

Schriftelijk tentamen Spraakherkenning en spraaksynthese (do. 23 december 2004, 13.30 – 16.30 u; zaal C.206)

Vermeld op iedere pagina je *naam*, je *studentennummer* en het *vervolgnummer* per pagina. Als je voor 16.30 u klaar bent, lever dan je tentamen in bij de surveillant en verlaat rustig de zaal. Vergeet niet ook de *onderwijsenquête* in te vullen en in te leveren.

Het cijfer voor dit schriftelijk tentamen bepaalt je eindcijfer voor dit college. Dat eindcijfer wordt echter pas aan het OWI doorgegeven wanneer je ook alle 8 practicumverslagen hebt voltooid en ze zijn goedgekeurd. Nadere informatie hierover vind je op de Blackboard site.

Beantwoord onderstaande **10** vragen zo *nauwkeurig*, *compleet* en *gedetailleerd* mogelijk, zonder in herhalingen te vervallen. Het kan geen kwaad relevante ervaringen vanuit het hoorcollege en de practicumopdrachten in je antwoorden te verwerken. Bij de beoordeling van dit tentamen wordt meer gelet op begrip dan op feitenkennis. Besteed niet te veel tijd aan afzonderlijke opgaven. Als een opgave te veel tijd vergt, probeer dan eerst een andere opgave.

Veel succes!

Rob van Son

A Spreken en verstaan - spraakklanken

1) Medeklinkers kunnen van elkaar worden onderscheiden door ze te klassificeren naar 'plaats van articulatie', 'manier van articulatie' en 'mate van stemhebbendheid'. Zo is een 'p' een stemloze labiale plofklank en 'h' een stemloze glottale fricatief.

a) Classificeer evenzo de klanken 'w', 'g', en 'ng' uit het woord *weging* (dwz. [w], [x], [ŋ])

b) Wat wordt er bedoeld met 'manier', 'plaats' en 'stem' in termen van de articulatie en het geluid (dwz., akoestiek) van de spraakklanken.

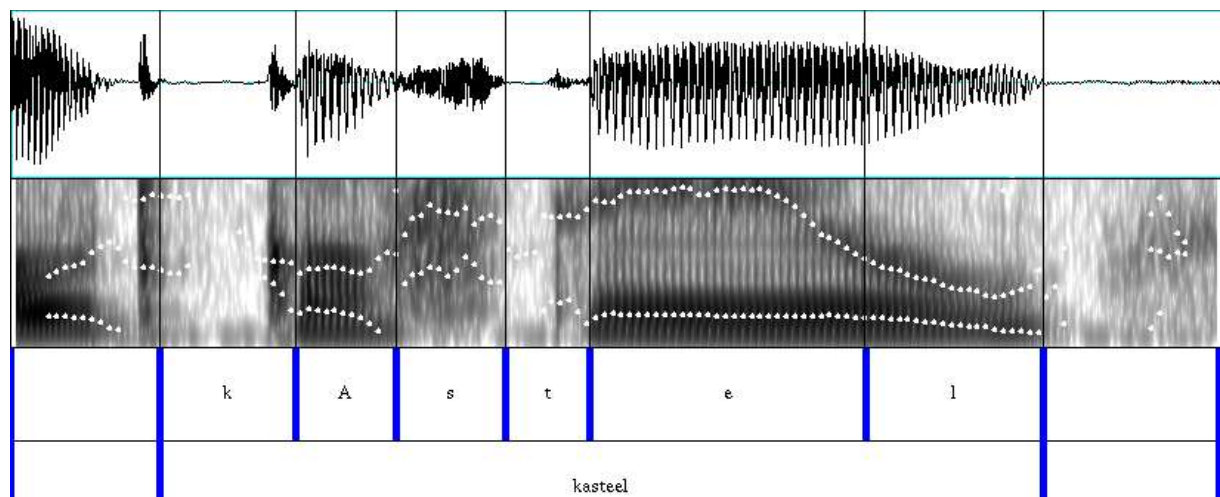
2) Wat zijn de meest opvallende verschillen tussen *geschreven tekst*, zoals 'natuurlijk was ik er', en *gesproken taal* /t y l @ k w A z @ k @ r/, onder andere in termen van

- de *vluchtigheid* van de informatie;

- de *variabiliteit* van de informatie, en

- het *discreet* of *continu* zijn van informatie?

- Licht ook toe welke *additionele* informatie (b.v. over akoestiek, emotie, e.d.) in het spraaksignaal zit, die in geschreven tekst ontbreekt.



