

Osnovna zgradba razpoznavalnika slovenskega tekočega govora UMB Broadcast News

Andrej Žgank, Tomaž Rotovnik, Mirjam Sepesy Maučec in Zdravko Kačič

Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru
Smetanova ul. 17, SI-2000 Maribor, Slovenija
andrej.zgank@uni-mb.si *http://www.dsplab.uni-mb.si*

Povzetek

V članku bomo predstavili zasnovo in osnovno zgradbo prvega slovenskega razpoznavalnika tekočega govora za domeno Broadcast News. Takšen razpoznavalnik govora je namenjen razpoznavanju govora v televizijskih (in radijskih) dnevnoinformativnih oddajah. Le-te predstavljajo zaradi zelo različnega akustičnega okolja in razlik v načinu govora zelo kompleksen problem na področju jezikovnih tehnologij. Sistem UMB Broadcast News smo zasnovali na slovenski govorni in tekstovni bazi BNSI Broadcast News. Za pripravo akustično homogenih zvočnih intervalov smo uporabili postopek akustične segmentacije na osnovi Gaussovih modelov. Kot akustične modele smo uporabili grafemske kontekstno odvisne HMM modele s 16 kombinacijami porazdelitev verjetnosti na stanje. Tvorili smo nabor različnih bigramskih in trigramskih jezikovnih modelov z 20k oz. 60k besedami v slovarju. Vrednotenje sistema UMB Broadcast News smo izvedli z bigramskim jezikovnim modelom ter slovarjem z 20k besedami. Dosegli smo 40,5% napako razpoznavanja besed na testnem naboru baze BNSI.

Basic Structure of the UMB Slovenian Broadcast News Transcription System

This paper presents basic structure and design scheme of the first Slovenian Broadcast News transcription system. Such system is used for large vocabulary continuous speech recognition of television (and radio) news shows. The acoustic environment and speaking style in Broadcast News speech corpora are heterogeneous and as such very complex for speech recognition. The UMB Broadcast News system was developed using the speech and text database BNSI Broadcast News. Acoustic homogeneous speech intervals were produced with the acoustic segmentation based on Gaussian models. Grapheme based context-dependent HMM acoustic models with 16 mixture probability density functions were applied. A set of different bigram and trigram language models with 20k and 60k words in the vocabulary was generated. The UMB Broadcast News system was evaluated with the bigram language model and 20k words in the vocabulary. The 40,5% word error rate was achieved on the BNSI evaluation set.

1. Uvod

Na področju avtomatskega razpoznavanja govora ločimo med sistemi različne kompleksnosti. Najtežjo nalogo tako predstavlja razpoznavanje tekočega spontanega govora z velikim slovarjem besed. V to kategorijo sodi tudi domena razpoznavalnikov govora Broadcast News (BN), kjer razpoznavamo govor v dnevnoinformativnih televizijskih oddajah. Razvoj na področju takšnih sistemov se je začel leta 1996 (Pallet, 2002) v okviru projekta ameriških organizacij DARPA in NIST. Danes v svetu najdemo razpoznavalnike govora iz domene BN, razvite za različne svetovne jezike (Gauvain et al., 2002; Woodland, 2002; Beyerlein et al., 2002). Kompleksnost samega razpoznavalnika govora je odvisna tudi od lastnosti jezika. Slovenski jezik sodi zaradi svojih lastnosti (pregibna narava, dvojina, relativno prost vrstni red besed v stavku,...) med težavnejše jezike za razpoznavanje govora, kar je ena izmed glavnih ovir za večji razvoj tega področja.

V članku bomo predstavili prvi razpoznavalnik slovenskega tekočega spontanega govora za domeno Broadcast News, ki je nastal na Univerzi v Mariboru¹. Sistem UMB Broadcast News je trenutno eden izmed najkompleksnejših slovenskih razpoznavalnikov govora. V prispevku bomo opisali zasnovo in osnovno strukturo sistema ter podali preliminarne rezultate razpoznavanja govora.

Osnovni jezikovni vir, ki smo ga uporabili pri razvoju UMB Broadcast News sistema, je slovenska baza BNSI Broadcast News (Zögling et al., 2003; Žgank et al., 2004/1; Žgank et al., 2005/1). Zaradi raznolikosti govornega materiala vključenega v bazo BNSI, se takšen razpoznavalnik govora sreča tako z branim govorom v študijskem okolju (npr. agencijske novice), kot tudi z narečnim spontanim govorom s šumom iz okolice ali glasbo v ozadju (npr. intervju na terenu).

