

# Prenos razpoznavalnika govora na nov jezik in vpliv jezikovnih skupin

Andrej Žgank, Zdravko Kačič, Bogomir Horvat

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

## Povzetek

V članku predstavljamo rezultate raziskave o uspešnosti prenosa obstoječih akustičnih modelov na nov jezik, ne da bi uporabili govor v novem jeziku za učenje ali adaptiranje akustičnih modelov. Prednost takšnega pristopa je v dejstvu, da se lahko na takšen način izognemo uporabi baze izgovorjav. Obstoječe večjezične akustične modele za slovenski, nemški in španski jezik smo zgradili z uporabo odločitvenih dreves na osnovi skupnih fonetičnih skupin. Tako pripravljene večjezične akustične modele smo uporabili za razpoznavanje italijanskega govora. Preslikavo italijanskih fonemov v obstoječe foneme smo izvedli s pomočjo IPA tabel, s katerimi smo iskali ekvivalentne foneme oziroma najboljši približek. Uporabili smo pet različnih načinov preslikave. S prenesenimi večjezičnimi akustičnimi modeli smo v najboljšem primeru dosegli 55.98% napako razpoznavanja besed, medtem ko je znašala napaka razpoznavanja besed italijanskega referenčnega razpoznavalnika govora 34.32%. Z dodatno analizo rezultatov smo preverili, kako jezikovne skupine, v katere spadajo posamezni jeziki v raziskavi, vplivajo na uspešnost prenosa razpoznavalnika govora na nov jezik.

## 1. Uvod

Z nenehnim razvojem tehnologije prihajajo vedno bolj v uporabo tudi jezikovne tehnologije (Johnston, et al., 1997) in z njimi avtomatsko razpoznavanje govora<sup>1</sup>. V primeru izgradnje govorno vodene aplikacije, ki je namenjena posredovanju informacij večjemu številu uporabnikov (razpoznavanje govora neodvisnega govorca), je velikokrat smiselno vključiti tudi podporo za avtomatsko razpoznavanje govora v različnih jezikih. Na takšen način lahko aplikacijo uporabljajo tudi tuje govoreči uporabniki.

V primeru, ko začnemo z izgradnjo sistema za avtomatsko razpoznavanje govora po standardnem postopku, potrebujemo za vsak posamičen vključen jezik samostojno bazo govornega materiala. Zagotovitev govorne baze za vsak posamezen jezik je v večini primerov časovno in finančno izredno potratno opravilo, še posebej v primerih, ko potrebujemo govorni material za izdelavo razpoznavalnika govora za neodvisnega govorca. Kot primer lahko navedemo, da je izdelava slovenske baze SpeechDat(II) 1000 FDB trajala približno dve leti (Kaiser in Kačič, 1998).

Pri izdelavi avtomatskega razpoznavalnika govora za nov jezik se je možno izogniti uporabi baze podatkov z govornim materialom s pomočjo prenosa obstoječih akustičnih modelov na nov jezik. Da bi lahko takšen prenos razpoznavalnika govora na nov jezik uspešno izvedli, zadostuje že, če imamo fonetični slovar v novem jeziku, ki ga je možno tvoriti tudi avtomatsko s pomočjo grafemske – fonemskih pravil. Uspešnost razpoznavanja govora v novem jeziku je možno dodatno izboljšati z adaptiranjem oziroma ponovnim učenjem akustičnih modelov v primeru, ko imamo na voljo tudi govorni material v novem jeziku (zadostuje že nekaj sto kratkih posnetkov - to je manj kot 1% celotne baze SpeechDat(II) (van den Heuvel, et al., 2001)).

Metode za prenos razpoznavalnika govora na nov jezik lahko v grobem razdelimo na dve skupini:

- Ekspertno vodene: prenos razpoznavalnika govora na nov jezik izvede razvijalec sistema sam, pri tem uporablja različne fonetične podatke (Ladefoged, 1993), npr. IPA tabele (IPA, 2002). Slabost opisanega pristopa je v tem, da metoda vsebuje tudi subjektivni vpliv izvajalca (Schultz, 2000).
- Podatkovno vodene: kriterije za prenos razpoznavalnika govora na nov jezik določimo na osnovi nabora posnetkov v novem jeziku, ki ga kasneje lahko uporabimo tudi za adaptacijo prenesenih akustičnih modelov (Schultz, 2000). Slabost te metode je, da je zahtevnejša kot ekspertno vodena metoda.

