

**VILNIAUS UNIVERSITETAS
KAUNO HUMANITARINIS FAKULTETAS**

INFORMATIKOS KATEDRA

Verslo informacinių sistemų studijų programa
Kodas 62103S138

JULIUS PAPLONSKIS

MAGISTRO BAIGIAMASIS DARBAS

**BALSIŲ IR PUSBALSIŲ DISKRIMINAVIMO, NAUDOJANT SPEKTRINĘ
ANALIZĘ, TYRIMAS**

Kaunas, 2007

**VILNIAUS UNIVERSITETAS
KAUNO HUMANITARINIS FAKULTETAS**

INFORMATIKOS KATEDRA

JULIUS PAPLONSKIS

MAGISTRO BAIGIAMASIS DARBAS

**BALSIŲ IR PUSBALSIŲ DISKRIMINAVIMO, NAUDOJANT SPEKTRINĘ
ANALIZĘ, TYRIMAS**

Leidžiama ginti _____

Magistrantas _____

Darbo vadovas dr. doc. Pranas Žvinys

Darbo įteikimo data _____

Registracijos Nr. _____

Kaunas, 2007

TURINYS

Paveikslų sąrašas.....	4
Lentelių sąrašas	5
Įvadas	6
1. Kalbos atpažinimas	8
1.1. Atpažinimo klasifikavimas (žodžių ir fonemų atpažinimas)	9
1.2. Literatūros apžvalga.....	10
1.3. Dirbtinių neuroninių tinklų naudojimas šnekos atpažinime	15
1.3. Lietuvių kalbos garsai ir jų požymiai.....	17
1.4. Balsių – pusbalsių panašumai ir skirtumai.....	22
2. Balsių ir pusbalsių diskriminavimo, naudojant spektrinę analizę, metodas	25
2.1. Filtravimas	26
2.2. Garsų atrinkimas pagal ribą (balsius ir pusbalsius)	27
2.3. Langavimas	28
2.4. Energijos skaičiavimas.....	28
2.5. Akustinis modeliavimas.....	29
2.6. Dirbtinio neuroninio tinklo struktūra.....	30
3. Balsių ir pusbalsių diskriminavimo, naudojant spektrinę analizę, tyrimas.....	31
Išvados	69
Literatūros sąrašas.....	70
Summary	73

PAVEIKSLŲ SĄRAŠAS

1 pav. Automatinio kalbos atpažinimo schema	8
2 pav. Algoritmas „sprogstamųjų balsių tarties atpažinimas“	11
3 Pav. Sistemos blokinė diagrama	12
4 Pav. Klausla paremtos sistemos blokine diagrama	13
5 pav. Kalbos garsai	17
6 pav. Lietuvių kalbos balsių ir pusbalsių artikuliacinė klasifikacija	24
7 pav. Balsių ir pusbalsių diskriminavimo, naudojant spektrinę analizę, metodas.....	25
8 pav. Lygiagrečių juostinių filtrų bankas	26
9 pav. Nepersidengiantis filtrų banko diapazonas	27
10 pav. Žodžio segmentavimas	27
11 pav. Langavimo grafinė schema	28
12 pav. Energijos apskaičiavimas juostiniuose filtruose	29
13 pav. Fonemų akustinio modeliavimo pavyzdys	29
14 pav. „Trijų sluoksnių daugiasluoksnis perceptronas“	30
15 pav. Klasifikavimo rezultatai nuo 0 iki 1000Hz kairiojo fragmento	32
16 pav. Neuroninio tinklo apsimokymas nuo 0 iki 1000Hz kairiojo fragmento	32
17 pav. Klasifikavimo rezultatai nuo 0 iki 1000Hz stacionaraus fragmento	33
18 pav. Neuroninio tinklo apsimokymas nuo 0 iki 1000Hz stacionaraus fragmento	34
19 pav. Klasifikavimo rezultatai nuo 0 iki 1000Hz dešinio fragmento	35
20 pav. Neuroninio tinklo apsimokymas nuo 0 iki 1000Hz dešinio fragmento	35
21 pav. Klasifikavimo rezultatai nuo 1000 iki 2000Hz kairiojo fragmento	36
22 pav. Neuroninio tinklo apsimokymas nuo 1000 iki 2000Hz kairiojo fragmento	37
23 pav. Klasifikavimo rezultatai nuo 1000 iki 2000Hz stacionaraus fragmento	38
24 pav. Neuroninio tinklo apsimokymas nuo 1000 iki 2000Hz stacionaraus fragmento	38
25 pav. Klasifikavimo rezultatai nuo 1000 iki 2000Hz dešinio fragmento	39
26 pav. Neuroninio tinklo apsimokymas nuo 1000 iki 2000Hz dešinio fragmento	40
27 pav. Klasifikavimo rezultatai nuo 2000 iki 3000Hz kairiojo fragmento	41
28 pav. Neuroninio tinklo apsimokymas nuo 2000 iki 3000Hz kairiojo fragmento	41
29 pav. Klasifikavimo rezultatai nuo 2000 iki 3000Hz stacionaraus fragmento	42
30 pav. Neuroninio tinklo apsimokymas nuo 2000 iki 3000Hz stacionaraus fragmento	43
31 pav. Klasifikavimo rezultatai nuo 2000 iki 3000Hz dešinio fragmento	44
32 pav. Neuroninio tinklo apsimokymas nuo 2000 iki 3000Hz dešinio fragmento	44
33 pav. Klasifikavimo rezultatai nuo 3000 iki 4000Hz kairiojo fragmento	45
34 pav. Neuroninio tinklo apsimokymas nuo 3000 iki 4000Hz kairiojo fragmento	46
35 pav. Klasifikavimo rezultatai nuo 3000 iki 4000Hz stacionaraus fragmento	47
36 pav. Neuroninio tinklo apsimokymas nuo 3000 iki 4000Hz stacionaraus fragmento	47
37 pav. Klasifikavimo rezultatai nuo 3000 iki 4000Hz dešinio fragmento	48
38 pav. Neuroninio tinklo apsimokymas nuo 3000 iki 4000Hz dešinio fragmento	49
39 pav. Klasifikavimo rezultatai nuo 4000 iki 5000Hz kairiojo fragmento	50
40 pav. Neuroninio tinklo apsimokymas nuo 4000 iki 5000Hz kairiojo fragmento	50
41 pav. Klasifikavimo rezultatai nuo 4000 iki 5000Hz stacionaraus fragmento	51
42 pav. Neuroninio tinklo apsimokymas nuo 4000 iki 5000Hz stacionaraus fragmento	52
43 pav. Klasifikavimo rezultatai nuo 4000 iki 5000Hz dešinio fragmento	53
44 pav. Neuroninio tinklo apsimokymas nuo 4000 iki 5000Hz dešinio fragmento	53
45 pav. Klasifikavimo rezultatai nuo 5000 iki 6000Hz kairiojo fragmento	54

46 pav. Neuroninio tinklo apsimokymas nuo 5000 iki 6000Hz kairiojo fragmento	55
47 pav. Klasifikavimo rezultatai nuo 5000 iki 6000Hz stacionaraus fragmento	56
48 pav. Neuroninio tinklo apsimokymas nuo 5000 iki 6000Hz stacionaraus fragmento	56
49 pav. Klasifikavimo rezultatai nuo 5000 iki 6000Hz dešinio fragmento	57
50 pav. Neuroninio tinklo apsimokymas nuo 5000 iki 6000Hz dešinio fragmento	58
51 pav. Klasifikavimo rezultatai nuo 6000 iki 7000Hz kairiojo fragmento	59
52 pav. Neuroninio tinklo apsimokymas nuo 6000 iki 7000Hz kairiojo fragmento	59
53 pav. Klasifikavimo rezultatai nuo 6000 iki 7000Hz stacionaraus fragmento	60
54 pav. Neuroninio tinklo apsimokymas nuo 6000 iki 7000Hz stacionaraus fragmento	61
55 pav. Klasifikavimo rezultatai nuo 6000 iki 7000Hz dešinio fragmento	62
56 pav. Neuroninio tinklo apsimokymas nuo 6000 iki 7000Hz dešinio fragmento	62
57 pav. Klasifikavimo rezultatai nuo 7000 iki 8000Hz kairiojo fragmento	63
58 pav. Neuroninio tinklo apsimokymas nuo 7000 iki 8000Hz kairiojo fragmento	64
59 pav. Klasifikavimo rezultatai nuo 7000 iki 8000Hz stacionaraus fragmento	65
60 pav. Neuroninio tinklo apsimokymas nuo 7000 iki 8000Hz stacionaraus fragmento	65
61 pav. Klasifikavimo rezultatai nuo 7000 iki 8000Hz dešinio fragmento	66
62 pav. Neuroninio tinklo apsimokymas nuo 7000 iki 8000Hz dešinio fragmento	67

LENTELIŲ SĄRAŠAS

1 lentelė Eksperimentų rezultatų lentelė (proc.)	68
--	----

ĮVADAS

Temos aktualumas

Įvairių šalių mokslininkai jau apie 50 metų aktyviai plėtoja veiklą, tirdami kalbos atpažinimą, kurdami programinę įrangą, skirta kalbos atpažinimui, tačiau iki šių dienų sukurtos programos vis dar negali prilygti tikrai „gyvų“ žmonių kalbai. Šiuo metu Lietuvoje nėra sukurtos nei vienos komercinės kalbos atpažinimo sistemos, kadangi su lietuvių kalbos įrašais yra atlikta labai mažai tyrimų.

Sparčiai vystantis kompiuterinėms technologijoms, mažėjant įrenginių dydžiui, didėja poreikis kalbos atpažinimo technologijoms.

Bendravimas kalba yra suprantamas žmonėms, tuo tarpu kompiuteriui toks bendravimo būdas yra pernelyg sunkus, nes kiekvieno diktoriaus balsas yra individualus, todėl kalbos atpažinimo sistemų atpažinimo tikslumas nepakankamas realizavimui. Atpažinimo uždavinys sprendžiamas dviem etapais: pirmame išskiriami požymiai, o antrame etape jie lyginami. Deja, dažniausiai atpažinimą pavyksta atlikti tik keletui šimtų, aiškiai, su pauzėmis pasakytų žodžių, kai kalba tas pats diktorius, pagal kurio balsą buvo suformuoti etalonai.

Kalbos atpažinimui taip pat gali būti atliekamas tyrimas, naudojant akustinių – artikuliacinių įvykių aptikimo metodus, kurie dar vadinami žiniomis grindžiamos kalbos atpažinimu. Šis metodas remiasi teorinėmis ir eksperimentinėmis fonetikos žiniomis.

Tyrimo problema yra ta, kad lietuvių kalboje dvi fonemų grupės – balsių ir pusbalsių, yra pakankamai sunku atskirti.

Lietuvių kalboje fonetinių atpažinimų tyrimai dar nėra išbaigti, todėl nėra pakankamai nei teorinės nei eksperimentinės medžiagos tam, kad būtų galima atskirti šių dviejų klasių garsus

Objektas. Balsių ir pusbalsių diskriminavimo tyrimas naudojant spektrinę analizę.

Tikslas. Atlikus balsių ir pusbalsių diskriminavimo tyrimą (naudojant spektrinę analizę), išskirti balsių bei pusbalsių signalų požymius. Nustatyti, kuriose dažnių juostose geriausiai atsiskiria balsių ir pusbalsių skirtumai.

Uždaviniai:

- Atlikti teorinės lietuvių kalbos fonetikos balsių ir pusbalsių akustinių ir artikuliacinių požymių analizę.
- Atlikti diskriminavimo arba atpažinimo metodų apžvalgą, naudojama balsiams ir pusbalsiams.

- Sukurti teorinį modelį balsių ir pusbalsių diskriminavimui, panaudojant spektrinės analizės požymius, o klasifikavimui naudoti dirbtinius neuroninius tinklus.
- Naudojantis sukurtu metodu atlikti eksperimentinius balsių ir pusbalsių klasifikavimo tyrimus.

Darbas sudarytas iš trijų pagrindinių skyrių: teorinio, analitinio ir eksperimentinio. Pirmoje darbo dalyje aprašoma tai kas yra jau pasiekta šioje srityje, nagrinėjama Lietuvos ir užsienio literatūra. Antroje darbo dalyje pateikiamas metodas eksperimentui, kuris bus atliekamas trečioje eksperimentinėje dalyje.

Balsių ir pusbalsių spektrinei analizei atlikti darbe pasiūlytas juostinis filtras su langavimu ir fonemų akustiniu modeliavimu panaudojant neuroninį tinklą buvo pasiektas 77,15% tikslumas.

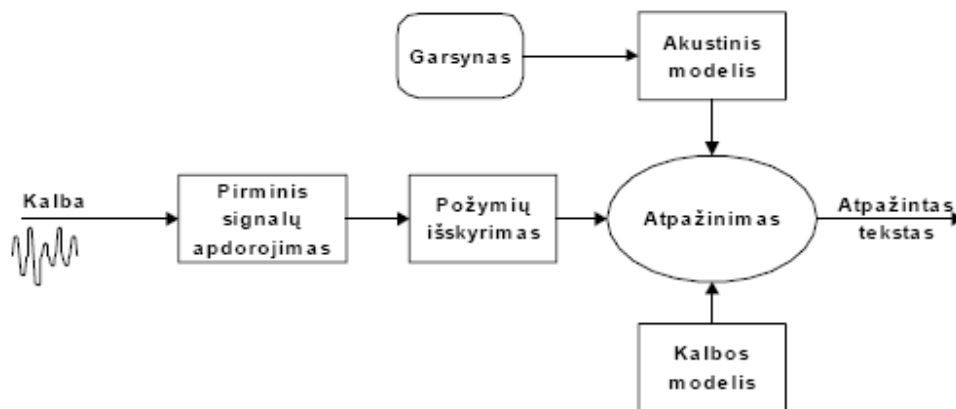
Bendra darbo apimtis 73 puslapiai, tame tarpe pateikiami 62 paveikslai ir 1 lentelė.

1. KALBOS ATPAŽINIMAS

Atpažinimo uždavinys sprendžiamas dviem etapais: požymių išskyrimo bei jų palyginimo. Kiekvienas kalbos žodis yra atskirai analizuojamas išskiriant pagrindinį toną, po to jis yra palyginamas su esamais duomenimis bazėje, ieškant minimalaus skirtumo. Deja, dažniausiai atpažinimą pavyksta atlikti tik keletui šimtų, aiškiai, su pauzėmis pasakytų žodžių, kai kalba tas pats žmogus, pagal kurio balsą buvo suformuoti etalonai.

Yra daugybė metodų, leidžiančių bendrauti su kompiuteriu, kuriais naudojas bene kiekvienas kompiuterio vartotojas: tai informacijos įvedimas klaviatūra, pele, jautriais paviršiais ir kitokiais elektroniniais prietaisais. Tačiau reikia pripažinti, kad patogiausias būdas žmogui bendrauti su kompiuteriu būtų kalba.

Tačiau galima pastebėti, kad bendravimas kalba yra suprantamas tarp žmonių, tuo tarpu kompiuteriui šiuo momentu tai yra sudėtinga, nes tai yra labai sudėtingas procesas, kuris priklauso nuo diktoriaus, jį supančios aplinkos ir t.t.



1 pav. Automatinio kalbos atpažinimo schema (K. Driaunys)

Kalbos atpažinimo principas pateiktas 1 pav. naudojamas daugelyje automatinio kalbos atpažinimo sistemų. Pagrindiniai kalbos atpažinimo principai būtų tokie:

1. kalbos nuskaitymas - diktoriaus sukeltas garsas (fonema, žodis) nuskaitymas mikrofono membranos dėka ir konvertuojamas į skaitmeninį pavidalą.

2. kalbos požymių išskyrimas – pagal matematinius skaičiavimus išskiriami naudingi požymiai, kurie apibrėžia: ar tai žodis, ar fonema, ar skienuo ir t.t. Galima pabrėžti tik tai, kad iki šiol nėra jokių požymių rinkinio, kuris galėtų tai identifikuoti.

3. kalbos atpažinimas – pagal išskirtus požymius vyksta atpažinimo etapas, kuriame priklausomai nuo uždavinio nustatoma, kokios fonemos, skiemenys, žodžiai ar jų sekos yra kalbos signale.

1.1. Atpažinimo klasifikavimas (žodžių ir fonemų atpažinimas)

Dabartines kalbos atpažinimo sistemas galima klasifikuoti į tokias kategorijas:

✓ priklausomos nuo diktoriaus ir nepriklausomos nuo diktoriaus

Priklausomos nuo diktoriaus sistemos tai tokios, kurios apsimoko pagal vieno diktoriaus sudiktuotus žodžius. Šios sistemos prisitaiko tik prie to diktoriaus, kuris sudiktuoja sistemos pateiktus žodžius.

Nepriklausomos nuo diktoriaus sistemos tai tokios, kurios sukurtos atpažinti kalbą nepriklausomai nuo diktoriaus. Tačiau jos sunkiau įgyvendinamos, nes jų klaidų tikimybė yra 3-5 kartus didesnė.

✓ Atskirai tariamų žodžių atpažinimas ir nepertraukiamos kalbos atpažinimas

Pirmuoju atveju kalbėtojas padaro pauzes tarp kiekvieno žodžio, skirtingai nei nepertraukiamame kalbos atpažinime (labai trumpos pauzės arba nėra visai). Todėl pavienio atpažinimo sistemas įgyvendinti yra lengviau.

✓ viso žodžio ir žodžio smulkesnių dalių atpažinimas

Pirmuoju atveju sistema yra išmokinta atpažinti visą žodį, tai naudinga, kai kuriama atpažinimo sistema balso komandoms. Efektyvesnis atpažinimas būtų atpažinti žodžių dalis – skiemenis arba net fonemas (Tan Keng Yan, Colin, 2000). Žodį būtų efektyviau atpažinti pagal jį sudarančias fonemas, nes, kaip žinoma, tas pats žodis gali būti kiekvieną kartą kitaip išstartas net ir to paties diktoriaus, priklausomai nuo aplinkybių

✓ viso žodžio ar fonemos atpažinimas pagal žodyno dydį

Nuo sistemos žodyno dydžio priklauso sistemos sudėtingumas. Sistemos žodyno dydis priklauso nuo sistemos paskirties,- jei sistema skirta bus atpažinti tik skaičiams, sistemai gali pakakti tik kelių dešimčių žodžių (skaičių pavadinimų). Kuriant sistemą, skirtą atpažinti rišlų tekstą, prireiktų dešimties tūkstančių žodžių žodyno.

