Robustna parametrizacija govora z uporabo postopkov za zmanjšanje vplivov aditivnega in konvolucijskega šuma pri avtomatskem razpoznavanju govora

Bojan Kotnik, Damjan Vlaj, Zdravko Kačič

Univerza v Mariboru
Fakulteta za elektrotehniko, računalništvo in informatiko
Smetanova ulica 17, 2000 Maribor
bojan.kotnik@uni-mb.si

Povzetek

V članku predstavljamo izboljšavo algoritma za robustno izločanje mel frekvenčnih kepstralnih značilk (MFCC) govornega signala z uporabo postopkov za zmanjšanje nivojev aditivnega in konvolucijskega šuma pri avtomatskem razpoznavanju govora. V fazi oknjenja govornega signala uporabimo hibridno Hamming-kosinusno okno. Sledita prehod v frekvenčni prostor s pomočjo hitre Fourierjeve transformacije ter postopek zmanjšanja nivoja aditivnega šuma s spektralnim odštevanjem, pri katerem za oceno spektra šuma uporabimo postopek glajenja spektra. V naslednji fazi izvedemo mel frekvenčno analizo signala, nelinearno transformacijo izhodov filtrov ter zaznavanje govornih okvirjev. Okvirjev, zaznanih kot šum oziroma tišino, ne izločimo iz nadaljnjega procesa, temveč jim le zmanjšamo magnitude izhodov filtrov. Nivo konvolucijskega šuma, ki je posledica karakteristike prenosnega kanala, zmanjšamo s pomočjo RASTA filtriranja časovnih trajektorij izhodov filtrskih bank. V zadnjem koraku za vsako okno generiramo vektor značilk govornega signala, sestavljen iz 12 MFCC elementov ter logaritma energije. Vrednotenje postopka za robustno parametrizacijo govora izvedemo s pomočjo orodja HTK in baz Auorora 2 in 3. Relativno izboljšanje uspešnosti razpoznavanja govora s predstavljenim algoritmom glede na referenčni postopek znaša 41.14 % za bazo Aurora 2 ter 45.06 % za baze Aurora 3.

