

## RECENZIJE I PRIKAZI

## EVROPSKA KONFERENCIJA O GOVORNOJ TEHNOLOGIJI

*J. Bakran i M. Stamenković*

Prisustvovali smo skupu o govornoj tehnologiji (European Conference on Speech Technology) održanom u Edinburghu od 2. do 4. rujna 1987. god. Govorna tehnologija (Speech Technology) termin je koji se odnosi na strojno, automatsko manipuliranje govornim signalom u uređajima za automatsko prepoznavanje i sintezu govora (APG i ASG). Prijavljeno je 250 radova iz 30 zemalja uključujući SAD, Japan, Kinu i SSSR. Izlagalo se istovremeno u četiri tematski odvojene sekcije. Radovi su izloženi usmeno i u okviru plakatne sekcije.

Za vrijeme savjetovanja organiziran je posjet fonetskom laboratoriju Odsjeka za lingvistiku Edinburškog sveučilišta i Centru za istraživanje tehnologije govora (CSTR — Centre for Speech Technology Research) gdje su održane demonstracije nekih njihovih postignuća. Klasičan fonetski laboratorij s početka stoljeća bitno je izmijenjen. Nismo vidjeli crni kimogaf ni palatografiju. Zastario je već i analogni višekanalni analizator za registraciju govornih pokreta i uvijek prisutan sonograf. Ključno mjesto opreme preuzima sistem kompjutera s perifernim elementi-

ma neophodnim za obradu govora (A/D i D/A pretvarači, procesori signala, digitalni filtri itd.) Pokazan nam je i višenamjenski program za analizu akustičkog signala AUDLAB razvijen na Edinburškom sveučilištu. Program impresionira svakog fonetičara. Omogućuje simultano promatranje višedimenzionalnih prikaza signala uz vrlo jednostavno rukovanje. Uključen je i spektralni prikaz sličan Key—spektrogramu.

Istovremeno s konferencijom organizirana je izložba opreme i proizvoda govorne tehnologije. Posebnu pažnju privukli su komercijalni sistemi za prepoznavanje povezanog govora tvrtke LOGICA (npr. sistem za praćenje teniskih utakmica). Nekoliko proizvođača ponudilo je sisteme za APG s rječnikom od 100 do 1000 riječi, s točnošću prepoznavanja 70—95% i vremenom odziva od 0.2 do nekoliko sekunda. Prepoznavanje se odnosi na definiranog govornika.

Prema broju članaka u vezi s govornom tehnologijom koje pojedine zemlje svijeta prikazuju na ICASSP (International Conferences in Acoustics, Speech and Signal Processing) možemo steći uvid u zamah istraži-

vanja. Sjedinjene su Američke Države najzastupljenije sa 53%, zatim zemlje Evrope sa 25% i Japan 12,5% članaka. U Evropi Francuska je najjača sa 31%, zatim Savezna Republika Njemačka 17%, pa Italija 12%, Velika Britanija 11%, Švedska 6% te Nizozemska sa 4% priloga. Na konferenciji je upućen poziv za sudjelovanje u osnivanju evropskog udruženja za govornu komunikaciju (European Speech Communication Association — ESCA), na žalost, samo za zemlje zapadne Evrope.

Teme članaka u teorijskom dijelu bile su: obrada govornog signala, sistemi za automatsko prepoznavanje govora, sinteza govora prema tekstu (Text to Speech), automatska identifikacija govornika, strojno prevodenje i ergonomske faktori u govornoj tehnologiji.

U području primjene članci obuhvaćaju sljedeće teme: telekomunikacije, uredsku opremu, vojne primjene, avijaciju, sistema za učenje, govornu kontrolu procesa i automatsko upravljanje, zdravstvo i primjenu govorne tehnologije kod proizvođa široke potrošnje.

