Razpoznava govora z usmerjeno nevronsko mrežo

Aleš Tetičkovič, Simon Klančnik
UM FERI Maribor, Smetanova 17, 2000 Maribor
ales.tetickovic@uni-mb.si, simon.klancnik@uni-mb.si

Abstract: ***The article describes voice recognition system, which will be elementary used in wheelchair. Wheelchair is designed for physically disabled person, who can not control the wheelchair with the joystick and voice is the only possible way to control machine. Speech recognition is composed from word isolation, LPC cepstral analysis, coefficient dimension reduction with SOM neural network and neural network for word recognition.***

I. Uvod

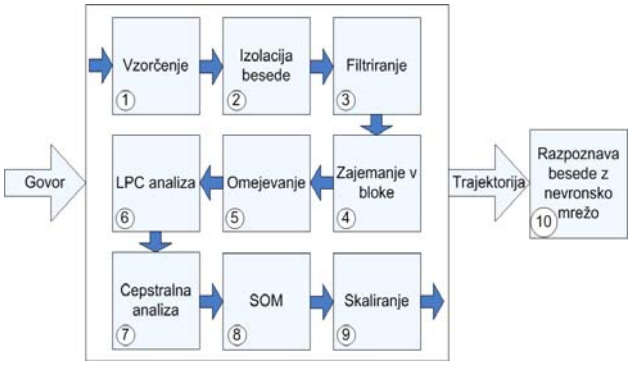
V vsakdanjem življenju je verbalna komunikacija ena najpomembnejših načinov komuniciranja med ljudmi. Razvoj tehnike je omogočil človeku, da je ustvaril razne naprave, ki mu služijo kot pomoč. Tako kot med ljudmi, je tudi med strojem in človekom potrebna komunikacija, ki pa postane najbolj učinkovita, če je mogoča v človeku najbolj vsakdanji obliki in sicer verbalni komunikaciji. Posebej pomembno rešitev predstavlja za ljudi, z raznimi telesnimi hibami, ki drugače niso sposobni komunicirati z napravami, a vendar so le te za njih velikega pomena.

Sistem za razpoznavo govora je bil razvit za uporabo na invalidskem vozičku, ki ga je mogoče krmiliti z govorom in je namenjen predvsem tetraplegikom, osebam hromim od vratu navzdol, ter osebam, ki ne kontrolirajo svojih gibov (spastiki, bolniki s cerebralno paralizo…). Sistem pa se lahko uporabi tudi na drugih lokacijah, kot so dvigala, vrata, na delovnih strojih za vnos podatkov in še na mnogih drugih področjih.

Na področju razpoznave govora so se v preteklosti oblikovale številne metode, tako glede obdelave signala (izolacija besed, analiza signala,..), kot tudi glede same prepoznave besed(skriti Markov model…), vendar nobena metoda ni popolna in ima vsaka določene slabosti. Za razpoznavo govora smo izbrali nevronske mreže, to pa predvsem zato, ker so dobro odporne na šume iz ozadja, ki so prisotni pri govornih ukazih in niso občutljive na spremembo barve in tona uporabnikovega glasu. Postavlja se tudi vprašanje glede same izbire metode za izolacijo besed, ki mora biti sposobna izolirati posamezne besede iz signala v katerem so prisotni različni šumi. Zato smo se osredotočili na dve metodi in sicer na metodo izračuna tekočega povprečja in t.i. »zero crossing« metodo in naredili kratko analizo njune učinkovitosti.

II. Razpoznava govora

Slika 2.1 prikazuje blokovno shemo, ki predstavlja princip delovanja razpoznave govora:



Slika 2.1: Blokovna shema razpoznave govora

Opis posameznih blokov:

1. vzorčenje govornega signala s frekvenco 8 kHz in 16 bitno ločljivostjo,
2. izolacija besede ali določitev začetka in konca izgovorjene besede,

3. filtriranje signala z visokoprepustnim filtrom,
4. zajemanje izolirane besede v bloke določene velikosti,
5. posameznemu bloku se odpravijo nezveznosti s Hamming-ovim oknom,
6. izračun 12 LPC koeficientov za vsak posamezni blok,
7. izračun 12 kepstralnih koeficientov iz LPC koeficientov za vsak posamezni blok,
8. redukcija 12 kepstralnih koeficientov na dva koeficienta s SOM nevronsko mrežo,
9. skaliranje koeficientov dobljenih iz SOM nevronske mreže na enotno dolžino,
10. razpoznava dobljenih skaliranih koeficienrov ali trajektorije z usmerjeno nevronsko mrežo.

