

Verslag EUROSPEECH/INTERSPEECH 2003 congres

1-4 september 2003 te Geneve en de COCOSDA meeting op 31 augustus 2003

Rob van Son

Instituut voor Fonetische Wetenschappen (IFA)/ACLC
Universiteit van Amsterdam

INLEIDING

Het achtste EUROSPEECH congres werd gehouden in het Centre International de Conférences de Genève (CICG) in Genève, Zwitserland. In het algemeen was de organisatie goed. Het gebouw was groot genoeg voor de 1100 deelnemers. Het was voor het eerst mogelijk om in het gehele gebouw met een standaard draadloze netwerkkaart te internetten. Dit ontlastte het aanwezige internetcafé zodat het zowaar mogelijk was om email te versturen en ontvangen. Bij eerdere congressen waren de wachttijden voor een plaatsje achter de computer zo lang dat email eigenlijk onmogelijk was. Een belangrijk minpunt was de onvoldoende ventilatie van de grote posterruimte in het souterrain. Het werd er in de loop van de dag erg warm en muf.

In de pauzes werden we voorzien van koffie/thee en sap met koekjes. Verder was er overal water verkrijgbaar. Na de eerste dag sloegen we het ontbijt (not-included) over aangezien de congresorganisatie ons rijkelijk voorzag van zoete broodjes.

Er vielen vrij veel presentaties uit. Soms wel twee in een sessie. De nationaliteit van de auteurs doet vermoeden dat hierbij, naast ontbrekende financiën, visumproblemen een rol kunnen hebben gespeeld.

COCOSDA

COCOSDA (www.cocosda.org) probeert een platform voor discussie en informatieuitwisseling te zijn op het gebied van spraak/taalbestanden en gereedschappen. Voorafgaande aan het eigenlijke congres was er op zondagmiddag 31 augustus een bijeenkomst gepland. Ik kwam pas op zondagmiddag aan en heb zo helaas het eerste uur gemis.

Tijdens de bijeenkomst werd er een overzicht gegeven over lopende en nieuwe programma's en de beschikbare corpora en databases in verschillende regio's. Naast de gewone problemen die er optreden bij het verzamelen van spraak-data, vnl geld en logistiek, zijn de beperkingen op de distributie van corpora een steeds terugkerend thema. Veel corpora worden verzameld door consortia van bedrijven en zijn commercieel. Deze corpora zijn moeilijk toegankelijk voor het onderzoek. Meestal hebben alleen de deelnemende onderzoeksinstituten direct toegang. Buitenstaanders moeten vaak aanzienlijke bedragen neertellen. Instellingen die zich toelagen op het distribueren van corpora, zoals ELRA en LDC, hebben overheadkosten die terugverdiend moeten worden met de verkoop van corpora. De politiek van LDC is om onderzoekers met een goed onderzoeksplan te helpen, ook als ze onvoldoende financiële middelen hebben (b.v., Oost-europese onderzoekers). Verder worden corpora die zichzelf terugverdiend hebben gratis ter beschikking gesteld.

Helaas heeft ELRA minder flexibele regels. De overhead van ELRA is vrij hoog zodat de prijzen van corpora die verspreid worden door ELRA erg hoog zijn. In de wandelgangen hoorde ik dat er nu met ELRA flink onderhandeld wordt over de verspreiding van het CGN. De gekozen licentie en de cont(r)acten met ELRA bemoeilijken bijvoorbeeld het beschikbaar maken van het CGN over het internet. Dit

dwingt onderzoekers de 100+ GB aan data fysiek op elk instituut te installeren. Dit kost niet alleen erg veel tijd, maar door enkel bugs in de CDrom's is dit ook niet triviaal.

Al deze overwegingen ondersteunen onze beslissing om het IFA corpus beschikbaar te stellen onder de GNU GPL licentie i.p.v. de licentie gebruikt voor het CGN.

In de discussietijd werd er gevraagd om bij een volgend Interspeech congres een special session voor COCODSA te organiseren. Dit wringt met de wens om ook deelnemers die niet voor EUROSPEECH geregistreerd zijn ter wille te zijn. Wel was het duidelijk dat COCODSA meer aan de weg moet timmeren op het gebied van de PR. Te weinig mensen weten van de bijeenkomsten af.

Er was nog een postersessie waarin verschillende nieuwe projecten gepresenteerd werden. Voor mij was vooral de constructie van een Japans prosodisch corpus met gebaren van belang. De opstellingen die voor de video-opnamen gebruikt zijn, zijn waarschijnlijk ook bruikbaar voor opnamen van gezichten in dialogen.

EUROSPEECH

Eurospeech 2003 trok ruim 1100 bezoekers. Er waren 1081 papers ingestuurd waarvan er 743 geaccepteerd waren. Er waren 3 plenary sessions, inclusief de acceptatierede van Jens Blauert na zijn ontvangst van de ISCA medal. Het congres zelf was georganiseerd rond 7 parallelsessies per slot van 2 uur: 4 oral- en 3 postersessies. De opening werd voorafgegaan en gevolgd door een optreden van Franstalige Jodelers. Ik zou het meer in de stijl van het congres gevonden hebben als er ook een machine-gestuurd jodeloptreden zou zijn geweest. De andere speeches waren meer PR dan informatie.

