

RESULTATRAPPORT: Lydstyring-prosjektet

Skrevet av: Magne Lunde

Sist oppdatert: 30.12.2010

Bakgrunn og mål

Sommeren 2007 gjennomførte MediaLT med støtte fra Norges Forskningsråd forprosjektet STEMmestyrte INTeraksjon (STEMINT, prosjektnr 176984/i40), som identifiserte behov for talestyring av PC. I det oppfølgende hovedprosjektet Stemmestyring i MULTimodal Dialog (SMUDI, prosjektnr. 183215/S10) jobbes det med å integrere norsk stemmestyring i multimodale grensesnitt. Som en del av dette ble det svensk/norske samarbeidsprosjektet VOMOTE initiert, støttet av Innovasjon Norge og Tillvekstværket (prosjektnr 145589), med fokus på å virkeliggjøre et norsk produkt for stemmestyring av PC. Med bakgrunn i disse prosjektene ble talegjenkjenningsproduktet VOMOTE lansert i oktober 2010.

Som en følge av arbeidet med norsk taleteknologi fikk MediaLT spørsmål fra fagmiljøer og brukere om mennesker med talevansker også kan benytte lyder og/eller utydelig språk til å styre teknologi. Vi fant lite forskning eller erfaring knyttet til dette, og MediaLT tok initiativet til forprosjektet Lydstyring. Hovedmålet i prosjektet var å:

”Undersøke mulighetene for lydstyring av PC, og legge grunnlaget for et utviklingsprosjekt, dersom dette er mulig og hensiktsmessig.”

Hovedmålet ble delt inn i følgende delmål:

1. Analysere internasjonal FoU status på området lydstyring.
2. Definere funksjonalitetsmatrise for lydstyring av PC.
3. Bestemme basisprofiler med nøkkelfunksjonalitet.
4. Vurdere mulig teknologi for lydstyring.
5. Etablere internasjonalt samarbeid.
6. Legge grunnlaget for et hovedprosjekt, dersom dette er hensiktsmessig.

Målgruppe

Hovedmålgruppen var personer som har talevansker og problemer med å benytte PC ved hjelp av standard brukergrensesnitt. Med andre ord innbefatter målgruppen alle mennesker som vil dra nytte av lydstyring fremfor andre interaksjonsformer. Aktuelle brukergrupper kan for eksempel være mennesker med Cerebral Parese, hørselshemmede med bevegelseshemminger og slagpasienter. Også fremmedspråklige med tilsvarende funksjonshemminger og behov faller inn under denne målgruppen.

Prosjektresultater

Prosjektresultatene samsvarer med oppsatte delmål:

Analysere internasjonal FoU: Referanser er samlet og lagt på prosjektets nettsider:

<http://www.medialt.no/lenkerreferanser/847.aspx>

Videre er FOU-situasjonen analysert i en egen statusrapport:

<http://www.medialt.no/statusrapport/1001.aspx>

Definere funksjonalitetsmatrise for lydstyring av PC: personas ble brukt som metode for å finne fram til hvilken funksjonalitet som typisk var ønsket i et lydstyringssystem. Personas vil si detaljerte beskrivelser av fiktive personer, som kan fungere som gode eksempler på hva som særpreger den aktuelle brukergruppen. Med bakgrunn i kunnskapen til SIKTE, Sunnaas og CP-foreningen ble i alt fire personas laget. Dette arbeidet tydeliggjorde at det er ønskelig med et bredt utvalg av tilgjengelig funksjonalitet. Dessuten ble det klart at brukerne av en lydstyringsløsning vil ha svært ulike behov og muligheter, og at en slik tilpasning derfor bør skje på individnivå. Dette støttes av resultater fra forskningen på området. Av denne grunn ble det vurdert som lite formålstjenlig å definere en funksjonalitetsmatrise. I stedet bør det legges til rette for en størst mulig grad av individuell tilpasning.

Bestemme basisprofiler med nøkkelfunksjonalitet: av samme grunn som i forhold til behovet for en funksjonalitetsmatrise, vurderte prosjektgruppen delmålet definering av basisprofiler som uhensiktsmessig. I stedet bør det tilstrebes å utvikle et system som gjør mulighetene for individuelle tilpasninger størst mulig. Med grunnlag i arbeidet med personas var det enighet i prosjektgruppen om at ønsket løsning kan beskrives på følgende måte:

- Brukeren gis tilgang til en stor "verktøykasse" med ulike funksjoner systemet er i stand til å utføre. For eksempel kan de funksjonene som inngår i talegjenkjenningsproduktet VOMOTE være en slik verktøykasse.
- Brukeren skal kunne benytte lyder/talekommandoer enten i tillegg til eller i stedet for de styringsmetodene han/hun i dag benytter for å styre PC-en. For eksempel kan løsningen brukes sammen med ulike bryterløsninger og øyestyring.