Signal dobljen iz mikrofona se vzorči s frekvenco 8 kHz in ločljivostjo A/D pretvorbe 16 bitov, saj sama DSP kartica vsebuje AUDIO CODEC AD535, ki ima vzorčevalno frekvenco fiksno določeno s frekvenco oscilatorja, ki je ni mogoče programsko spreminjati.

Iz zajetega signala je potrebno izolirati izgovorjeno besedo ali določiti začetek in konec besede. Pri tem smo se osredotočili predvsem na dve metodi izoliranja besede:
1. izolacija besede po metodi izračuna tekočega povprečja,
2. izolacija besede po metodi »zero crossing« algoritma.

Za zmanjšanje vpliva šumov nizkih frekvenc, bomo signal izolirane besede filtrirali z digitalnim visokoprepustnim FIR filtrom. Za to smo se odločili zato, ker sam govor ne vsebuje zelo nizkih frekvenc. Nizke frekvence nam samo vnašajo motnje v sam razpoznavalnik besede.

Celotno razpoznano besedo razdelimo v bloke dolžine 330 vzorcev ali 41 ms. Posamezne bloke pa med sabo zamikamo za 110 vzorcev ali 13 ms, tako da se bloki med sabo prekrivajo. S tem pri izračunu LPC in kepstralnih koeficientov vsak posamezni odtipek upoštevamo trikrat.

Vsak posamezni blok pa omejimo s Hamming-ovim oknom, ki nam odpravi nezveznosti v posameznem bloku.

Za vsak posamezni blok se izračuna vektor z dvanajstimi LPC kepstralnimi koeficienti (blok 6 in 7) po Durbinovi metodi [1] in rekurzivnih izrazih, ki jih je razvil Furui [2]. Celoten postopek bo podrobneje opisan kasneje.

Ker pa je količina podatkov še vedno preobsežna za razpoznavalnik z nevronsko mrežo, smo reducirali dobljene LPC kepstralne koeficiente iz vektorja 12 členov v vektor z 2 členoma s samoorganizirajočo nevronsko mrežo (SOM) [3].

Število vhodov v usmerjeno nevronsko mrežo je fiksno določeno, zato je potrebno trajektorijo, ki jo dobimo iz SOM nevronske mreže skalirati na fiksno dolžino. Dolžina trajektorije, ki jo dobimo iz SOM-a je odvisna od dolžine izolirane besede, zato jo skaliramo na dolžino 100×2 člena.

Skalirana trajektorija je sedaj pripravljena za razpoznavo z usmerjeno nevronsko mrežo. Trajektorijo 2×100 sedaj razporedimo v en vektor dolžine 200 elementov, ki predstavljajo vhode v nevronsko mrežo. Glede na trenutni vhodni vektor se izračunajo izhodi iz posameznih nivojev nevronov. Rezultat razpoznave predstavljajo izhodi iz nevronov izhodnega nivoja.

III. Izolacija besede

Za samo razpoznavo govora, je potrebno iz signala izločiti izgovorjeno besedo. Potrebno je določiti začetek in konec besede iz signala. Za izolacijo besed obstaja več različnih metod. Mi se bomo osredotočili predvsem na dve metodi izoliranja besede. Obe metodi bomo med sabo primerjali s posnetim signalom, ki bo vseboval izgovor različnih besed.

Prva metoda, ki jo bomo opisali je metoda izračuna tekočega povprečja signala in primerjavo rezultata izračuna s pragovno