1.2. Literatūros apžvalga

1) Sprogstamųjų priebalsių, automatiniam klasifikavimui, akustiniai - fonetiniai bruožai (Ahmed M. Abdelatty Ali, Jan Van der Spiegel, Paul Mueller. 2001)

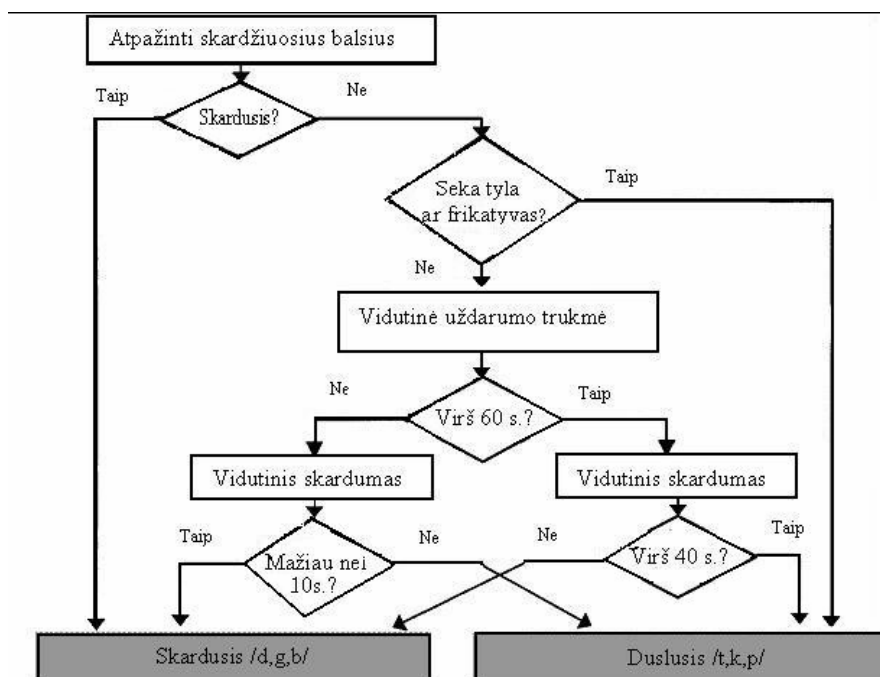
Šiame straipsnyje aprašomos sprogstamųjų priebalsių akustinės – fonetinės charakteristikos. Mišriu būdu sujungiamos ekspertų žinios ir statistinė analizė, tam, kad geriau suprasti įvairių statistinių ir dinaminių bruožų, atpažinimo procese, vaidmenį. Autoriai susikoncentruoja ties sprogstamųjų balsių klasifikavimu. Išplečiant sprogstamųjų balsių klasifikavimą aptariamos segmentavimo ir klasifikavimo sistemos.

Ši sistema yra suprojektuota naudojant 10 žmonių nenutrūkstamą kalbėjimą (penkių vyrų ir penkių moterų) iš TIMIT duomenų bazės ir tada išbandyta ant 1200 sprogstamųjų balsių, atrinktų iš 60 skirtingų asmenų, kalbančių 7 Amerikos anglų dialektais, nenutrūkstamos kalbos (nenaudotų kūrimo fazėje) iš TIMIT duomenų bazės.

Buvo naudojami i trys balso atpažinimui požymiai:

1. skardumo uždarumas (baigiant tarimą);
2. skardumas pradedant kalbėti;
3. uždarumo trukmė.

Šie trys bruožai yra sujungti algoritmu, pavaizduotu žemiau esančiame grafike, norint sugeneruoti skardumo nutarimą. Skardumo uždarumo aptikimas atliekamas matuojant santykį tarp žemo dažnio reikšmės energijos (iki 450 Hz) per 20 uždarumo intervalo sekundžių ir maksimalios reikšmės per visą kalbą. Jei santykis viršija tikrą pradžią (pasiektą statistiškai, naudojant histogramų analizę), sprogstamasis priebalsis yra apgalvotai uždaras. Trukmė (VOT ir uždarumo trukmė) yra matuojama naudojant ribas, sugeneruotas segmentacijos ir klasifikavimo sistemas tam, kad pažymėti įvairius sprogstamųjų priebalsių segmentus.

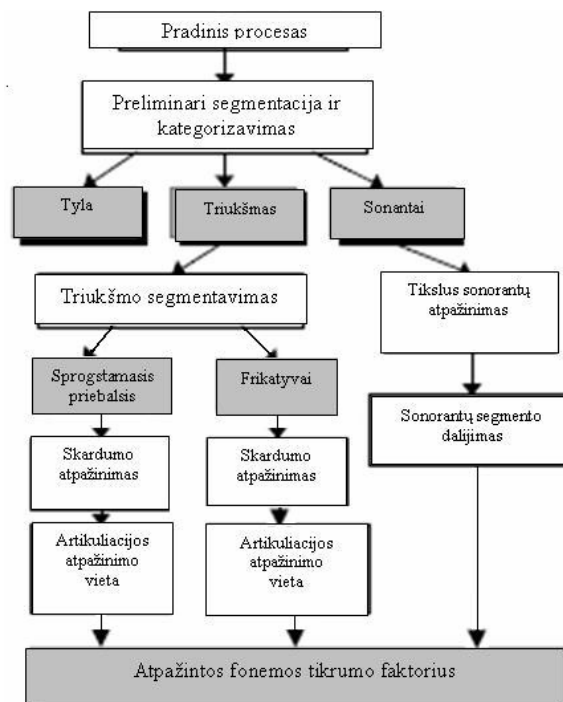


2 pav. Algoritmas „sprogstamųjų balsių tarties atpažinimas“ (Ahmed M. Abdelatty Ali, Jan Van der Spiegel, Paul Mueller. 2001)

2) Akustiniais – fonetiniiais bruožais paremta sistema, automatiniam fonemų atpažinimui nenutrūkstamoje kalboje (Ahmed M. Abdelatty Ali, Jan Van der Spiegel, Paul Mueller, Gavin Haentjens and Jeffrey Berman. 1999)

Šis darbas yra susijęs su naujos sistemos modeliavimu ir įdiegimu. Aprašoma sistema, kuri segmentuoja, skirsto į kategorijas ir atpažįsta nenutrūkstamos kalbos fonemas. Koncentruojamasi ties fonemų klasėmis (frikatyvine sprogstamąja) norint jas atskirti (išskirti) iš aplinkos triukšmo. Tai kuriama tam, kad fonemų atpažinimo sistemoje sumažinti klaidų skaičių ir informacijos praradimą.

Sistemos blokinė diagrama pavaizduota paveiksle (3 Pav.),



3 Pav. Sistemos blokinė diagrama (Ahmed M. Abdelatty Ali, Jan Van der Spiegel, Paul Mueller, Gavin Haentjens and Jeffrey Berman. 1999)

Pradinis procesas yra klausa paremta sistema. Pradinio proceso išėja galima segmentavimo ir klasifikavimo sistemai, kuri naudoja sekančius bruožus:

1. Visą energiją;
2. Spektrinę gravitacijos centrą (SCG);
3. Triukmę;
4. Žemo, vidutinio ir aukšto dažnio energiją;
5. Formančių perėjimus;
6. Tylos atpažinimą;
7. Kalbėjimo atpažinimą
8. Energijos dažnumo pokytį įvairaus dažnio diapazonuose;
9. SCG dažnumo pokytį;
10. Didžiausio (garsiausio) triukšmo dažnumą;
11. Didžiausio (garsiausio) triukšmo dažnumo pokytį.
12. Nulinio perėjimo dažnį.

Naudojant aukščiau išvardintus bruožus taisyklėmis paremtame algoritme, kuris detalizuoja skardžiuosius ir friktyvinius priebalsius, skardume ir tyloje, buvo pasiektas 92 procentų tikslumas

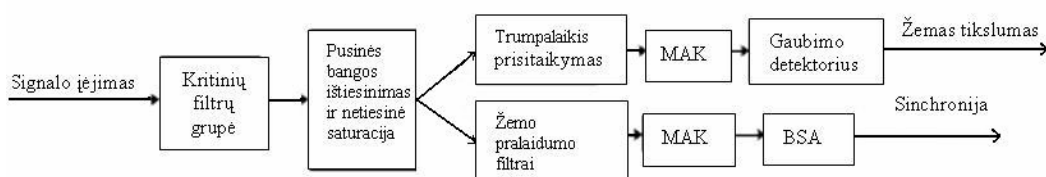
(4% keitimų, 3 % įterpinių ir 1 % išbraukimų), kai buvo tikrinama 30 žmonių besitęsianti kalba, 6 Amerikos anglų kalbos dialektais iš TIMIT (daugiau negu 7000 fonemų iš 300 sakinių).

3) Akustiniai – fonetiniai požymiai grindžiamo signalo apdorojimas ir panaudojimas automatiniam kalbos atpažinimui (Ahmed M. Abdelatty Ali.)

Nenutrūkstamos kalbos segmentavimas ir skirstymas į kategorijas tam, kad išskirstyti kalbos signalus į balsinguosius, friktyvinius, sprogstamuosius priebalsius ir pauzes.

Naudojantis klausa paremta sistema, pateikta (4 pav.), akustiniai fonetiniai požymiai buvo ištirti automatinio friktyvinių ir sprogstamųjų priebalsių atpažinimui. Užduotis buvo padalinta į tris pagrindines dalis:

1. Besitęsiančios šnekos segmentavimas, tam kad klasifikuoti kalbos signalus į skardžiuosius, friktyvinius (ir afrikatas), sprogstamuosius ir tylėjimą;
2. Friktyvinių priebalsių klasifikavimą;
3. Sprogstamųjų priebalsių klasifikavimą.



MAK - mechaninio augimo kontrolė; BSA - bendras sinchronizavimo atpažinimas

4 Pav. Klausa paremtos sistemos blokine diagrama

Algoritmo segmentavimas ir klasifikavimas buvo vystomas tam, kad išrinkti friktyvus (bei afrikatas) iš besitęsiančios kalbos. Algoritmas buvo tikrinamas 30 asmenų kalbėsenoje (nenaudotų projektavimo (lavinimo) procese), 6 dialektų iš TIMIT duomenų bazės. Buvo pasiektas 92 % fonemų atpažinimo tikslumas.

4) Kalbos atpažinimas per fonetinių bruožų skiemonavimą (Simon King, Paul Taylor, Joe Frankel, Korin Richmond 2000)

Aprašomas darbas ties dvejomis automatinėmis atpažinimo sistemomis, kurios vadinamos *Espresso-P* ir *Espresso-A*. *Espresso-P* remiasi fonologiniais bruožais, o *Espresso-A* artikuliaciniais

parametrais. Taip pat aptariamas linijinis dinaminis sisteminis modelis, naudojamas artikuliacinių parametru, automatiškai įvertintų reikšmių, atpažinimui.

5) Garsų klasifikavimas panaudojant sprendimų medžius(G. Daunys, D. Balbonas, 2005)

Nagrinėjama šnekos signalo analizė, panaudojant fonemų diferencinius požymius. Analizuojami parametrai, kurie reikalingi klasifikuojant šnekos intervalus į fonemų klases. Aptariamos naujesnės publikacijos, kuriuose pateikiami požymiai, leidžiantys klasifikuoti fonemas pagal artikuliacijos vietą. Diskutuojama apie galimybę sukurti metodiką lietuvių kalbos šnekos segmentavimą ir fonemų atpažinimą.

6) Hierarchine fonemų struktūra grindžiamo LTDIGITS fonemų klasifikavimo tyrimas(K. Driaunys, V. Rudžionis, P. Žvinys, 2005)

Šios grupės tyrimai orientuoti į kalbos atpažinimo tikslumo pagerinimą per geresnį fonetinių vienetų diskriminavimą (arba per geresnius fonetinių vienetų diskriminavimo metodus) kaip vieną iš perspektyviausių kalbos atpažinimo pagerinimo alternatyvų. Šiame darbe pateikiamas struktūrinis fonemų klasifikavimo algoritmas. Toks metodas yra grindžiamas prielaida, kad fonetinių vienetų požymių erdvėje yra pakankamai informacijos užfiksuoti ir išnaudoti struktūrinės kalbos signalo (būdingas tik tam tikros klasės fonemoms) savybes, kurios HMM (hidden Markov model) modelyje išnaudojamos nepakankamai.

7) Antros eilės požymių sistema šnekos signalo segmentavimo taškų atpažinimui (A. Raškinis, G. Raškinis, D. Kuliešienė, 2005)

Darbe pristatoma operatorių sistema ir jų naudojimo vienas po kito nuoseklumo taisyklės, kurios drauge sudaro gramatiką, generuojančią antros eilės požymių kalbą, aprašančią akustinį kontekstą mokymo pavyzdžio aplinkoje. Automatinio šnekos signalo segmentavimo sistemos, pagrįstos loginio mokymo metodais, segmentavimo taškus nustato loginių taisyklių pagalba, kurios gaunamos mokymo proceso metu, apibendrinant mokymui pateiktų dviejų klasių – artefaktų ir tikrųjų segmentavimo taškų pavyzdžių duomenis. Kiekvienas mokymo pavyzdys - tai vektorius, kurio komponentės - tai eksperto sukonstruotų to pavyzdžio aplinką charakterizuojančių požymių (antros eilės požymių) reikšmės. Antros eilės požymių sistema yra konstruojama kaip aibė operatorių, nustatančių tam tikrus santykius tarp duotojo mokymo pavyzdžio laiko momento ir

garso signalą charakterizuojančių parametrų trekų (pirmos eilės požymių) ypatingų laiko momentų.

8) Diktoriaus skaitomo rišlaus teksto garsynas (VDU-RTG) (A. Raškinis, G. Raškinis, J. Kapočiūtė Informacinės technologijos 2005)

Pranešime pristatomas naujas VDU sukurtas diktoriaus skaitomo rišlaus teksto bendrinės lietuvių šnekos garsynas (RTG), skirtas lietuvių šnekos technologijų mokslo tyrinėjimams. Jį sudaro apie 17,5 val. dviejų diktorių (vyro ir moters) šnekos įrašai. Juose yra 114130 (33645 skirtingų) žodžių. Tekstai parinkti iš 3 skirtingų žanrų. Garsynas anotuotas sakiniais, parengta tipinė dokumentacija. Garsynas naudojamas VDU atliekant šnekos atpažinimo ir šnekos sintezės mokslo tyrimo darbus.

1.3. Dirbtinių neuroninių tinklų naudojimas šnekos atpažinime

Dabartinės automatinio kalbos atpažinimo sistemos dažniausiai naudojami paslėptais Markovo modeliais HMM (angl. *Hidden Markov Model*), tiesinės prognozės koeficientai taip pat dažnai naudojami.

Bendros neuroninių skaičiavimų savybės:

Apsimokymas. Tinklai gali būti apmokomi formuoti ryšius tarp bet kokių įėjimo ir išėjimo šablonų. Tai panaudojama apmokant tinklą klasifikuoti šnekos pavyzdžius į fonemų kategorijas.

Apibendrinimas. Tinklas ne tik įsimena apmokymui naudojamus duomenis, bet tuo pačiu išmoksta esminius šablonus, kad galėtų apibendrinti apmokymo duomenis į naujus pavyzdžius. Tai yra būtina šnekos atpažinime, nes akustiniai modeliai niekada nėra tiksliai tokie patys.

Netiesiškumas. Tinklai gali skaičiuoti netiesines, beparametres įėjimo funkcijas, tai leidžia atlikti sudėtingas, komplikotas duomenų transformacijas. Tai yra naudinga šnekos atpažinime, nes šneka yra netiesinis procesas.

Tvirtumas. Tinklai yra tolerantiški tiek fiziniams nuostoliams, tiek triukšmingiems duomenims; be to, triukšmingi duomenys gali padėti tinklams formuoti geresnes taisykles. Tai naudinga savybė, kadangi šnekos šablonus iškraipo triukšmas.

Vienarūšiškumas. Tinklai siūlo vienarūšę skaičiavimo paradigmą, kuri lengvai integruoja apribojimus iš įvairių įėjimo tipų. Tai leidžia lengvai naudoti elementarius ir diferencijuotus šnekos įėjimus arba kombinuoti akustinius ir vaizdinius signalus multimodalinėje sistemoje.

Lygiagretumas. Tinklai yra lygiagretūs gamtoje, taigi jie yra tinkami įgyvendinant paralelinius skaičiavimus. Tai leistų pasiekti labai greitą tiek šnekos, tiek kitokių duomenų apdorojimą (J. Tebelski, 1995).

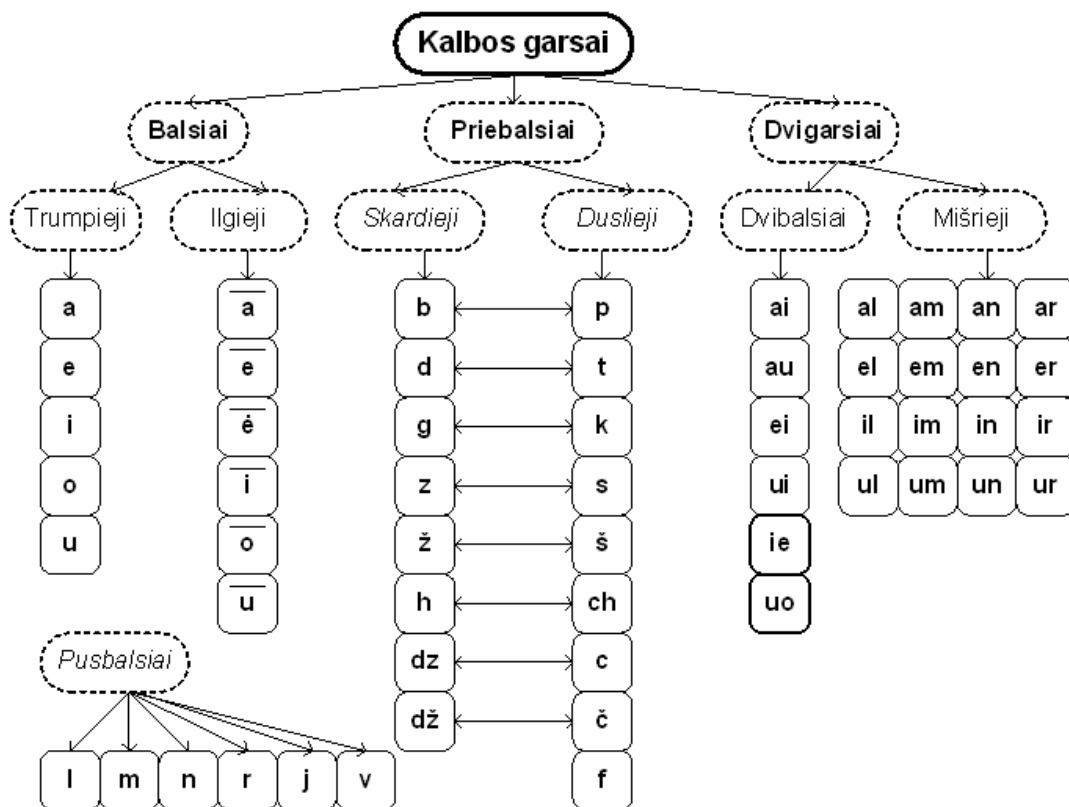
Taigi dirbtiniai neuroniniai tinklai taikomi prognozavimo, klasifikavimo/atpažinimo, modeliavimo, kontrolės problemoms spręsti. Jie vis plačiau naudojami dėl šių pagrindinių priežasčių:

➤ Neuroninis tinklas yra galingas modeliavimo aparatas, įgalinantis modeliuoti ypač sudėtingas funkcijas. Bendru atveju neuroninis tinklas yra netiesinė struktūra. Ilgą laiką buvo naudojami tiesiniai modeliavimo metodai, turintys gerai žinomas optimizavimo strategijas. Tais atvejais, kada tiesinės aproksimacijos nepakanka, tiesiniai modeliai nėra priimtini. Neuroniniai tinklai padeda išvengti dimensiškumo problemos, išskylančios modeliuojant daugelio kintamųjų funkcijas tiesiniais metodais (M. Filipovič, 2003).