Teorijski radovi obrađuju uglavnom razvoj novih modela kodiranja govornog signala, te mehanizme prepoznavanja i sinteze govora. Za digitalnu reprezentaciju govornog signala najčešće se upotrebljava tehnika LPC (Linear Predictive Coding) koja omogućuje redukciju širine kanala za prijenos govora na svega 800 bita u sekundi. Ista tehnika koristi se i za sintezu govora. Za razliku od prethodnih skupova posvećenih govornoj tehnologiji, gdje se uglavnom govorilo o prepoznavanju izoliranih riječi jednog govornika, edinburška konferencija protekla je u znaku sistema APG za povezani govor i tehnike adaptacije na više govornika. Pre-

poznavanje povezanog govora omogućila je, s jedne strane, paralelna arhitektura mikroprocesora (SPC—Speech Processing Chip i DSP—Digital Signal Processing Chip), a s druge, razvijeni matematički mehanizmi DTW (Dinamic Time Warping) i HMM (Hidden Markov Model). DTW algoritima uskladuju se vremenski pomaci i nelinearno normalizira signal, čime se smanjuje utjecaj različitog trajanja iste riječi. HMM model tretira govorne segmente kao Markovljev proces gdje vjerojatnost prijelaza na iduće stanje zavisi od prethodnog, i trenutno je jedan od najefikasnijih modela APG za izolirane riječi.

Primjena govorne tehnologije posljednjih nekoliko godina brzo se širi i očekuje se da će za idućih 5 godina 10% svjetske industrije na bilo koji način koristiti njezine proizvode. Sistemi APG istražuju se od 1950. godine, a primjenjuju se tek od 1970. Unatoč ogromnom broju istraživača i institucija koje se bave razvojem sistema APG, relativno je malo krajnjih korisnika. U industriji se sistem APG najčešće koristi za kontrolu proizvoda gdje kontrolor ima zauzet pogled i ruke. U Americi pošta razvrstava pakete pomoću sistema APG. Za avijaciju APG sistemi naročito se intenzivno istražuju čak i u internacionalnoj suradnji, ali tek od 1984. izvan laboratorijskih uvjeta — tj. u avionu, što je razumljivo zbog sigurnosnih razloga: pouzdanost sistema važnija je nego kod drugih primjena. Uz to su i uvjeti vrlo zahtjevni: prepoznavanje je otežano zbog velike buke i utjecaja psihološkog i fiziološkog stresa na glas govornika. U Evropi se Francuska, Engleska i SR Njemačka spremaju za uvođenje javnih telefonskih govornica bez bročjanika (govorno biranje brojeva).

Prva istraživanja na planu prepoznavanja govora motivirana su namjerom da se napravi pisani stroj koji bi pokretao govor. U studenom 1985. godine »Dragon Systems« (HMM model) demonstrirao je stroj za pisanje (2000 riječi za jednog govornika — Speaker Dependent Speech Recognition) koji je funkcionirao u približno realnom vremenu pomoću IBM PC/AT. Uz dodatne datoteke sistem se proširuje do 20000 riječi. Referat J. Baker o stanju govorne tehnologije u SAD vjerojatno je jedan od prvih koji nije ručno pisan, već diktiran. Očekuje se da će govorni pisani stroj (Talkwriter) u idućih 10 godina biti jedan od najtraženijih proizvoda govorne tehnologije. Posebno mjesto u sistemima APG zauzimaju primjene vezane za velike baze podataka s govornim ulazom. Još jedno od područja primjene pomoć je invalidnim osobama, i to za razne svrhe: glasom upravljana pomagala, komunikacija itd.

