UiT Norgga árktalaš universitehta

Åarjelsaemien soptsestimmiesyntese

Katri Hiovain-Asikainen, Maja Lisa Kappfjell, Sjur Nørstebø Moshagen, Thomas Brevik Kjærstad, Ina Therés Andrea Sparrock jïh orre dorjeldhgïele Aanna

Buerie aerede gaajhkesh dovnesh jïh buerie båeteme dan bæjhkoehtæmman. Daan biejjien, rahkan asken **5. biejjien**, mijjieh, Divvun diekie båateme **Tråantese**, åarjelsaemien gïelesynteesem dijjide vuesiehtidh.

Soptsestimmiesynteese lea dïrrege mij tekstide tjamhki låhka. Dïhte tekstem soptsestæmman dorje, jïh nimhtie almetjh maehtieh tekstide goltelidh mah åarjelsaemiengïelesne tjaalasovveme.

Divvunen bieleste libie dan aavosne åadtjodh dam dijjide vuesiehtidh. Sïjhtijibie dellie vuesiehtidh guktie maahta dam dorjeldhgïelem åtnose vaeltedh, mïsse maahta aevhkine sjïdtedh jïh guktie libie buektiehtamme dam darjodh.

Velkommen til lansering av sørsamisk talesyntese.

Talesyntese er en funksjon som leser opp tekst på din digitale enhet. Den gjør altså tekst om til tale.

I denne lanseringa vil vi vise hvordan talesyntese kan brukes, til hva og hvordan vi har jobbet for å lage den.

Divvun

Divvun lea dåehkie Nöörjen arktihke universiteetesne (UiT:esne), mij saemien gïeleteknologijine barka, Gïeleteknojne, saemien gïelejarngine. Divvun gïeleteknologijes dïrregh evtede, orrestahta jïh reerie, maam saemien gïeleseabradahke daarpesje, vuesiehtæmman staeriedimmieprogramme, grammatihkestaeriedimmie, båaloebuertieh, baakoegærjah jïh daelie jis aaj soptestimmiesynteese. Tjïelte- jïh dajvedepartemente beetnegh Divvunasse laevehte.

Divvun jobber med samisk språkteknologi på UiT i samarbeid med Giellatekno. Divvun utvikler praktiske språkteknologiske hjelpemidler, som stavekontroll, grammatikkontroll, tastaturer, ordbøker og taleteknologi. Divvun finansieres av Kommunal- og distriktsdepartementet.

Vielie nuepieh

- Gïelelohkehtimmie
- Lohkeme- jïh tjaelemelohkehtimmie
- Dåarjege tjelmiehtadtjide jïh pleeside
- Tjamki nehtebielide lohkedh
- Saernieh goltelidh mearan maam joem jeatjah dorje.
- Tjoejenassh Tiktok-videjovide
- Gosse maam joem åvtese buektedh, dellie såemies boelhkh

maehtieh saemiengïelesne årrodh, jïlhts eah jïjtjh saemien maehtieh.

- Healsoeviehkievierhtieh mah saemiestieh
- Dubbejovveme saernieh / voiceover Ođđasat-saernine
- Universelle hammoehtimmie

- Lese- og skriveopplæring
- Støtte for blinde og svaksynte
- Høytopplesing frå nettsider
- Lytte til samiske nyheter mens man gjør noe annet
- Ved presentasjon kan man ha med lydfiler, selv om man ikkje kan språket selv.
- Hjelpemidler innen helse
- Dubbing av nyheter i f.eks. Ođđasat
- Støtte til skrivingen - høre gjennom teksten
- Universell utforming (tilgjengelig for alle slags brukere)