3) Hierboven zie je het oscillogram en het spectrogram met de eerste twee formanten (witte stippen) van het woord *kasteel*, uitgesneden uit een lopende zin. Onderaan zijn de fonemen en hun grenzen weergegeven.

a) Fonemen worden geklassificeerd op grond van manier van articulatie, plaats van articulatie en stemhebbendheid. Zo is het foneem /n/ een alveolaire nasaal, het foneem /v/ een stemhebbende, labiale fricatief. Geef van ieder foneem in het woord *kasteel* de manier en plaats van articulatie en de stemhebbendheid.

b) Geef de asbijschriften van het bovenstaande plaatje.

c) Beschrijf de karakteristieke acoustische eigenschappen van de fonemen uit *kasteel* zoals die geïllustreerd worden in de golfvorm, het spectrogram en de formanten in het bovenstaande plaatje.

d) Beschrijf hoe je de plaats van de foneemgrenzen zou bepalen rond de /A/, /s/, en de /t/ in *kasteel*. Dat wil zeggen, op welke akoestische eigenschappen zou jij letten?

e) Welke lettergreep van *kasteel* is beklemtoond, en waaraan kun je dat zien in het bovenstaande plaatje.

zie nieuwe pagina

B Spraaksynthese

4) Het *zinsaccent* wordt eerst en vooral gerealiseerd via *toonhoogtebewegingen* op de beklemtoonde lettergrepen.

a) Geef van onderstaande zin aan welke lettergrepen beklemtoond zijn, en welk(e) woord (en) jij zinsaccent zou geven.

In welke zaal wordt het tentamen afgenomen?

b) Teken vervolgens een *gestileerde toonhoogtecontour* die bij deze classificaties past.

c) Geef tevens van iedere klinker aan of de *duur* korter, langer, dan wel ongeveer gelijk aan de gemiddelde duur van die klinker zal zijn, en geef aan waarom je dat denkt.

5) Bij Nextens en Fluent Dutch en bij veel andere tekst-naar-spraak synthesesystemen wordt gebruik gemaakt van difonen als bouwstenen voor het genereren van spraak.

a) Welke difonen zijn nodig voor het genereren van de uiting: *Daar ook!*.

b) Welke difonen heb je nog extra nodig om ook de uiting *rook maar!* te kunnen genereren?

6) Beschrijf in het kort wat er zoal nodig is om een gedrukte-tekst-op-papier om te zetten in verstaanbare synthetische spraak. Geef waar mogelijk voorbeelden voor iedere stap. Zeg tenminste iets over:

- *text preprocessing*;

- *grafeem-foneemconversie* via uitspraaklexicon en/of regels;

- *prosodiecontrole* (spreektempo, woordklemtoon, zinsaccent, intonatie, frasering), en

- feitelijke *signaalgeneratie*.

- Geef tenslotte aan hoe de *natuurlijkheid* en de *verstaanbaarheid* van een spraaksynthesesysteem kan worden geëvalueerd? Geef tenminste twee voorbeelden van mogelijk *stimulusmateriaal* om zo'n evaluatie uit te voeren.

zie nieuwe pagina

C Spraakherkenning

7) Beschrijf in het kort wat er zoal nodig is om een spraakherkenner te bouwen voor door de telefoon ingesproken tekst. De herkenner moet lopende spraak van willekeurige bellers herkennen. De herkende tekst wordt gebruikt voor het opzoeken van informatie uit een taak-gerichte database. Zeg tenminste iets over:

- Training en trainingsmateriaal voor het akoestische gedeelte (HMM foneemmodellen)
- Structuur van het taalmodel in relatie tot de taak (bigram of anderszins)
- Trainingsmateriaal voor het taalmodel
- Evaluatie van de herkenner

8) De performance van een spraakherkenner wordt vaak uitgedrukt in 'word error rate (WER)', zijnde de som van het aantal substituties (S), inserties (I) en deleties (D), gedeeld door het totaal aantal woorden in de referentie (W), maal 100%.

$$WER = \frac{(S + I + D)}{W} * 100$$

Als de woordsequentie van de ingesproken letterreeks was: *NMLKJIHGF*, wat is dan, uitgaande van een optimale match,

a) de WER bij een herkende output: *NKMKJIIGFN*, en

b) wat bij een herkende output: *LKJJIH*?

Licht je antwoorden toe en geef aan waar er Substituties, Inserties en Deleties hebben plaatsgevonden.