➤ Neuroniniai tinklai apmokomi iš pavyzdžių. Turint duomenų pavyzdžius ir naudojant mokymo algoritmus, neuroninis tinklas pritaikomas prie duomenų struktūros. Nors neuroninio tinklo vartotojas turi turėti tam tikrų įgūdžių, kaip parinkti ir paruošti duomenis, parinkti neuroninį tinklą ir jo struktūrą, tačiau norint sėkmingai naudoti neuroninius tinklus pakanka mažiau žinių nei naudojant, pavyzdžiui, tradicinius statistinius metodus (M. Filipovič, 2003).

1.3. Lietuvių kalbos garsai ir jų požymiai

Kalbos garsai, kitaip vadinama fonetika,- yra kalbos garsų mokslas; kalbotyros šaka, tirianti kalbos garsų susidarymą ir jų akustines (girdimąsias) ypatybes.



5 pav. Kalbos garsai (šaltinis <http://ualgiman.dtiltas.lt/kalbos%20garsai.html>)

Fonetikos objektas yra ne tik garsai (balsiai, priebalsiai, pusbalsiai, dvigarsiai), bet ir jų derinimo dėsniai, skiemenys, kirtis, priegaidė, intonacija.

Kalbos garsinės sandaros nagrinėjimo aspektai yra keli: **artikuliacinis**, **akustinis**, **funkcinis** ir **ortoepinis**. Pagal tai fonetika (plačiąja prasme) gali būti skirstoma į **artikuliacinę** ir **akustinę** fonetiką (siaurąja prasme), **fonologiją** (tiria garsų funkcijas kalboje), **ortoepiją** (taisyklingos tarties mokslas). Be abejo, šios dalys yra reliatyviai savarankiškos, jos gali būti atskiros disciplinos. Tačiau praktiniais sumetimais lietuvių bendrinės kalbos fonetinę sistemą tikslingiausia, kad ir glaustai, apibūdinti visais atžvilgiais. (Lietuvių kalba ir literatūros istorija, http://ualgiman.dtiltas.lt/fonetikos_savoka.html)

Artikuliacinė fonetika. Ši fonetikos mokslo dalis nagrinėja kalbos padargų veiklą. Kalbos padargai yra valdomi centrinės nervų sistemos impulsų. Kadangi psichomotoriniai procesai mažai ištirti, artikuliacinė fonetika iš esmės grindžiama periferinių kalbos padargų veiklos analize.

Pagrindiniai kalbos padargai yra balso stygos, liežuvis (jo priešakinė, vidurinė, užpakalinė dalis), lūpos, priešakiniai dantys, dantenos, kietasis ir minkštasis gomurys. Kalbos garsų susidarymui svarbios burnos, nosies, ryklės ertmės. Į jas oras patenka iš plaučių pro bronchus, trachėją, gerklas. Tačiau pastarieji kvėpavimo organai nelemia garsų skiriamųjų požymių.

Tariant balsius oro srovė virpina suartėjusias balso stygas (taip atsiranda balsas), bet ryklės ir burnos ertmėje nesutinka tikrų kliūčių ir išeina pro burną palyginti laisvai. Todėl balsiai vadinami nekliūtiniais garsais. Tiesa, kai kurių balsių artikuliacijai būdinga tam tikra sąsmauka.

Pavyzdžiui, tariant $\bar{u}(u)$, liežuvio užpakalinė dalis būna pakilusi aukštyje prie minkštojo gomurio, lūpų anga susiaurėjusi. Vis dėlto tarpai tarp kalbos padargų dar nėra tokie siauri, kad kliudytų orui išeiti be trinties.

Artikuluojant priebalsius, kurioje nors burnos vietoje susidaro kliūtis: aklina uždaruma arba bent plyšys. Pro tą kliūtį balsinga ar nebalsinga oro srovė turi skverbtis su didesne jėga, negu tariant balsius. Pavyzdžiui, tardami priebalsį **p**, iš pradžių lūpas sučiaupiame ir užtveriamė kelią oro srovei, o po to šią uždarumą susprogdiname. Tardami **s**, tarp priešakinės liežuvio dalies ir beveik sukąstų dantų paliekame siaurą plyšelį, pro kurį skverbiasi oras. Dėl šių priežasčių priebalsiai vadinami kliūtiniais garsais.

Savitas priebalsių poklasis – pusbalsiai – pagal artikuliacijos būdą užima lyg tarpinę padėtį tarp balsių ir kitų priebalsių. Pavyzdžiui, tariant **m**, suspaustos lūpos sudaro uždarumą, bet oras gali laisvai išeiti pro nosį, nes minkštasis gomurys būna nusileidęs žemyn. Tariant **L**, liežuvio priešakinė dalis būna prisispaudusi prie priešakinių dantų, bet tarp liežuvio ir šoninių dantų lieka laisvi tarpai.

Priebalsių (neišskiriant pusbalsių) artikuliacijos vietą galima tiksliai nurodyti. Pavyzdžiui, **p** – abilūpis, **s** – liežuvio priešakinis, dantinis, **š** – liežuvio priešakinis, alveolinis, **j** – liežuvio vidurinis, gomurio vidurinis, **k** – liežuvio užpakalinis, gomurio užpakalinis. Balsių taip negalėtume apibūdinti. Sakysim, balsio **y** (**į**) artikuliacija yra priešakinė, bet susidarymo vieta nėra apibrėžta, nes liežuvio priešakinė ir vidurinė dalis priartėjusi ir prie priešakinės, ir prie vidurinės gomurio dalies. Vadinasi, priebalsiai yra **lokalizuoti**, o balsiai **nelokalizuoti**.

Balso stygų virpėjimas ar nevirpėjimas – tai skiriamasis skardžiųjų ir dusliųjų garsų požymis. Be to, virpesių dažnis ir amplitudė pareina nuo kirčio, priegaidės, intonacijos.

Akustinė fonetika. Jos objektas – kalbos padargų sukelti virpesiai. Akustinę fonetiką dar galima būtų skirti į dvi dalis: fizinę ir psichinę. Fizinė akustinė fonetika nagrinėja fizinius kalbos virpesius, o psichoakustinė fonetika – tų virpesių jutimą, suvokimą. Kalbotyroje dažniausiai remiamasi fizinės akustikos duomenimis. Tačiau neretai vartojami ir psichoakustiniai terminai, pvz.: tembras, tono aukštis ir pan.

Tariant balsius, balso stygų sukelti periodiniai oro virpesiai – vadinamasis pagrindinis tonas – sužadina ryklės, burnos, kartais ir nosies ertmių rezonansą. Rezonansiniai tonai ir lemia balsių kokybę – tembrą. Taigi balsiai yra toningiausi, skambiausi garsai.

Artikuluojant priebalsius oro srovė skverbiasi pro tam tikras kliūtis ir sukelia ne periodinius virpesius, kuriuos suvokiame kaip triukšmą. Virpesių pobūdis ir kiekis nėra vienodas. Duslieji priebalsiai susideda tik iš neperiodinių virpesių, o skardieji priebalsiai – iš pagrindinio tono ir neperiodinių virpesių. Pusbalsiai yra gana toningi garsai su nedideliu triukšmo priedu.

Daugelį akustinių požymių galima nustatyti iš oscilogramų. Oscilografu užrašytų kreivių pasikartojantys žemų dažnių svyravimai rodo pagrindinį toną, amplitudė – intensyvumą, ilgis – trukmę. Oscilogramose atsispindi ir kokybės ypatybės, tik jas gana sunku analizuoti. Į sudedamuosius virpesius kalbos garsai suskaidomi spektrografu. Balsių (iš dalies ir pusbalsių) spektrogramose išryškėja būdingieji dažniai, vadinami formantėmis. Svarbiausios yra dvi pirmosios formantės, sutrumpintai žymimos F1 ir F2. Pirmoji formantė (F1) 250 - 900 Hz, o antroji formantė (F2) – 400 - 2400 Hz diapazone. Triukšmo spektras yra platus, be ryškesnių formančių.

Fonologija. Jos tikslas – nustatyti fonetinių elementų funkcionavimą kalboje. Skiriami dvejopi elementai: segmentiniai ir supersegmentiniai arba kitaip prozodiniai elementai.

Kiekvieną posakį galime įsivaizduoti kaip tiesę, kurią skirstome segmentais. Mažiausias funkciškai savarankiškas tiesinis kalbos segmentas vadinamas fonema. Fonema yra abstraktus vienetas, kuris kalboje realizuojamas kaip konkretus garsas. Funkcinį garsų savarankiškumą rodo tai, kad jie padeda skirti žodžius. Pavyzdžiui, žodžiai **tarė** ir **darė** fonetiškai skiriasi pirmaisiais priebalsiais. Vadinasi, **t** ir **d** funkcionuoja kaip skirtingos fonemos. Jos atlieka skiriamąją funkciją.

Garsai kalboje funkcionuoja nevienodai. Balsiai eina skiemens pagrindu, o priebalsiai tik prisišlieja prie balsių. Lietuvių bendrinėje kalboje skiemens pagrindą gali sudaryti ir mišrieji dvigarsiai, kurių antrieji dėmenys yra **l**, **m**, **n**, **r**. Tai rodo, jog pusbalsiai funkciškai artimi balsiams. Fonologijos darbuose pusbalsinės fonemos paprastai apibūdinamos kaip turinčios ir

balsių, ir priebalsių požymių arba neigiamai – kaip nepriklausančios nei balsių, nei priebalsių klasei.

Pagrindinę fonologinės sistemos dalį sudaro fonemos, vartojamos grynuosiuose lietuviškuose žodžiuose arba senuosiuose skoliniuose, visiškai prisitaikiusiuose prie lietuviškų žodžių modelio. Šiais žodžiais pirmiausia turime remtis apibūdinami fonemų junginius, jų vartojimą. Bendrinėje kalboje yra ir naujesnių skolinių, kurių fonemos papildo sistemą. Tai vadinamieji šalutiniai kalbos elementai. Nagrinėdami lietuvių bendrinės kalbos fonologinę sandarą, negalima atsižvelgti į visas jaustukų, ištiktukų garsines ypatybes. Šiuose, dažniausiai ekspresyvinę funkciją atliekančiuose kalbos vienetuose, rastume balsių ir priebalsių, šiaip jau nebūdingų lietuviškiems žodžiams, pvz.: op, fui, oho, cha. Čia pasitaiko net ir labai keistos darybos garsų, pvz.: lūpinis, virpamasis priebalsis, kuriuo stabdomi arkliai, lūpinis, čiulpiamasis garsas, kuriuo šaukiami šunys, raginami arkliai, nuostabą reiškiantis švilptelėjimas, prasmingas krenkštelėjimas ir pan., jų net neįmanoma tiksliai užrašyti.

Nevienodomis sąlygomis realizuotos fonemos kiek skiriasi. Pavyzdžiui, žodyje **tarė** – **t** nesulūpintas, o žodyje **tūpė** – **t** sulūpintas. Tai tos pačios fonemos variantai. Jie funkciškai nėra savarankiški, nes priklauso nuo konkrečių aplinkybių; sakykime, nesulūpintas *t* vartojamas prieš nelūpinį balsį, o sulūpintas *t* – prieš lūpinį balsį. Taigi fonemų variantai negali savarankiškai atlikti skiriamosios funkcijos. Kalbėdami paprastai nė nepastebime fonemų varijavimo, nes jis tiesiogiai nėra susijęs su žodžių reikšme. Tam, kad pajustume fonemų variantų skirtumus, turime specialiai sukaupti dėmesį ir analizuoti jų tarimą.

Drauge su segmentiniais elementais kalboje vartojami ir kiti, supersegmentiniai elementai: kirtis, priegaidė, intonacija. Jie apima stambesnius kalbos vienetus: žodžius, skiemenis, frazes, ant jų yra tarsi uždedami, užklojami. Supersegmentiniai elementai susiję su tarimo gaida. Pagal tai jie patys ir juos išreiškiantys fonetiniai požymiai dar vadinami prozodiniais. Funkciškai savarankiški prozodiniai elementai yra prozodemos. Jos, kaip ir fonemos, paprastai atlieka skiriamąją funkciją. Pavyzdžiui, žodžius *giria*: *giria* padeda diferencijuoti kirtis, žodžius *šauk* : *šauk* – priegaidė, frazės *Pavasaris. Pavasaris?* – intonacija. Kai kuriuose darbuose intonacijos prozodemos vadinamos intonemomis.

Fonologiniai vienetai, ypač prozodemos, atlieka ir daugiau funkcijų: išskiriamąją, atskiriamąją ir kt.

Prozodemos dėl įvairių fonetinių aplinkybių varijuoja. Sakykime, žodžio *kitas*, išarto tvirtinamąją ir klausiamąją intonacija, prozodiniai požymiai gerokai skiriasi, bet vis tiek kirčiuotas

yra pirmasis skiemuo. Tai tos pačios kirčio prozodemos variantai. Kirčio, priegaidės, intonacijos prozodemų variantai nėra pakankamai ištirti ir susisteminti, todėl apie juos mažai ką galima pasakyti.

Tie požymiai, kurie padeda skirti fonemas arba prozodemas, vadinami distinktyviais (skiriamaisiais, diferenciniais, relevantiniais).

Ortoepija. Tai fonetikos mokslo dalis, kurios uždavinys – nustatyti tarties normas, numatyti galimus jų pažeidimus.

Bendrinės kalbos normos nustatomos pagal aiškųjį tarties stilių. Šiam stiliui būdingas lėtokas kalbos tempas, ryški garsų artikuliacija. Aiškioju stiliumi dažniausiai kalbama viešose vietose: paskaitose, per radiją, televiziją ir pan. Buitinėje kalboje, greičiau šnekant, tartis neretai būna ne tokia aiški, garsų požymiai blankesni, kai kurie elementai gali būti net visai praleidžiami. Tai vadinamasis šnekamasis stilius.

Dėl tam tikrų istorinių, socialinių, kultūrinių sąlygų lietuvių bendrinė tartis susiformavo vakarų aukštaičių kauniškių šnektos pagrindu. Tačiau nereikia manyti, jog bendrinė kalba yra vienos šnektos kopija. Kai kuriomis fonetinėmis ypatybėmis bendrinė tartis yra nutolusi nuo vakarų aukštaičių kauniškių tarties. Be to, esama ir nenusistovėjusių bendrinės tarties dalių. Tarties normas padeda įtvirtinti mokykla, radijas, televizija, teatras, spauda.

Rūpinantis taisyklinga bendrine tartimi, nuolat susiduriama su normų pažeidimais – tarties klaidomis. Kad galėtume sėkmingiau lavinti tartį, turime žinoti tarties klaidų priežastis, paplitimą, mokėti vertinti klaidų didumą, gajumą.

Tarties klaidų atsiranda dėl tarmių, svetimų kalbų, rašybos įtakos bei kitų priežasčių. Lietuvių bendrinės kalbos ir tarmių fonetinės sistemos skiriasi. Mokantis bendrinės tarties, iš pradžių bandoma verstis gimtosios kalbinės sistemos priemonėmis. Sakysim, rytų aukštaičiai, kirčiuotus negalimus balsius *i*, *u* taria pusilgius. Todėl šių vietų žmonės, šnekėdami ir bendrine kalba, dažnai ilgina kirčiuotus trumpuosius balsius. Kirčiuoti trumpieji balsiai (netgi žodžio gale) būna ilginami, pavyzdžiui, ir dėl rusų kalbos įtakos. Nes rusų kalboje bet kuris kirčiuotas balsis yra pailgėjęs.

Mūsų rašyba nėra gryna fonetinė. Nemaža žodžių rašome vadovaudamiesi morfologiniu, istoriniu, tradiciniu principu. Dėl rašybos ir tarties neatitikimo gali atsirasti klaidų. Sakysim, žodyje *iššoko* pagal morfologinį principą rašomos dvi raidės *šš*, nors bendrinėje kalboje dvigubi priebalsiai netartini. Dažnas, veikiamas rašto, stengiasi išlaikyti abu priebalsius *šš*.

Tarties klaida nelygu klaidai. Vienos yra didelės, kitos – mažos. Didelės tarties klaidos yra tos, kurias darant pažeidžiama bendrinės kalbos fonologinė sistema: painiojamos fonemos,

prozodemos, nesilaikoma jų kaitos dėsnų ir pan. Pavyzdžiui, lietuvių bendrinėje kalboje trumpieji ir ilgieji balsiai yra skirtingos fonemos, nes padeda skirti žodžius, plg.: *lupa* : *lūpa*, *lygus* : *lygūs*. Kas vietoj kirčiuoto trumpojo *u* taria *ū* arba vietoj nekirčiuoto ilgojo *ū* taria trumpąjį *u*, tas pažeidžia bendrinės kalbos sistemą – daro stambių tarties klaidų.

Daug menkesnės klaidos – netiksliai tariami bendrinės kalbos fonemų ar prozodemų variantai. Pavyzdžiui, daugelis rytų aukštaičių taria ne tokius uždarus kirčiuotus *o*, *é*, kokie turėtų būti bendrinėje kalboje. Šiek tiek atviresnis jų tarimas nekenkia fonologinei sistemai, nes vienos fonemos nepakeičiamos kitomis.

Sunkiausia išmokti tų fonetinių elementų, kurių iš viso nėra gimtojoje tarmėje. Pavyzdžiui, daugeliui žemaičių sunku išmokti bendrinės kalbos balsių *o*, *é*, nes jų šnekėsenoje šie balsiai paprastai nevartojami. Dar sunkiau taisyti mažąsias klaidas – bendrinei kalbai nepriimtinius fonemų, prozodemų variantus. Mat gimtojoje tarmėje dažniausiai reikiamų variantų iš viso nėra, be to, variantų skirtumus daug sunkiau išgirsti. Pavyzdžiui, rytų aukštaičių šnekėsenoje kitokių balsių *o*, *é* ir nerasime. Gimtojoje tarmėje neturimų fonetinių elementų reikia mokytis labai atkakliai, beveik taip pat, kaip ir neįprastų užsienio kalbos garsų.

Jei gimtojoje tarmėje yra reikiamų garsų ir jie vartojami tokioje pat pozicijoje, kaip ir bendrinėje kalboje, taisyklingos tarties mokytis nėra sunku. Pavyzdžiui, žodžio *daug* gale žemaičiai taria skardųjį priebalsį *g*, bet jie lengvai pasako šį žodį ir pagal bendrinės kalbos normą – su dusliuoju priebalsiu *k*, nes tokį garsą vartoja kitų žodžių gale, pvz.: *būk*, *šauk*. (Lietuvių kalba ir literatūros istorija, http://ualgiman.dtiltas.lt/fonetikos_savoka.html)

1.4. Balsių – pusbalsių panašumai ir skirtumai

Kalbos garsų sistemoje yra išskiriamos dvi pagrindinės garsų grupės – balsiai ir priebalsiai. Balsių akustinį pagrindą sudaro tonai su nedideliu kiekiu šlamesių. Tonai - tai tam tikro dažnumo periodiškai pasikartojantys virpesiai. Šlamesiai yra netolygus, neperiodiški virpesiai.