Utvikling av talesyntese

- Basert på: 1) et **korpus** med taleopptak og tilpassede tekster, og: 2) **maskinlæring**
  - I korpuset det må være ~ 8–10 timer taleopptak for å få bra resultat
- Prosjektet er utvikla med åpen kildekode – materiale og kode blir publisert
  - Teknisk side (maskinlæring) er basert på FastPitch, som er åpen kildekode og finns tilgjengelig på GitHub: https://fastpitch.github.io/
  - Samme metode kan brukes med alle språk: Divvun har lansert nye talesynteser og for nord- og lulesamisk i 2024
- Utvikling krever kompetanse fra mange felt: lingvistikk, språkteknologi, informatikk, fonologi/fonetikk, lydteknologi, programvareutvikling

Utvikling av talesyntese

- Det finns visse utfordringer når man jobber med minoritetsspråk
- Sammenlignet med større språk er det vanskeligere å finne materiale (lyd og tilpasset tekst) som trengs til prosjektet, men selve den teknologiske implementeringa er den samme som for de store språkene
- På grunn av de få ressursene som finnes for de samiske språkene, må vi sørge for å gjøre god bruk av det materialet vi har
  - Mulighet å bruke arkivmateriale med forskjellige opptaksforhold ved å forbedre og normalisere lyd, som i prosjektet med «Aanna»
  - Transkriberere med høyt nivå av språkkunnskaper i sørsamisk og selve transkriberinger av lydfiler er veldig viktigt – tekstene må stemme svært nøyaktig overens med det som blir sagt på lydmaterialer for å få gode resultater i talesyntese

Utviklingsprosess 1: Lydmateriale og transkribering

- Anna Jacobsen-arkivmateriale
  - Opptak gjort mellom 1989-1993
  - Digitalisert fra kassetter/CD-er (NRK)
  - Materialet inneholder ulike sjangre, bl. a., Bibel, radiosendinger med nåtidsdokumentasjon, eventyr (samisk og oversatt), fortellinger fra AJ sitt liv
  - Hele lydsamlingen ble lyttet igjennom for å finne opptak som passer til talesyntesen
    - Deler med ødelagt lyd, for mye støy og med andre talere enn AJ ble ikke inkludert i syntesekorpuset
- Tekster til lydmaterialet
  - Eksisterende tekster tilpasset lyd (*Don jih daan bijre* – tekstversjon av radiosendinger) eller helt manuell transkribering om det ikke fantes ferdige tekster
  - Transkriberingsarbeidet ble gjennomført i samarbeid med Arkivverket/Samisk Arkiv og Divvun; Thomas, Ina Theres, Maja Lisa
    - Transkribering i ortografisk tekst

Utviklingsprosess 2: Prosessering

- Lydfiler og tekster er prosessert til et talekorpus:
  - Lydfiler ble renset ved å kutte alle unødvendige pauser og bakgrunnsstøy osv.
  - Tekstene ble tilpasset lydfilene: hvis taleren gjentok noe eller leste noe annerledes enn i teksten, ble tekstene redigert slik at de gjenspeilte det som ble sagt
  - Lyden ble så ytterligere renset (fjerning av støy og ekko) og nivåtilpasset slik at hele materialet ble konsistent i kvalitet og lydstyrke
    - Programmer som ble brukt: Resemble-enhance og Audacity – åpen kildekode
  - Alle lydfiler ble konvertert til felles format (16-bit/22.5 kHz .wav, mono)
  - Siste steg var å dele opp lange lydfiler til individuelle setninger og å samle setninger til en maskinlesbar tabell

Utviklingsprosess 3: Trening

- Total varighet av prosessert og filtrert lydkorpus er ~ 10,5 timer, 4670 setninger
- For å få konsistent lyd og høg kvalitet ut fra talesyntesen, har vi delt opp hele korpuset i 2 "stemmer" i treningssteg, basert på opptakskvalitet
  - 1: opptak med bra kvalitet ~ 2 timer
  - 0: opptak med litt dårligere kvalitet ~ 8,5 timer
  - Bare stemme **1** brukes til å generere syntetisk tale
  - Andre opptak (stemme 0) er med i syntesemodellen på bakgrunnen for å bygge generell informasjon om sørsamisk uttale

