

Norges første kunstige barnestemme

Torbjørn Nordgård

Lingit AS

Universitetet i Nordland



Hvorfor barnestemmer?

- Viktigste motivasjon er funksjonshemmede som er avhengig av talesyntese i sin kommunikasjon med omverdenen



- Mer naturlig for kommunikasjonssituasjonen at stemmen er tilpasset barnets alder



Hva er adapsjon?

- "Eliciting limited data (< 2 mins) from the subject should sufficiently allow adaptation of an existing synthesizer to his voice." (Anumanchipalli & Black, 2010).
- Å modifisere stemmen til en gitt taler, A, slik at den høres mer ut som en annen taler, B.
- Eller, modifisere en "gjennomsnittsstemme" derivert fra flere talere slik at den høres mer ut som en ny taler
- Hvis man lykkes, kan det rulles ut mange stemmer ved gjenbruk av omfattende opptak av en eller flere stemmer
 - En moderne kunstig stemme har typisk 5 - 10 tusen innleste setninger
 - De adapterte stemmene er basert på små mengder opptak (noen hundre)



Adapasjon og talegjenkjenning

- I systemer for talegjenkjenning / diktering må systemet tilpasses brukeren
- Det er essensielt samme teknikker som brukes til adaptasjon innen talesyntese
- Gjenkjenning: Hva kan man gjøre for at modellene skal tilpasses brukeren av systemet?
- Adaptert talesyntese: Hva kan man gjøre for at modellene som "masterstemmen" er basert på, skal ligne den nye stemmen?



Valg av masterstemme

- Lingit har opptak til tre stemmer:
 - Isak
 - Inger
 - T16 (kvinnestemme som ikke er implementert i ny stemme)
- Inger peker seg ut som en naturlig kandidat
 - Kvinnestemmer "ligner mer" på barnestemmer
- ... Men Isak har en del viktige positive egenskaper:
 - Bedre intonasjon
 - Bedre forståelighet
- Man kan lage en stemme basert på en modell av alle opptakene (sannsynlig utfall)



Eksperimenter så langt

- Vi har laget en masterstemme basert på 500 setninger fra Fondat2-materialet
- Opptakene til 3 kvinnelige talere og 1 mannlig taler er slått sammen til en taleruavhengig masterstemme
- Deretter er har vi brukt 50 setninger fra en annen stemme og laget en adaptert stemme
- Resultatet er ikke så ille, synes vi



Masterstemme



Adaptert stemme



Eksempler fra datagrunnlaget

- Barneopptakene
 - NorDubb Studios
- Den adapterte stemmen ...?
- ... det er den vi jobber med nå



Prosjektfremdrift høsten 2011 (anslag)

	September	Oktober	November	Desember
Tilrettelegge masterstemme				
Studier/utredning adaptasjon				
Opptak av to barn				
Implementasjon				
Alfaversjon				
Testing				
Betaversjon				



Testing

- Betaversjonen som skal være ferdig ved utgangen av 2011, skal testes vinteren og våren 2012
- MediaLT har ansvar for utprøvingen
 - Teknisk rådgivning fra Lingit
 - Haukåsen skole (Oslo), Bretvedt (Oslo) og Møller kompetansesenter (Trondheim)
 - Barn med mangelfull stemme eller lesevansker
- Testfasen forventes å være ferdig i mai 2012
- Deretter evaluering av stemmen
- En konklusjon kan være at teknologien er lovende, men at vi trenger mer data og kunnskap om adaptasjon ...



Leveranse

- Etter testing i januar - april lages en endelig versjon i mai - juni 2012
- Mulige "showstoppere":
 - Tekniske / praktiske problemer
 - Inger/Isak tok nærmere ca to årsverk hver
 - En ny stemme på 3-4 måneder ...
 - Gjentatte opptreninger kan forsinke leveransene



Prosjektdeltakere fra Lingit

- Dyre Meen
 - Analyse, programmering, implementasjon
- Jørn Almberg
 - Kontroll av innlest manuskript
- Torbjørn Nordgård
 - Administrasjon
 - Ressursarbeid ved behov



Andre deltakere

- MediaLT (prosjekteier)
- Bretvedt, Statped Vest, Trøndelag kompetansesenter
- Universitetet i Oslo
- Språkrådet
- Haukåsen skole
- Abilia
- Norsk Lyd- og Blindeskriftsbibliotek
- Cerebral Parese-foreningen



Marked?

- Begrenset
- Ikke mulig uten offentlig støtte til utvikling
 - NFR er viktigste bidragsyter
- Støtteordninger er nødvendig også etter at produktet er ferdig
 - Vedlikehold
 - Tilpasninger til ulike produkter
 - Rette feil
 - Implementere støtte på ulike plattformer
 - ...



Hvorfor ønsker Lingit å arbeide med adapsjon?

- Adapsjon er faglig interessant
 - Utfordrende
 - Åpner for teknologiske nyvinninger
 - Tilrettelegger for nye produkter
 - Muliggjør "spesialsøm" for kunder/brukere som ønsker det
- Barnestemme er faglig utfordrende
 - Akustiske avstander mellom masterstemme og barnestemme(r)
 - Opptakssituasjonen for unge innlesere
 - Intonasjon
 - Evaluering av kompetente og interesserte brukere/veiledere

