å opprette en personlig profil, om du vet at gevinsten er økning i gjenkjenningsrate fra xx% til omkring yy%?

Liker du å bruke talegrensesnittet? Hvis ja: fortsetter du å bruke det om antall korrekte svar er...

- 9 av 10
- 4 av 5
- 2 av 3
- 1 av 2
- Annet:

Hva likte du godt:

Hva var du minst fornøyd med:

Er du vant med PDA?

Hvilken mobiltelefon(-er) er du vant til?

Alder:

Kjønn:

Kommer fra (grovt sett):

10.7 TESTLEDERENS BRUKERTESTEN

NOTERINGSSKJEMA I

HEMMELIG NOTERINGSSKJEMA

Testbruker nr:

PRATEMÅTE (Stamming, nøling, fylte pauser, restart, høyt og tydelig/lavt og snøvlete...)

DIALEKT/SOSIOLEKT-KJENNETEGN (Rulle-r/skarre-r, tykk/tynn L...)

GENERELT AVVIKENDE OPPFØRSEL (urolig...)

PRATEVOLUM

NÆR/FJÆRN

FRA

MIKROFON

10.8 TESTLEDERENS INTRODUKSJONEN

HUSKELAPP

FOR

INTRO

”Kameratintroduksjon”

Brukeren skal ha foran seg: ”Unormale begrensninger brukeren må vite”
”Bussholdeplasser”
Turistkart over Oslo

- Min nye PDA har et program for bussinfo i Oslo
- Det smarte: både bruke tale og trykking – og det kan du også gjøre samtidig!
- taledelen er litt fin på det. Snakk ”*litt pent*”, så går det bra.

Trykking. Øvre høyre subkart. A og B i samme rute. *Hvis man vil angi til-holdeplassen først, trykk ”T”.*

Tale. ”*Neste buss fra Telenor Fornebu til Majorstuen*”

Peik&Preik: Øvre høyre subkart: A. ”neste buss *til Majorstuen, herifra* (subkart øverst til høyre).”

La forsøkspersonen teste for noen minutter.

Kontrollør at brukeren har fått med seg:

- Bare bussholdeplassnavn er lagret i applikasjonen
- hjem: Harald Hardrådes plass, i Gamlebyen.
- ”neste buss” eller ”fra...til”.

Spørsmål?

Mine hemmelige triks:

(For å se navnet på en busstasjon uten å starte en spørring: pek på en holdeplass og dra penna bort før du slipper.)

10.9 GRAMATIKK

```
grammar      ENU_mustGrammar;
language     nob_standard;
conceptset{<Action>, <GOF>, <Buss_stasjon>, <Buss_stasjon2>, <Temporal>, <Parallel_Ref>, <Confirmation>, <Filler>}
<Action>= hjelp           {Help}
| avslutt                 {avslutt}
| zoomut                  {zoom_out}
| zoom ut                 {zoom_out}
| startigjen              {home}
| startpånytt             {home}
| nysessjon               {home}
| hovedkartet             {zoom_out};
<Confirmation>= ja        {Yes}
| nei                     {No};
<GOF>= bussholdeplasser   {Buss_stasjon}
| buss stasjoner          {Buss_stasjon}
| holdeplasser            {Buss_stasjon};
<Buss_stasjon>= huitfeldts {Huitfeldts gate}
| observatoriegata        {Observatoriegata}
| bryggetorget            {Bryggetorget}
| vika atrium              {Vika Atrium}
| klingenberg             {Klingenberg}
| bankplassen             {Bankplassen}
| jernbanetorget          {Jernbanetorget}
| kongens gate            {Kongens gate}
| stortorvet              {Stortorvet}
| bussterminalen grønland {Bussterminalen Gronland}
| tøyengata               {Toyengata}
| politihuset             {Politihuset}
| oslo                    {Oslo gate}
| kampen                  {Kampen}
| galgeberg               {Galgeberg}
| harald hardrådes        {Harald Hardraades plass}
| sankt halvards          {St. Halvards plass}
| munkegata               {Munkegata}
| lapsetorvet             {Lapsetorvet}
| solli                   {Solli}
| sofies                  {Sofies plass}
| slottsparken            {Slottsparken}
| nationalteateret        {Nationalteateret}
| sankthanshaugen         {St.Hanshaugen}
| stensberggata           {Stensberggata}
| apotekergata            {Apotekergata}
| møllerveien             {Møllerveien}
| brugata                 {Brugata}
| ankertorget             {Ankertorget}
| tøyen kirke             {Tøyen kirke}
| tøyen                   {Tøyen}
| munchmuseet             {Munchmuseet}
| lakkegata skole         {Lakkegata skole}
```