Pri razvoju sistema UMB BN smo izhajali iz predhodnih razpoznavalnikov slovenskega govora (Žgank et al., 2001; Rotovnik, 2004), ki so bili razviti za bazo SNABI (Dreo, 1995) ter iz izkušenj pridobljenih v okviru razvoja demonstracijskega sistema za podnaslavljanje televizijskih oddaj v živo (Žgank et al., 2004/2).

V nadaljevanju članka bomo v drugem poglavju opisali različne slovenske jezikovne vire, ki smo jih uporabili pri razvoju. Osnovno zgradbo sistema UMB BN bomo opisali v tretjem poglavju. Rezultate razpoznavanja govora z bazo BNSI bomo predstavili v četrtem poglavju. Zaključek s smernicami za prihodnje delo bomo podali v zadnjem – petem – poglavju.

2. Jezikovni viri

Razpoložljivost primernih jezikovnih virov je osnova za razvoj vsakega razpoznavalnika govora. Za učenje akustičnih modelov tako potrebujemo transkribiran govorni material, za razvoj jezikovnih modelov pa tekstovne korpuse.

Učenje akustičnih modelov za sistem UMB BN smo izvedli z govornim korpusom slovenske baze BNSI Broadcast News (Žgank et al., 2004/1). Baza vsebuje 36

¹ Delo je bilo delno financirano s strani Agencije za raziskovalno dejavnost Republike Slovenije po pogodbi št. P2-0069.

ur transkribiranega govornega materiala iz obdobja 1999-2003, zajetega iz dnevnoinformacijskih oddaj RTV Slovenija. Med razvojem osnovne zgradbe razpoznavalnika govora je potreben precejšen delež ročnega dela za pripravo in uskladitev vseh jezikovnih virov. Zato smo zaradi poenostavitve dela uporabili samo polovico razpoložljivega govornega korpusa baze BNSI. Učni korpus je tako vseboval 15 ur govornega materiala, testni nabor pa približno 1,5 ure govora.

Tabela 1 prikazuje statistiko virov, ki smo jih uporabili pri učenju jezikovnih modelov. **BNSI-Speech** označuje transkripcije 69 oddaj iz osnovnega in razširjenega učnega govornega korpusa baze BNSI (Žgank et al., 2005/1). Oznaka **BNSI-Text** predstavlja tekstovni korpus baze BNSI (Žgank et al., 2005/1), ki vsebuje scenarije dnevnoinformativnih oddaj RTV Slovenija iz obdobja 1998-2004. Iz tekstovnega korpusa smo izločili tiste mesečne sklope scenarijev, v katerih se nahajata oddaji iz testnega nabora. **Večer** je korpus člankov časopisa Večer, ki smo jih zbrali v obdobju od leta 1998 do leta 2001. Poudariti velja, da sta prva dva jezikovna vira predstavnika govornega jezika, slednji pa sodi v skupino jezikovnih virov pisanega jezika. Najpomembnejše lastnosti govornega jezika, ki jih v pisanem jeziku ne zasledimo so: ponovitve, popravki, slovnično neujemanje in svobodni vrstni red besed. Vse našteje pojave zasledimo v korpusu **BNSI-Speech**, nekatere med njimi tudi v korpusu **BNSI-Text**. Naš cilj je razpoznavanje govornega jezika, zato sta iz tega stališča pomembnejša korpusa **BNSI-Speech** in **BNSI-Text**. Po drugi strani je uspešnost statističnega modeliranja odvisna od velikosti učnega korpusa, zato se kot pomemben vir znanja izkaže tudi korpus Večer.

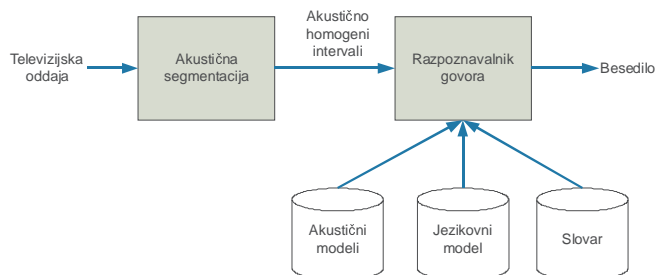
Korpus	Število stavkov	Število besed	Štev. različnih besed
BNSI-Speech	30k	573k	51k
BNSI-Text	614k	11M	175k
Večer	12M	95M	736k

Tabela 1: Statistika korpusov, uporabljenih pri gradnji jezikovnega modela.