Velik broj članaka odnosi se na sisteme transformacije pisanog teksta u sintetički govor (Text to Speech). Problem uključuje jezičnu i fonetsku obradu prije same proizvodnje zvuka. Dio radova odnosi se samo na preobliku grafema u fonem, a dio rješava problem alofona i asimilacija. Uz sintezu se kao aktualna tema pojavljuju stil i dijalekatske razlike. Više ne zadovoljava kakva-takva razumljivost jer sinteza ne predstavlja više čudo, već je masovno primjenljiva. Vrlo kvalitetan prirodan govor danas se može sintetizirati samo na razini riječi ili sintagme, ali ne i na razini cijele rečenice, odnosno teksta. Za kvalitetnu je sintezu osim fonetskog bitan i lingvistički aspekt, osobito semantika teksta. Francuska je jedna od prvih zemalja koja već sada posjeduje eksperimentalni govorni sistem vezan na nacio-

nalnu računarsku mrežu. Na konferenciji je bio prikazan dio njihovog sistema razvijenog u CNET-u (Centre National d'Etudes des Telecommunications) koji sintetizira govor na osnovi neakcentiranog teksta.

Početna istraživanja iz sinteze i prepoznavanja govora bila su vezana za problem telekomunikacija, teoriju informacija i prepoznavanja uzoraka bez posebnog iskorištavanja fonetskog znanja o govornom zvuku. Međutim, doseg je takvog pristupa ograničen i on nikada ne može dovesti do konačnog rješenja — prepoznavanja povezanog govora. Zato se sve više govori o KBSR (Knowledge Based Speech Recognition) sistemima, tj. o nastojanju da se u postojeće formalne mehanizme ugradi fonetsko znanje o segmentaciji i prepoznavanju govornog niza. Teži se sisteme APG i ASG, hardverski i softverski, što više približiti načinu na koji čovjek generira i percipira zvuk, odnosno govor.

Među fonetskim temama nalazimo nekoliko priloga o važnosti i načinu programiranja trajanja govornih segmenata. Vrijeme se pojavljuje kao parametar i u mehanizmima prepoznavanja jer može signalizirati sintaksičku strukturu, položaj akcenta i pomoći diskriminaciji fonema.

Stalno su aktualne diskusije o relevantnim akustičkim karakteristikama pojedinih kategorija glasova — onima koje se efikasno mogu koristiti za automatsku detekciju. Kao prilog fundamentalnim fonetskim istraživanjima prikazan je jedan sistem analize artikulacijskih pokreta koji za razliku od röntgenskog i ultrazvučnog nije »agresivan«. Na jezik se pričvršćuje minijaturna zavojnica — »odašiljač«, a na fiksiranim mjestima oko glave nalaze se prijemnici. Kompjutor izračunava položaj

u prostoru i prikazuje ga na monitoru.

Dopustimo malu usporedbu — zamislimo li stanje fonetske tehnologije (odnosno znanosti) kao utrku, ne može se samo reći da je zagrebačka (jugoslavenska) fonetika »u zaostatku«, ovoliko ili onoliko krugova (godina), već da ona, na žalost, još uopće nije u trci. I sada bi se mogla uključiti, ali za to bi se trebalo dobro pripremiti i ući u krug. Poteškoća je u visokoj »kotizaciji« koju pojedinci ne mogu platiti. Pri tome treba stalno imati na umu da je govorna tehnologija ostavljena na milost i nemilost svakom pojedinom narodu, odnosno određenom govornom području. Moramo biti svjesni da su za gradnju APG i ASG sistema najkompetentniji fonetičari, lingvisti (uz podršku električara i informatičara) upravo onog govornog područja za koje se takvi sistemi projektiraju! Paradoksalno je da hrvatski standar-

dni govor još nema opisane ni formantske frekvencije. Danas to nema više smisla raditi metodama iz četrdesetih i pedesetih godina kada postoje daleko precizniji i efikasniji sistemi.

Fonetika je u svom velikom dijelu znanstvena disciplina koja se služi metodama »prirodnih« znanosti. Nepostojanje odgovarajućih instrumenata zaustavlja stvarni rad i rezultira bijegom u nedovoljno zasnovanu spekulaciju.

#### LITERATURA

1. »European Conference on Speech technology«, izdavač: J. Lever i M.A. Jack, Edinburgh 1987, vol. I/II