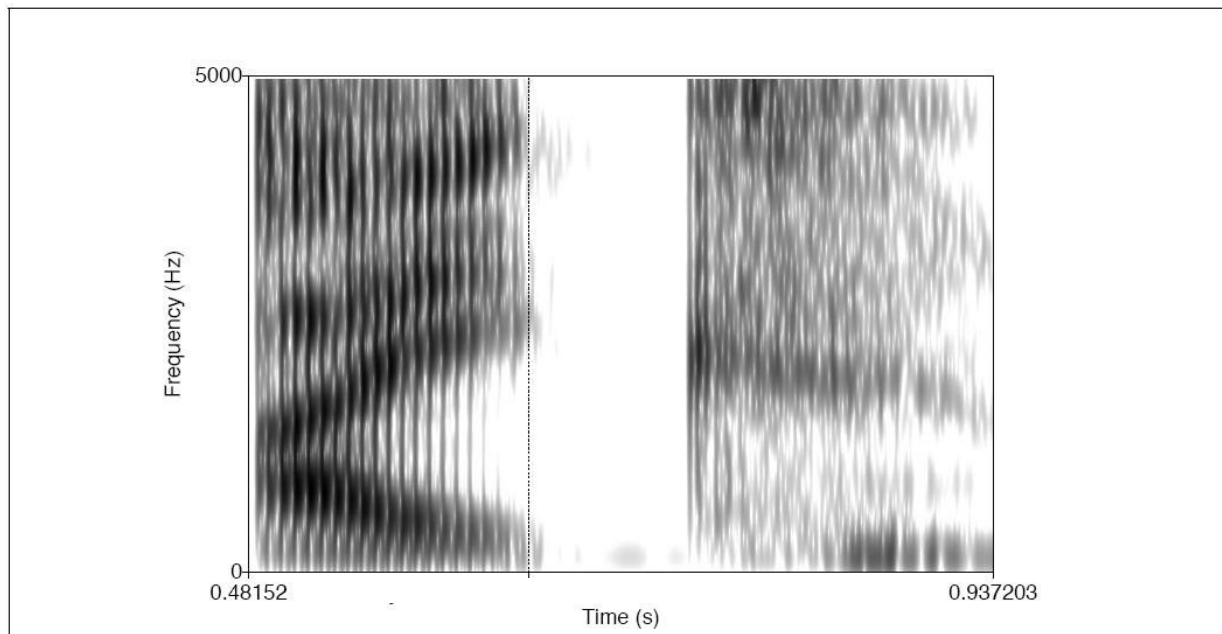


Tentamen Taal- en Spraaktechnologie

27 augustus 2010 11.0-13.00 uur, Ruppert 1.36

1. Het onderstaande spectrogram toont een woord met twee fonemen



- Welke twee typen fonemen zijn het?
- Wat zijn de zwarte balken links? En waarom zouden ze uit elkaar lopen?
- Links zijn in het spectrogram verticale lijntjes zichtbaar, wat betekent elk lijntje?
- Is de spreker een man of een vrouw? Leg uit.

2. a) Vul de onderstaande transitie matrix A aan (S=Start, E=Finish; non-emitting):

	1	2	3	4	F
S	0.5	0.3	?	0	0
1	0.3	?	0.1	0.1	0
2	0	0.2	0.5	0.1	?
3	0	0	0.1	?	0.5
4	0	0	0	0.6	?

- Wat is het kenmerkende van deze architectuur?
- Wat is de kans op de statereeks **2 2 3 4** ?
- Per state kun je nu twee observaties doen: Warm of Koud. Vul de B-matrix aan:

s	P(W s)	P(K s)
1	0.3	?
2	0.4	?
3	0.2	?
4	0.6	?

- Wat is de kans op de observatiereeks **Warm, Koud, Warm, Warm** voor de statereeks uit c)?
 - Wat is de meest waarschijnlijke statereeks voor de observatiereeks **Warm, Koud**?
 - Wat is de totale kans dat dit model de observatiereeks **Warm, Koud** verklaart?
3. Hoe kun je met de voorwaartse en de terugwaartse variabele de kans uitrekenen dat je op tijdstip t in een bepaalde state j zit (gegeven het hmm)?

4. Correctie van typfouten in een woord kan gebaseerd zijn op Bayesiaanse inferentie volgens

$$c^* = \operatorname{argmax}_{c \in C} P(t|c) P(c)$$

waar t het fout getypte woord is, c een kandidaatwoord en c^* de beste kandidaat.

- Leg de formule uit
 - In de formule is een term weggelaten omdat die niet nodig is. Welke term, en waarom?
 - Hoe zou je aan $P(t|c)$ en $P(c)$ kunnen komen?
5. Bij *unit selection* in spraaksynthese wordt de keuze bepaald door *join costs* en *target costs*.
- Leg uit wat met deze kostenfactoren wordt bedoeld, en welke aspecten er in mee worden gewogen.
 - Met welke methode kunnen we deze kosten over een te synthetiseren woord of zin minimaliseren?

6. Geef aan wat het verschil is tussen een deterministische en een non-deterministische FSA, gegeven een toestand en een alfabetsymbool.

Verschillen deterministische vs non-deterministische FSA's wat betreft de klasse van talen die door die machines worden herkend?

- Zo ja, welke soort machine herkent dan de grootste klasse van talen?
- Zo nee, verklaar waarom niet.

7. Een subklasse van de finite state transducers zijn de ***sequentiële*** transducers. Wat is de kenmerkende eigenschap van een sequentiële transducer? Waarom is deze subklasse van transducers van belang?

8. Maak een transducer die, gegeven de infinitief van een regelmatig werkwoord in het Nederlands, het voltooid deelwoord maakt.

Voorbeeld. INPUT: *werken*, OUTPUT: *gewerkt*

Gebruik de notatie van het practicum:

macro(voltdeelw, ...).

Bouw de transducer op als de ***compositie*** van een aantal eenvoudige stappen: uitgang weghalen, 't kofschip regel toepassen, (ver|ont)dubbeling van (mede)klinkers, enzovoort.

Je transducer moet werken voor de volgende gevallen:

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