- Brukeren skal kunne velge, sammen med fagmiljø/foresatte/kontaktpersoner, hvilke lyder/kommandoer han/hun vil si til PC-en.
- Brukeren skal kunne velge, sammen med fagmiljø/foresatte/kontaktpersoner, hvilke funksjoner PC-en skal gjøre og knytte dette opp mot en lyd/talekommando.
- Hva brukeren vil være i stand til å gjøre, avhenger av hvor mange ulike lyder brukeren er i stand til å lage. Kan hun/han lage ti ulike lyder, kan hun/han plukke ut de ti mest ønskede funksjonene og knytte lydene til disse funksjonene.
- Ønsker brukeren å endre de funksjonene som er valgt, kan hun/han enkelt gjøre dette ved å mappe en eller flere lyder opp mot de nye funksjonene hun/han ønsker (men dette betinger da at like mange av de tidligere valgte funksjonene må velges bort).

Vurdere mulig teknologi for lydstyring: Et av de sentrale spørsmålene i forhold til verdien av en lydstyringsløsning er hvor mange lyder en typisk bruker er i stand til å lage, samt om hun/han vil klare å reprodusere disse lydene godt nok til at systemet makter å skille lydene fra hverandre. Siden brukergruppen er svært differensiert, er det ikke mulig å gi noe klart svar på dette spørsmålet, men det ble vurdert som hensiktsmessig å gjøre opptak av lyder fra typiske brukere, for å kunne vurdere de teknologiske mulighetene for lydgykjennning. Lydopptak ble derfor planlagt. Parallelt fortsatte imidlertid arbeidet med å kartlegge internasjonalt arbeid på området, og som et ledd i dette arbeidet deltok prosjektleder Miriam Nes Begnum på konferansen International Computer Conference for Handicapped People (ICCHP) i juli 2010. Her ble kontakt etablert med Foad Hamidi fra universitetet i Yorku i Canada. På konferansen presenterte Hamidi paperet: "CanSpeak: A Customizable Speech Interface for People with Dysarthric Speech". Siden Hamidi og hans kollegaer allerede hadde gjennomført lydopptak, ble resultatene fra hans arbeid vurdert som tilstrekkelige for vårt formål og de planlagte lydopptakene ble skrinlagt.

Testene ved Yorku University ble gjort med fire personer og et vokabular på 47 ord. Uten tilpasning av vokabularet lå gjenkjenningen mellom 30 og 56 %. Blant personer uten talevansker var resultatet 94 prosent. Med tilpassing økte gjenkjenningsraten radikalt til 84,3 %. De desidert beste resultatene ble oppnådd der familie, lærere, pleiepersonell eller talespesialister ble involvert. Det å kun basere seg på brukeren selv ved definering av hensiktsmessige fraser ga minimal bedring. Derimot ble gjenkjenningsraten doblet, ved å få med miljøet rundt brukerne. Ved individuelle tilpasninger av systemet vil derfor vurderinger av uttalevansker fra talespesialister kunne være verdifull input.

Brukertestene i SMUDI-prosjektet viste at funksjonshemmedes utfordringer knyttet til mikrofonbruk var langt større enn vi hadde forutsett. På grunn av at en lydstyringsløsning rettet seg inn mot bevegelsehemmede med talevansker, forsterket dette utfordringene ytterligere, og av denne grunn fant vi det nødvendig å inkludere mikrofontesting i de teknologiske vurderingene av løsningen. Resultatene fra testingen er samlet i rapporten: Test av mikrofoner i prosjektene SMUDI og Lydstyring (<http://www.medialt.no/test-av-mikrofoner-i-prosjektene-smudi-og-lydstyring/1003.aspx>). Denne testingen viste tydelig at det var behov for ytterligere arbeid på fagfeltet, og prosjektet Smidige mikrofon og bryterløsninger

(<http://www.medialt.no/smidig-smidige-mikrofon-og-bryterloesninger/966.aspx>, prosjektnr 208103) ble derfor startet den 1. september 2010.

I prosjektet prøvde vi å finne en eksisterende lyd-gjenkjenner, men ut fra de undersøkelsene vi har foretatt nasjonalt og internasjonalt tyder mye på at hensiktsmessige løsninger ikke finnes. Funnene fra undersøkelsen er oppsummert i statusrapporten:

<http://www.medialt.no/statusrapport/1001.aspx>. Vår konklusjon er derfor at: Skal det lages en løsning for lydstyring av teknologi, må det utvikles en spesifikk lyd-gjenkjenner. Gjennom vårt samarbeid med det svenske selskapet Veridict ble kontakt etablert med det svenske selskapet Hørselsbroen. I samarbeid med Hørselsbroen ble både kostnader og oppgaver i et utviklingsprosjekt av en lyd-gjenkjenner definert, og dermed var det teknologiske grunnlaget lagt for et hovedprosjekt.