Kiekvienas balsis turi vieną pagrindinį toną ir tam tikrą skaičių nevienodo aukštumo šalutinių tonų, priklausančių nuo balso vamzdžio (įgerklio, ryklės ir burnos ertmės) formos. Tie šalutiniai tonai vadinami būdingais balsio tonais. Nuo šių tonų visumos priklauso balsio kokybė, kuria vienas balsis skiriasi nuo kito, pvz.: *ā* nuo *ō*, *é* nuo *ī*, *ō* nuo *u* ir t. t.

Visų balsių artikuliacijai būdinga:

- a) vieno artikuliacijos židinio nebuvimas;

- b) silpna iškvepiama iš plaučių oro srovė;
- c) burnos atvirumas, pro kurią oro srovė iškvepiama be kliūčių;
- d) balso stygų veikimas. (G. Daunys)

Iš visų skardžiųjų priebalsių išsiskiria *l, m, n, r*, vadinami sonantai. Jie sudaro tarpinę grupę tarp balsių ir priebalsių.

I balsius jie panašūs šiais bruožais:

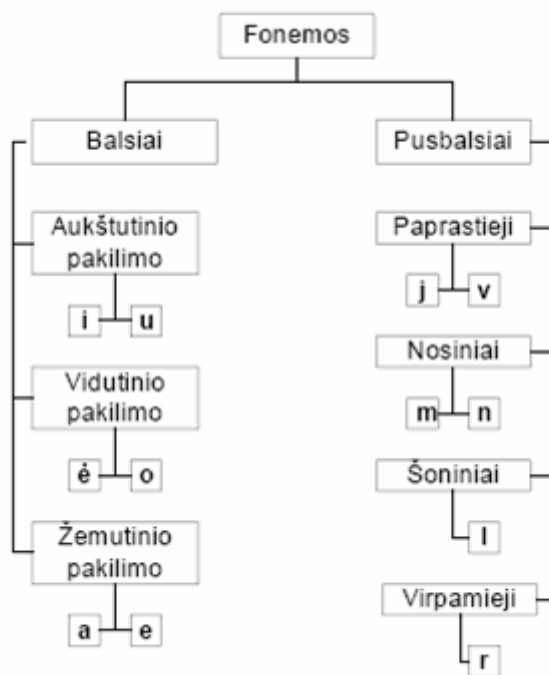
- a) daug silpnesne, negu kitų priebalsių, iškvepiama oro srove ir nedideliu šlamesių kiekiu;
- b) nepaprastu toningumu - lietuvių literatūrinėje kalboje jie visi yra skardūs ir neturi atitinkamų dusliųjų porų;
- c) kaip ir balsiai, jie niekada neduslėja, o priebalsiai prieš juos niekada neskardėja, pvz.: *dvelkti, išrovė* (ne *ižrovė*);
- d) eidami mišriųjų dvigarsių antraisiais dėmenimis, sonantai tariami su tono aukštumo moduliacijomis ir įvairiose fonetinėse pozicijose yra nevienodo ilgumo.

Tarpiniai tarp sonantų ir kitų priebalsių yra *v, v'* ir *j*.riebalsiai *v, j* visur, išskyrus žodžio galą, ir *v'* turi gana didelį kiekį šlamesių. Be to, jie neturi tono moduliacijos. Prieš juos, kaip ir prieš balsius bei sonantus, priebalsiai neskardėja, pvz.: *išjojo* (ne *ižjojo*), *atvarė* (ne *advarė*), *išvirė* (ne *ižvirė*). (G. Daunys)

Kiekvieną šnekamosios kalbos sistemos garsą (fonemą) galima apibūdinti šiais aspektais: artikuliaciniu, akustiniu (Lietuvių kalba I, 1995). Žemiau pabandyčiau trumpai apžvelgti kalbos garsų požymius artikuliaciniais ir akustiniais aspektais.

Artikuliacinis aspektas apibūdina kalbos padargų padėtį ir veiklą tariant garsus, arba, kitaip tariant, yra tiriami kalbos padargų judesiai ir jų įtaka garsų savybėms. Artikuliacinius balsių ir priebalsių skirtumus ryškiausiai atskleidžia balso trakto forma. Tariant balsius oro srovė virpina suartėjusias balso stygas, bet ryklės ir burnos ertmėje nesutinka tikrų kliūčių ir išeina pro burną palyginti laisvai. Artikuliuojant priebalsius, kurioje nors burnos vietoje susidaro kliūtis: aklina uždaruma arba bent ankštuma (plyšys). Pro tą kliūtį balsinga ar nebalsinga oro srovė turi skverbtis su didesne jėga, negu tariant balsius. Savitas priebalsių poklasis – pusbalsiai – pagal artikuliacijos būdą užimantys tarpinę padėtį tarp balsių ir kitų priebalsių. Pavyzdžiui, tariant *m*, suspaustos lūpos sudaro uždarumą, bet oras gali laisvai išeiti pro nosį; tariant *l*, liežuvio priešakinė dalis būna prispaudusi prie priešakinių dantų, bet tarp liežuvio ir šoninių dantų lieka laisvi tarpai (Pakerys,

1986). Balso stygų virpėjimas ir nevirpėjimas – tai skiriamasis skardžiųjų ir dusliųjų garsų požymis. Be to, virpesių dažnis ir amplitudė priklauso nuo kirčio, priegaidės ir intonacijos. Lietuvių kalbos fonemų (be dvibalsių, dvigarsių ir priebalsių) artikuliacinė klasifikacija pateikta 6 pav. (K. Driaunys 2006)



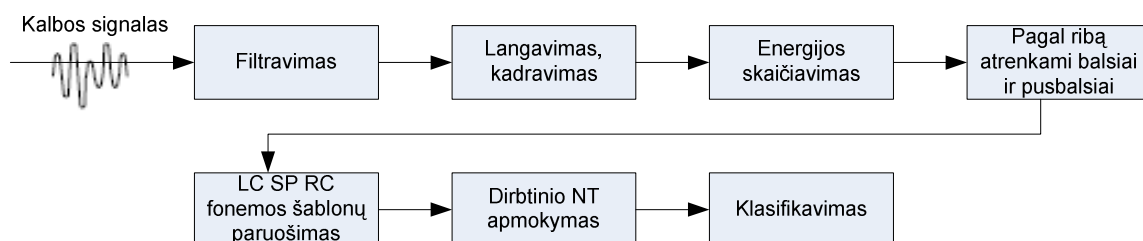
6 pav. Lietuvių kalbos balsių ir pusbalsių artikuliacinė klasifikacija (K. Driaunys 2006)

Akustiniu aspektu garsai yra tyrinėjami kaip kalbos padargų sukelti virpesiai. Akustinius balsių ir priebalsių skirtumus lemia šių garsų grupių artikuliacijos ypatybės. Pagrindinis akustinis skirtumas – balsingumas (Lietuvių kalba I, 1995). Balsiai yra balsingi garsai, t.y. juos formuoja periodiškai oro virpesiai – vadinamasis pagrindinis tonas.

Tiriant lietuvių kalbą, eksperimentinė fonetika nemažai dėmesio skiria balsių savybėms, tačiau priebalsių akustinės savybės nėra detalios išnagrinėtos. Fonemų savybės taip pat priklauso nuo kalbėjimo stiliaus, emocijų bei kitų įvairių aplinkybių. Galima išskirti du pagrindinius kalbėjimo stilius: aiškųjį tarties stilių ir šnekamąjį stilių. (Pakerys, 1995). Nustatinėjant lietuvių kalbos fonemas paprastai yra remiamasi aiškioju stiliumi ir tai reikėtų išskirti kaip neigiamą savybę, kadangi taikomiesiems kalbos atpažinimo paketams dažniausiai reikia atpažinti šnekamojo stiliaus žodžius. (K. Driaunys. 2006)

2. BALSIŲ IR PUSBALSIŲ DISKRIMINAVIMO, NAUDOJANT SPEKTRINĘ ANALIZĘ, METODAS

Automatinio kalbos atpažinimo sistemos dažniausiai remiasi pavyzdžiais grįstais metodais. Šio metodo pateikto (7 pav.) esmė yra ta, kad paruošiami duomenys (fonemos), kurie yra klasifikuojami į dvi dalis tai yra balsius ir pusbalsius, apmokius neuroninį tinklą.



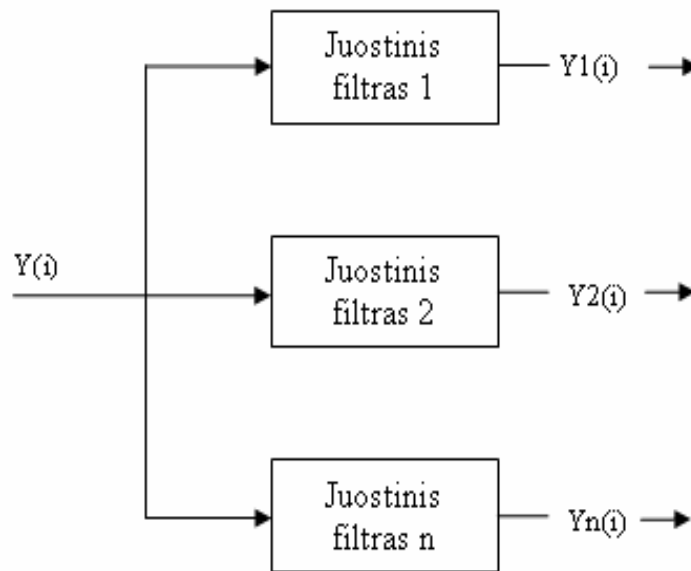
7 pav. Balsių ir pusbalsių diskriminavimo, naudojant spektrinę analizę, metodas

Fonemų atpažinimas – tai šnekos atpažinimas, kurio metu priima sprendimą neuroninis tinklas, kokią fonemų klasę (šiuo atveju balsis ar pusbalsis) labiausiai atitinka atpažinta fonema.

Prieš atliekant kalbos požymių išskyrimą ir apskaičiavimą yra atliekami paruošiamieji kalbos signalo apdorojimo veiksmai eiliškumas yra matomas 7 paveiksle, tai būtų signalo filtravimas, jo kadravimas, energijos paskaičiavimas, pagal ribas balsių ir pusbalsių atrinkimas, fonemų šablonų paruošimas, dirbtinio neuroninio tinklo apmokymas ir fonemų klasifikavimas ir balsius ir pusbalsius. Išsamiau aprašyti etapai sekančiuose skyreliuose.

2.1. Filtravimas

Kalbos signalo požymiai gauti spektrinės analizės būdu yra plačiai naudojami. Viena iš pagrindinių plataus šio metodo paplitimo priežasčių yra tai, kad atliekant spektrinę analizę galima išgauti svarbias akustines kalbos signalo charakteristikas įvairiose dažnių juostose. Kaip žinoma, kalbos signalas yra nestacionarus procesas ir spektrinė kalbos signalų analizė remiasi prielaida, kad kalbos signalą galima suskaidyti į trumpus intervalus, kuriuose signalas tampa stacionarus arba kvazistacionarus (Rudžionis, 1998).



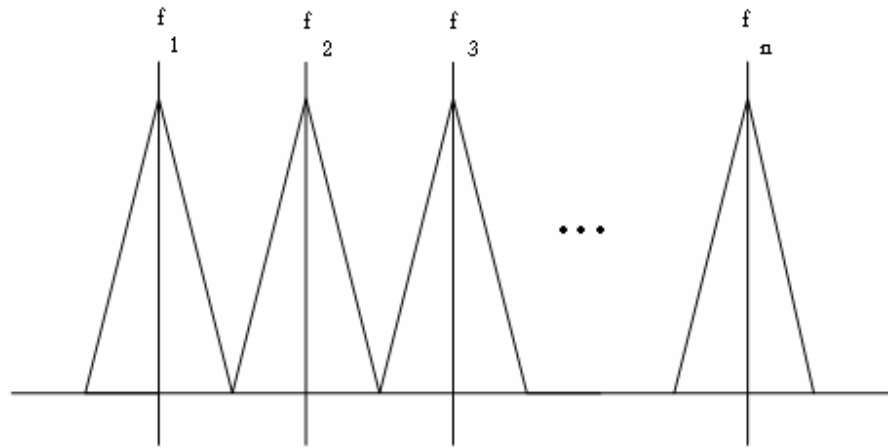
8 pav. Lygiagrečių juostinių filtrų bankas

Trumpo intervalo spektrinei analizei bus naudojamas filtrų bankų metodas.

Bus naudojama 80 juostinių filtrų sistema, imant 100 Hz pralaidumo juostas, signalo spektras gausis nuo 0 Hz iki 8000 Hz.

Naudojant filtrų banko metodą, kalbos signalas $y(i)$ yra praleidžiamas per iš n juostinių filtrų sudarytą filtrų banką (8 pav.), kuris yra nubraižytas (9 pav.).

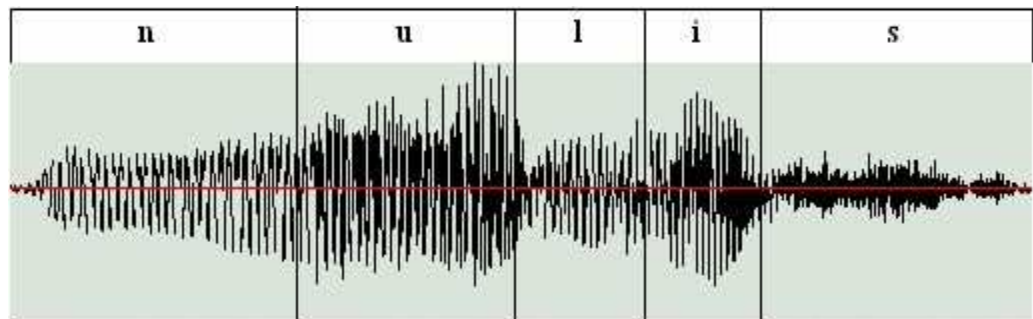
Tokiu būdu n -tojo juostinio filtro, kurio centrinis dažnis f_n , išėjime yra kalbos signalo $y(i)$ energija, o visų n filtrų energijos tikslina trumpalaikį signalo spektrą.



9 pav. Nepersidengiantis filtrų banko diapazonas

2.2. Garsų atrinkimas pagal ribą (balsius ir pusbalsius)

Fonemos, balsių ir pusbalsių, buvo paimtos iš lietuviško garsyno LTDIGITS. Kaip pavyzdys galėtų būti (10 pav.)



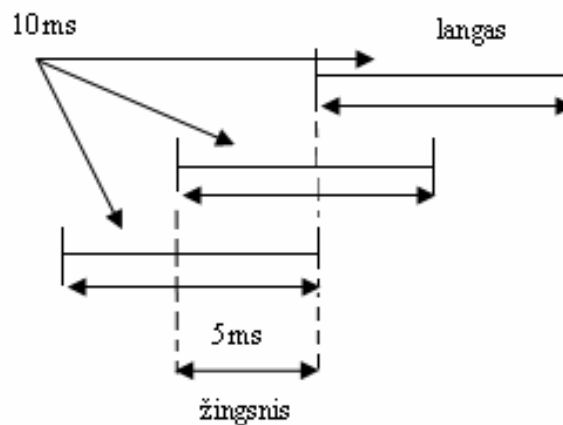
10 pav. Žodžio segmentavimas

Pagal (10 pav.) galima spręsti, kad žodis buvo susegmentuotas fonemomis ir iš jų buvo atrinkti balsiai ir pusbalsiai. Šiuo atveju balsiai būtų u ir i , pusbalsiai l ir n .

2.3. Langavimas

Filtruotas signalas yra suskaidomas į persidengiančių kadru (langu) seką. Signalo langu formavimas nusakomas dviem parametrais: lango ilgiu ir lango postūmiu arba žingsniu. Lango ilgio ir žingsnio pasirinkimas priklauso nuo atpažinimo sistemos darbe naudojamų metodų; ilgis buvo pasirinktas 10ms, o lango žingsnis 5ms.

Langu persidengimas yra naudojamas siekiant efektyviau išnaudoti iš dviejų gretimų langu gautą informaciją, su tikslu panaudoti energijos apskaičiavimui.



11 pav. Langavimo grafinė schema

2.4. Energijos skaičiavimas

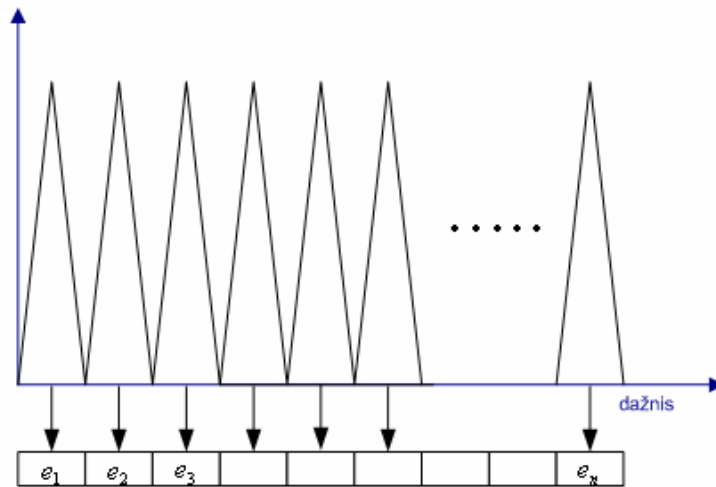
Sulangavus signalą iš gautų duomenų bus apskaičiuojama lango (freimo) energija pagal formulę:

$$E = \sum_i^N s(i)^2$$

Kur N lango ilgis, s - signalas

s(i) – i-toji signalo reikšmė.

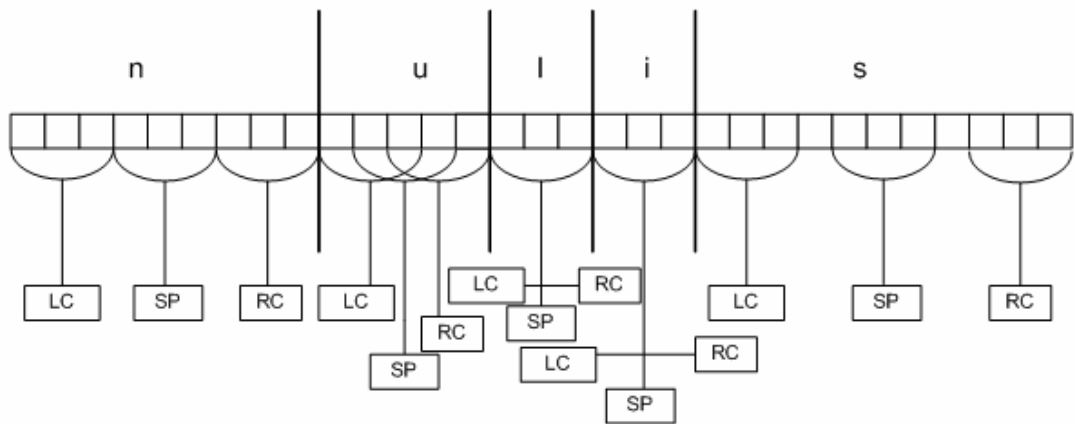
Šiuo atveju toks energijos apskaičiavimas grafiškai turėtų atrodyti taip:



12 pav. Energijos apskaičiavimas juostiniuose filtruose

2.5. Akustinis modeliavimas

Jei fonema yra ilgesnė nei užduota etalono trukmė, šiuo atveju fonema (S), tuomet dalis akustinio kalbos signalo yra ignoruojama vadovaujantis prielaida, kad šiuose fragmentuose esanti informacija neturi esminės įtakos atpažinimui.



