

Herkansing 1^e Deeltentamen Spraakherkenning en -synthese

Rob van Son

10-13 uur, 20 december 2007
GEBP/P2.27

Dit is een herkansing van het eerste deeltentamen. Je moet deze opgaven *alleen* maken als je mee wil doen aan de herkansing.

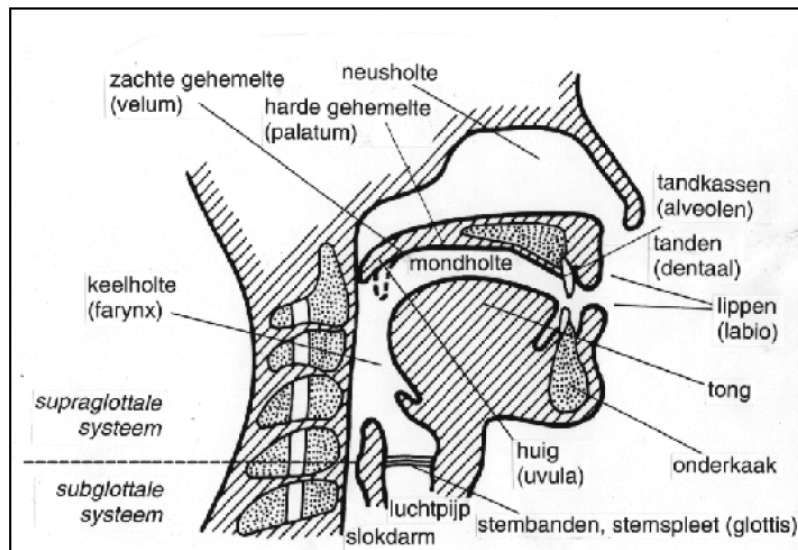
Vermeld op iedere pagina je naam, je studentnummer en het volnummer per pagina. Gebruik voor elke opgave (1-3) een apart vel. Als je voor 13.00 uur klaar bent, lever dan je tentamen in bij de surveillant en verlaat rustig de zaal.

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

Beantwoord onderstaande vragen zo nauwkeurig, compleet en gedetailleerd mogelijk, zonder in herhalingen te vervallen. De vragen zijn zo gesteld dat een kort antwoord volstaat. Reken op 5 minuten gemiddeld per sub-onderdeel als je het antwoord weet. Een erg lang antwoord is bijna zeker een verspilling van tijd. Het kan geen kwaad relevante ervaringen vanuit het hoorcollege en de practicumopdrachten in je antwoorden ter 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



Opgave 1): Een dwarsdoorsnede van een menselijk hoofd

1) Spreken en verstaan

1.a) Spraakorganen

Geef aan de hand van de figuur aan welke organen van belang zijn voor het spreken en wat hun functie is. Hou het kort.

1.b) Formanten

Welk orgaan of welke organen zijn verantwoordelijk voor de hoogte van de formanten? Hoe verandert dit orgaan of veranderen deze organen de formanten? Wat zijn de verschillen tussen de formanten van mannen, vrouwen en kinderen?

1.c) Pitch accent

Beschrijf aan de hand van de bovenstaande afbeelding hoe een lettergreep geaccentueerd wordt.

1.d) Nasalen

Beschrijf aan de hand van de bovenstaande afbeelding hoe een nasaal wordt gevormd. Wat maakt nasalen anders dan andere medeklinkers? Wat zou men verstaan onder een *genasaliseerde klinker* zoals die voorkomen in bijvoorbeeld het Frans of Portugees?

1.e) Articulatie

Welke bewegingen zijn nodig om het woord “rook” (/ro:k/) uit te spreken? (: geeft aan dat het hier om een lange klinker gaat)

1.f) Reductie

Onder Reductie wordt het verschijnsel verstaan dat klanken minder precies uitgesproken worden wanneer ze onbeklemtoond zijn of “voorspelbaar” zijn in spontane spraak. Een gereduceerde *klinker* is korter, minder luid en heeft formanten die dichterbij die van de /ə/ liggen. Men neemt aan dat de oorzaak van reductie is dat de spreker minder moeite doet om alles goed uit te spreken. Beschrijf hoe de articulatie van een gereduceerde “aa”, /a:/, uit het woordje “daad”, da:d, zou kunnen verschillen van die van een precies uitgesproken, ofwel citatie, vorm.

2) Automatische Spraaksynthese

2.a) Spraaksynthesesystemen

Er zijn vijf methoden voor spraaksynthese behandeld in het college:

- Articulatorische synthese
- Diphone concatenation
- Regel- of formantsynthese
- Unit selection
- hmm synthese

Geef van tenminste vier methoden aan welke spraakgegevens verzameld moeten worden om de synthese uit te kunnen voeren. Geef ook een relatieve indicatie hoeveel data nodig zijn voor een goede synthese op een schaal van 1 (minst) tot 5 (meest). Je mag meerdere systemen dezelfde score geven.