V članku bomo predstavili nadaljevanje našega dela (Imperl, et al., 2002; Žgank, et al., 2001/1; Žgank, et al., 2001/2) na področju večjezičnega razpoznavanja govora. V primeru, kadar želimo prenesti razpoznavalnik govora na nov jezik, se je izkazala za upravičeno uporaba večjezičnih akustičnih modelov, saj le-ti pokrivajo obsežnejši akustični prostor kot pa enojezični akustični modeli (Schultz, 2000). V opisanem eksperimentu so bili zato uporabljeni obstoječi večjezični (za slovenski, nemški in španski jezik) akustični modeli. S pomočjo IPA tabel (IPA, 2002) smo takšen razpoznavalnik govora prenesli v italijanski jezik, ne da bi uporabili učenje z italijanskimi posnetki. Dobljene rezultate smo primerjali z rezultati prenosa razpoznavalnika govora na norveški jezik (Žgank, et al., 2001/1) in s tem preučili vpliv jezikovnih skupin na uspešnost prenosa sistema za avtomatsko razpoznavanje govora.

V drugem poglavju sledi kratek opis tvorjenja večjezičnih akustičnih modelov. Uporabljene baze podatkov so opisane v tretjem poglavju. Četrto poglavje vsebuje opis izgradnje sistema in predstavitev italijanskega enojezičnega razpoznavalnika govora, ki je služil kot referenca. Prenos večjezičnega sistema na nov jezik je opisan v petem poglavju. Rezultate in njihovo analizo vsebuje šesto poglavje, zaključek pa je podan v sedmem poglavju.

<sup>1</sup>Delo je bilo finančno podprto s strani Ministrstva za šolstvo, znanost in šport Republike Slovenije, po pogodbi PP-0796/99-0796-503.

## 2. Večjezični trifonski akustični modeli

V opisani raziskavi smo za prenos razpoznavnika govora na italijanski jezik uporabili večjezične kontekstno odvisne (trifonske) akustične modele, ki smo jih tvorili za slovenski, nemški in španski jezik. Večjezični trifonski akustični modeli so bili zgrajeni iz enojezičnih monofonskih akustičnih modelov. Pri tem je bilo uporabljeno združevanje akustičnih modelov na osnovi binarnega odločitvenega drevesa (Young, et al., 1994). Fonemi iz vseh treh jezikov so bili na osnovi svojih fonetičnih lastnosti (Ladefoged, 1993; Bezljaj, 1960) združeni v skupne večjezične fonetične skupine. Vsakemu fonemu je bila dodana še oznaka jezika, tako da smo lahko med seboj razločevali tudi identične foneme v različnih jezikih. Te fonetične skupine so bile uporabljene za tvorjenje vprašanj v odločitvenih drevesih.

Da bi omogočili tudi vključevanje dodatnih jezikov v akustične modele, smo se v odločitvenem drevesu izognili uporabi posebnih vprašanj za vrsto jezika. Uporaba odločitvenega drevesa pri tvorjenju trifonskih akustičnih modelov omogoča kasnejše dodajanje novih, "nevidenih" trifonov k obstoječim trifonom, kar je izredno pomembno pri metodi prenosa razpoznavnika govora. Število "nevidenih" trifonov med postopkom prenosa na nov jezik znaša nekje med 30% in 60% celotnega števila trifonov v fonetičnem slovarju razpoznavnika govora (Žgank, et al., 2001/1).

## 3. Govorna baza podatkov

Za izdelavo večjezičnega razpoznavnika govora so trenutno ene izmed najbolj primernih baze iz serije SpeechDat (Höge, et al., 1997). Serija baz SpeechDat(II) trenutno pokriva več kot 20 različnih evropskih jezikov. Ostale baze iz družine SpeechDat (SALA, SpeechDat East) pa pokrivajo še dodatne jezike. Baze SpeechDat so bile zasnovane za razvoj sistemov za telefonsko razpoznavanje govora (Höge, et al., 1997). Posnete so bile na osnovi enakih specifikacij (van den Heuvel, et al., 2001). Klicanci so uporabljali stacionarne ali mobilne telefone, odvisno od tipa posamezne baze. Vsebujejo izgovarjave 1000 ali 5000 različnih govorcev. V naši raziskavi smo uporabili sledeče baze iz serije SpeechDat(II):

- slovenski 1000 FDB SpeechDat(II) - SL (Kaiser in Kačič, 1998),
- nemški 1000 FDB SpeechDat(II) - DE,
- španski 1000 FDB SpeechDat(II) - ES,
- italijanski 1000 FDB SpeechDat(II) - IT.