tøyen stasjon	{Toyen stasjon}
sofienberg	{Sofienberg}
majorstuen	{Majorstuen}
anton schjøths	{Anton Schjots gate}
adamstuen	{Adamstuen}
marienlyst	{Marienlyst}
vestre aker	{Vestre Aker kirke}
ullevål sykehus	{Ullevaal sykehus}
falck ytters	{Falck Ytters plass}
aleksander kiellands	{Aleksander Kiellands plass}
evald ryghs	{Evald Ryghs gate}
arkitekt rivertz	{Arkitekt Rivertz plass}
lovisenberg	{Lovisenberg}
vøyenbrua	{Voyenbrua}
carl berners	{Carl Berners plass}
fagerheimsgata	{Fagerheimsgata}
christies	{Fagerheimsgata}
københavnsgata	{Kobenhavnsgata}
rosenlundsgata	{Vogts gate}
torshovsparken	{Torshovsparken}
torshov	{Torshov}
åsegata	{Aasengata}
agathe grøndhahls	{Agathe Grondhahls gate}
sinsenterrassen	{Sinsenterrassen}
telenor	{Telenor Fornebu};
<Buss_stasjon2>= til huitfeldts	{Huitfeldts gate}
til observatoriegata	{Observatoriegata}
til bryggetorget	{Bryggetorget}
til vika atrium	{Vika Atrium}
til klingenberg	{Klingenberg}
til bankplassen	{Bankplassen}
til jernbanetorget	{Jernbanetorget}
til kongens gate	{Kongens gate}
til stortorvet	{Stortorvet}
til bussterminalen grønland	{Bussterminalen Gronland}
til tøyengata	{Toyengata}
til politihuset	{Politihuset}
til oslo	{Oslo gate}
til kampen	{Kampen}
til galgeberg	{Galgeberg}
til harald hardrådes	{Harald Hardraades plass}
til sankt halvards	{St. Halvards plass}
til munkegata	{Munkegata}
til lapsetorvet	{Lapsetorvet}
til solli	{Solli}
til sofies	{Sofies plass}
til slottsparken	{Slottsparken}
til nationalteateret	{Nationalteateret}
til sankthanshaugen	{St.Hanshaugen}
til stensberggata	{Stensberggata}
til apotekergata	{Apotekergata}
til møllerveien	{Møllerveien}
til brugata	{Brugata}
til ankertorget	{Ankertorget}
til tøyen kirke	{Toyen kirke}
til tøyen	{Toyen}
til munchmuseet	{Munchmuseet}

til lakkegata skole	{Lakkegata skole}
til tøyen stasjon	{Tøyen stasjon}
til sofienberg	{Sofienberg}
til majorstuen	{Majorstuen}
til anton schjøths	{Anton Schjots gate}
til adamstuen	{Adamstuen}
til marienlyst	{Marienlyst}
til vestre aker	{Vestre Aker kirke}
til ullevål sykehus	{Ullevaal sykehus}
til falck ytters	{Falck Ytters plass}
til aleksander kiellands	{Aleksander Kiellands plass}
til evald ryggs	{Evald Ryggs gate}
til arkitekt rivertz	{Arkitekt Rivertz plass}
til lovisenberg	{Lovisenberg}
til vøyenbrua	{Vøyenbrua}
til carl berners	{Carl Berners plass}
til fagerheimsgata	{Fagerheimsgata}
til christies	{Fagerheimsgata}
til københavngata	{Kobenhavngata}
til rosenlundsgata	{Vogts gate}
til torshovsparken	{Torshovsparken}
til torshov	{Torshov}
til åsengata	{Aasengata}
til agathe grøndhahls	{Agathe Grøndhahls gate}
til sinsenterrassen	{Sinsenterrassen}
til telenor	{Telenor Fornebu} ;
<Temporal>= når	{Bussrute}
beskrivelse	{Bussrute}
hvaslags	{Bussrute}
info	{Bussrute}
neste	{Bussrute}
forrige	{Bussrute}
neste buss	{Bussrute}
jeg vil reise	{Bussrute}
buss	{Bussrute}
fra	{Fra}
til	{Til};
<Parallel_Ref>= ogher	{Parallell}
herog	{Parallell}
ogdenne	{Parallell}
ogfordenne	{Parallell}
ogdette	{Parallell}
ogfordette	{Parallell};
<Filler>=hvaer	{Garbage}
hvorer	{Garbage}
hvilken	{Garbage}
gimeg	{Garbage}
vis	{Garbage}
finn	{Garbage}
går	{Garbage}
her	{Garbage}
hit	{Garbage}
gate	{Garbage}
plass	{Garbage};