Slovar razpoznavalnika smo sestavili tako, da je vseboval 20k in 60k najpogostejših besed v korpusih govornega jezika: **BNSI-Speech** in **BNSI-Text**.

3. Arhitektura sistema UMB Broadcast News

Sistem za razpoznavanje tekočega govora UMB Broadcast News je zasnovan na statističnem modeliranju govora. Blokovna shema osnovnih gradnikov sistema je predstavljena na sliki 1.



Slika 1: Osnovna zgradba sistema UMB Broadcast News za razpoznavanje govora.

3.1. Akustična segmentacija

Naloga akustične segmentacije je iz vhodnega zvočnega signala izrezati akustično homogene dele, ki so primerni za razpoznavanje govora. Tukaj lahko upoštevamo različne kriterije za doseganje homogenosti. Nekateri najpogostejši kriteriji so: zvok/tišina, govor/glasba/šum, spol govorca, združevanje govorcev v skupine, ...

Osnovno zasnovo sistema UMB BN za razpoznavanje govora smo zaenkrat pripravili za vključitev segmentacije na osnovi prvih treh kriterijev (Žgank, 2006). V prvem koraku segmentacije tako izločimo iz zvočnega signala dele, ki vsebujejo tišino. Sledi določitev akustičnega ozadja govornega signala, ki je posebej namenjeno identifikaciji govora z glasbo v ozadju. Le-ta namreč zelo oteži razpoznavanje govora. V tretjem koraku določimo spol govorca, kar omogoča uporabo različnih akustičnih modelov za oba spola. Ker količina govornega materiala, ki smo jo trenutno uporabili za učenje akustičnih modelov, ne omogoča kvalitetnega modeliranja ob uporabi drugih dveh kriterijev, smo za vrednotenje razpoznavalnika govora (tabela 4) uporabili samo klasifikacijo zvok/tišina.

Modul za segmentacijo smo zasnovali na osnovi Gaussovih modelov (GMM). Učenje GMM-ov je potekalo na 9 oddajah iz nabora baze BNSI. Glede na razpoložljivost zvočnega materiala za posamezno kategorijo, ter glede na dosežene rezultate segmentacije, smo uporabili različno kompleksnost modelov GMM – najpreprostejši model je vseboval 4 Gaussove porazdelitve verjetnosti, najkompleksnejši pa 512.

3.2. Učenje akustičnih modelov

V sistemu UMB BN smo zasnovali akustične modele na osnovi tri stanjskih levo-desnih prikritih modelov Markova s kombinacijami zveznih Gaussovih porazdelitev verjetnosti. Vektor značilik vsebuje 12 mel kepstralnih koeficientov in energijo, ter njihove prve in druge odvode.

Akustične modele smo zasnovali na grafemski osnovni enoti (Žgank in Kačič, 2005/2). V osnovnem naboru je bilo 27 akustičnih modelov. Z uporabo grafemov smo se izognili dodatni napaki, ki bi jo v sistem vnesla grafemsko-fonemska pretvorba. Za učenje prikritih modelov Markova smo uporabili prosto dostopno orodje HTK (HTK, 2006).

V prvem koraku učenja smo tvorili kontekstno neodvisne akustične modele na osnovi inicializacije z globalnimi vrednostmi. Z dobljenimi akustičnimi modeli smo izvedli postopek prisilne poravnave in izločili 0,63% neustreznih transkripcij. V drugem koraku smo ponovili postopek učenja. Sedaj smo inicializacijo izvedli na

osnovi ločenih vrednosti za vsak posamezni model. Ponovili smo postopek prisilne poravnave, ter kot neustrezne izločili dodatnih 0,47% transkripcij.