De organisatie had gevraagd om voorstellen voor special sessions over afzonderlijke onderzoeksgebieden. De onderzoekers konden deze sessies in overleg vullen. Dit werkte erg goed. In tegenstelling tot de eerdere poging tijdens ICSLP 2002 in Denver, waar de special sessions regelmatig ontaarden in gezellige onderonsjes, waren de sessies van goede kwaliteit. Toch waren er nog voordrachten in deze special sessions die duidelijk 'invited' waren en het nivo van de gewone congresbijdragen bij lange na niet haalden. Dit gold met name voor de Special Session 'Time is of Essence'. De laatste Special Session, 'Towards a roadmap for speech technology' was een regelrechte teleurstelling waar niets nieuws werd verteld. Al met al gaven de special sessions toch aan in welke richtingen het onderzoek zich beweegt.

De special session 'Towards Synthesizing Expressive Speech' is een goed voorbeeld. Naast deze overzichtssessie behandelden een vrij groot aantal bijdragen tijdens dit congres onderzoek naar mogelijkheden om spraaksynthese te voorzien van "emotie" en "gevoel" in de ruimste zin van het woord. Het onvermogen om een toepasselijke spraakstijl te gebruiken begint steeds meer de belemmerende factor in de spraaksynthese te worden. De gebruikelijke stemmen zijn geschikt voor "neutraal" nieuws. Maar toepassingen richten zich op voorleesmachines voor kinderen en 'entertainment' waar "expressiviteit" alles is.

De technieken om stemmen aan te passen aan verschillende spreekstijlen worden nu ontwikkeld op verschillende laboratoria. Een aansprekend voorbeeld is de bijdrage van Salor et al. (*PWeDf-5: A system of voice conversion based on adaptive filtering and line spectral frequency distance optimization for Text-to-Speech synthesis*) over het veranderen van een stem naar een andere stem. Zo kan de stem van een acteur die een buitenlandse film insprekt voor een lokale (Turkse) markt zo verdanderd worden

dat hij lijkt op die van de oorspronkelijke acteur. Dit onderbouwt de waarschuwingen (o.a., vanuit het OGI) dat bandopnamen van Osama bin Laden of Saddam Hoesein door iedereen gemaakt kunnen zijn.

Een andere trend is de opkomst van het chinese spraakonderzoek. Bijdragen over de verschillende chinese "dialecten" waren talrijk (~30), en er was een special session ingeruimd over Chinese Spoken Language Processing. Het Chinees volgt hierin het Koreaans. Vreemd genoeg waren er vrij weinig bijdragen over Indiase talen. Slecht 3 abstracts noemen specifiek Hindi en geen een, b.v., Tamil. Er waren er zelfs meer over het Thai (5 abstracts) en het Nederlands (6). Dit is vreemd gezien het grote aantal deelnemers dat oorspronkelijk uit India komt (alleen al 17 met een adres in India). De meeste Indiase onderzoekers werken blijkbaar in de Verenigde Staten en doen onderzoek aan andere (westerse) talen.

EIGEN BIJDRAGEN

Mijn eigen posterpresentatie was op dinsdagochtend. Ik stond naast een poster van Louis ten Bosch, een beetje achteraf. Dit betekende dat ik eigenlijk alleen belangstellenden kreeg die specifiek op mijn poster of een van de buurposters afkwamen. Ik had genoeg bezoekers om me ruim 2 uur, de duur van een sessie, aan het praten te houden. Ik ben tevreden over de reacties en heb enkele nuttige suggesties gekregen voor verder onderzoek. Met name ga ik meer aandacht geven aan interacties tussen factoren.

Dinsdagmiddag was de Education Arena waar ik een demonstratie van Praat heb gegeven. In totaal duurde de presentaties meer dan 4 uur en ik had op het eind weinig stem meer over. Ik heb ongeveer 20 CDroms met Praat software en demonstratiescripts uitgedeeld. Dit komt bovenop de Praat software die op de CDrom van de Education Arena zelf gepubliceerd was. Als demonstratie waren de sexe-veranderingen, vrouwenstem naar mannen- of kinderstem, en de bron-filter decompositie (een bijdrage van Masahiko Komatsu uit Japan) het meest populair. Het lijkt vreemd dat ik nog steeds Praat demonstraties geef, aangezien de meeste bezoekers Praat al schijnen te kennen. Maar toch zijn er veel gebruikers die zijn blijven hangen bij een oude versie of Praat enkel voor zeer beperkte taken gebruiken. Een belangrijke wens was om Praat beter te kunnen koppelen aan andere programma's, zoals TCL, Python of Perl.

De huidige demonstratie moest op zeer korte termijn samengesteld worden. Aangezien dit een jaarlijks terugkerend fenomeen lijkt te zijn, is het raadzaam om structureel demonstratiescripts te verzamelen. Deze moeten algemeen beschikbaar komen zodat ook andere Praatgebruikers hier nut van hebben. Ik heb al contact opgenomen met een vrijwilliger op de PraatUsers mailing list voor het beschikbaar maken van een Praat Script archief. Hier kunnen dan voortaan demonstratiescripts en andere nuttige scripts opgeslagen worden zodat we in de toekomst een groter aanbod aan demonstraties zullen hebben.