Tisoč govorcev iz posamične baze podatkov smo razdelili v učni nabor (800 govorcev) in testni nabor (200 govorcev). Stavke, ki so neprimerni za učenje akustičnih modelov (van den Heuvel, et al., 2001), smo izločili iz obeh naborov. Učni nabor za posamičen jezik je vseboval več kot 20.000 posnetkov, tako v obliki izoliranih ali povezanih besed, kot v obliki tekočega govora. Da smo se izognili vplivu jezikovnega modela, smo v testnem naboru uporabili izolirane besede. Na takšen način smo lahko bolj

zanesljivo ovrednotili uspešnost prenosa akustičnih modelov na nov jezik, saj smo izločili vpliv jezikovnega modela (Young, 1996).

## 4. Izgradnja razpoznavnika govora

Vse potrebne akustične modele smo zgradili s skripto "refrec0.9" (Refrec homepage, 2002), ki je nastala v okviru projekta COST 249 "SpeechDat task force" (Johansen, et al., 2000; Lindberg, et al., 2000), tako da smo lahko primerjali dobljene rezultate s predhodno objavljenimi rezultati.

Uporabili smo spremenjen modul za tvorjenje akustičnih značilk (Haunstein in Marschall, 1995) in s tem izboljšali robustnost sistema za razpoznavanje govora. Akustične značilke so vsebovale 24 mel kepstralnih koeficientov (uporabljenih je bilo 30 filtrov na mel frekvenčni skali) in energijo ter prve in druge odvode. Akustični modeli so bili zasnovani na osnovi levo desnih prikritih modelov Markova. Posamezen model je imel tri stanja, končni akustični modeli pa so imeli 32 Gaussovih porazdelitev verjetnosti na stanje (Johansen, et al., 2000). Najprej so bili zgrajeni enojezični akustični modeli za vsak uporabljen jezik. Slovenski, nemški in španski enojezični akustični modeli so služili za izgradnjo večjezičnih akustičnih modelov, italijanski enojezični modeli pa kot referenca.

Italijanski enojezični akustični modeli so na testnem naboru fonetično uravnoveženih izoliranih besed (nabor W1-4 (van den Heuvel, et al., 2001)) dosegli napako razpoznavanja besed 34.32%. V fonetičnem slovarju italijanskega razpoznavnika govora je bilo 538 različnih besed, kar sistem uvršča med razpoznavnike govora s srednjim slovarjem besed. Testni nabor za italijanski jezik je vseboval 711 posnetkov.

## 5. Prenos večjezičnega razpoznavnika na italijanski jezik

V prvem koraku prenosa na nov jezik je potrebno preslikati italijanski fonetični slovar tako, da namesto italijanskih fonemov vsebuje foneme iz obstoječih jezikov (SL, DE, ES). To preslikavo smo izvedli s pomočjo IPA tabele (IPA, 2002). Vsakemu italijanskemu fonemu smo poiskali njegov ekvivalent v obstoječem jeziku. V primeru, kadar ekvivalenten fonem ni obstajal, smo poiskali najbolj primerno zamenjavo med vsemi obstoječimi fonemi. Odločili smo se, da bomo uporabili pet različnih načinov preslikanja italijanskega fonetičnega slovarja:

- italijanščina v slovenščino (IS),
- italijanščina v nemščino (ID),
- italijanščina v španščino (IE),
- italijanščina v optimalnega (IM),
- italijanščina v vse tri jezike (IP).

Primer originalnega italijanskega fonetičnega slovarja in vseh petih načinov preslikanja je podan v tabeli 1.