V tretjem prehodu smo korakoma nadaljevali z učenjem kontekstno neodvisnih akustičnih modelov. V nadaljevanju smo tvorili notranje-besedne kontekstno odvisne akustične modele, s katerimi se praviloma doseže boljši rezultat kot z kontekstno neodvisnimi modeli. Ker se posledično drastično poveča število prostih parametrov akustičnih modelov, ki jih je potrebno oceniti, smo uporabili postopek vezave stanj z odločitvenim drevesom (Young et al., 1994). S pomočjo tega postopka vezemo stanja, ki so si akustično dovolj podobna med seboj in tako združimo razpoložljivi učni material. Inicializacijo odločitvenih dreves smo izvedli s podatkovno tvorjenimi grafemskimi razredi (Žgank et al., 2005/3). Število Gaussovih porazdelitev verjetnosti na stanje v kontekstno odvisnih akustičnih modelih smo korakoma povečevali do 16. Po končanem postopku učenja smo akustične modele pretvorili v format, ki ga podpira razpoznavnik govora razvit na Univerzi v Mariboru.

3.3. Učenje jezikovnih modelov

Zgradili smo standardne bigramske in trigramske jezikovne modele. Uporabili smo Good-Turingovo glajenje. V model smo vključili vse bigrame oz. trigrame, tudi tiste, ki so se pojavili samo enkrat. Posledično so nastali relativno obsežni jezikovni modeli. Tabela 2 prikazuje velikosti jezikovnih modelov, učenih na različno velikih učnih besedilnih zbirkah. Jezikovni model **LM1** je učen le na korpusu **BNSI-Speech**, **LM2** na korpusih **BNSI-Speech** in **BNSI-Text** ter **LM3** na vseh treh korpusih: **BNSI-Speech**, **BNSI-Text** in korpusu **Večer**. V jezikovnem modelu **LM3** imajo vsi trije korpusi enak vpliv, ki smo ga v nadaljevanju želeli uravnovežiti glede na značilnosti BN korpusa. Zgradili smo tri komponente jezikovnega modela: prvo komponento na korpusu **BNSI-Speech**, drugo na korpusu **BNSI-Text** in tretjo na korpusu **Večer**. Optimalno razmerje med komponentami smo poiskali tako, da smo v iterativnem postopku iskali optimalne interpolacijske koeficiente (ki dajo najmanjšo perpleksnost interpoliranega modela). Nastali model smo poimenovali **LM4**. Vse omenjene postopke učenja jezikovnih modelov smo izvajali ločeno za slovar velikosti 20k besed in slovar velikosti 60k besed.

Jezikovni model	20k		60k	
	2-grami	3-grami	2-grami	3-grami
LM1	186k	298k	244k	380k
LM2	1,403M	3,542M	2,004M	4,640M
LM3	5,250M	18,623M	9,037M	28,307M
LM4	5,250M	18,623M	9,037M	28,307M

Tabela 2: Velikosti jezikovnih modelov, pri slovarjih 20k in 60k besed.

3.4. Razpoznavnik govora

Za razpoznavnik govora smo uporabili statistični pristop zasnovan na Bayesovem odločitvenem pravilu, ki vsebuje naslednje komponente: akustični analizador, iskalni algoritem, stohastična modela (akustični in jezikovni model). Akustični analizador izvaja kratkočasovno spektralno analizo govornega signala.

Akustični model zajema trenutno akustično in časovno karakteristiko govornega signala in skupaj z jezikovnim modelom podaja osnovna jezikovna vira za iskalni algoritem. Iskalni algoritem določi besedni niz neznane dolžine na osnovi največje aposteriori verjetnosti. Za naš razpoznavnik govora smo uporabili sinhroni iskalni algoritem (niz akustičnih vektorjev značilke se procesira od začetka do konca govornega signala), ki za zmanjšanje računske zahtevnosti vključuje Viterbijev aproksimacijo. Za zmanjšanje iskalnega prostora smo uporabili drevesno obliko slovarja, katerega značilnost so skupna vozlišča za enake začetne foneme besed. Sama izvedba slovarja vpliva na velikost iskalnega prostora. Ker smo v razpoznavniku govora uporabili bigramske jezikovne modele je bilo potrebno za vsako besedo iz slovarja tvoriti kopije dreves slovarja za ohranitev zgodovine hipotez v iskalnem prostoru. V korenih dreves se nato izvaja rekombinacija na besednem nivoju. Upošteva se samo najverjetnejša hipoteza, ki pride v določenem časovnem okviru v koren, vse ostale se zavrnejo. Zaradi vzporedne obdelave vseh možnih hipotez iskalnega prostora, smo v vsakem časovnem okviru izključili tiste hipoteze, katerih verjetnost je bila za določen prag slabša od trenutno najboljše hipoteze. Tako imenovano snopovno omejevanje dodatno zmanjša aktivni iskalni prostor, torej del iskalne mreže, ki jo je potrebno preiskati, za nekaj desetkrat in s tem pohitri proces razpoznavanja govora. Za učinkovitejše snopovno omejevanje smo v iskalni prostor predčasno vključili verjetnosti jezikovnega modela. Tako imenovan pogled naprej jezikovnega modela v vsako skupno vozlišče drevesnega slovarja postavi najboljšo verjetnost za vse možne besede, ki lahko nastanejo iz danega vozlišča. Za dodatno zmanjšanje iskalnega prostora smo v vsakem časovnem okviru omejili število aktivnih vozlišč, katerim se določajo (izračunavajo) nove poti v iskalni mreži. Omejevanje smo izvedli na dva načina. V prvem primeru smo aktivne modele sortirali po najboljšem rezultatu in pri tem ohranili samo N najboljših. V drugem primeru smo nadaljevali samo N najboljših delnih hipotez, ki so se v trenutnem času končale v zadnjem stanju in zadnjem vozlišču dreves.