vrednostjo, ki jo bomo eksperimentalno določili. Za signal, ki ga dobimo iz mikrofona računamo vsoto zadnjih 1000 vzorcev in vsoto delimo z številom vzorcev v našem primeru 1000. Dobljen rezultat nato primerjamo s pragovno vrednostjo in če je tekoče povprečje večje od pragovne vrednosti, je bil zaznan začetek besede. Ker pa je uporabljeno ogromno vrednosti vzorcev za izračun tekočega povprečja, dodamo k začetku besede še predhodnih 500 vzorcev, ki so bili uporabljeni za izračun tekočega povprečja v trenutku, ko je bil zaznan začetek besede. Ko pa tekoče povprečje pade pod pragovno vrednost počakamo še naslednjih 2000 vzorcev in spet primerjamo vrednost tekočega povprečja s pragovno vrednostjo. In če je sedaj tekoče povprečje manjše od pragovne vrednosti, je bil zaznan konec besede. Zato od izolirane besede odštejemo teh 2000 vzorcev, ter dodatnih 500 vzorcev iz istega razloga, kot pri zaznavi začetka besede. To zakasnitev 2000 vzorcev smo uporabili zaradi presledka med zlogi, ki jih vsebujejo nekatere besede (npr. beseda levo). Ta presledek med zlogi povzroči, da za nekaj časa tekoče povprečje pade pod pragovno vrednost, čeprav beseda še ni bila v celoti izgovorjena. Pragovno vrednost je potrebno določiti eksperimentalno, saj na vrednost tekočega povprečja vpliva uporabljen mikrofon in ojačanje mikrofonskega ojačevalnika. Ta postopek je pomanjkljiv, saj je pragovna vrednost konstantna in ne upošteva motenj, ki jih povzroča hrupno okolje.

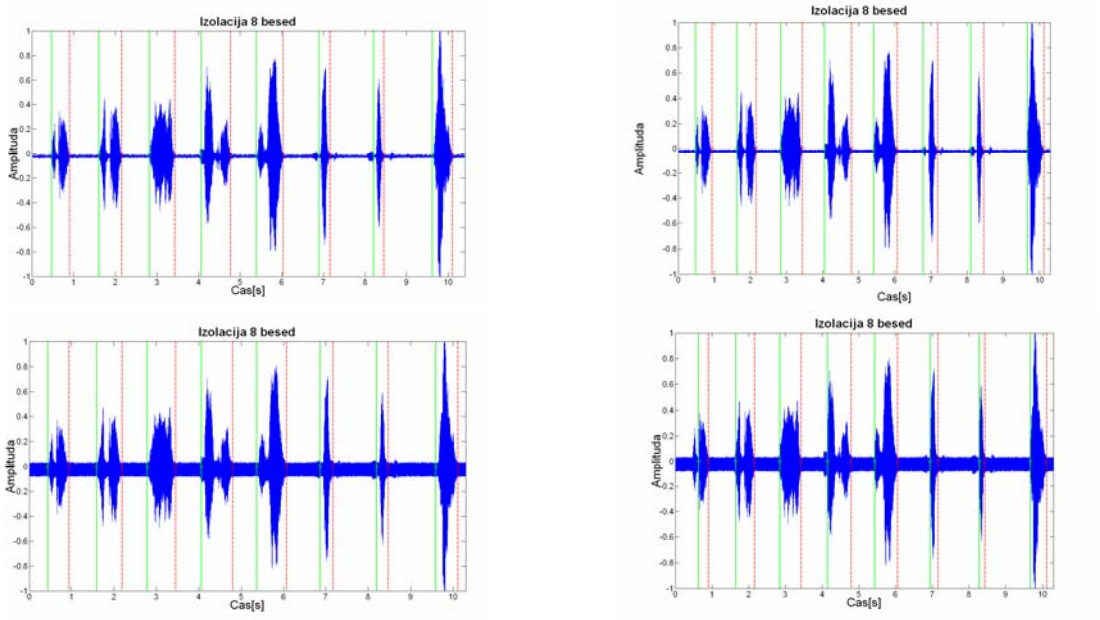
Druga metoda izoliranja besede se imenuje »zero crossing« metoda [4]. Ta metoda temelji na izračunu števila prehodov signala čez nič v posameznem bloku in računanju vsote amplitud v posameznem bloku signala. Po izračunu števila prehodov čez nič in vsote amplitud posameznega bloka signala, se izračunajo pragovne vrednosti. Te vrednosti se izračunajo iz začetnih blokov signala, ko še ni bila izgovorjena beseda in je v signalu samo šum. Spodnja pragovna vrednost števila prehodov čez nič se izračuna tako, da izračunamo srednjo vrednost prehodov čez nič. Izračuna se tudi spodnja pragovna vrednost amplitude signala, in sicer tako, da se izračuna srednja vrednost amplitude. Zgornjo pragovno vrednost amplitude dobimo tako, da srednjo amplitudno vrednost množimo s faktorjem, ki naj ima vrednost nekje 3. Ko je izračun vseh pragovnih vrednosti končan, lahko začnemo z iskanjem izgovorjenih besed iz posameznih blokov signala. Začetek besede iščemo tako, da vsoto amplitud treh blokov deljenih z tri primerjamo z zgornjo pragovno vrednostjo amplitude. Prekoračitev te vrednosti pomeni začetek besede, vendar ta začetek besede še malo zamikamo v levo, dokler se ne doseže spodnja pragovna vrednost amplitude. Na koncu se natančen začetek določi s primerjavo števila prehodov čez nič posameznih blokov okoli trenutnega začetka besede s pragovno vrednostjo števila prehodov čez nič. Na obraten način se po določitvi začetka besede določi tudi konec besede.