V IM načinu preslikave smo najprej pogledali, če ima italijanski fonem svoj ekvivalent v španskem naboru

Slovar	Primer
Org.	castello: k a s t E l l o
IS	castello: SL-k SL-a SL-s SL-t SL-E SL-l SL-o
ID	castello: DE-k DE-a DE-s DE-t DE-E: DE-l DE-o:
IE	castello: ES-k ES-a ES-s ES-t ES-e ES-l ES-o
IM	castello: ES-k ES-a ES-s ES-t SL-E ES-l ES-o
IP	castello: SL-k SL-a SL-s SL-t SL-E SL-l SL-o castello: DE-k DE-a DE-s DE-t DE-E: DE-l DE-o: castello: ES-k ES-a ES-s ES-t ES-e ES-l ES-o

Tabela 1: Primer fonetičnega zapisa italijanske besede v različnih načinih preslikave fonetičnega slovarja

fonemov. Za takšno izbiro smo se odločili, ker oba jezika pripadata romanski jezikovni skupini in sta si tako najbolj sorodna. V primeru, če v španskem naboru fonemov nismo našli ekvivalenta italijanskemu fonemu, smo ekvivalent dalje iskali med nemškimi in slovenskimi fonemi. Šele v primeru, ko tudi tukaj nismo našli ekvivalenta italijanskemu fonemu, smo namesto ekvivalenta začeli iskati najbolj podoben fonem iz nabora vseh treh jezikov. V primeru, ko je bil uporabljen IM način preslikave fonetičnega slovarja, se lahko pojavijo trifoni, ki jih sestavljajo fonemi iz različnih jezikov. Ta primer preslikave fonetičnega slovarja ima običajno največje število novih "nevidenih" trifonov. IP način preslikave fonetičnega slovarja je izveden na takšen način, da ima vsaka italijanska beseda v slovarju fonetični zapis v vsakem od treh obstoječih jezikov. Na ta način se sicer za trikrat poveča iskalni prostor v dekodirniku, vendar prepustimo izbiro med različnimi načini preslikave razpoznavaniku govora, ki se odloča na osnovi podatkov.

V drugem koraku prenosa sistema avtomatskega razpoznavanja govora na nov jezik je potrebno poiskati vse novo nastale trifone, ki jih še ni v obstoječih akustičnih modelih. S pomočjo ponovne uporabe istih odločitvenih dreves, kot so že bila uporabljena za tvorjenje večjezičnih akustičnih modelov, so bili novi trifoni dodani že obstoječim trifonom.

## 6. Analiza rezultatov prenosa razpoznavalnika govora

Enak nabor testnih stavkov tistemu, ki smo ga uporabili za vrednotenje uspešnosti razpoznavanja govora z enojezičnimi italijanskimi akustičnimi modeli (referenčna vrednost), smo sedaj uporabili za vrednotenje prenesenih večjezičnih akustičnih modelov. Rezultati razpoznavanja govora za ta primer so podani v tabeli 2. Različni načini preslikanja italijanskega jezika v obstoječe jezike so označeni na enak način kot v tabeli 1. Uspešnost razpoznavanja govora smo vrednotili na osnovi napake razpoznavanja besed, ki je podana v odstotkih.

Kot lahko vidimo v tabeli 2, smo dobili najboljši rezultat (55,98%) v primeru, ko je bilo uporabljeno preslikanje italijanskih fonemov v španske foneme (IE). Tako rekoč enako dober rezultat (56,12%) je bil dosežen tudi z uporabo paralelne metode (IP), ko je vsaka italijanska beseda v fonetičnem slovarju imela fonetični zapis v vseh treh uporabljenih jezikih (glej tabelo 1). Na takšen

Način preslikanja	NRB (%)
IS	83,41
ID	66,67
IE	55,98
IM	58,09
IP	56,12

Tabela 2: Napaka razpoznavanja besed (NRB) za italijanski jezik z različnimi načini preslikave italijanskega fonetičnega slovarja

način je razpoznavalnik govora sam izbral optimalen način preslikave italijanskega fonetičnega slovarja. Upoštevati je potrebno tudi slabost takšnega načina preslikave, to je trikrat daljši čas razpoznavanja govora pri enakem testnem naboru. Če primerjamo ta rezultata z enojezičnim italijanskim razpoznavalnikom, ki je služil kot referenca (34,32% NBR), lahko vidimo, da znaša razlika v napaki razpoznavanja besed nekaj več kot 20% absolutno. Upoštevati moramo, da preneseni večjezični akustični modeli niso bili adaptirani ali doučeni z italijanskim govornim materialom. Če bi uporabili še postopek adaptacije akustičnih modelov na majhnem naboru posnetkov, bi lahko razliko med enojezičnim sistemom in prenesenim sistemom dodatno zmanjšali (Žgank, et al., 2001/2).