4. Rezultati razpoznavanja govora

Pred samim vrednotenjem razvitega sistema UMB Broadcast News bomo podali statistiko uporabljenih akustičnih in jezikovnih modelov, ki kaže na kompleksnost razvitega sistema.

V naboru sta bila pred vezavo stanj 16.902 kontekstno odvisna grafemska akustična modela. Po vezavi stanj je ostal 21,5% neodvisnih akustičnih modelov, ki so imeli skupaj približno 175k Gaussovih porazdelitev verjetnosti. Skupna velikost datoteke z akustičnimi modeli je znašala 86,6 MB.

Jezikovni model	20k		60k	
	2-gram	3-gram	2-gram	3-gram
LM1	504	459	837	821
LM2	419	372	641	575
LM3	344	285	471	390
LM4	292	234	400	316

Tabela 3: Perpleksnosti jezikovnih modelov.

Delež besed izven slovarja (OOV) v naboru testnih stavkov pri slovarju z 20k besedami je 12.34% in pri slovarju s 60k besedami 5.44%. Tabela 3 prikazuje perpleksnosti nabora testnih stavkov z uporabo različnih jezikovnih modelov. Prvi trije modeli LM1-3 kažejo, da z večanjem učnega korpusa perpleksnost testnih stavkov pada.

Rezultate razpoznavanja govora bomo podali za velikost slovarja 20k najpogostejših besed v tekstovnih korpusih in bigramski jezikovni model LM4. Beležili smo napako razpoznavanja besed (NRB), hitrost razpoznavanja (večkratnik realnega časa, CPU: P4, 2,4 GHz) in velikost iskalnega prostora, izraženega s povprečnim številom aktivnih modelov. Povprečno število aktivnih modelov smo izračunali tako, da smo najprej v vsakem časovnem okviru za vsak testni stavek pred procesom omejevanja beležili število aktivnih modelov. Vsoto modelov za posamezni stavek smo normalizirali s številom okvirov. Skupno vsoto normaliziranih aktivnih modelov posameznih stavkov smo povprečili z velikostjo testne množice.

Eksperiment	NRB[%]	Hitrost	Št. aktivnih modelov
B20	40,5	42,3	26620

Tabela 4: Rezultat razpoznavanja z besednimi modeli pri velikosti slovarja 20k enot.

Z velikostjo slovarja 20k besed smo dosegli napako razpoznavanja 40,5 % (tabela 4). Podrobnejša analiza je pokazala, da imajo največji vpliv na napako razpoznavanja manjkajoče besede, ki jih ni v slovarju. 20% vseh zamenjanih besed predstavljajo besede, ki so izpeljane iz skupne leme in so si fonetično zelo podobne. Analiza rezultatov razpoznavanja nakazuje na uporabo tehnik razpoznavanja primernih za pregibne jezike (npr.: uporaba podbesednih enot razpoznavanja). V primerjavi z enakovrednim najboljšim predhodnim razpoznavnim sistemom, kjer smo uporabili govorno bazo SNABI, smo zmanjšali napako razpoznavanja za absolutno 12,8%.

Z omejevanjem iskalnega prostora smo vplivali na hitrost razpoznavanja in ga optimirali glede na najboljši rezultat razpoznavanja. Dosežena hitrost razpoznavanja je bila 42,3-kratna vrednost realnega časa, pri povprečju 26620 aktivnih modelov na časovni okvir.