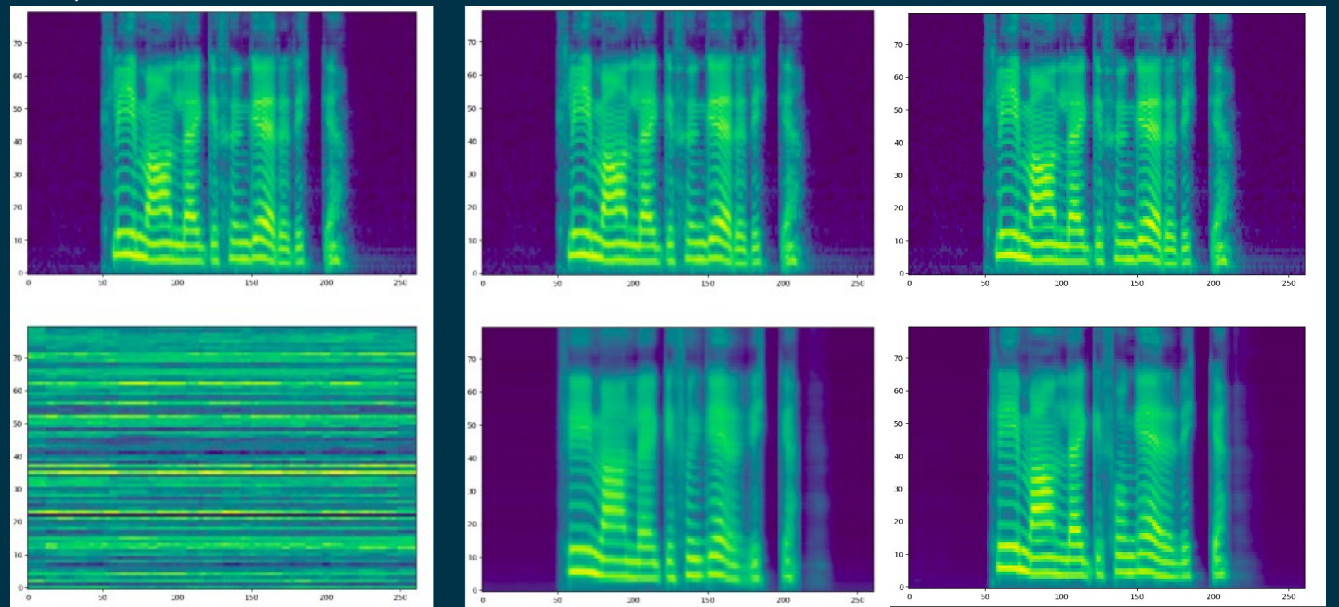
```
sorsamisk_-_114_01_-_Saemieh_gellien_laakan_barkeminie_007_020.wav|Fïerhte almetje sæjhta daejredh gie lea. Sæjhta gaavnedh dam maam daaroen baakojne identiteetine gåhtjoejibie.|1
sorsamisk_-_114_01_-_Saemieh_gellien_laakan_barkeminie_007_021.wav|Jeatjhlaakan aaj datamasjinah maehtieh almetjasse barkoem liehtestehtedh.|1
sorsamisk_-_114_01_-_Saemieh_gellien_laakan_barkeminie_007_022.wav|Guktie lïeremisnie, lohkehtæjjaj jïh learohki gaskem jallh joekehth skovli gaskem.|1
sorsamisk_-_114_01_-_Saemieh_gellien_laakan_barkeminie_007_023.wav|Dïhte jis lij mijjide åarjelsaemide joekoen hijven, guhth dan bårrode libie årroeminie.|1
sorsamisk_-_114_01_-_Saemieh_gellien_laakan_barkeminie_009_000.wav|Saemiedigkie mij lea dïhte? Dagke aa gee aa? Ammes girtiehtæjja aate.|0
sorsamisk_-_114_01_-_Saemieh_gellien_laakan_barkeminie_009_001.wav|Nimhtie Gåstan elkien Jávva, Jon Gustavsene tjaala gihtjie, daaroenplaeresne maam Kla Klassekampine gåhtjoeh.|0
sorsamisk_-_114_01_-_Saemieh_gellien_laakan_barkeminie_009_002.wav|Nov sih såamesh, gosse dam plaerien nommem guvlieh dellie dallah ussjedieh ij leah dïhte seatadimmesne.|0
sorsamisk_-_114_01_-_Saemieh_gellien_laakan_barkeminie_009_003.wav|Nimhtie hov juhkoe mijjieh almetjh sinsitniem båasi jallh gi gïerti sïjse tsaekiejibie.|0
sorsamisk_-_114_01_-_Saemieh_gellien_laakan_barkeminie_009_004.wav|Dïhte dan aelhkie, dellie ibie daarpesjh vielie maam dan bïjre ussjedalledh.|0
```