Welke systemen zijn gebaseerd op automatische verwerking van spraakcorpora en welke op het handmatig evalueren van spraakgegevens?

2.b) Tekst in, Spraak uit: Bewerkingsstappen

We onderscheiden ruwweg 6 stappen in de Tekst-naar-Spraak generatie.

- Tekstnormalisatie
- Duurbepaling
- Letter/klank conversie
- Intonatiegeneratie
- Accentplaatsing
- Spraaksynthese

Beschrijf het doel van ieder van deze stappen. Hou het zo kort mogelijk. Accentplaatsing en Intonatiegeneratie hebben allebei betrekking op de toonhoogte. Waarom zijn ze dan toch gesplitst en worden eerst de accenten bepaald, en dan pas de intonatie?

2.c) Eigenschappen van een goede stem

Voor het aanmaken van een nieuwe “stem” voor een difoon of unit selection TTS synthese moeten er nieuwe spraakopnamen gemaakt worden. Geef de eigenschappen waarop gelet moet worden bij de selectie van een goede spreker en bij het maken van de opname.

2.d) Evaluatie

Bij de evaluatie van automatische Text naar Spraaksynthese wordt vaak gebruik gemaakt van teksten die geselecteerd zijn uit een groot corpus op specifieke kenmerken, zoals:

- random
- minimum word frequency based
- overall word frequency based
- overall trigram frequency based

Geef voor ieder van deze selectiecriteria aan wat ermee bedoeld wordt en wat er getest wordt.

2.e) Blizzard Challenge

De “Blizzard Challenge” is een wedstrijd tussen TTS systemen. De volgende quote gaat over de “Blizzard Challenge 2005”

The key aspect of the Blizzard Challenge [2] is a common dataset shared between participants. Removing the variability of the data itself allows for a much closer comparison of the voices generated from the data. There are number of issues which must be answered before we can provide such a set. The issues include: what size and shape should the data be; who will collect and release it; what additional data may also be used.

Leg in het kort uit wat het doel was van de “Blizzard Challenge”. Geef in het kort, puntsgewijs, aan hoe de hierboven weergegeven aspecten hieraan bijdragen. (ga niet in op de details)

3) Voorbeelden van nieuwe spraaksynthese

Studentenprojecten over het maken van een TTS voor een nieuwe taal.

3.a) Een nieuwe taal

Om een TTS voor een nieuwe taal te maken, moet informatie verzameld worden. Welke informatie over de taal is nodig om een TTS te maken?

Welke digitale hulpmiddelen moeten er beschikbaar zijn?

Ga niet in op alle details, maar het moet wel duidelijk zijn wat je bij elkaar moet zoeken om aan de slag te kunnen.

3.b) Friese TTS

Een studente maakt een Friese spraaksynthese die werkt met de Nederlandse difonen. Dit leverde spraak op met een niet zo beste kwaliteit. Beschrijf in het kort hoe je eenvoudig een Friese difoon set zou kunnen maken met Festival/Nextens. Negeer de uiteindelijke codering van de golfvorm (bijvoorbeeld met MBROLA).

3.c) Friese TTS, “phrasing” en accenten

De schrijfwijze van het Fries lijkt erg op die van het Nederlands. Bij het maken van de Friese TTS wil de studente geen taalkundige analyses uitvoeren. Toch moet ze een manier vinden om lange zinnen op te splitsen in kortere stukken die eenvoudig uitgesproken kunnen worden. Wat kan ze het best doen?

In elke zinsdeel moeten één of meer woorden geaccentueerd worden. Wat voor een strategie kan ze gebruiken om woorden uit te kiezen om te accentueren? Wat voor een informatie over het Fries moet ze daarvoor zien te krijgen?

3.d) Latijnse TTS, klinkers

Twee studenten gebruiken het *eSpeak* programma om een Latijnse TTS te maken. De standaardversie van *eSpeak* gebruikt “difonen” die gemaakt worden met een hele simpele formantsynthese. Dat wil zeggen, ze moeten afzonderlijke klinkers en medeklinkers definiëren, en dan plakt *eSpeak* ze aan elkaar. Welke gegevens van de Latijnse *klinders* moeten de studenten hebben om deze synthese te laten werken?

3.e) Latijnse TTS, synthese

Het blijkt ook mogelijk te zijn om binnen eSpeak een MBROLA difoonset te gebruiken. De studenten maken ook een versie van hun synthese die een bestaande Latijnse MBROLA difoonset gebruikt. Deze set klinkt veel beter dan de eerdere formantsynthese.

Waarom is dat?

Met de MBROLA set is de synthese wel minder flexibel geworden. Wat kan niet met de MBROLA set, wat wel met de formantsynthese kan?