Etablere internasjonalt samarbeid: Med bakgrunn i analysen av den internasjonale FOU-situasjonen på fagområdet utkrystalliserte to miljøer seg: Yorku University og Hørselsbroen. Samarbeid ble etablert med begge disse miljøene; Yorku University som forskningspartner og Hørselsbroen som utviklingsmiljø. Et godt fundament for videre internasjonalt samarbeid er derfor lagt.

Legge grunnlaget for et hovedprosjekt: Høsten 2010 var de fleste byggsteinene i et hovedprosjekt på plass, men signaler fra fagmiljøene gjorde at vi var usikre på hvor stort behovet var for lydstyring av teknologi. Derfor bestemte vi oss for å gjennomføre en målrettet undersøkelse blant fagmiljøene og brukerne. Resultatene fra undersøkelsen er oppsummert i rapporten: Oppsummering spørreundersøkelse lydstyring (<http://www.medialt.no/oppsummering-spoerreundersoekelse-lydstyring/982.aspx>). Med bakgrunn i resultatene fra spørreundersøkelsen var det enighet i prosjektgruppen om at det foreløpig ikke var grunnlag for å gå videre med et hovedprosjekt, fordi mye tyder på at antall brukere som vil ha glede av en slik løsning er relativt liten.

FoU-oppgaver og sentrale miljøer

Prosjektet ble ledet av MediaLT. I tillegg har de mest sentrale miljøene vært NAV SIKTE, Frambu, Universitetet i Oslo, CP-foreningen og Sunnaas spesialpedagogiske senter.

De sentrale oppgavene i prosjektet har vært å analysere hva som tidligere har vært gjort av FOU-innsats på området, finne fram til hva som er de sentrale FOU-utfordringene, undersøke hvilken funksjonalitet/hvilket innhold som målgruppen ønsker at løsningen skal ha, avklare de teknologiske utfordringene og mulighetene og undersøke hvor stort behovet er for lydstyring av teknologi.

Prosjektgjennomføring og ressursbruk

Ressursbruken i prosjektet er noe mindre enn budsjettet. Dette skyldes at arbeidsinnsatsen til samarbeidspartnerne i prosjektet har vært mindre enn antatt. Noe som igjen har sammenheng med at de planlagte lydopptakene ble skrinlagt, fordi slike opptak allerede var gjennomført ved Yorku University i Canada. Dessuten ble også arbeidsinnsatsen redusert noe på grunn av at det ikke ble satt i gang et arbeid med en søknad om et hovedprosjekt. Prosjektperioden ble forlenget med to måneder. Det var behov for mer tid til å undersøke om det var tilstrekkelig grunnlag for å sette i gang et forsknings- og utviklingsprosjekt med målsetning om å ta fram en lydstyringsløsning.

Betydning/nytteverdi

Dette prosjektet har frambrakt ny kunnskap om hva det vil innebære å utvikle en løsning som gjør det mulig å lyd styre teknologi, samt om hvilke grupper som vil ha nytte av en slik løsning og hvor stort behovet er. Selv om arbeidet med lyd styring av teknologi ikke vil bli videreført, vil både fagmiljøene, brukergruppen og MediaLT ha nytte av denne kunnskapen i sitt videre arbeid med fagfeltet teknologi for funksjonshemmede.

Utnyttelse av resultater og videre arbeid

MediaLT vil gjøre seg nytte av resultatene i prosjektet i sitt videre arbeid med teknologi for funksjonshemmede generelt og med norsk taleteknologi spesielt. Den foreløpige konklusjonen er at målgruppen er for liten til å forsvare en større satsing på lyd styring, men dette er en konklusjon som fortløpende bør revurderes. For det første vil det forskningsmessig være interessant å videreføre arbeidet, fordi vi ennå ikke har kunnskap om hvilken nytte målgruppen vil ha av en slik løsning. Videre vil det også være interessant å se på andre anvendelsesområder for lyd styring, for eksempel i forhold til varsling for hørselshemmede og i forhold til kommunikasjon med dyr. Dessuten om disse anvendelsesområdene kan kombineres med lyd styring for mennesker med talevansker, og om de tre anvendelsesområdene samlet sett kan forsvare et større utviklingsprosjekt.