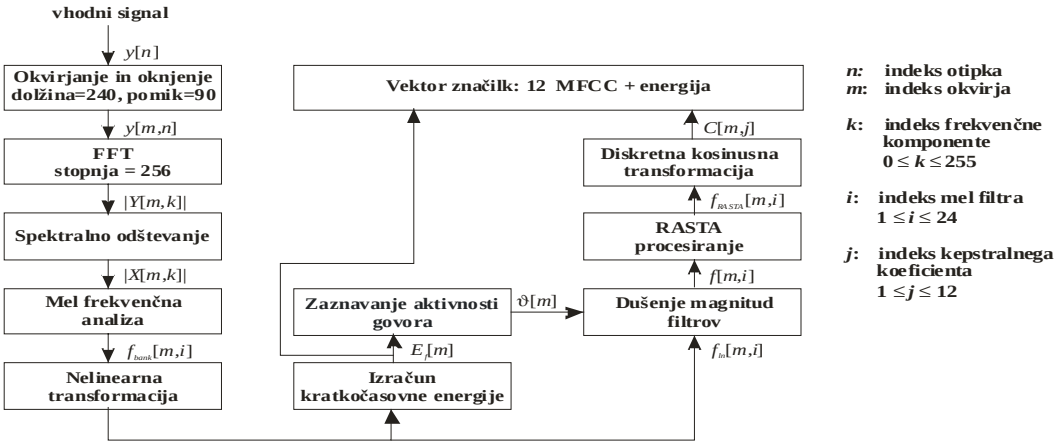
1. Uvod

Postopek izločanja značilk je zelo pomemben v procesu avtomatskega razpoznavanja govora in ima velik vpliv tudi na samo učinkovitost sistema, saj so vsi nadaljnji koraki v postopku avtomatskega razpoznavanja govora odvisni od načina ter kakovosti parametrizacije govornega signala. Parametrizacija vhodnega govornega signala se lahko spreminja zaradi več dejavnikov. Če na primer isti govorec izgovori isto besedo v različnih šumnih okoljih (na primer v restavraciji, na letališču, v avtomobilu med vožnjo…) pri uporabi kanalov različnih prenosnih karakteristik (na primer fiksna telefonija, ISDN, GSM…), potem ima lahko rezultirajoč šumni signal različne značilnosti (Junqua in Haton, 1996). Aditivni ali konvolucijski šumi lahko močno zmanjšajo uspešnost razpoznavanja govora, kar je najpogosteje posledica neujemanja akustičnih modelov, ki so bili učeni v drugačnem okolju od tistega, v katerem ob uporabi sistema zajemamo govorni signal, ter nezmožnosti akustičnih modelov, da natančno opišejo s šumom okvarjen govorni signal (Hermansky in Morgan, 1994). Danes najpogostejše metode za parametrizacijo govora temeljijo na mel frekvenčnih kepstralnih koeficientih (MFCC), ki izkoriščajo tudi lastnosti človekovega slušnega ter zaznavnega sistema. Z namenom doseganja večje robustnosti značilk govornega signala glede na različna šumna okolja ter karakteristike prenosnih kanalov predstavljamo izboljšan algoritem za izločanje MFCC vektorjev značilk.

Navadno uporabimo v fazi oknjenja govornega signala kot privzeto simetrično Hammingovo okno, vendar so eksperimentalni rezultati pokazali, da lahko izboljšanje rezultatov razpoznavanja dosežemo z uporabo predlaganega nesimetričnega hibridnega Hamming-kosinusnega okna. Posebno pozornost smo v nadaljevanju posvetili zmanjšanju nivojev aditivnega in konvolucijskega šuma. Zmanjšanje nivoja aditivnega šuma dosežemo z uporabo spektralnega odštevanja, pri katerem uporabimo metodo spektralnega glajenja kot način za ocenitev spektra šuma. Sledi standardni postopek mel frekvenčnega filtriranja s pomočjo prekrivajočih se filtrov trikotne oblike. Izvedemo tudi nelinearno transformacijo magnitud izhodov filtrov. Zaznavanje aktivnosti govora, ki temelji na magnitudah izhodov mel filtrov, je podrobneje opisano v poglavju 2.4. Magnitude izhodov mel filtrov tistih okvirjev, ki so zaznani kot šumni, dušimo. Konvolucijski šum, ki je posledica prenosne karakteristike kanala, lahko učinkovito odstranimo s pomočjo RASTA filtriranja časovnih trajektorij magnitud izhodov filtrov. Proces filtriranja je podrobneje predstavljen v poglavju 2.6. V zadnjem koraku generiramo za vsako okno vhodnega signala vektor značilk, sestavljen iz 12 MFCC koeficientov ter logaritma energije okvirja. Eksperimente, ki bodo podrobneje opisani v poglavju 3, smo izvedli s programskim orodjem HTK in bazami Aurora 2 in Aurora 3. Diskusijo rezultatov bomo podali v poglavju 4, zaključke pa bomo strnili v poglavju 5.

2. Postopek izločanja značilk

V idealnem primeru bi lahko problem robustnosti sistema avtomatskega razpoznavanja govora rešili v modulu za izločanje značilk govornega signala in bi s tem odpravili potrebo po dodatnih, aplikacijsko odvisnih akustičnih učnih podatkih. Slika 1 prikazuje blokovno shemo predlaganega postopka za parametrizacijo govora, ki zajema fazo predprocesiranja s hibridnim Hamming-kosinusnim oknom, stopnjo za zmanjšanje nivoja aditivnega šuma s spektralnim odštevanjem, mel frekvenčno filtriranje, nelinearno transformacijo, zaznavanje aktivnosti govora na osnovi ocene razmerja signal-šum, stopnjo za zmanjšanje magnitud izhodov mel filtrov tistih okvirjev, ki so zaznani kot šumni, RASTA filtriranje ter postopek za generiranje MFCC koeficientov na osnovi diskretne kosinusne transformacije.



