

Opdateret Fagprojekt Plan - Voice Conversion

Gustav Larsen - 180820
Lukas Leindals - s183920
Peter Grønning - s183922

May 2020

1 Revuderet Tidsplan

- Sværere at implementere metoderne end forventet
- Hvad gør vi med data?
- Main fokus er bare at få voice conversion til at fungere
- AutoVC er nogenlunde succesfuld, men mangler lidt
- StarGAN er påbegyndt som en anden metode
- Metode afsnittet skal skrives færdigt/muligvis cuttes i her og der

Oprindeligt forventede vi at have to state-of-the-art VC modeller AutoVC og StarGAN færdigt implementerede på nuværende tidspunkt. Denne forventning var baseret på, hvad vi troede omfanget af source-code var. Det viste sig eksempelvis, at store dele af AutoVC ikke var nemt tilgængeligt, herunder et Speaker Encoder modul og en træningsfunktion, som hhv. måtte findes andetsteds fra og implementeres selv. Det har medført et uventet ekstra arbejde med PyTorch, hvilket har været en stejl læringskurve, idet vi ikke har arbejdet rigtigt med sådanne Deep Learning moduler tidligere i vores uddannelse. Desuden har specifikationerne for data håndteringen (lydsignal til mel spektrogram) været uklare, hvilket har ført til en lang og ikke færdig afsluttet parametertuning, så modellen kan spille sammen med vocoderen anvendt i AutoVC.

Ved siden af overraskende umedgørlige modeller kommer den tunge teori, som de bygger på. Meget af den underliggende teori er nyt og vanskeligt stof, som har krævet meget tid at sætte sig ind i. Vi må også tilkendegive, at den forståelse og forklaringsgrad af teorien, vores vejleder ønsker, ligger over vores tidlige forventninger. Det har medført et møjsommeligt arbejde, som har været værdifuld for vores forståelse, men samtidig har gjort, at flere deadlines måtte overskrides.

Grundet diverse copyright love etiske årsager har det også været svært at skaffe test data, hvilket vi stadig kæmper med.

Vores optimistiske forventninger om at kunne forbedre conversion i form af at reducere tid og data i processen har vi måtte skyde ned. Vi besidder ikke de rette forudsætninger til dette, så vi ændrer fokus til bare at få det implementeret i stil med litteraturen og gøre det brugbare på det datasæt, vi nu måtte finde.

Det er lykkedes os, at implementere AutoVC og en dertilhørende træningsfunktion, og det virker lovende. Vi mangler dog, at få de rette spektrogram specifikationer, før vi kan bruge AutoVC's vocoder. StarGAN er undervejs og det næste skridt er at teste, hvorvidt det er rigtigt implementeret.