Skoraj tako dobre rezultate kot IE in IP preslikanje je dalo IM preslikanje. Kljub temu da IM način preslikanja italijanskih fonemov privede do kršenja pravil fonotaktike (Ladefoged, 1993), je razpoznavalnik govora dosegel 58,09% napako razpoznavanja besed. Slabše kot prvi trije opisani načini preslikanja sta se odrezala ID in IS način preslikanja, ki sta dosegla 66,67% in 83,41% napako razpoznavanja besed. Na osnovi primerjave dobljenih rezultatov je možno sklepati, da je bil slabši rezultat v teh dveh primerih preslikanja dosežen zaradi večje različnosti med italijanskim ter nemškim in slovenskim jezikom.

Rezultate razpoznavanja italijanskega govora, ki smo jih dobili z IP načinom preslikanja italijanskih fonemov, smo podrobneje razčlenili in tako raziskali vpliv jezikovne skupine na uspešnost prenosa razpoznavalnika govora na nov jezik. Rezultati te analize so predstavljeni v tabeli 3.

Jezik	Izbrani	Pravilni
SL	9,84%	32,86%
DE	43,04%	34,64%
ES	47,12%	54,93%

Tabela 3: Porazdelitev izbranih jezikovnih različic fonetičnega slovarja v primeru IP preslikave in odstotek izbranih variant, ki so bile pravilno razpoznane

Kot lahko vidimo v tabeli 3, je razpoznavalnik govora na osnovi akustičnih značilik največkrat izbral špansko različico fonetičnega zapisa italijanske besede. Od tega je bilo 54,93% izbranih besed tudi pravilno razpoznanih. Za slovensko oziroma nemško različico fonetičnega zapisa italijanskih besed se je razpoznavalnik odločil v 9,84% in 43,04%. Tudi uspešnost razpoznavanja izbranih besed

je bila v obeh primerih nižja kot v primeru španskega fonetičnega zapisa.

Podoben rezultat, odvisen od jezikovnih skupin, smo dobili tudi, ko smo enak nabor večjezičnih akustičnih modelov (SL, DE, ES) prenašali na norveški jezik (Žgank, et al., 2001/2). V primeru norveškega jezika je bilo največ izbranih besed z nemško različico fonetičnega zapisa. Oba jezika, tako norveški kot nemški, sodita v germansko jezikovno skupino. Na jezika iz ostalih dveh skupin - slovanske in romanske - je prišlo občutno manjše število izbranih različic fonetičnega zapisa. Tudi uspešnost razpoznavanja izbranih različic fonetičnega zapisa v neidentični jezikovni skupini je bila nižja kot pa pri primeru nemške različice fonetičnega zapisa.

Na osnovi rezultatov, predstavljenih v tabeli 3 in članku (Žgank, et al., 2001/2), lahko zaključimo, da ima jezikovna skupina pomemben vpliv pri prenosu razpoznavnika govora na nov jezik. Najboljši rezultat razpoznavanja govora z novim jezikom lahko pričakujemo takrat, kadar foneme iz novega jezika preslikamo v foneme obstoječega jezika, ki sodi v isto jezikovno skupino. Pri tem moramo dodatno upoštevati, da so bili vsi eksperimenti izvedeni na večjezičnem naboru akustičnih modelov. Takšni večjezični akustični modeli so zaradi vezave stanj z odločitvenim drevesom učeni na širšem naboru govornih posnetkov, kot bi bili podobni enojezični akustični modeli. Takšno učenje na širšem naboru govornih posnetkov delno vodi k "posplošenju" akustičnih modelov, kar omogoča doseganje boljših rezultatov pri prenosu na nov jezik. Učinek posplošenja akustičnih modelov je bil opažen že v (Imperl, et al., 2002). Stopnja posplošenja akustičnih modelov je delno odvisna od števila in sorodnosti jezikov, vključenih v večjezične akustične modele. Postavimo lahko hipotezo, da bi z vključitvijo dovolj velikega števila jezikov v večjezične akustične modele dobili "univerzalen" nabor akustičnih modelov, ki bi enako dobro pokrival nove jezike iz različnih jezikovnih skupin.

## 7. Zaključek

V predstavljeni raziskavi smo preučili možnost prenosa obstoječih večjezičnih akustičnih modelov na italijanski jezik, ne da bi uporabili učenje ali adaptacijo na italijanskem govornem materialu. Doseženi rezultati razpoznavanja italijanskega govora so se približali referenčnim vrednostim, dodatno pa bi jih bilo možno izboljšati z adaptacijo na majhnem naboru italijanskega govornega materiala. V raziskavi smo tudi ugotovili, da jezikovne skupine, v katere spadajo uporabljeni jeziki, pomembno vplivajo na rezultate razpoznavanja govora s prenesenimi akustičnimi modeli.