Slika 1: Shema predlaganega postopka za robustno parametrizacijo govornega signala.

2.1. Oknjenje in izračun FFT

Vhodni šumni govorni signal $y[n]$ razstavimo v prekrivajoče se okvirje dolžine 30 ms (dolžina L = 240 otipkov pri frekvenci vzorčenja 8 kHz). Interval pomikanja okvirjev znaša 11.25 ms (90 otipkov). Vsak okvir pomnožimo s funkcijo okna. Pri izbiri primernega okna smo pozorni predvsem na tip in dolžino okna. Tipične dolžine oken so v območju 20 do 30 ms. Množenje okvirja s funkcijo okna v časovnem prostoru predstavlja v frekvenčnem prostoru konvolucijo med spektrom signala, zajetega v okvirju, in spektrom okna. V idealnem primeru naj ima spekter okna čimbolj ozkopasovni osrednji vrh, stranski vrhovi pa naj imajo čim nižjo magnitudo. Če imajo na primer stranski vrhovi višjo magnitudo, prihaja v širšem frekvenčnem področju do povprečenja vplivov zajetih frekvenčnih komponent, kar vodi v popačenje frekvenčne slike analiziranega signala. V predstavljenem postopku za izločanje značilk uporabimo hibridno Hamming-kosinusno okno (ITU-T, 1996).

V naslednjem koraku izvedemo postopek hitre Fourierjeve transformacije (FFT) stopnje N = 256, pri čemer vhodno zaporedje $y[m,n]$ dopolnimo z ustreznim številom otipkov z vrednostjo nič.

2.2. Spektralno odštevanje na osnovi glajenja spektra

S pomočjo spektralnega odštevanja zmanjšamo nivo aditivnega šuma v vhodnem govornem signalu. Postopek je podrobneje opisan v (Kotnik et al., 2002). Glavna prednost omenjenega postopka pred osnovnim spektralnim odštevanjem (Boll, 1979) je v tem, da za določitev spektra šuma ne potrebuje eksplicitnega zaznavanja aktivnosti govora, saj šum ocenjujemo na statističen način, in ne le v intervalih, kjer govor ni prisoten. Dobra lastnost spektralnega odštevanja na osnovi glajenja spektra je tudi manjši nivo dodanega glasbenega šuma v izhodnem signalu $|X[m,k]|$, kar je sicer glavna slabost osnovnega spektralnega odštevanja (Boll, 1979).

2.3. Mel frekvenčna analiza in nelinearna transformacija

Podrobnejši opis algoritma je podan v (Kotnik et al., 2001). Uporabili smo *StMel* = 24 polovično prekrivajočih se filtrov trikotne oblike, katerih centralne frekvence so po mel frekvenčni skali ekvidistančno razporejene. Izhod $f_{bank}[m,i]$ vsakega izmed filtrov smo logaritmirali:

$$f_{\ln}[m,i] = \ln\left(1 + f_{bank}[m,i]\right). \quad (1)$$

2.4. Zaznavanje aktivnosti govora

Modul za zaznavanje aktivnosti govora opravi za vsak okvir vhodnega govornega signala odločitev, ali okvir m vsebuje govor ($\vartheta[m] = 1$) ali tišino oziroma samo šum ($\vartheta[m] = 0$), s pomočjo razmerja med signalom in šumom, ki ga v trenutnem okvirju primerja s pragovno vrednostjo. Razmerje med signalom in šumom trenutnega okvirja m ustreza razliki med ocenama kratkočasovne in počasi spreminjajoče se spektralne energije. Oceno počasi spreminjajoče se spektralne energije posodobimo vsakič, ko modul za zaznavanje aktivnosti govora odloči, da okvir ne vsebuje govora. Oceno kratkočasovne spektralne energije $E_f[m]$ izračunamo za vsak okvir po enačbi:

$$E_f[m] = \sum_{i=1}^{StMel} f_{\ln}[m,i], \quad (2)$$

Oceno kratkočasovne spektralne energije $E_f[m]$ uporabimo za posodabljanje ocene počasi spreminjajoče se spektralne energije $E_m[m]$ s pomočjo enačbe:

Če $(E_f[m] - E_m[m-1]) < a$ **Potem**

$$E_m[m] = E_m[m-1] + \frac{E_f[m] - E_m[m-1]}{b} \quad (3)$$

Sicer

$$E_m[m] = E_m[m-1]$$

Vrednosti praga a = 10 in faktorja b = 100 smo določili na osnovi analiz baz Aurora 2, 3 in slovenske baze SpeechDat 2. Po ocenitvi kratkočasovne $E_f[m]$ in počasi spreminjajoče se $E_m[m]$ spektralne energije lahko modul za zaznavanje aktivnosti govora odloči, ali okvir vsebuje govor, pomešan s šumom, ali samo šum.

2.5. Dušenje magnitud filtrov

Odločitev $\vartheta[m]$, dobljena v modulu za zaznavanje aktivnosti govora, je uporabljena v modulu za dušenje magnitud okvirjev. Logaritmi magnitud mel-frekvenčnih kanalov $f_{\ln}[m,i]$ okvirja m so duženi, če je okvir m označen kot šumen. Kriterij za dušenje okvirjev predstavimo z naslednjo enačbo:

$$f[m,i] = \begin{cases} f_{\ln}[m,i], & \vartheta[m] = 1 \\ \psi \cdot f_{\ln}[m,i], & \vartheta[m] = 0 \end{cases}, \quad (4)$$

kjer je faktor dušenja ψ = 0.03 in je določen izkustveno. Predstavljen princip dušenja okvirjev se razlikuje od principa odstranjevanja okvirjev, kjer so okvirji v celoti odstranjeni iz nadaljnjega postopka procesiranja vektorjev značilk (Andrassy et al., 2001).

2.6. Normalizacija kanala s pomočjo RASTA procesiranja

Stacionarna konvolucijska popačenja, ki nastanejo kot posledica prenosne karakteristike kanala, se v domeni logaritma energije spektra kažejo kot aditiven šum. Če predpostavimo stacionarnost prenosnega kanala ali karakteristike mikrofona, lahko pokažemo, da vsako konvolucijsko popačenje spremeni srednjo vrednost časovne trajektorije logaritma energije spektra. Namesto odštevanja srednje vrednosti je v (Hermansky in Morgan, 1994) vpeljana metoda za odpravljanje popačenja kanala, znana kot RASTA (RelAtive SpecTrAl) procesiranje. V članku predlagamo časovno RASTA procesiranje logaritma magnitud izhodov mel filtrov.

2.7. Postopek izgradnje vektorja značilk MFCC

Zaradi uporabe diagonalne kovariančne matrike tako v procesu učenja akustičnih modelov kot tudi pri razpoznavanju je potrebno komponente signala $f_{RASTA}[m,i]$, i = 1, 2, ..., *StMel* med sabo dekorelirati. Ta postopek izvršimo z diskretno kosinusno transformacijo (Kotnik et al., 2001). Kot rezultat dobimo 12 mel frekvenčnih kepstralnih koeficientov $C[m,j]$, kjer je j = 1, 2, ..., 12. V procesu zaznavanja okvirjev smo predhodno izračunali še energijo okvirja $E_f[m]$, ki jo dodamo kot trinajsti element v vektorju statičnih značilk. Dinamične značilke, sestavljene iz prvih in drugih odvodov statičnih značilk, računamo interno v procesu učenja akustičnih modelov oziroma pri razpoznavanju.

3. Eksperimenti in rezultati

Za vrednotenje učinkovitosti predstavljenega algoritma smo uporabili govorne baze Aurora 2 in Aurora 3 ter orodje HTK za učenje in razpoznavanje s prikritimi modeli Markova.