Slika 3.1 prikazuje rezultate izoliranja besed z obema postopkoma. Testirala se je izolacija besede s signalom z majhnim šumom in signalom, kateremu je bil dodan šum. Na levi strani so prikazani rezultati metode izoliranja besede z računanjem tekočega povprečja, na desni pa rezultati metode števila prehodov čez nič »zero crossing«. S polno črto so označeni začetki besed, s prekinjeno črto pa konci posameznih besed.

Iz primerjave je razvidno, da obe metodi dokaj uspešno razpoznata začetke in konce besed, vendar metoda prehoda čez nič bolj točno določi začetke in konce besed.

IV. LPC cepstralna analiza

Sedaj je potrebno izolirano besedo kodirati v obliko, ki je primernejša za razpoznavo. Pri tem smo uporabili kodiranje zvoka z Durbinovo metodo [1] in rekurzivnimi izrazi, ki jih je razvil Furui [2]. Ideja LPC (»Linear predictive coding«) analize temelji na aproksimaciji zvočnega signala kot linearne kombinacije predhodnih zvočnih vzorcev.



Slika 3.1: Odzivi obeh metod izolacije besed

Z minimizacijo vsote kvadratov razlike na končnem intervalu med dejanskim zvokom in linearno aproksimiranim lahko določimo edinstvene koeficiente. LPC analiza zagotavlja robustno, zanesljivo in natančno metodo ocenitve parametrov, ki karakterizirajo linearno variabilen sistem, kot je področje zvoka.

Kot je že bilo omenjeno, je potrebno izolirano besedo razdeliti na bloke (okvirje) z 330 vzorci, katerih začetki so med sabo zamaknjeni za 110 vzorcev. Za vsak posamezni blok se izračunajo LPC kepstralni koeficienti v naslednjih korakih:

1. Izračun avtokorelacijskih koeficientov z naslednjo enačbo:

$$r(m) = \sum_{n=0}^{N-1-m} x(n)x(n+m)$$
$$m = 1...13;\ N = 330 \qquad (1)$$

Kjer x predstavlja trenutni vhodni vzorec, m predstavlja število avtokorelacijskih koeficientov in N predstavlja število vzorcev v enem bloku (okvirju).

2. Ko so izračunani vsi avtokorelacijski koeficienti, je potrebno izračunati LPC koeficiente po naslednji enačbi:

$$a = -B^{-1} \cdot y \qquad (2)$$

Kjer je a vektor, ki vsebuje dvanajst LPC koeficientov, B je Toeplitz matrika dimenzije 12 × 12 in y je vektor zadnjih dvanajstih r(2:13) avtokorelacijskih koeficientov. Za izračun inverza matrike je bil uporabljen postopek, ki zahteva matriko dimenzije 12 × 24. Torej ima matrika enkrat več stolpcev kot Toeplitz matrika B. Toeplitz matriko pa dobimo tako, da v njo vstavimo prvih dvanajst r(1:12) avtokorelacijskih koeficientov v takšnem zaporedju, da ima matrika po diagonalah iste avtokorelacijske koeficiente. Zaradi varčevanja s pomnilniškim prostorom DSP procesorja, je pri izračunu inverza matrike uporabljena enotna matrika B dimenzije 12 × 24, ki v prvih dvanajstih stolpcih vsebuje Toeplitz matriko vektorja r(1:12), v drugih dvanajstih stolpcih pa je enotska matrika I.. Postopek izračuna inverza matrike se ponavlja tako dolgo, da ne dobimo v prvih 12 stolpcih enotske matrike I in v drugih 12 stolpcih inverza matrike, osnovne matrike, saj velja da je $B*B^{-1} = I$.