V prihodnje bi bilo potrebno raziskati odvisnost posplošenja akustičnih modelov od števila v modele vključenih jezikov. Da bi se lahko izognili subjektivnemu vplivu, ki je slabost metode prenosa s pomočjo IPA tabel, bi bilo potrebno za prenos obstoječih akustičnih modelov na nov jezik uporabiti kakšno podatkovno vodeno metodo prenosa.

## ZAHVALA

Predstavljeni raziskava je bila opravljena v okviru bilateralnega sodelovanja z Republiko Nemčijo. Za pomoč

med obiskom v Centralnih raziskovalnih laboratorijih podjetja Siemens AG se zahvaljujemo dr. Josefu G. Bauerju.

## 8. Literatura

- F. Bezlaj, 1960. Osnove fonetike. Ljubljana, Slovenija.
- A. Haunstein, E. Marschall. 1995. Methods for improved speech recognition over the telephone lines. V: *Proc. ICASSP '95*.
- H. van den Heuvel, L. Boves, A. Moreno, M. Omologo, G. Richard, E. Sanders. 2001. Annotation in the SpeechDat Projects. *International Journal of Speech Technology*, 4(2):127 – 143.
- H. Höge, H. Tropsch, R. Winski, H. van den Heuvel, R. Haeb-Umbach. 1997. European speech databases for telephone applications. V: *Proc. ICASSP '97*, strani 1771 – 1774, Munich.
- B. Imperl, Z. Kačič, B. Horvat, A. Žgank, 2002. Clustering of Triphones Using Phoneme Similarity Estimation for the Definition of a Multilingual Set of Triphones. *Speech Communications*, v tisku.
- IPA, 2002. <http://www2.arts.gla.ac.uk/IPA/ipa.html>
- F. T. Johansen, N. Warakagoda, B. Lindberg, G. Lehtinen, Z. Kačič, A. Žgank, K. Elenius, G. Salvi. 2000. The COST 249 SpeechDat Multilingual Reference Recogniser. V: *Proc. LREC'2000*, Athens.
- D. Johnston, Sorin, C., Gagnoulet, C., Charpentier, F., Canavesio, F., Lochschmidt, B., Alvarez, J., Cortazar, I., Tapias, D., Crespo, C., Azevedo, J., Chaves, R., 1997. Current and experimental applications of speech technology for telecom services in Europe. *Speech Communication*, Volume 23, Issues 1-2, oktober 1997, strani 5-16.
- J. Kaiser, Z. Kačič. 1998. Development of the Slovenian SpeechDat database. V: *Speech Database Development for Central and Eastern European Languages*, Granada.
- P. Ladefoged, 1993. *A Course in Phonetics*. Harcourt Brace College Publishers.
- B. Lindberg, F. T. Johansen, N. Warakagoda, G. Lehtinen, Z. Kačič, A. Žgank, K. Elenius, G. Salvi. 2000. A noise robust multilingual reference recogniser based on SpeechDat(II). V: *Proc. ICSLP 2000*, Beijing, Kitajska.
- Refrec home, 2002. <http://www.telenor.no/fou/prosjekter/-taletek/refrec/>
- T. Schultz, 2000, *Multilinguale Spracherkennung - Kombination akustischer Modelle zur Portierung auf neue Sprachen*. Doktorska disertacija.
- S. Young, Odell, J., Woodland, P., 1994. Tree-based State Tying for High Accuracy Acoustic Modelling. V: *Proc. ARPA Human Language Technology Conference*, Plainsboro, ZDA.
- S. Young, 1996. Large vocabulary continuous speech recognition: A review. *IEEE Trans. on Signal and Audio Processing*, SAP-4(1):31 – 44.
- A. Žgank, B. Imperl, F.T. Johansen, Z. Kačič, B. Horvat, 2001. Crosslingual Speech Recognition with Multilingual Acoustic Models Based on Agglomerative and Tree-Based Triphone Clustering. V: *Proc. Eurospeech 2001*, Aalborg, Danska
- A. Žgank, B. Imperl, F.T. Johansen, Z. Kačič, B. Horvat,

2001. Crosslingual Adaptation of Multilingual Triphone  
Acoustic Models. V: Proc. MSLP 2001, Aalborg, Dansk