13 pav. Fonemų akustinio modeliavimo pavyzdys

Naudojant tokį fonemos modeliavimo algoritmą, galima sumažinti apdorojamos informacijos kiekius, klasifikavimui naudojant tik stacionarią dalį, tačiau neprarandant svarbios kontekstinės informacijos. Jei fonemos žymė po klasifikavimo yra priskiriama su žema tikimybe, papildomai galima įvesti kontekstinių dalių klasifikavimą, tikintis pataisyti padėtį.

2.6. Dirbtinio neuroninio tinklo struktūra

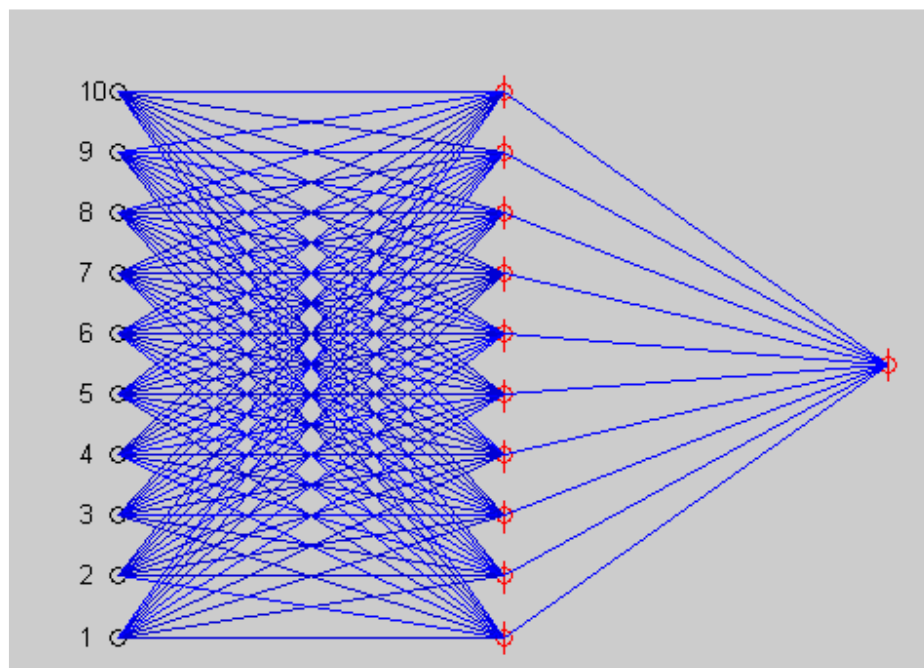
Balsių ir pusbalsių atpažinimui pasirinktas vieno paslėpto sluoksnio daugiasluoksnis perceptronas.

Paprastasis perceptronas yra paprasčiausias gyvo organizmo neurono matematinis modelis. Vienasluoksnis perceptronas yra grupė paprastųjų perceptronų, naudojančių bendrą požymių vektorių.

Dirbtinio neuroninio tinklo struktūros su daugiau nei vienu paslėptu sluoksniu nebuvo pasirinktos dėl šių priežasčių:

- Yra įrodyta, kad bet kokia funkcija, kuri gali būti paskaičiuota naudojant daugiasluoksnį perceptroną su daugeliu paslėptų sluoksnių taip pat gali būti paskaičiuota su daugiasluoksnio perceptronu, turinčiu tik vieną paslėptą sluoksnį, jeigu šiame sluoksnyje yra pakankamas paslėptų neuronų skaičius;

- Tyrimai parodė, kad neuroniniams tinklams su daugeliu paslėptų neuronų sluoksnių mokymo laikas smarkiai padidėja (M. Filipovič, 2003).



14 pav. „Trijų sluoksnių daugiasluoksnis perceptronas“

Kaip matome pateiktame 14 paveiksle, perceptrono įėjimo sluoksnį darbe sudaro 10 požymių vektorių, paslėptąjį sluoksnį taip pat sudaro 10 neuronų, o išėjimo sluoksnį sudaro vienas išėjimas.

3. BALSŲ IR PUSBALSŲ DISKRIMINAVIMO, NAUDOJANT SPEKTRINĘ ANALIZĘ, TYRIMAS

Šiame skyriuje eksperimentais bus siekiama rasti išsiskiriančius balsų ir pusbalsų požymius spektrinės analizės pagalba, t.y. nustatyti, kurioje dažnių juostoje kuri stacionarioji dalis duos geriausius rezultatus balsų ir pusbalsų atpažinime, neuroninio tinklo pagalba.

Buvo atlikti 24 eksperimentai. Eksperimentas atliekamas „*Matlab 6.5*“ programiniu paketu.

Duomenų paruošimas

Signalas filtruojamas juostinių filtrų pagalba, kurių yra 80 imant 100 Hz pralaidumo juostas, naudojantis Lietuvių kalbos LTDIGITS garsyno duomenimis, imamos 60 diktorių – 30 moterų ir 30 vyrų jau atfiltruotos fonemos t.y. balsiai ir pusbalsiai. Kiekvieno filtro išėjime signalas buvo suskaidytas į 10ms ilgio langus, su 5ms žingsniu, taip gaunant efektyvesnius duomenis energijos apskaičiavimui. Apskaičiavus energiją atliekamas akustinis modeliavimas, kuriame išskiriamos trys dalys: kairysis kontekstas, stacionarioji dalis ir dešinysis kontekstas; šių dalių ilgis imtas 40ms. Dirbtinis neuroninis tinklas yra apmokomas nuo 11 diktoriaus iki 30 diktoriaus.

Eksperimentas Nr.1

Duomenų paruošimas

Apmokymui:

Dažnis: 0 Hz – 1000 Hz

Diktoriai: moterys nuo 011 iki 030
vyrai nuo 011 iki 030

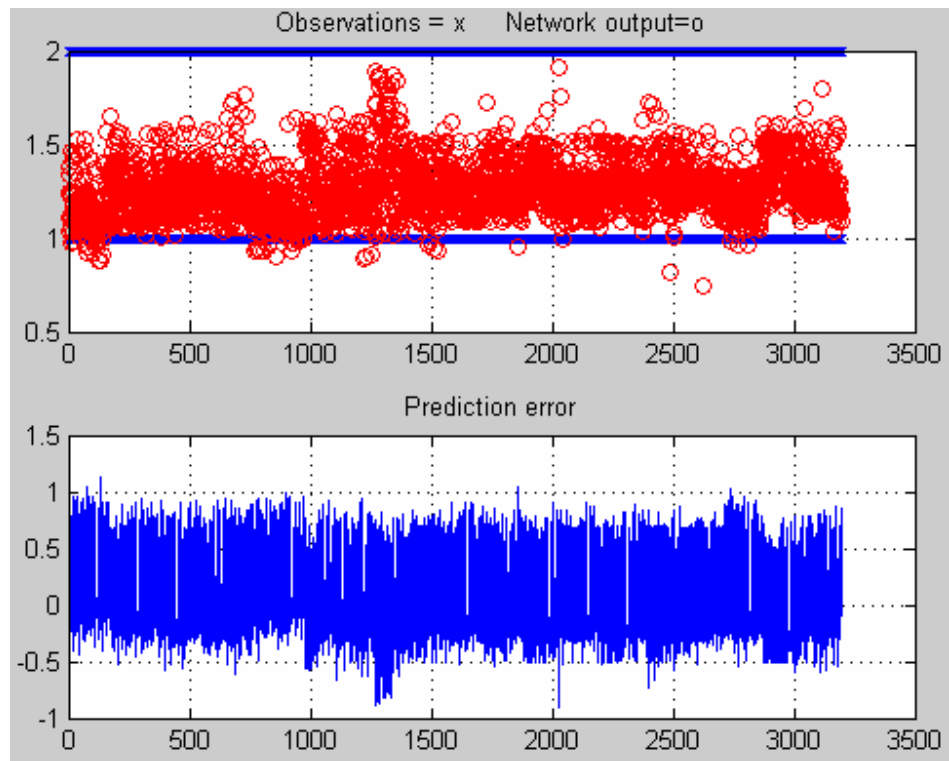
Fonemos fragmentas LC (kairysis)

Testavimui:

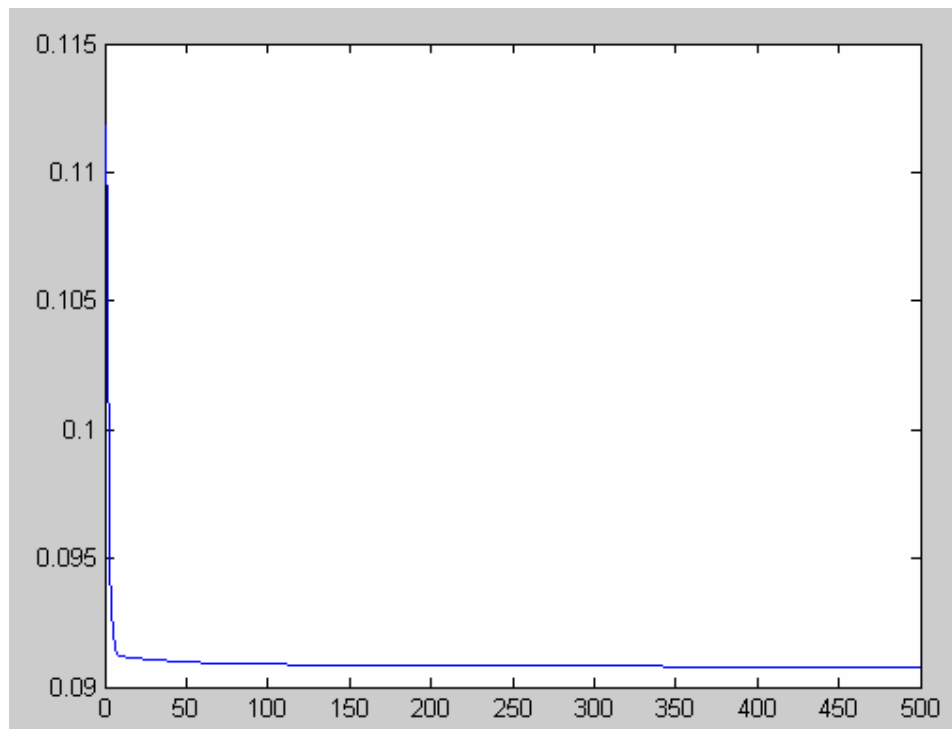
Dažnis: 0 Hz – 1000 Hz

Diktoriai: moterys nuo 001 iki 010
vyrai nuo 001 iki 010

Fonemos fragmentas LC (kairysis)



15 pav. Klasifikavimo rezultatai nuo 0 iki 1000Hz kairiojo fragmento



16 pav. Neuroninio tinklo apsimokymas nuo 0 iki 1000Hz kairiojo fragmento

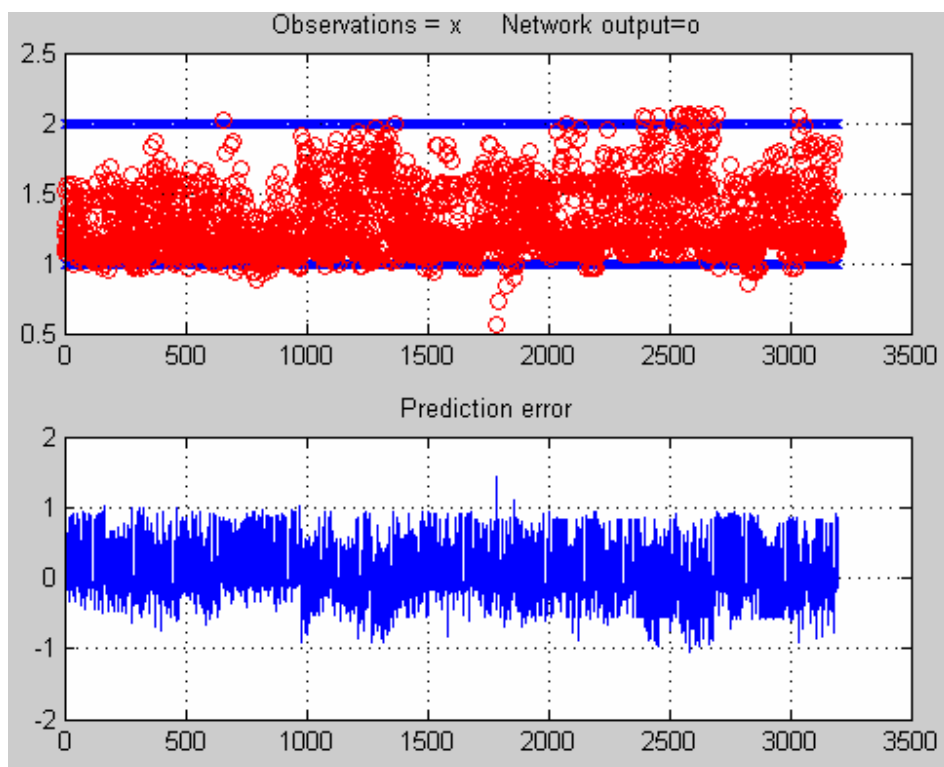
Rezultatas

Paėmus fonemos fragmentą LC nuo 0 iki 1000Hz, buvo gautas 73.08% atpažinimo tikslumas

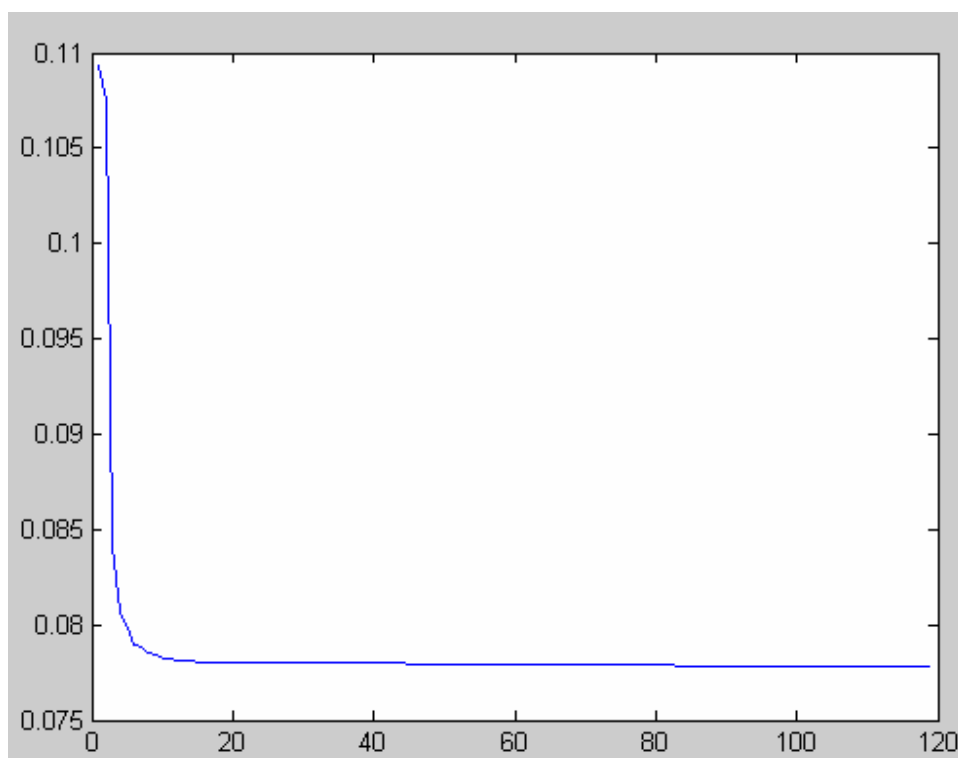
Eksperimentas Nr.2
Duomenų paruošimas

Apmokymui:
Dažnis: 0 Hz – 1000 Hz
Diktoriai: moterys nuo 011 iki 030
vyrai nuo 011 iki 030
Fonemos fragmentas SP
(stacionarioji - vidurinė)

Testavimui:
Dažnis: 0 Hz – 1000 Hz
Diktoriai: moterys nuo 001 iki 010
vyrai nuo 001 iki 010
Fonemos fragmentas SP
(stacionarioji - vidurinė kairysis)



17 pav. Klasifikavimo rezultatai nuo 0 iki 1000Hz stacionaraus fragmento



18 pav. Neuroninio tinklo apsimokymas nuo 0 iki 1000Hz stacionaraus fragmento

Rezultatas

Paėmus fonemos fragmentą SP nuo 0 iki 1000Hz, buvo gautas 75.56% atpažinimo tikslumas

Eksperimentas Nr.3

Duomenų paruošimas

Apmokymui:

Dažnis: 0 Hz – 1000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrų nuo 011 iki 030

Fonemos fragmentas RC

(dešinysis)

Testavimui:

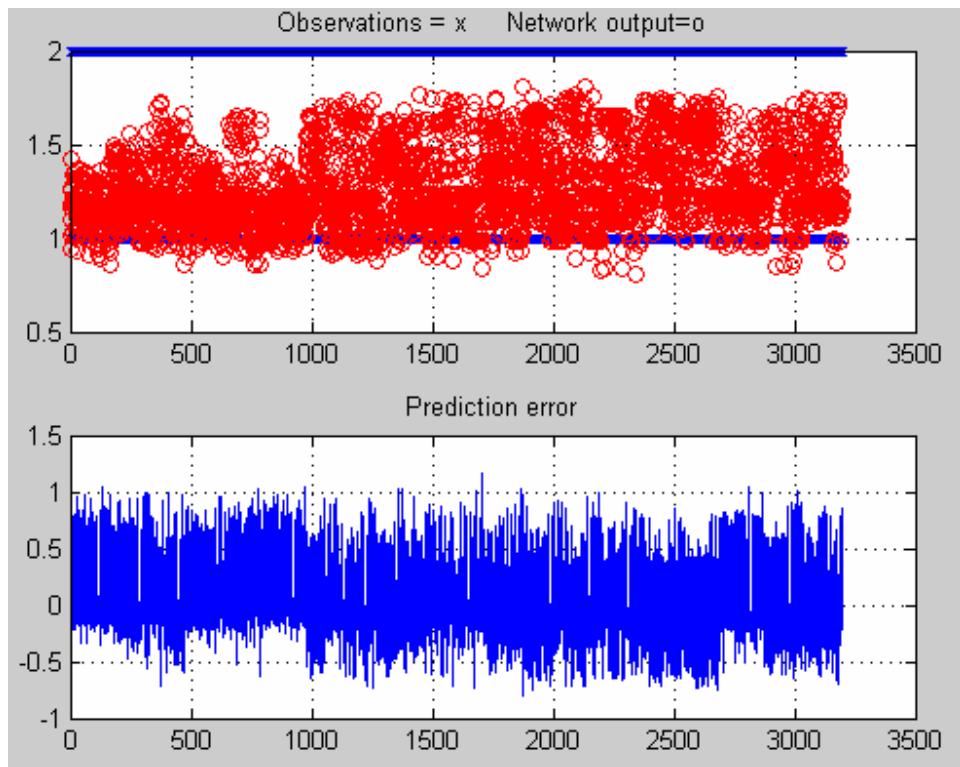
Dažnis: 0 Hz – 1000 Hz

Diktoriai: moterys nuo 001 iki 010

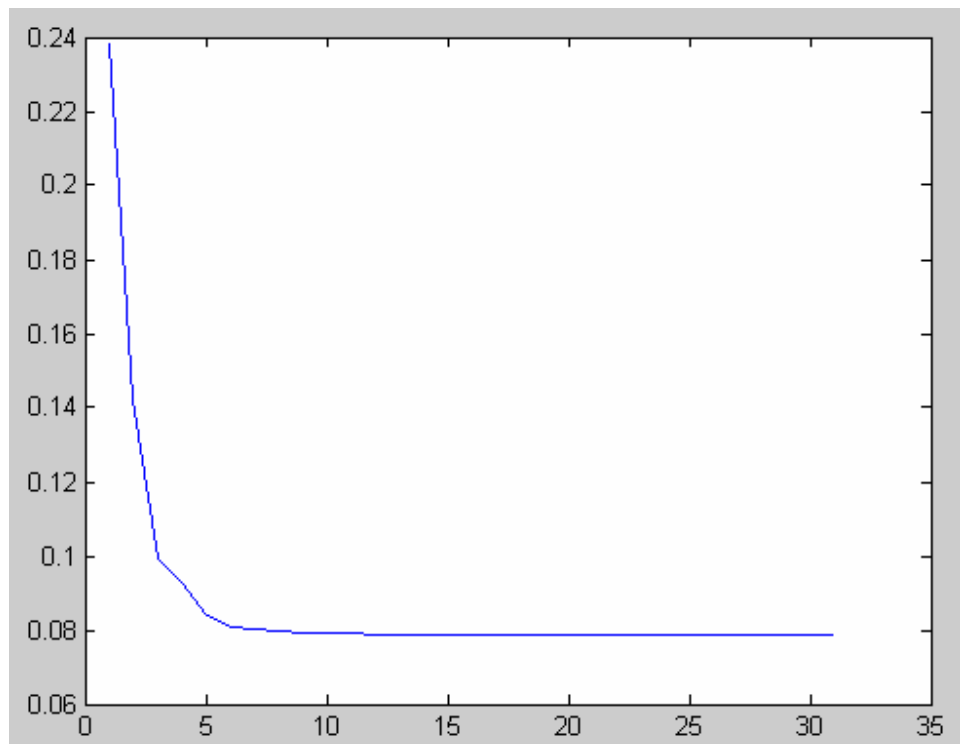
vyrų nuo 001 iki 010

Fonemos fragmentas RC

(dešinysis)



19 pav. Klasifikavimo rezultatai nuo 0 iki 1000Hz dešinio fragmento



20 pav. Neuroninio tinklo apsimokymas nuo 0 iki 1000Hz dešinio fragmento

Rezultatas

Paėmus fonemos fragmentą RC nuo 0 iki 1000Hz, buvo gautas 75.31% atpažinimo tikslumas

Ekspertas Nr.4

Duomenų paruošimas

Apmokymui:

Dažnis: 1000 Hz – 2000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrai nuo 011 iki 030

Fonemos fragmentas LC (kairysis)

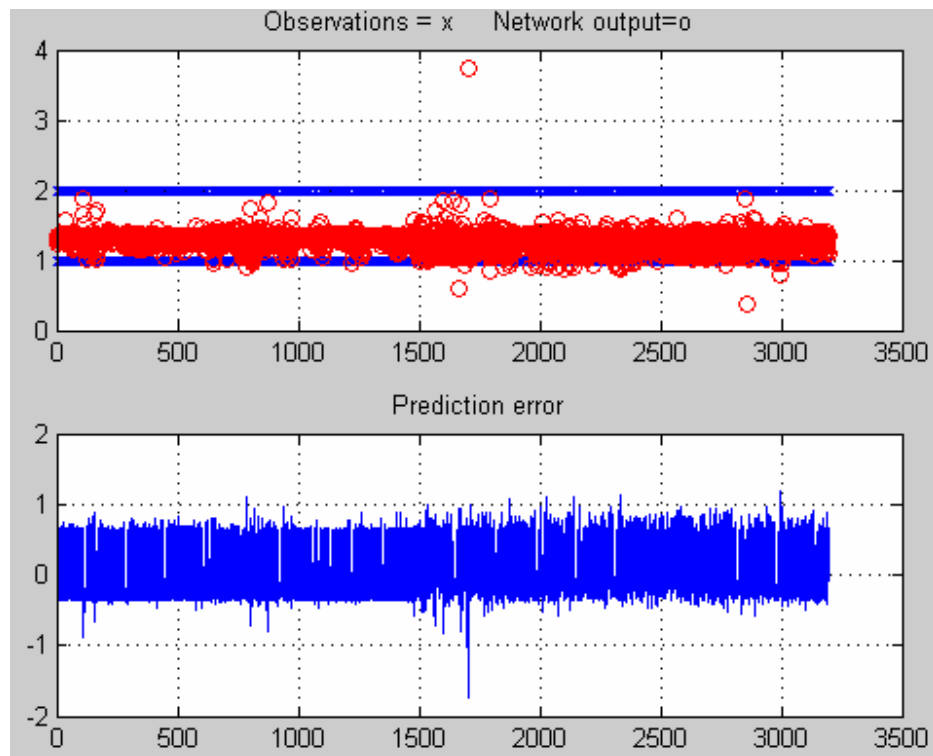
Testavimui:

Dažnis: 1000 Hz – 2000 Hz

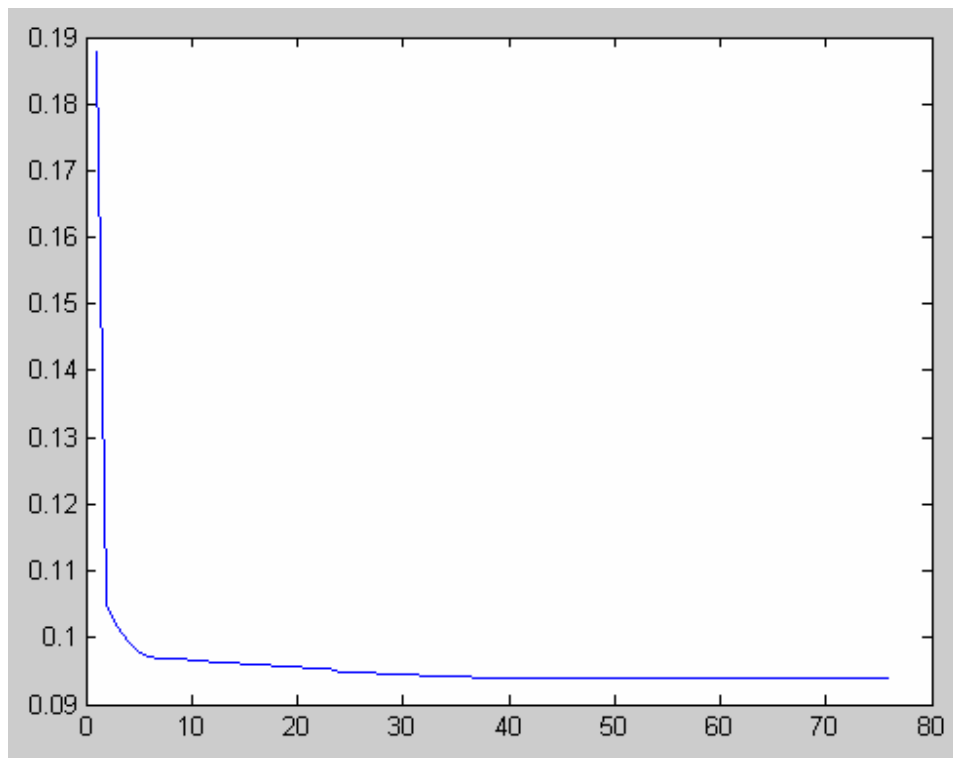
Diktoriai: moterys nuo 001 iki 010

vyrai nuo 001 iki 010

Fonemos fragmentas LC (kairysis)



21 pav. Klasifikavimo rezultatai nuo 1000 iki 2000Hz kairiojo fragmento



22 pav. Neuroninio tinklo apsimokymas nuo 1000 iki 2000Hz kairiojo fragmento

Rezultatas

Paėmus fonemos fragmentą LC nuo 1000 iki 2000Hz, buvo gautas 71.99% atpažinimo tikslumas

Eksperimentas Nr.5

Duomenų paruošimas

Apmokymui:

Dažnis: 1000 Hz – 2000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrų nuo 011 iki 030

Fonemos fragmentas SP

(stacionarioji - vidurinė)

Testavimui:

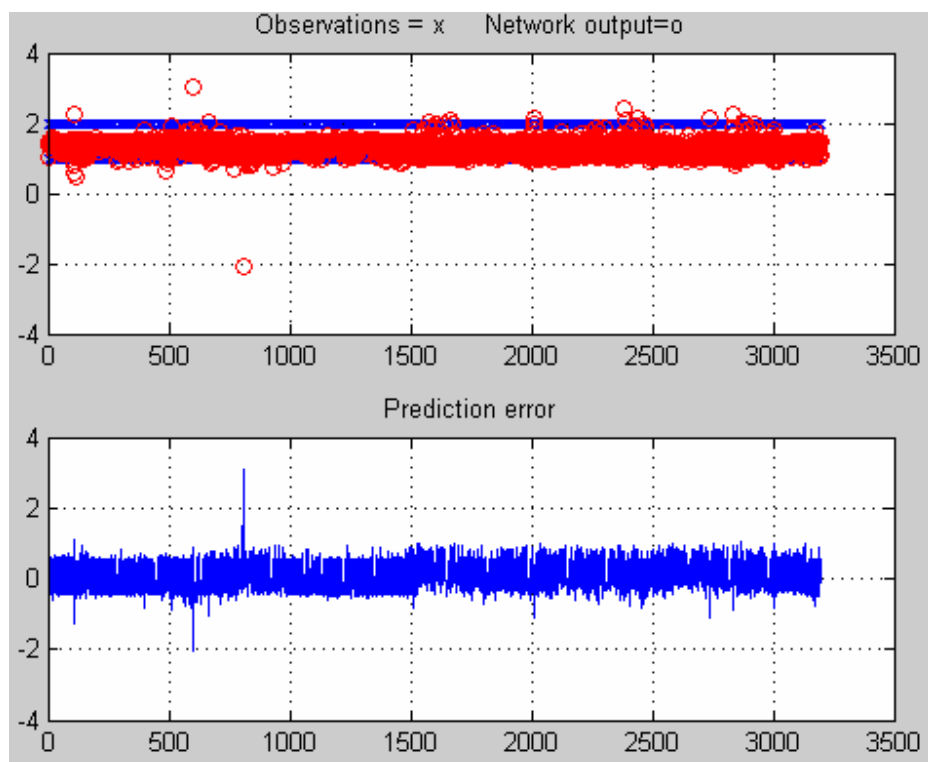
Dažnis: 1000 Hz – 2000 Hz

Diktoriai: moterys nuo 001 iki 010

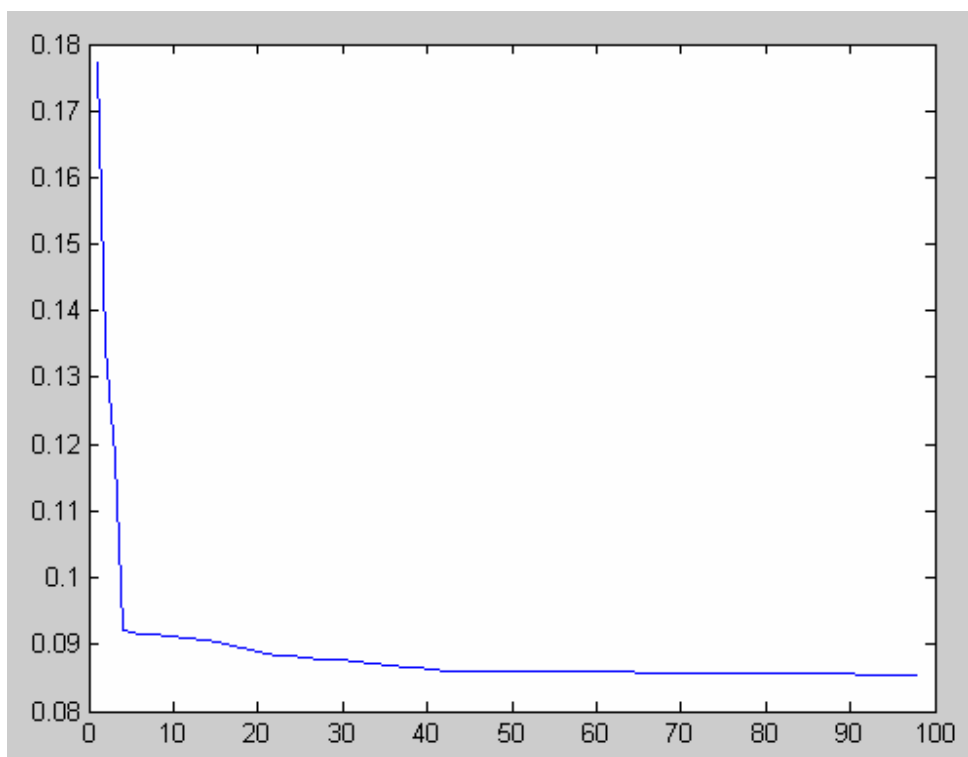
vyrų nuo 001 iki 010

Fonemos fragmentas SP

(stacionarioji - vidurinė kairysis)



23 pav. Klasifikavimo rezultatai nuo 1000 iki 2000Hz stacionaraus fragmento



24 pav. Neuroninio tinklo apsimokymas nuo 1000 iki 2000Hz stacionaraus fragmento

Rezultatas

Paėmus fonemos fragmentą SP nuo 1000 iki 2000Hz, buvo gautas 72.08% atpažinimo tikslumas.

Eksperimentas Nr.6

Duomenų paruošimas

Apmokymui:

Dažnis: 1000 Hz – 2000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrai nuo 011 iki 030

Fonemos fragmentas RC

(dešinysis)

Testavimui:

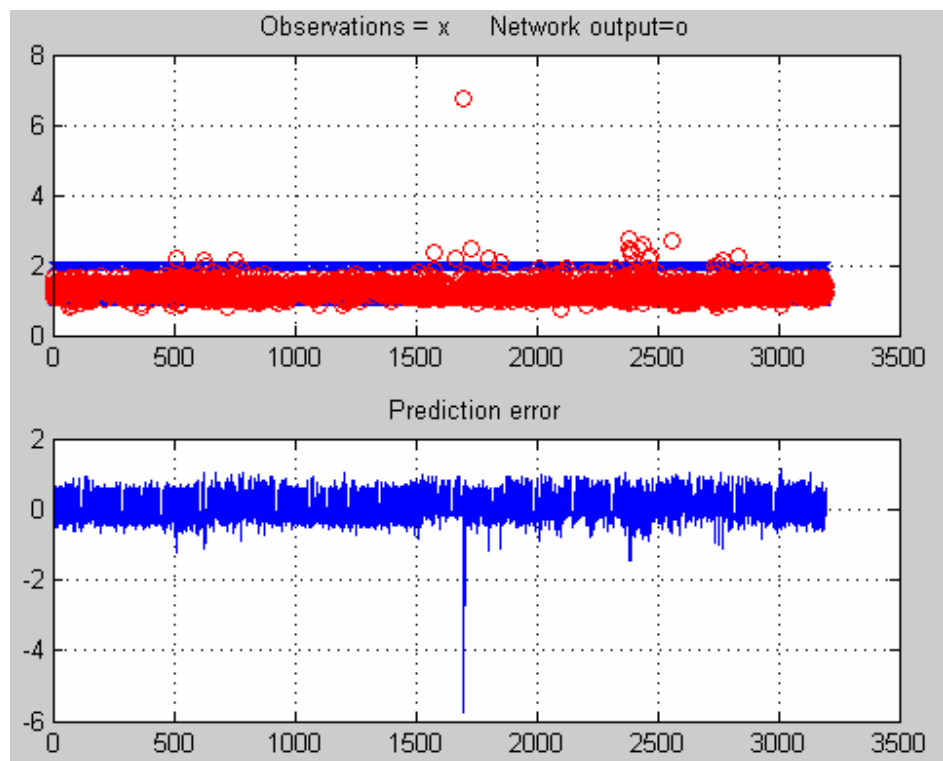
Dažnis: 1000 Hz – 2000 Hz

Diktoriai: moterys nuo 001 iki 010

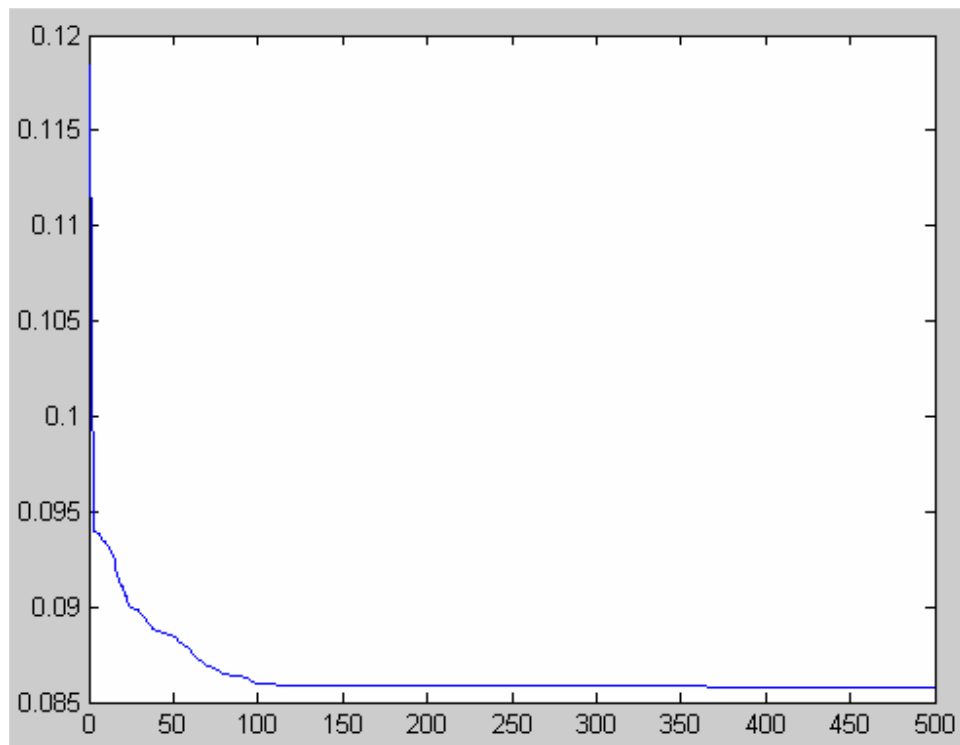
vyrai nuo 001 iki 010

Fonemos fragmentas RC

(dešinysis)



25 pav. Klasifikavimo rezultatai nuo 1000 iki 2000Hz dešinio fragmento



26 pav. Neuroninio tinklo apsimokymas nuo 1000 iki 2000Hz dešinio fragmento

Rezultatas

Paėmus fonemos fragmentą RC nuo 1000 iki 2000Hz, buvo gautas 74.74% atpažinimo tikslumas.