9) In het MATIS project vragen gebruikers informatie over treinreizen op door een formulier op het computerscherm in te vullen met, naar keuze, spraak, muisklikken of door tekst in te typen. Gebruikers kunnen voor ieder veld of keuzemogelijkheid in het formulier zelf bepalen hoe ze het willen invullen. MATIS geeft naast spraakuitvoer ook altijd de al ingevulde velden van het formulier op het scherm weer. Het spraakherkennings en -synthese gedeelte van MATIS is gelijk aan dat van het OVIS project (uit: *A spoken dialogue system for public transport information*).

a) MATIS werkt met een microfoon aan de computer, OVIS werkte over de telefoon zonder beeldscherm. Wat moeten de onderzoekers doen om de OVIS spraakherkenner geschikt te maken voor gebruik in het MATIS project?

b) De spraak invoer van Matis kan door de gebruiker aan- en uitgezet worden door op een knop op het formulier te drukken. Een belangrijk aspect van MATIS is dat gebruikers het systeem op willekeurige momenten kunnen onderbreken met een muisklik. Beschrijf wat jij denkt dat er moet veranderen aan het dialoog management systeem van OVIS om het geschikt te maken voor MATIS. Er zijn vele oplossingen mogelijk, kies er één uit.

zie nieuwe pagina

10) Hieronder zijn uit het boek van Jurafsky & Martin (2000) allereerst 4 woordmodellen voor 4 Amerikaanse woorden (the; on; need; I) weergegeven. In de daarna weergegeven tabel zijn de bigramwaarschijnlijkheden voor een aantal woordcombinaties weergegeven. Tenslotte is een stukje van een min of meer volledig HMM-herkenningsmodel weergegeven. (Bij de volgende opgaven moet u de berekeningen opschrijven. Het gaat *niet* om de uitkomst maar om de opbouw van de berekening!)

a) Het onderstaande model bevat ook een (bigram) taalmodel. Een taalmodel geeft de kans op een gegeven zinnetje. Bereken de kansen op de volgende deelzinnetjes (let op, dit is telkens het begin van een zin, er volgen nog meer woorden die hier niet gegeven zijn):

The need I...

I need the...

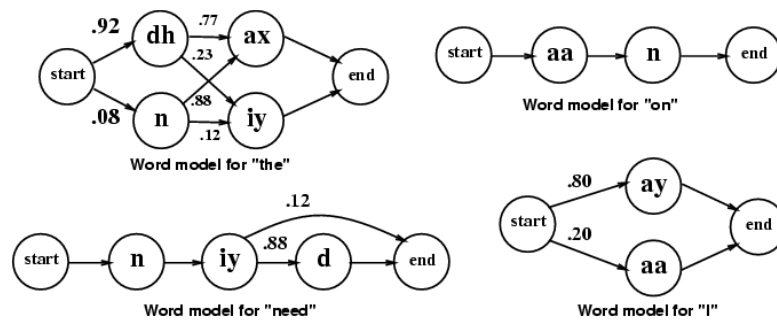
b) Waarom staan er geen getallen in het woordmodel van "on"?

c) Geef de totale kans van de foneemstring: "# aa n iy aa n "

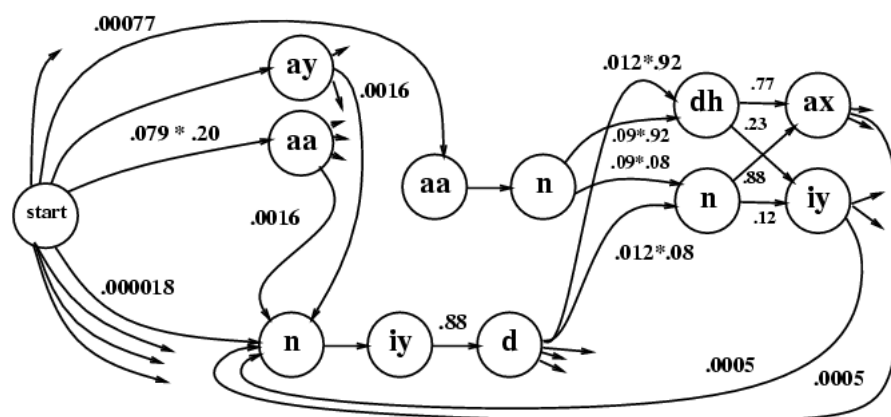
(let niet op een eventuele "betekenis" van de uiting, enkel op fonemen en woorden)

Schrijf dit ook uit in termen van enkel het taalmodel, dwz., met weglating van de foneem-foneem overgangswaarschijnlijkheden.

d) Wat is de meest waarschijnlijke transcriptie in woorden van deze foneemstring?



l need	0.0016	need need	0.000047	# Need	0.000018
l the	0.00018	need the	0.012	# The	0.016
l on	0.000047	need on	0.000047	# On	0.00077
l l	0.039	need l	0.000016	# l	0.079
the need	0.00051	on need	0.000055		
the the	0.0099	on the	0.094		
the on	0.00022	on on	0.0031		
the l	0.00051	on l	0.00085		



Dit is het einde van dit tentamen. Als je zeker weet dat je niets meer aan je antwoorden kunt toevoegen, lever dan je papieren in bij de docent.

Vergeet ook niet de *onderwijsenquête* in te vullen en in te leveren.

Uiterlijk medio januari kunnen jullie op Blackboard de uitslag van dit tentamen vinden onder Assignments, Tentamenresultaten.