

Schriftelijk tentamen Spraakherkenning en spraaksynthese (do. 10 juli 2003, 9.30 u; zaal P.227)

Vermeld op iedere pagina je *naam*, je *studentennummer* en het *vervolgnummer* per pagina. Als je voor 12.30 u klaar bent, lever dan je tentamen in bij de docent 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 7 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. Besteed in ieder geval voldoende tijd en aandacht aan de twee belangrijkste vragen 6 en 7. Het kan geen kwaad relevante ervaringen vanuit het hoorcollege en de practicumopdrachten in je antwoorden te verwerken. Dit tentamen is meer een test op begrip dan op feitenkennis. Veel succes!

Louis Pols

A Spreken en verstaan - spraakklanken

- 1) Medeklinkers kunnen van elkaar worden onderscheiden door ze te rubriceren naar 'manier van articulatie', 'plaats van articulatie' en 'mate van stemhebbendheid'. Zo is een 't' een stemloze alveolaire plofklank en 'v' een stemhebbende labiale fricatief.
- Classificeer evenzo een 'd', 'f', en 'm';
 - licht de bovengenoemde begrippen '*manier*', '*plaats*' en '*stem*' nader toe.

- 2) Wat zijn de meest opvallende verschillen tussen *geschreven tekst*, zoals b.v. 'op een gegeven moment', en *gesproken taal* /xem ε nt/, 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) - *Beschrijf* van een drietal fonemen, namelijk de stemloze fricatief /s/, de open lange klinker /a/ en de labiale stemloze plofklank /p/ in het woordje 'Saab', de wijze waarop ze door de mens worden geproduceerd (longen, stembanden, vorm mondkanaal, stand lippen).
- Geef schematisch weer hoe het *oscillogram* voor dit woord er uit ziet.
 - Geef daarin de *foneemgrenzen* aan en geef een indicatie van alle *segmentduren* in milliseconden.
 - Teken ook een *gestileerd spectrogram* van dit woord, incl. asbijschriften.

zie ommezijde!!

B Spraaksynthese

4) *Beklemtoonde* lettergrepen in een woord zijn veelal *langer, luider en nadrukkelijker gesproken* dan onbeklemtoonde. Evenzo wordt *zinsaccent* eerst en vooral gerealiseerd via *toonhoogtebewegingen*.

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

Wat is de precieze betekenis van deze Nederlandse zin?

- Teken vervolgens een *gestileerde toonhoogtecontour* die bij deze classificaties past.
- 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 Fluent Dutch en bij veel andere tekst-naar-spraak synthesesystemen wordt gebruik gemaakt van difonen als bouwstenen voor het genereren van spraak.

- Welke difonen zijn nodig voor het genereren van de uiting: *Halt, taxi*.
- Welke difonen heb je nog extra nodig om ook het woord *halter* 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. 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.

C Spraakherkenning

7) Beschrijf in het kort wat er zoal nodig is om een onbekende, via de telefoon ingesproken tekst, van een willekeurige spreker, woord voor woord te herkennen. Zeg tenminste iets over:

- *spraakinvoer en digitalisatie*;
- *parametrische representatie* iedere 10 ms;
- toetsen tegen getrainde *foneem- en woordmodellen* (veelal HMM), en
- het gebruik van een *taalmodel* (veelal bigram).

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, maal 100%. Als de woordsequentie van de ingesproken getallenreeks was: 98765432, wat is dan, uitgaande van een optimale match,

- de WER bij een herkende output: 197644321, en
- wat bij een herkende output: 875532?

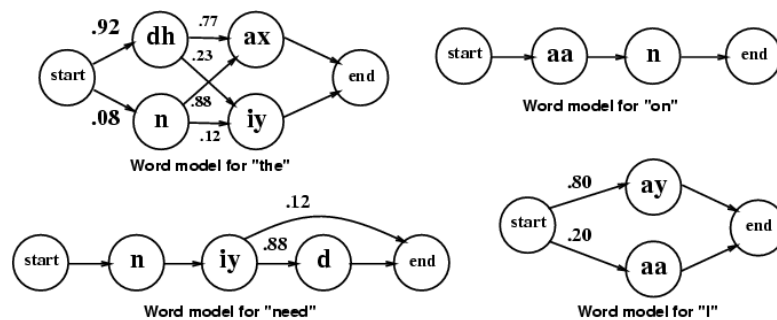
Licht je antwoorden toe.

9) Bij de subsidie-instantie NWO is zojuist een nieuw onderzoekprogramma IMIX gestart. IMIX staat voor Informatieve Multi-modale Informatie-extractie. Licht toe wat je denkt dat voor zo'n project nog meer nodig is dan eenvoudige gesproken-woordherkenning en tekst-naar-spraaksynthese.

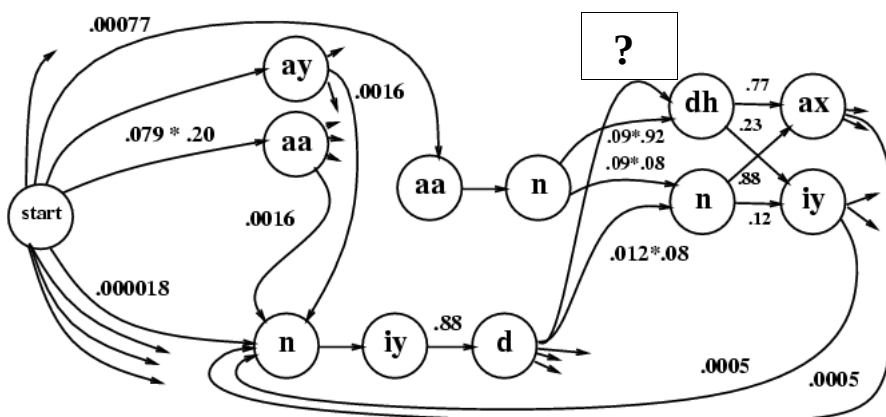
zie nieuwe pagina

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.

- Beschrijf hoe je dit model moet interpreteren.
- Geef tevens aan wat
 - de betekenis is van de getallen $0.079 * 0.20$
 - en van het getal 0.23 in het woordmodel voor “the”.
- Geef tenslotte aan welk getal of getallencombinatie moet staan bij de met een vraagteken gemarkeerde pijl.



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 augustus kunnen jullie op Blackboard de uitslag van dit tentamen vinden onder Assignments, Tentamenresultaten.