

Tentamen Taal- en Spraaktechnologie

29 juni 2011 13.15-15.00, ADD 22, zaal 0.03

1. De bandbreedte van een telefoon is 300-3000 Hz. Waarom klinkt het geluid toch heel natuurlijk?

2. a) Geef een transitie matrix voor 4 states (nummer 1 t/m 4, plus non-emitting beginstate S en eindstate F), waarbij je niet terug kunt, en alle kansen vanuit een state gelijk zijn. Je kunt niet meteen van S naar F.
b) Wat is de kans op de statereeks 2 3 3 4?

Per state kan een klinker V of een medeklinker C gegenereerd worden. De kans op een klinker is in states 1 t/m 4 achtereenvolgens 0.2, 0.8, 0.7, 0.2.

- c) Wat is de meest waarschijnlijke statereeks voor de uitvoer C V?
d) En welke kans is er dan op C V?

3. Druk $P(O|\lambda)$ op drie manieren uit:

- a) in de voorwaartse variabele α op het einde van de observatiereeks
b) in de achterwaartse variabele β aan het begin van de observatiereeks
c) in zowel de voorwaartse als achterwaartse variabele op een willekeurige positie daar tussenin.

4. Wat zijn de belangrijkste beperkingen die in POS tagging worden gesteld aan

- a) de prior probability van een tagsequentie
b) de likelihood van de woordreeks, gegeven een tagsequentie

5. Noem zes verschillende gevallen waarin cijfers een aparte interpretatie behoeven bij de voorbewerking voor spraaksynthese. Beschrijf kort je aanbevolen uitwerking.

6. Twee woorden x en y komen als volgt in een tekst voor:

er zijn 20 zinnen 'we zagen x y gisteren', 70 zinnen 'we hebben erg van x y genoten' en 10 zinnen 'heb je x y al gezien?'

- a) Wat zijn het gemiddelde en variantie in dit geval (via het meten van de offset)?
b) Is x y een collocatie?

7. Wat is de nulhypothese H_0 in de context van collocatie extractie voor twee woorden w_1 and w_2 ?

8. Voor deze oefening gebruik je de notatie van FSA Utilities.

Maak een transducer die de infinitief van een regelmatig werkwoord in het Nederlands omzet in het voltooid deelwoord.

Voorbeeld: *werken* -> *gewerkt*

Testgevallen: *haten, zweten, hakken, strekken, maken, straffen, keffen, krassen, grazen, matsen, kuchen, kapen, kappen, verven, zweven.*

Bouw je hoofddefinitie op als een cascade (compositie) van eenvoudige omzetteren. Denk aan de regel van 't kofschip voor de keus tussen uitgang d of t, korte versus lange klinkers, stemloze versus stemhebbende medeklinkers aan het eind, etc.

macro(vtdlw, .. o .. o ..)

9. Wanneer noem je een finite state **herkenner** non-deterministisch?
Wanneer noem je een finite state **transducer** non-deterministisch?

Geef aan wat het verschil is tussen herkenners en transducers wat betreft determiniseerbaarheid.