3. Izračun kepstralnih koeficientov iz LPC koeficientov:

Prva dva LPC kepstralna koeficienta dobimo z naslednjima enačbama:

$$c(1) = -a(1)$$
$$c(2) = -a(2) + \frac{a(1)^2}{2} \qquad (3)$$

Ostale LPC kepstralne koeficiente pa izračunamo z naslednjo enačbo:

$$k = (1 : n-1)$$
$$nkn = \frac{n-k}{n} \qquad (4)$$
$$c(n) = -a(n) - \sum_{k=1}^{n} a(k) \cdot c(n-k) \cdot nkn(k)$$

Kjer gre n od 3 do želenega števila kepstralnih koeficientov m, k je vektor, ki vsebuje n elementov in ima vrednosti od 1 do n-1, nkn je vektor z n členi. m predstavlja število LPC kepstralnih koeficientov.

Rezultat LPC kepstralne analize je, da se vsaka beseda pretvori v zaporedje točk, kjer vsaka točka pripada LPC kepstralnemu prostoru dimenzije 12. Posamezne točke so med sabo zamaknjene za 13 ms (zamik med posameznimi bloki). LPC kepstralni algoritem pretvori nihanje zračnega tlaka v diskretizirano trajektorijo v LPC kepstralnem prostoru.

V. SOM nevronska mreža

S samoorganizirajočo nevronsko mrežo (SOM) izvedemo dodatno redukcijo prostora trajektorije dobljene iz LPC kepstralne analize. SOM nam dobljeni 12 dimenzionalen prostor iz LPC kepstralne analize reducira na 2 dimenziji ob ohranitvi dovolj informacij trajektorije za dobro razpoznavo z usmerjeno nevronsko mrežo.

SOM spada med nevronske mreže s tekmovalnim načinom učenja, ki spada med nenadzorovane postopke učenja nevronske mreže (NM). Pri tem načinu učenja nam ni potrebno podati niti želenih izhodov niti ocen, ampak samo vhode v NM, torej nam ni potrebno določiti preslikave, ampak jo določi mreža sama. S SOM lahko izvajamo preslikave, ki transformirajo signalni vzorec poljubnih dimenzij v eno ali dvodimenzionalno polje. SOM se samodejno prilagajajo tako, da se podobni vhodni objekti povezujejo s topološko bližnjimi nevroni v SOM, kar pomeni, da se nevroni, ki so si v SOM sosednji, podobno odzivajo na podobne vhode. Medtem ko nevroni, ki so med sabo bolj oddaljeni, reagirajo različno na podobne vhode. Položaj nevrona, ki najbolj ustreza vhodnemu vektorju lahko določimo s primerjavo njegovih uteži in vhodnega vektorja, ki se imenuje tudi zmagovalni oz. prevladujoči nevron. V procesu učenja se v največji meri spremenijo uteži zmagovitega nevrona, medtem ko se uteži sosednjih nevronov spremenijo v manjši meri. Izhod iz SOM predstavljajo pozicijo zmagovalnega nevrona. Postopek iskanja zmagovalnega nevrona in spremembe uteži ali učenja SOM-a je povzet po Kohonenu [3].

Na kakovost redukcije dimenzij vpliva konfiguracija nevronske mreže. Potrebno je izbrati ustrezno število nevronov v mreži. Za našo aplikacijo smo se odločili, da bomo uporabili 256 nevronov, ki smo jo razporedili v dvodimenzionalno mrežo 16×16. Število vhodov posameznega nevrona pa mora ustrezati velikosti vektorja LPC kepstralnih koeficientov. SOM nevronsko mrežo učimo z LPC kepstralnimi koeficienti različnih besed. S korakom učenja pa se naj vrednosti učnih konstant zmanjšujejo. Ker učenje SOM nevronske mreže spada med nenadzorovane postopke učenja, po končanem učenju ni ustrezne metode kontroliranja kakovosti naučene SOM nevronske mreže. Preprosta metoda kontrole učenja SOM nevronske mreže je, da si na graf izrišemo pozicije posameznih zmagovalnih nevronov za različne besede. Pozicija nevrona v dvodimenzionalni SOM mreži je določena z dvema koordinatama, ki si ju lahko predstavljamo kot X in Y vrednosti kartezičnega koordinatnega sistema. Če se odzivi različnih besed med sabo razlikujejo in so si odzivi enakih besed podobni, lahko rečemo, da je bilo učenje SOM nevronske mreže uspešno.