Utviklingsprosess 3: Trening

- Maskinlæring er en stor del av talesyntesen: maskinen lærer seg sammenhengen mellom bokstaver og lyd ved å gjøre millioner av beregninger fra korpuset = trening av en **syntesemodell**
- Treningsprosessen – hvordan lærer datamaskinen seg å snakke samisk:
  - En algoritme regner akustiske egenskaper for alle bokstaver i treningsmaterialet, går gjennom hele materialet hundrevis av ganger, og oppdaterer modellen med visse intervaller – da blir modellen finjustert og syntesen begynner å høres mer og mer ut som en menneskestemme
  - Treningsprosessen er veldig krevende for datamaskinen, derfor blir prosessen vanligvis utført på en "supercomputer", en stor datamaskin med veldig mye kapasitet og databehandlingskraft, som man kan styre fra egen datamaskin. Vi har brukt en norsk supercomputer, Sigma2, og det tar ~ 1 uke for å kjøre treningen
  - Visualisering fra treningsprosessen viser hvordan datamaskinen lærer seg tale ved å "lytte på" tale fra et menneske (1 -- 100 -- 500 epoker)

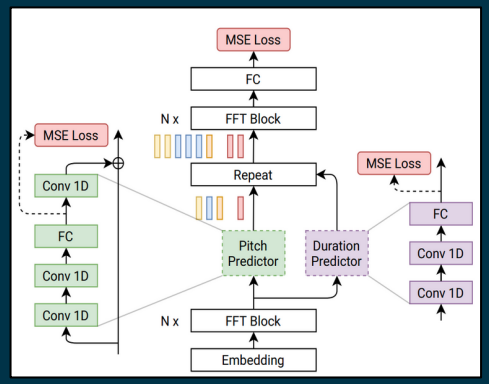
Menneske

Syntesemodell

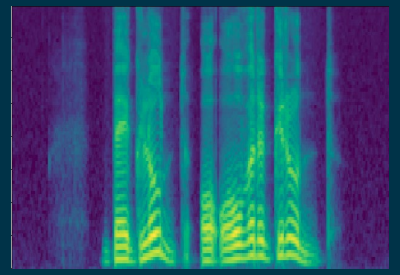


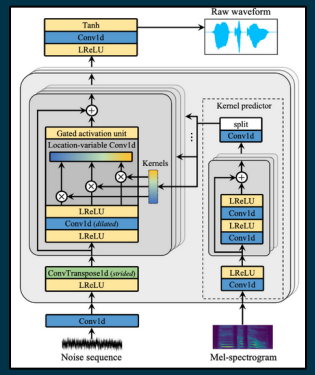
Utviklingsprosess 4: Syntesemotoren

“Båetieh mov gaavelohke, Båetieh mov aaltoe, miesieh, sarvah, båetieh mov gaavelohke.”