5. Zaključek

V članku smo predstavili zasnovo in osnovno strukturo prvega slovenskega sistema za razpoznavanje tekočega govora v domeni Broadcast News. Dosegli smo vzpodbudne rezultate razpoznavanja govora.

V nadaljnjem delu bomo korakoma povečevali velikost uporabljene učne govorne baze in tako vključili segmentacijo zvočnega signala glede na druga dva kriterija, ter povečali slovar razpoznavalnika govora. Postopoma bomo sistemu UMB Broadcast News dodajali tudi nove module, s katerimi lahko pričakujemo dodatno izboljšanje kvalitete razpoznavanja govora.

6. Literatura

Beyerlein, P., Aubert, X., Haeb-Umbach, R., Harris, M., Klakow, D., Wendenmuth, A., Molau, S., Ney, H., Pitz,

- M., and Sixtus, A., (2002). Large vocabulary continuous speech recognition of Broadcast News - The Philips/RWTH approach, *Speech Communication*, Volume 37, Issues 1-2, 109-131.
- Dreo, D., (1995). Slovene speech data base SNABI. *Dialog Man - Machine: second International Workshop*, Maribor, Slovenija.
- Gauvain, J., Lamel, L., Adda, G., (2002). The LIMSI Broadcast News Transcription System. *Speech Communication*, Volume 37, Issues 1, 89-108.
- HTK domača stran, <http://htk.eng.cam.ac.uk>.
- Pallett, D. S. (2002). The role of the National Institute of Standards and Technology in DARPA's Broadcast News continuous speech recognition research program. *Speech Communication*, Vol. 37, Issues 1-2, 1:3--14.
- Rotovnik T., Avtomatsko razpoznavanje govora za pregibni jezik z velikim slovarjem besed z uporabo podbesednih modelov osnova - končnica, *Doktorska disertacija, Univerza v Mariboru*, 2004.
- Zögling Markuš, A., Žgank, A., Rotovnik, T., Sepesy Maučec, M., Vlaj, D., Hozjan, V., Kotnik, B., (2003). Spoken Language Resources at University of Maribor. *Proc. of 10th International Workshop Advances in Speech Technology 2003*, Maribor, Slovenija.
- Woodland, P.C., (2002). The development of the HTK Broadcast News transcription system: An overview, *Speech Communication*, Volume 37, Issues 1-2, 47-67.
- Young, S., Odell, J., Woodland, P., (1994). Tree-based State Tying for High Accuracy Acoustic Modelling. *Proc. ARPA Human Language Technology Conference Plainsboro*.
- Žgank, A., Kačič, Z., and Horvat, B., (2001). 'Large vocabulary continuous speech recognizer for Slovenian language, *Proc. Text, speech and dialogue : 4th international conference, TSD 2001*, Železna Ruda, Češka, *Lecture notes in Artificial Intelligence*, Vol. 2166, 242-248, Springer.
- Žgank, A., Rotovnik, T., Sepesy Maučec, M., Verdonik, D., Kitak, J., Vlaj, D., Hozjan, V., Kačič, Z., Horvat, B., (2004/1). Acquisition and annotation of Slovenian broadcast news database. *Fourth international conference on language resources and evaluation, LREC 2004*, Lizbona, Portugalska.
- Žgank, A., Rotovnik, T., Verdonik, D., Kačič, Z., (2004/2). Baza Broadcast News za slovenski jezik (BNSI) in sistem za razpoznavanje tekočega govora. Zbornik konference Jezikovne tehnologije 2004, Ljubljana, Slovenija.
- Žgank, A., Verdonik, D., Zögling Markuš, A., Kačič, Z., (2005/1). BNSI Slovenian Broadcast News database - speech and text corpus. *Proc. Interspeech 2005*, Lizbona, Portugalska.
- Žgank, A., Kačič, Z., (2005/2). Primerjava treh tipov akustičnih osnovnih enot razpoznavalnika slovenskega govora. *Elektrotehniški vestnik*, 2005, Ljubljana, Slovenija.
- Žgank, A., Horvat, B., Kačič, Z., (2005/3). Data-driven generation of phonetic broad classes, based on phoneme confusion matrix similarity. *Speech Communication*, vol. 47, issue 3, 379--393, november 2005.
- Žgank, A., (2006), Acoustic Segmentation for Slovenian Broadcast News Transcription System. *Proc. of 13th International Workshop Advances in Speech Technology 2006*, v tisku, Maribor, Slovenija.