VI. Usmerjena nevronska mreža

Ko dobimo iz SOM nevronske mreže vse reducirane LPC kepstralne koeficiente, gredo te vrednosti v razpoznavalnik z usmerjeno nevronsko mrežo [5], ki razpozna katera beseda je bila izgovorjena. Koeficiente dobljene iz

SOM mreže je potrebno normalizirati na določeno dolžino, saj je število vhodov v nevronsko mrežo fiksno določeno. Na število vhodov torej vpliva dolžina izgovorjenih besed, torej mora normalizacija SOM koeficientov potekati tako, da je normalizirani vektor koeficientov daljši od skupnega števila SOM koeficientov. S tem ne izgubimo dodatnih informacij o izgovorjeni besedi.

Postopek učenja, izračuna izhodov in sigmoidnih funkcij usmerjene NM je povzet po Dobnikarju [5]. Osnovni gradnik usmerjene nevronske mreže je umetni nevron, ki predstavlja poenostavljen model biološkega nevrona. Sam umetni nevron pa ni primeren za uporabo pri tako zapletenih primerih, kot je razpoznava govora. Zato posamezne nevrone združujemo v enonivojske ali večnivojske NM, ki so sestavljene iz vhodnega nivoja (vhodi v NM), skritih nivojev in izhodnega nivoja, katerega dolžina je odvisna od dolžine vektorja tarče. V našem primeru, je dolžina vektorja tarče in s tem število nevronov v izhodnem nivoju odvisno od števila besed, ki jih bomo z NM lahko razpoznali.

Postopek učenja je dokaj preprost, pomembna pa je pravilna izbire učilne konstante in število nevronov v skritem nivoju. Število nevronov v izhodnem nivoju in velikost vektorja tarče je v našem primeru določena z številom besed, ki jih želimo razpoznati. Vektor tarče pa je za vsako besedo drugačen. Pri razpoznavi besede pa izhode iz nevronov primerjamo s tarčami posameznih besed in tarča besede, ki je najbližje vrednostim trenutnega izhoda, je razpoznana beseda. Ker so členi vektorjev tarč določeni samo z vrednostmi 0 in 1, je v izhodnem nivoju uporabljena sigmoidna funkcija[5], ki ima zvezen izhod med 0 in 1. V skritem nivoju pa je uporabljena sigmoidna funkcija, ki ima zvezen izhod med -1 in 1.

VII. Zaključek

Iz primerjave rezultatov izolacije besed z metodo računanja tekočega povprečja in metodo »zero crossing«, smo ugotovili, da obe metodi dokaj dobro zaznata začetke in konce besed, vendar jih metoda prehoda čez nič določi bolj natančno. Kot smo že omenili je to posledica tega, da je pragovna vrednost pri izračunavanju tekočega povprečja konstantna in je določena eksperimentalno in sicer tako, da je metoda sposobna izolirati besede tako v mirnem, kot tudi v čimbolj hrupnem okolju. Iz tega vidimo, da bi ta metoda, ki ima tako določeno pragovno vrednost kar hitro zatajila, če bi jo uporabljali v zelo hrupnem okolju. Pri t.i. »zero crossing« metodi, pa se pragovni vrednosti spreminjata glede na motnje, ki so prisotne v signalu in zato je uporaba te metode bolj učinkovita in tudi za našo uporabo bolj zanimiva. Razpoznava govora je delo celotne projektne skupine, avtorja članka sva implementirala izolacijo besed z računanjem tekočega povprečja in LPC kepstralno analizo na DSP kartici in je pripravljena za shranjevanje izračunanih kepstralnih koeficientov v datoteko. Učenje nevronske mreže bo potekalo na osebnem računalniku, sama nevronska mreža pa je implementirana v programskem okolju .NET in je že primerna za razpoznavo govora. Zero crossing metodo pa sva zaenkrat implementirala v MATLABU kot m-funkcijo in testirala njeno delovanje, v prihodnosti pa jo bo potrebno implementirati za delovanje na DSP kartici, saj se je pokazala kot učinkovita metoda za izolacijo besed.

VIII. Literatura

[1] L. Rabiner, G. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, 1993.

[2] S. Furui, *Digital Speech Processing, Synthesys and Recognition,* Marcel Dekker Inc., 1989.

[3] T. Kohonen, *Self-Organization and Associative Memory,* Springer-Verlang, Berlin, 1984.

[4] S. MacAlpine, J. P. Slavinsky, N. Bharani, A. Virani, *Speaker Verification,* http://www.owlnet.rice.edu/~elec301/Projects99/.

[5] A. Dobnikar, *Nevronske mreže*, Didakta, Radovljica, 1990.

Mentor študentskega projekta: Gregor Pačnik