Eksperimentas Nr.7

Duomenų paruošimas

Apmokymui:

Dažnis: 2000 Hz – 3000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrų nuo 011 iki 030

Fonemos fragmentas LC (kairysis)

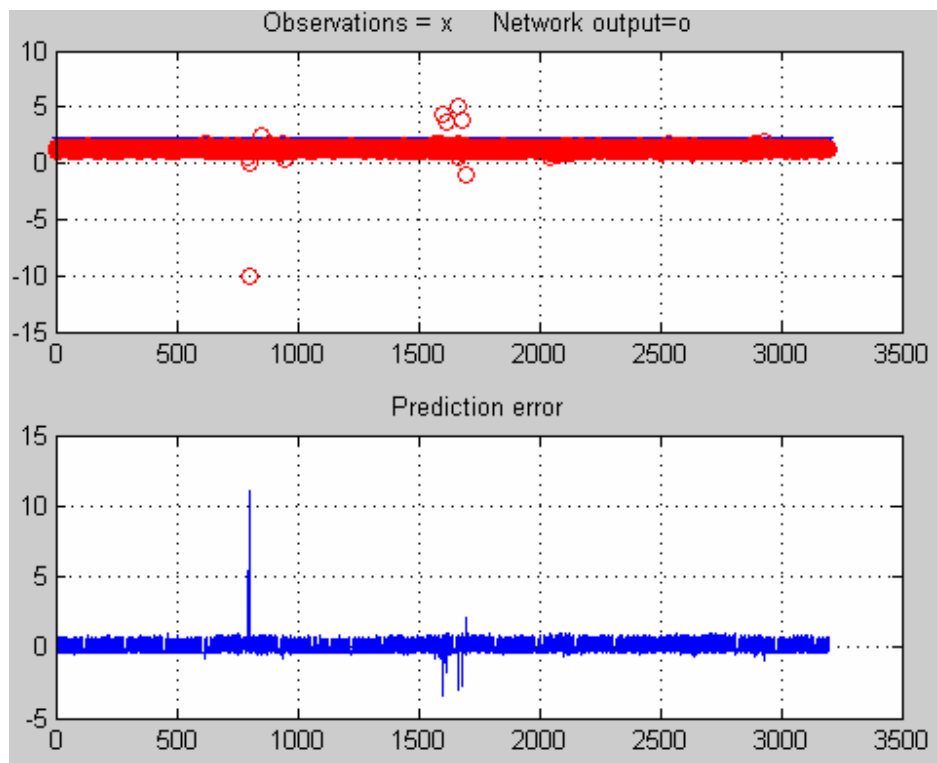
Testavimui:

Dažnis: 2000 Hz – 3000 Hz

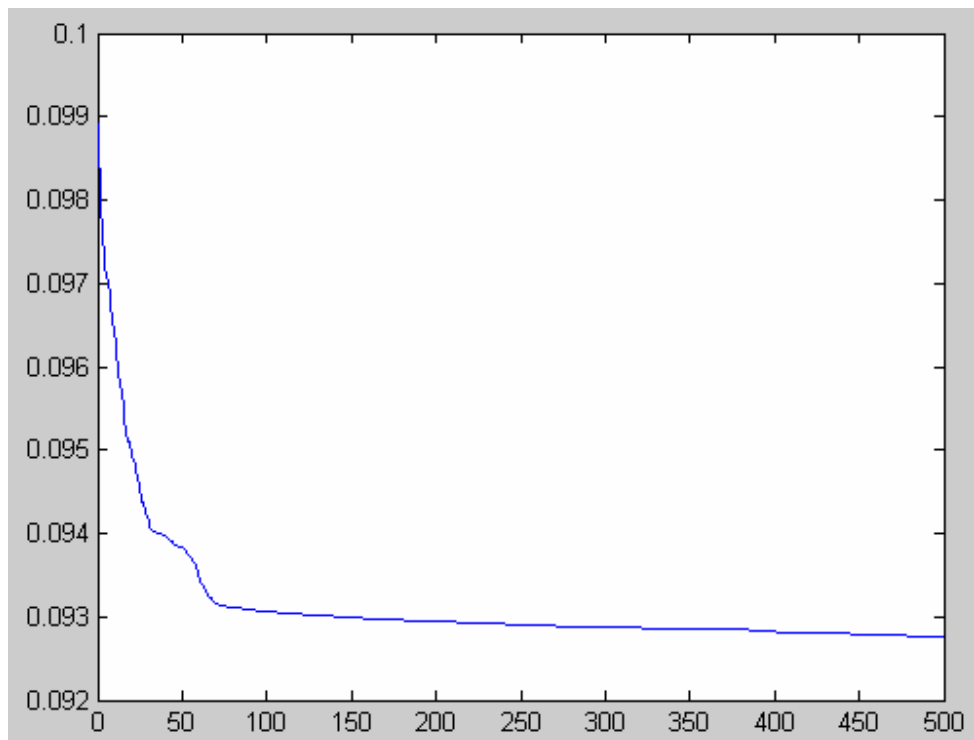
Diktoriai: moterys nuo 001 iki 010

vyrų nuo 001 iki 010

Fonemos fragmentas LC (kairysis)



27 pav. Klasifikavimo rezultatai nuo 2000 iki 3000Hz kairiojo fragmento



28 pav. Neuroninio tinklo apsimokymas nuo 2000 iki 3000Hz kairiojo fragmento

Rezultatas

Paėmus fonemos fragmentą LC nuo 2000 iki 3000Hz, buvo gautas 71.74% atpažinimo tikslumas.

Eksperimentas Nr.8

Duomenų paruošimas

Apmokymui:

Dažnis: 2000 Hz – 3000 Hz

Diktoriai: moterys nuo 011 iki 030
vyrai nuo 011 iki 030

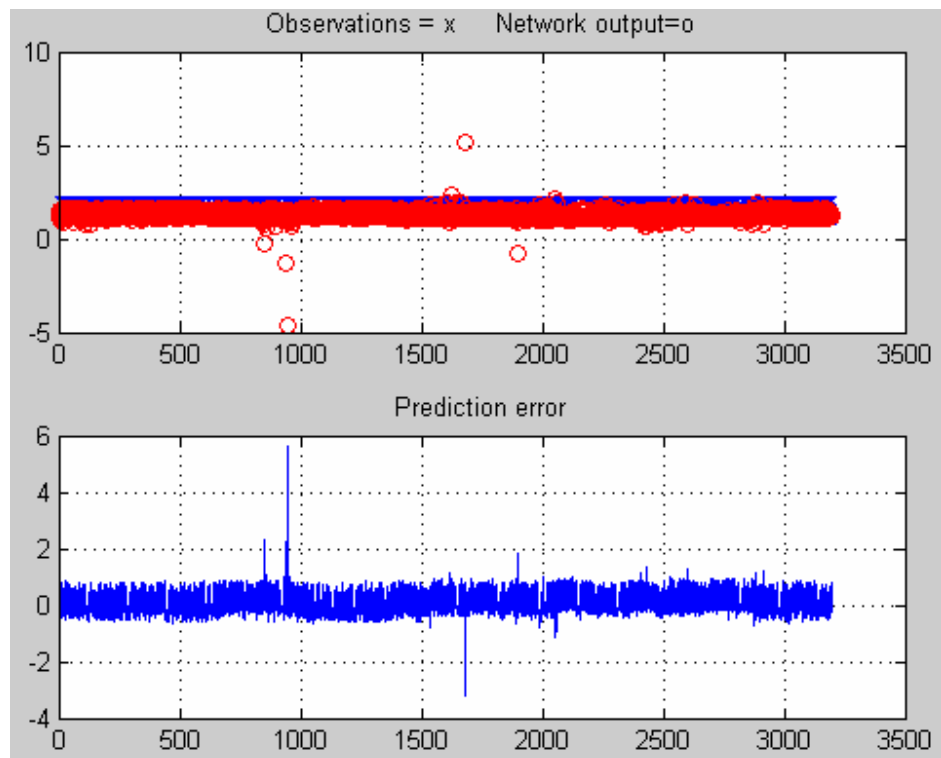
Fonemos fragmentas SP
(stacionarioji - vidurinė)

Testavimui:

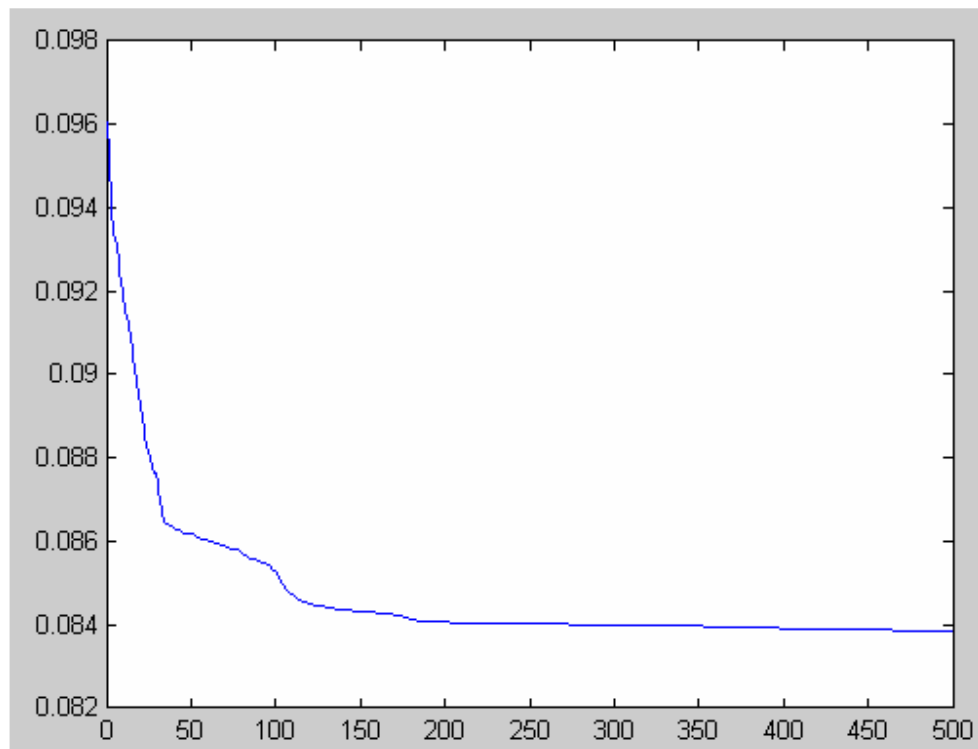
Dažnis: 2000 Hz – 3000 Hz

Diktoriai: moterys nuo 001 iki 010
vyrai nuo 001 iki 010

Fonemos fragmentas SP
(stacionarioji - vidurinė kairysis)



29 pav. Klasifikavimo rezultatai nuo 2000 iki 3000Hz stacionaraus fragmento



30 pav. Neuroninio tinklo apsimokymas nuo 2000 iki 3000Hz stacionaraus fragmento

Rezultatas

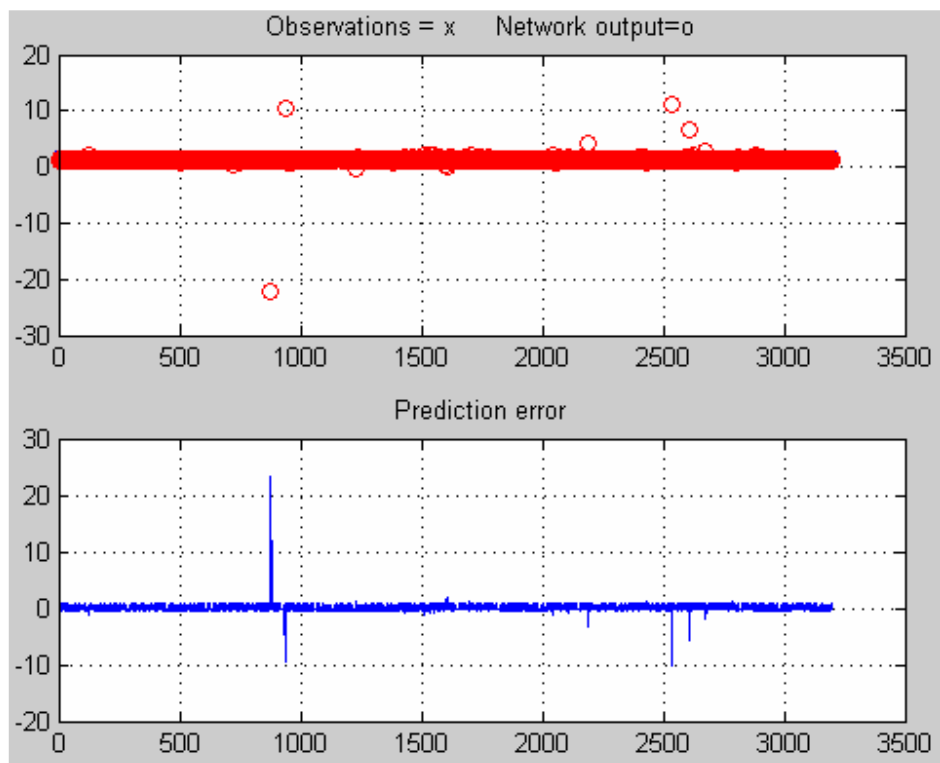
Paėmus fonemos fragmentą SP nuo 2000 iki 3000Hz, buvo gautas 74.90% atpažinimo tikslumas.

Eksperimentas Nr.9

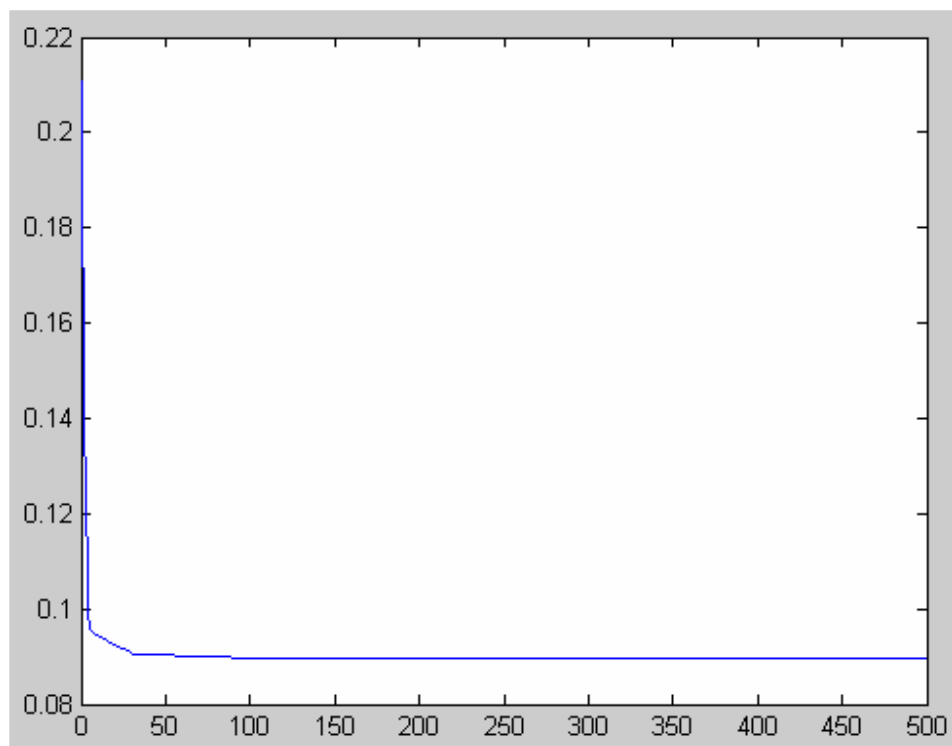
Duomenų paruošimas

Apmokymui:
 Dažnis: 2000 Hz – 3000 Hz
 Diktoriai: moterys nuo 011 iki 030
 vyrai nuo 011 iki 030
 Fonemos fragmentas RC
 (dešinysis)

Testavimui:
 Dažnis: 2000 Hz – 3000 Hz
 Diktoriai: moterys nuo 001 iki 010
 vyrai nuo 001 iki 010
 Fonemos fragmentas RC
 (dešinysis)



31 pav. Klasifikavimo rezultatai nuo 2000 iki 3000Hz dešinio fragmento



32 pav. Neuroninio tinklo apsimokymas nuo 2000 iki 3000Hz dešinio fragmento

Rezultatas

Paėmus fonemos fragmentą RC nuo 2000 iki 3000Hz, buvo gautas 72.24% atpažinimo tikslumas.