3.1. Baza Aurora 2

Govorna podatkovna baza Aurora 2 je izpeljanka studijske baze TI-Digits, ki vsebuje izgovarjave različno dolgih nizov povezanih števk (Hirsch in Pearce, 2000). V bazi Aurora 2 je studijskim posnetkom iz TI-Digits kontrolirano dodanih osem vrst šumov iz različnih okolij pri razmerjih signal-šum (SNR) 20 dB, 15 dB, 10 dB, 5 dB, 0 dB in -5dB. Definirana sta dva načina učenja akustičnih modelov. Pri *učenju s čistim govorom* (UČG) so uporabljeni le posnetki, ki jim predhodno ni bil dodan šum, medtem ko pri *učenju s šumnim govorom* (UŠG) akustične modele učimo s signalom z dodanimi štirimi vrstami šumov pri različnih SNR. Baza Aurora 2 ima definirane tri testne načine. Pri *Testu A* so za tvorjenje posnetkov uporabljeni enaki šumi kot v primeru učenja s šumnim signalom, kar pomeni, da je v tem primeru ujemanje učnega in testnega okolja akustičnih modelov dobro. Pri *Testu B* so uporabljeni štirje drugačni šumi kot v primeru učenja s šumnim signalom, zato ta test najbolje simulira uporabo sistema v realnem šumnem okolju, saj je v tem primeru ujemanje učnega in testnega okolja akustičnih modelov slabo. Govorni material v *Testu C* je dodatno filtriran tako, da čimbolj natančno ponazarja popačitve govornega signala, ki ga prenašamo preko prenosnega kanala (na primer preko GSM omrežja).

3.2. Baze Aurora 3

Baze Aurora 3 sestavlja govorni material finske, španske, nemške in danske baze SpeechDat-Car (AU/225/00 et al., 2000). Izgovarjave nizov števk so posnete v avtomobilu med vožnjo z dvema mikrofonoma, pri čemer je en mikrofon prostoročni, drugi pa je nameščen blizu govorčevih ust. Baza ima glede na ujemanje učnih in testnih okolij definirane tri učno-testne načine: *dobro ujemanje* (DU), *srednje neujemanje* (SN), *veliko neujemanje* (VN). Pri načinu DU je 70 % izgovarjav celotne baze uporabljenih za učenje, preostalih 30 % pa za testiranje. V tem primeru učni nabor pokriva vso spremenljivost testnega okolja. Pri načinu SN so tako za učenje kot tudi za razpoznavanje uporabljene le izgovarjave, posnete s prostoročnim mikrofonom, medtem ko so v načinu VN za učenje uporabljeni posnetki z bližnjim mikrofonom, razpoznavanje pa poteka s posnetki prostoročnega mikrofona.

3.3. Konfiguracija razpoznavalnika govora

Učinkovitost predstavljenega algoritma primerjamo z rezultatom avtomatskega razpoznavanja govora s standardiziranim algoritmom za izločanje značilk, predstavljenim v (ETSI ES 201 108, 2000). Konfiguracija postopka učenja akustičnih modelov ter razpoznavanja je enaka kot v (Hirsch in Pearce, 2000). Na ta način dosežemo primerljivost tudi z rezultati drugih raziskovalcev na področju robustne parametrizacije govora. Uporabili smo torej orodje za delo s prikritimi modeli Markova HTK (Young, 2001), šestnajststanjske celobesedne akustične modele s kombinacijo treh Gaussovih funkcij gostote verjetnosti ter z diagonalno kovariančno matriko. Tabela 1 prikazuje povzetek rezultatov, doseženih s predstavljenim algoritmom za robustno parametrizacijo govora.