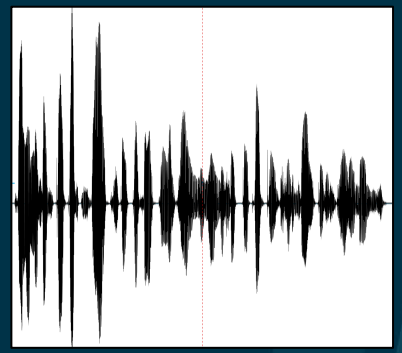


SYNTESEMODELL





VOCODER



1) Brukeren sender inn tekst som skal syntetiseres.

2) Syntesemodellen genererer et mel-spektrogram fra fonemsekvensen til teksten med de prosodiske trekkene (intonasjon, uttrykk). Mel-spektrogrammet er en tidsfrekvensrepresentasjon av talesignalet.

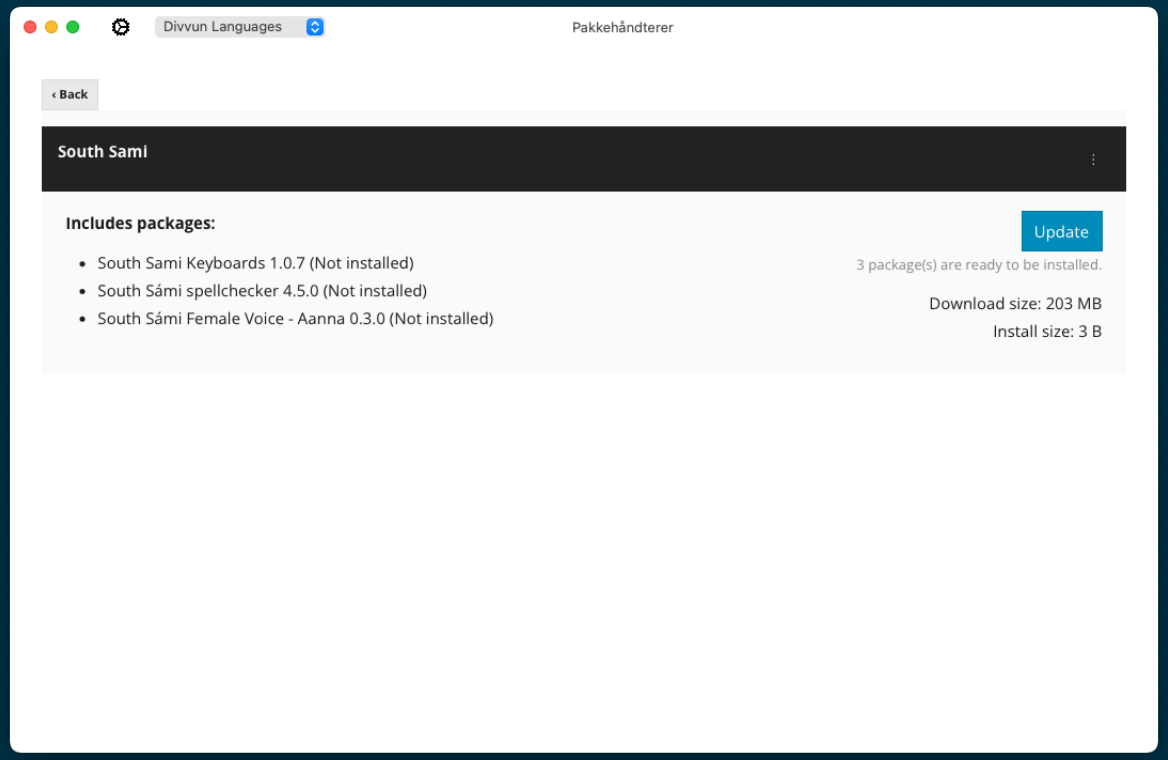
3) En vocoder er en talemodell som leser mel-spektrogrammet og konverterer det til et lydsignal som spilles av som syntetisk tale. Vi har brukt NeMo UnivNet som vocoder– det er en åpen modell trent med store mengder av varierende taledata

