

Schriftelijk tentamen Spraakherkenning en spraaksynthese
(do. 25 augustus 2005, 9.30 – 12.30 u; Euclides P0.16)

Vermeld op iedere pagina je *naam*, je *studentnummer* en het *vervolgnummer* per pagina. Als je voor 12.30 u klaar bent, lever dan je tentamen in bij de surveillant en verlaat rustig de zaal.

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 **8** 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 'b', 'oe', en 'l' uit het woord *boel* (dwz. [b], [u], [l])
- b) Beschrijf heel kort voor de drie klanken uit deel **a** welke articulator gebruikt wordt en waar in de mond-keelholte de constrictie gemaakt wordt bij het spreken. Geef, waar nodig, andere belangrijke aspecten van de klankproductie van deze drie klanken aan.
- 2) Wat zijn de meest opvallende verschillen tussen *geschreven tekst*, zoals 'op een gegeven moment', en *gesproken taal* /O p @ n g e m E n t/, onder andere in termen van
- hoe *vluchtig* de informatie is;
 - hoe *variabel* de weergave is, en
 - in welke mate de informatie *discreet* of *continu* is?
- 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 de vijf fonemen, in het woordje 'markt', de wijze waarop ze door de mens worden geproduceerd (longen, stembanden, vorm spraakkanaal, stand lippen).
- Geef schematisch weer hoe het *oscillogram* voor dit woord er uit ziet.
 - Geef daarin de *foneemgrenzen* aan en geef een indicatie van relatieve *segmentduren* in milliseconden.
 - Teken ook een *gestileerd spectrogram* van dit woord, incl. asbijschriften. Het gestileerde spectrogram moet de belangrijkste aspecten van deze vier klanken weergeven.

zie volgende 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.

We vlogen om vier uur 's morgens naar Kreta.

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, 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 het woord: *ochtendtelevisie!*

b) Welke difonen heb je nog extra nodig om ook de uiting *mocht een theelepels vies zijn!* te kunnen genereren?

6) Om een *digitale-tekst-in-unicode* om te zetten in verstaanbare *synthetische spraak* zijn meerdere stappen nodig. We onderscheiden onder andere

- *text preprocessing*; voor o.a., afkortingen, getallen, tijden en geldbedragen.
- *grafeem-foneemconversie* via uitspraaklexicon en/of regels;
- *prosodiecontrole* (spreektempo, woordklemtoon, zinsaccent, intonatie, frasering)
- *feitelijke signaalgeneratie*

Deze stappen kunnen ieder aanleiding geven tot specifieke problemen. De kwaliteit van het totale systeem en van de deelsystemen kan geëvalueerd worden op:

- *natuurlijkheid*
- *verstaanbaarheid*
- *bruikbaarheid*

Geef voor ieder van deze criteria een uitgewerkt voorbeeld van mogelijk *stimulusmateriaal* en bijbehorende *taak* om zo'n evaluatie uit te voeren. Kies voor iedere evaluatie een van de stappen in de Tekst-naar-Spraak omzetting

zie nieuwe pagina

C Spraakherkenning

7) 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: 31205252183, wat is dan, uitgaande van een optimale match,

a) de WER bij een herkende output: 020622196, en

b) wat bij een herkende output: 243612907

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

8) 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 maar gebruikt een goede microfoon ipv. een telefoon (uit: *A spoken dialogue system for public transport information*).

a) Om de herkenner in MATIS te trainen is ander spraak- en tekstmateriaal nodig dan voor OVIS. Waarin zal het tekstmateriaal verschillen? Waarin het spraakmateriaal?

b) De spraak invoer van Matis kan door de gebruiker aan- en uitgezet worden door op een knop op het formulier te “klikken”. Een belangrijk aspect van MATIS is dat gebruikers het systeem op willekeurige momenten kunnen onderbreken met een muisklik. Geef een schematische weergave van een mogelijk dialoog management van OVIS en wat daar moet veranderen voor het MATIS systeem? Hou het kort en beperk je tot de relevante delen.

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.

Ik neem zo snel mogelijk contact met jullie op over het behaalde resultaat.