4. Diskusija rezultatov

Tabela 1 vsebuje absolutne rezultate razpoznavanja govora, dosežene s predstavljenim algoritmom v obliki deleža napačno razpoznanih besed (WER – "word error rate"), ter relativno izboljšanje rezultatov razpoznavanja glede na referenčni postopek za izločanje MFCC značilk govornega signala. Razvidno je, da dobimo tako pri bazi Aurora 2 kot pri bazi Aurora 3 konstantno izboljšanje rezultatov razpoznavnja, kljub temu da so šumne izgovarjave pri bazi Aurora 2 umetno sprocesirane, medtem ko je baza Aurora 3 posneta v realnem šumnem okolju. Predstavljeni postopek se izkaže še posebej učinkovit v primeru neujemanja učnega in testnega okolja, kar je razvidno iz visokega relativnega izboljšanja 62.02% v načinu UČG pri bazi Aurora 2 in 61.03 % v načinu VN pri bazi Aurora 3. Eksperimenti z bazo Aurora 3 so tudi pokazali, da je uporaba postopka zaznavanja aktivnosti govora ter dušenja šumnih okvirjev zelo pomemben člen v

Tabela 1: Primerjava doseženih rezultatov razpoznavanja s predlaganim algoritmom glede na referenčni postopek za izločanje MFCC značilk na bazah Aurora 2 in Aurora 3.

Rezultati na bazi Aurora 2

Referenca - delež napačno razpoznanih besed				
	Test A	**Test B**	**Test C**	**Celotno**
UŠG	11.93%	12.78%	15.44%	**12.97%**
UČG	41.26%	46.60%	34.00%	**41.94%**
Povprečje	**26.59%**	**29.69%**	**24.72%**	**27.46%**

Dosežen delež napačno razpoznanih besed				
	Test A	**Test B**	**Test C**	**Celotno**
UŠG	9.25%	9.84%	9.46%	**9.53%**
UČG	17.12%	16.59%	17.61%	**17.00%**
Povprečje	**13.18%**	**13.21%**	**13.54%**	**13.26%**

Relativno izboljšanje				
	Test A	**Test B**	**Test C**	**Celotno**
UŠG	14.36%	21.49%	29.56%	**20.25%**
UČG	59.37%	68.19%	55.01%	**62.02%**
Povprečje	**36.87%**	**44.84%**	**42.28%**	**41.14%**

Rezultati na bazah Aurora 3

Referenca - delež napačno razpoznanih besed					
Baza	**Finska**	**Španska**	**Nemška**	**Danska**	**Povprečje**
DU (x40%)	7.26%	7.06%	8.80%	12.72%	**8.96%**
SN (x35%)	19.49%	16.69%	18.96%	32.68%	**21.96%**
VN (x25%)	59.47%	48.45%	26.83%	60.63%	**48.85%**
Celotno	**24.59%**	**20.78%**	**16.86%**	**31.68%**	**23.48%**

Dosežen delež napačno razpoznanih besed					
Baza	**Finska**	**Španska**	**Nemška**	**Danska**	**Povprečje**
DU (x40%)	3.69%	3.65%	6.99%	7.24%	**5.39%**
SN (x35%)	10.94%	7.60%	14.06%	22.12%	**13.68%**
VN (x25%)	18.90%	11.85%	13.09%	30.83%	**18.67%**
Celotno	**10.03%**	**7.08%**	**10.99%**	**18.35%**	**11.61%**

Relativno izboljšanje					
Baza	**Finska**	**Španska**	**Nemška**	**Danska**	**Povprečje**
DU (x40%)	49.17%	48.30%	20.57%	43.08%	**40.28%**
SN (x35%)	43.87%	54.46%	25.84%	32.31%	**39.12%**
VN (x25%)	68.22%	75.54%	51.21%	49.15%	**61.03%**
Celotno	**52.08%**	**57.27%**	**30.08%**	**40.83%**	**45.06%**

algoritmu za robustno parametrizacijo govora, saj dobimo v nasprotnem primeru znatno večji delež napak zaradi vrinjanja besed. Uspešnost uporabe RASTA procesiranja je razvidna iz relativno visokega izboljšanja rezultatov razpoznavanja (29.56 % relativno glede na onovni MFCC postopek) pri načinu UŠG in *Testu C* baze Aurora 2. RASTA uspešno odpravi vpliv prenosne funkcije kanala na govorni signal.