Utviklingsprosess 5: tekstprosessering

- Synteseprogrammet må vite og hvordan alle siffer, forkortninger og akronymer skal sies, så de må være skrevet fullt ut med bokstaver i teksten som sendes til synteseprogrammet, f.eks.:
  - 5. -> femte (ikke *fem punkt*)
  - NRK -> /ænn ærr kåå/
  - NAV -> /na:v/ (ikke /ænn aa vee/)
- Dette er gjort med Divvun sine språkteknologiske, regelbaserte verktøyene
- Reglene for tekstprosessering ble lagat manuellt av vår sørsamisk lingvist (Maja Lisa Kappfjell) -> systemen bruker maskinlæring i lyd + regelbasert i tekst

Hvordan tar man i bruk stemmene

- Talesyntesen blir installert via Divvun Manager (https://divvun.org/)
  - Automatisk oppdatering i bakgrunnen for nye versjoner



Bruk: skjermleser

- Stemmen kan brukes som skjermleser både på macOS og Windows
- Leser opp tekst, vinduselement og andre ledetekster i grensesnittet
- På Windows vil man typisk bruke en tredjeparts-skjermleser
  - Et godt gratis-alternativ er NVDA (https://www.nvaccess.org/)

Egen teknologi & åpen kildekode

- All kode brukt i systemet er vår egen eller åpen
- Vi har full kontroll på alle deler, fra opptak til distribusjon og installering
- Kan enkelt oppdatere talesyntesen, med automatisk distribusjon:
  - Stemmene
  - Tekstprosessering
  - Feilrettinger
- Kan enkelt legge til nye stemmer og nye språk

Vil bare bli bedre

- Mer data og kontinuerlig arbeid med tekstprosessering og feilretting
  - ⇒ bedre syntese
- Viktig med tilbakemelding – vi kan ikke rette feil vi ikke vet om
- Kontakt:
  - Facebook (facebook.com/Divvun)
  - Instagram (instagram.com/divvun.no)
  - Twitter (twitter.com/divvun)
  - GitHub (github.com/giellalt)
  - Zulip (med GitHub-konto) (giella.zulipchat.com)
  - E-post: feedback@divvun.no

Daate barre aalkove

- Daate barre saemien gïeleteknologijen aalkove
- Jïenebh smaarehtjïerth/dialekth
- Jïenebh gïelh
- Divvun-dåehkie aaj voejhkeleminie maasjijne edtja buektiehtidh tjaeledh dam maam almetjih åarjelsaemiengïelesne jiehtieh

Mijjide, Divvunisnie, soptsestimmieteknologije vihkele, mejnie sijhtijiibie barkedh. Mijjieh libie joe daan jaepien noerhtesaemien jïh julevsaemien soptsestimmiesynteesem bæjhkoehtamme, jïh daan biejjien jis åadtjoejidie åarjelsaemien soptsestimmiesynteesine åahpenidh. Mijjieh libie aaj soptsestimmie-dabteminie barkeminie. Dïhte hov nimhtie, daatamasjijne tjaala maam datne jeahtah, dovne noerhtesaemien-, julevsaemien-jïh åarjelsaemien.

For Divvun er taleteknologi et viktig satsningsområde, og i år har vi allerede publisert talesyntese for nord- og lulesamisk. Nå lanserer vi en sørsamisk stemme. Divvun er også godt i gang med arbeidet med talegjenkjenning, dvs. at datamaskinen skriver det man sier, for nord-, lule- og sørsamisk.

Gæjhtoe!

Mijjieh hohkesjibie dorjeldhgïele «Aanna» viehkine, aavojne jïh aevhkine åarjelsaemien seabradahkese sjædta

Vi håper syntesestemme "Aanna" kan komme til nytte og glede for det sørsamiske språksamfunnet