Gruppe 09

Gjennomførelse:

Vi installerte pip for python 2.7. Deretter installerte vi modulene pyttsx, pyaudio og SpeechRecognition.Vi fikk
deretter begge programmene til å kjøre, med både text to speech og speech to text. For å få speech to text til å fungere måtte vi endre på …

Speech to text:

Det første steget i speech synthesis, blir vanligvis kalt normalization eller pre-processing. Det handler om å krympe ned de mange måtene man kan lese en tekst inn til den som passer best. Det handler om å fikse teksten slik at datamaskinen gjør mindre feil når den leser det opp. (1)

Preprosessering takler også homographs, som er ord som blir sagt forskjellig i forhold til hva de mener. “Read” kan bli uttalt “red” eller “reed”. “I read the book” kan være vanskelig for en speech synthesizer å forstå. (1)

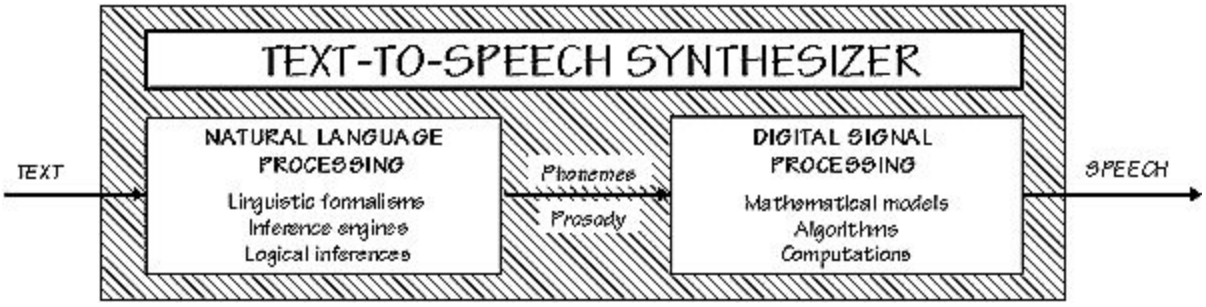
Text to speech:

Phenomener brukes av datamaskinen for å lage lyder som oppfattes som ord. Det finnes ca. 40 phonomer tilsammen i det engelske alfabet. Konteksten i setningen er en utfordring for datamaskinen. Prosody handler om meningen på teksten, hvem som snakker og følelsene som vil fremføres som pårivker lyden og tonen. (1)

“Formant” benytter seg av 3-5 viktige frekvenser som vokalen til mennesker genererer og kombinerer det med å lage lyden til snakking eller synging. Den kan dermed generere hvilke som helst lyder, fremfor å basere seg på tidligere innlest lyd. Ved særlige vanskelige ord/lyder er dette en bra metode. Formant hørtes derimot ikke like naturlig som en vanlig “concatenation” synthesizer som baserer seg på tidligere ord. (1)

Formant programmer er vanligvis mindre siden de ikke trenger en helt database å hente lyder fra. (3)

Eksempel på en enkelt text to speech synthesizer modell:



En NLP, Natural Language Processing Module brukes til å lage en fonetisk oversettelse av teksten, og sammen med en valgt rytme og tone, og en Digital Signal Processing module (DSP stansformerer den det symbolske informationen den får til ord lest høyt. (2)

Ved hjelp av visse algoritmer og formaliseringer kan man kutte ned prosessen og gjøre den mer effektiv, selv om dette legger visse restriksjoner på hva som blir uttalt, men det løser problemet med limitert minne i en real-time situasjon. (2)

Kilde:

1. http://www.explainthatstuff.com/how-speech-synthesis-works.html
2. http://tcts.fpms.ac.be/synthesis/introtts_old.html
3. https://en.wikipedia.org/wiki/Speech_synthesis