5. Zaključek

V članku smo predstavili učinkovit postopek za robustno izločanje mel frekvenčnih kepstralnih značilk govornega signala v najrazličnejših šumnih okoljih. Kot alternativo splošno uporabnemu Hammingovemu oknu smo v fazo predprocesiranja govornega signala vpeljali hibridno Hamming-kosinusno okno. V naslednjem koraku smo s pomočjo metode spektralnega odštevanja, pri kateri smo spekter šuma ocenjevali statistično, zmanjšali nivo aditivnega šuma. Sledila sta postopka mel frekvenčne analize vhodnega signala ter nelinearna transformacija magnitud izhodov filtrov. Nato smo vpeljali postopek zaznavanja aktivnosti govora z računanjem razmerja signal-šum na osnovi logaritma magnitud izhodov filtrov ter izvedli dušenje magnitud v okvirjih, zaznanih kot šum ali tišina. Vpliv konvolucijskih popačenj smo zmanjšali z RASTA procesiranjem. V zadnjem koraku smo generirali vektor značilk s 13 elementi - 12 mel frekvenčnih kepstralnih koeficientov ter logaritem energije okvirja. Vrednotenje učinkovitosti postopka za izločanje smo izvedli z bazami Aurora 2 in Aurora 3 ob uporabi orodja HTK. Skupno reltivno izboljšanje rezultatov razpoznavanja opisanega algoritma glede na referenčni MFCC postopek znaša 41.14 % pri bazi Aurora 2 in 45.06 % pri bazi Aurora 3.

6. Literatura

Andrassy, B., Vlaj, D., in Beaugeant, C., 2001. Recognition Performance of the Siemens Front-end with and without Frame Dropping on the Aurora 2 Database. *Eurospeech 2001 Proceedings*, pp. 193 – 196, Aalborg, Danska.

AU/225/00, AU/271/00, AU/273/00, AU/378/00, 2000. Finnish, Spanish, German, Danish Databases for ETSI STQ Aurora WI008 Advanced DSR Front-End Evaluation: Description and Baseline Results.

Boll, S.F., 1979.Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Trans. Speech and Audio Proc.*, ASSP-27,2, 113-120.

ETSI ES 201 108, 2000. ETSI standard document 2000. Speech Processing, Transmission and Quality aspects (STQ), Distributed Speech Recognition, Front-end Feature Extraction Algorithm, Compression Algorithm. *ETSI ES 201 108 v1.1.1* (2000-02).

Hermansky, H., Morgan, N., 1994. RASTA Processing of Speech. *IEEE Trans. Speech and Audio Proc.*, Vol. 2, No. 4.

Hirsh, H. G., in Pearce, D., 2000. The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Conditions. *ISCA ITRW ASR2000,* Pariz, Francija.

ITU-T, 1996. Coding of Speech at 8kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP). *ITU-T Recommendation G.729*.

Junqua, J. C. in Haton, J. P., 1996. Robustness in Automatic Speech Recognition. *Kluwer Academic Publishers*, Norwell, Massachusetts, ZDA.

Kotnik, B., Kačič, Z., in Horvat, B., 2001. A Multiconditional Robust Front-End Feature Extraction with a Noise Reduction Procedure Based on Improved Spectral Subtraction Algorithm. *Eurospeech 2001 Proceedings*, pp. 197 – 200, Aalborg, Danska.

Kotnik, B., Kačič, Z., in Horvat, B., 2002. Predstavitev učinkovitega postopka za robustno avtomatsko razpoznavanje govora. *Elektrotehniški vestnik*, št.1, letn. 69, 69-74.

Young, S., 2001. HTK Book - Version 3.1. *Cambridge University Engineering Department*, Cambridge, Velika Britanija.