Ekspertas Nr.10

Duomenų paruošimas

Apmokymui:

Dažnis: 3000 Hz – 4000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrų nuo 011 iki 030

Fonemos fragmentas LC (kairysis)

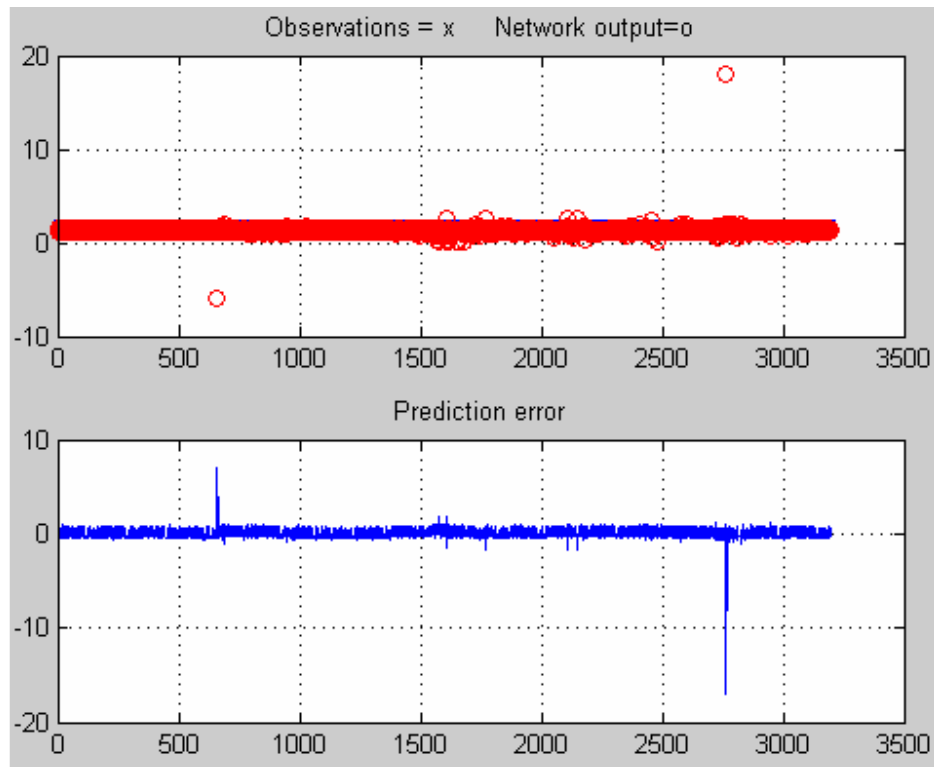
Testavimui:

Dažnis: 3000 Hz – 4000 Hz

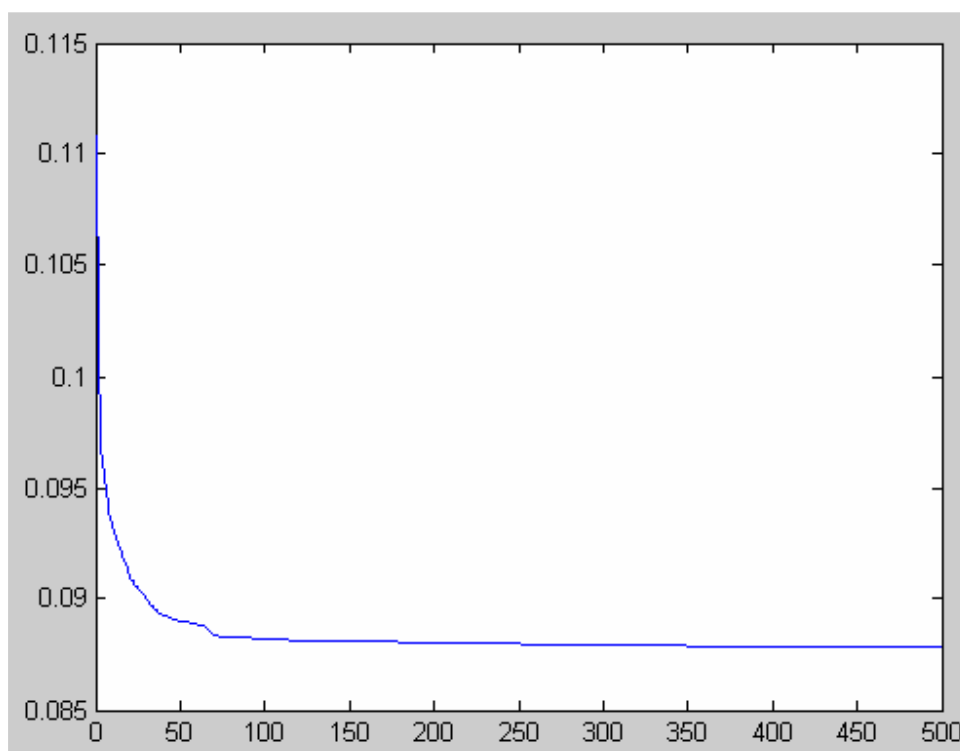
Diktoriai: moterys nuo 001 iki 010

vyrų nuo 001 iki 010

Fonemos fragmentas LC (kairysis)



33 pav. Klasifikavimo rezultatai nuo 3000 iki 4000Hz kairiojo fragmento



34 pav. Neuroninio tinklo apsimokymas nuo 3000 iki 4000Hz kairiojo fragmento

Rezultatai

Paėmus fonemos fragmentą LC nuo 3000 iki 4000Hz, buvo gautas 73.43% atpažinimo tikslumas.

Eksperimentas Nr.11

Duomenų paruošimas

Apmokymui:

Dažnis: 3000 Hz – 4000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrų nuo 011 iki 030

Fonemos fragmentas SP

(stacionarioji - vidurinė)

Testavimui:

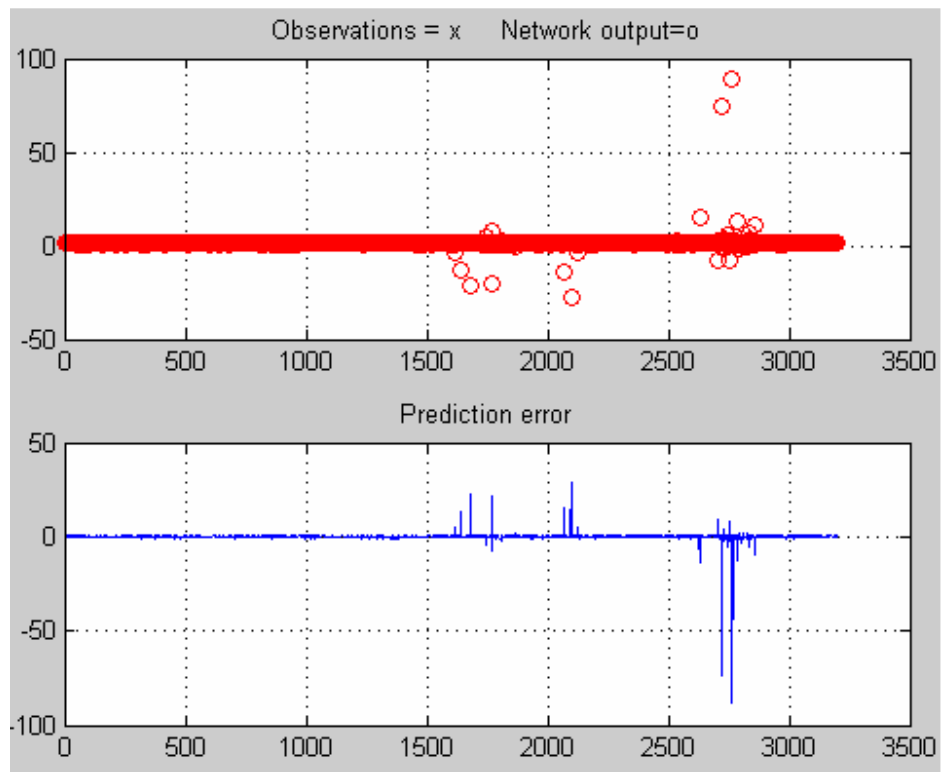
Dažnis: 3000 Hz – 4000 Hz

Diktoriai: moterys nuo 001 iki 010

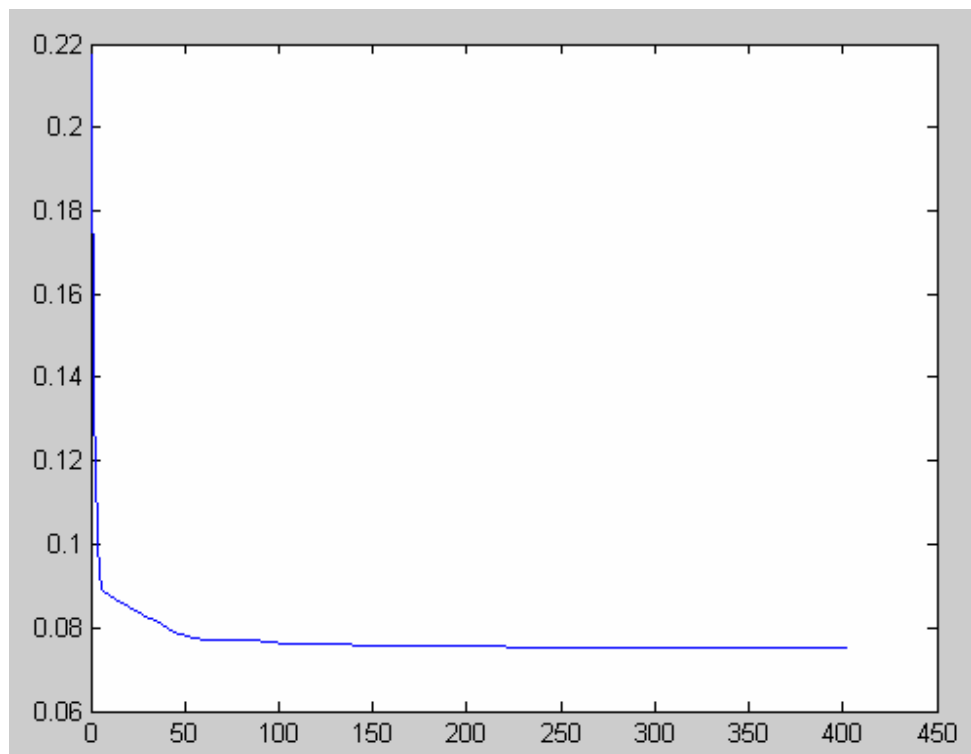
vyrų nuo 001 iki 010

Fonemos fragmentas SP

(stacionarioji - vidurinė kairysis)



35 pav. Klasifikavimo rezultatai nuo 3000 iki 4000Hz stacionaraus fragmento



36 pav. Neuroninio tinklo apsimokymas nuo 3000 iki 4000Hz stacionaraus fragmento

Rezultatas

Paėmus fonemos fragmentą SP nuo 3000 iki 4000Hz, buvo gautas 75.37% atpažinimo tikslumas.

Ekspertas Nr.12

Duomenų paruošimas

Apmokymui:

Dažnis: 3000 Hz – 4000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrai nuo 011 iki 030

Fonemos fragmentas RC

(dešinysis)

Testavimui:

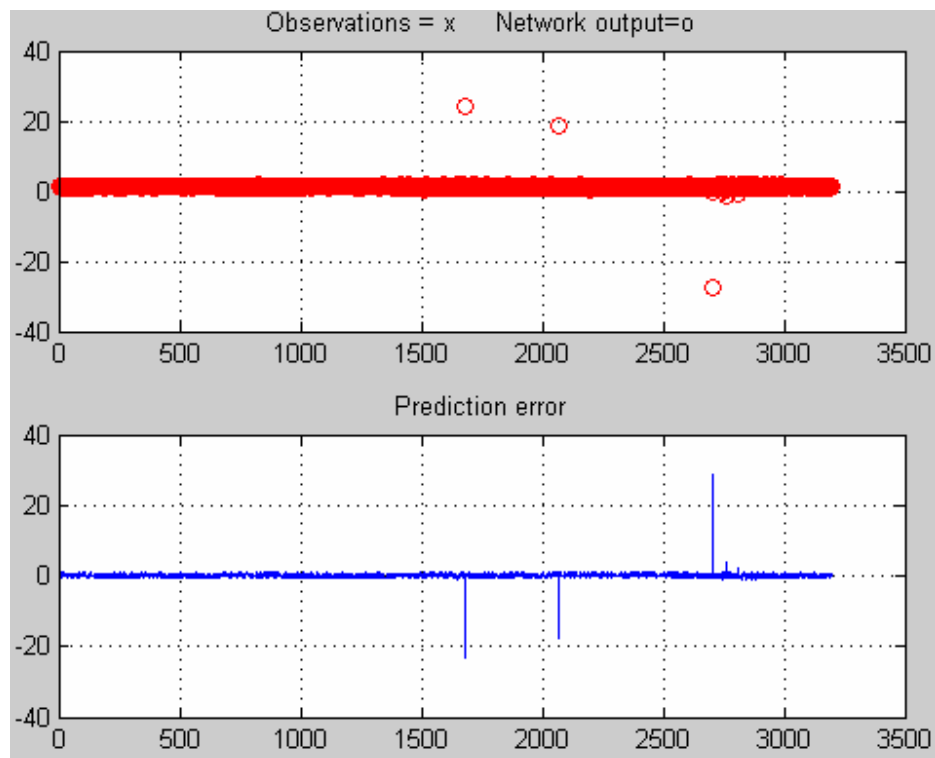
Dažnis: 3000 Hz – 4000 Hz

Diktoriai: moterys nuo 001 iki 010

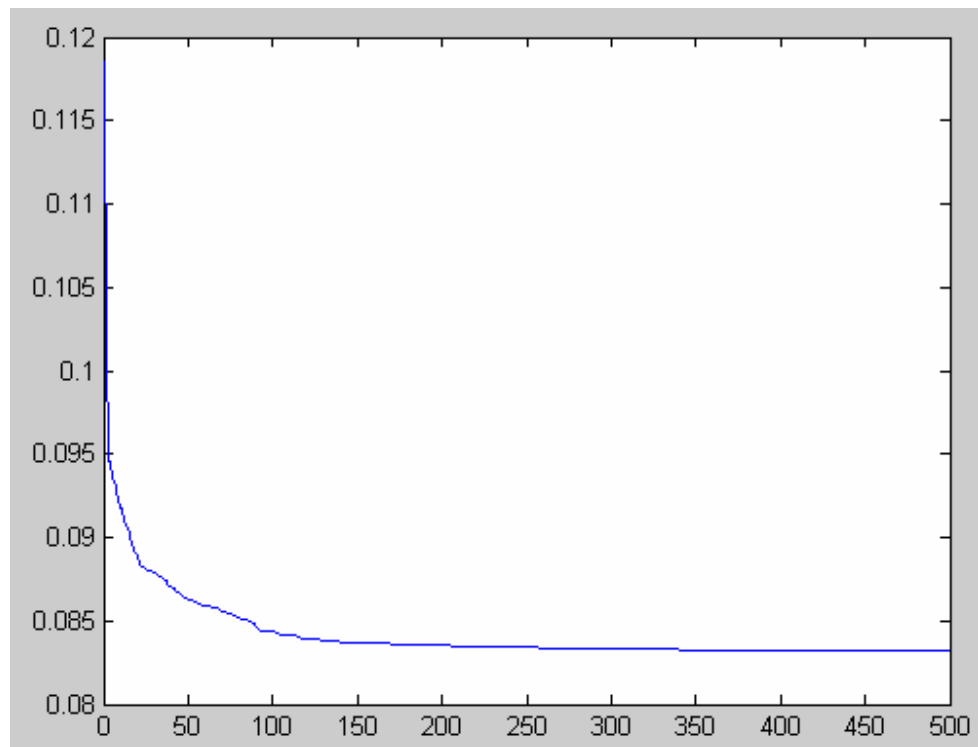
vyrai nuo 001 iki 010

Fonemos fragmentas RC

(dešinysis)



37 pav. Klasifikavimo rezultatai nuo 3000 iki 4000Hz dešinio fragmento



38 pav. Neuroninio tinklo apsimokymas nuo 3000 iki 4000Hz dešinio fragmento

Rezultatas

Paėmus fonemos fragmentą RC nuo 3000 iki 4000Hz, buvo gautas 74.74% atpažinimo tikslumas.

Ekspertas Nr.13

Duomenų paruošimas

Apmokymui:

Dažnis: 4000 Hz – 5000 Hz

Diktoriai: moterys nuo 011 iki 030
vyrai nuo 011 iki 030

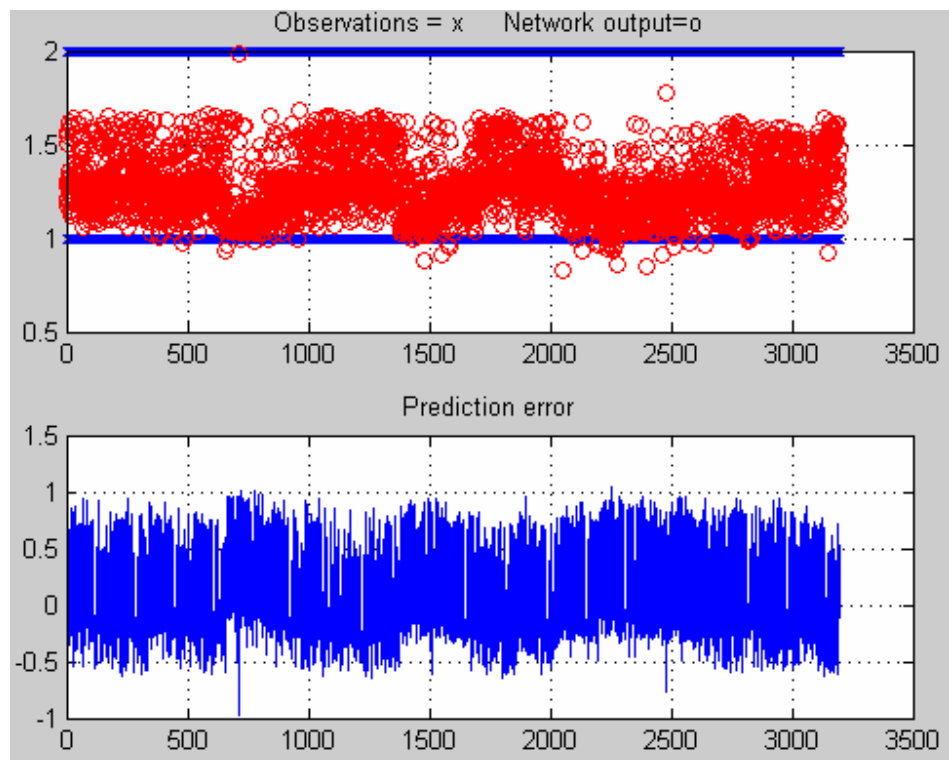
Fonemos fragmentas LC (kairysis)

Testavimui:

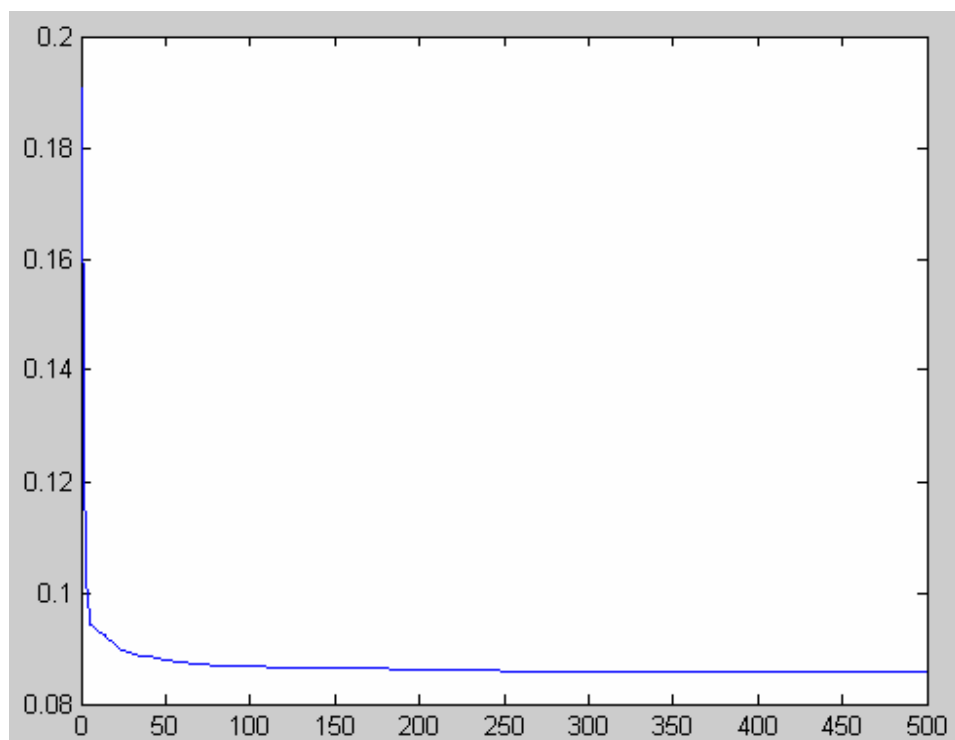
Dažnis: 4000 Hz – 5000 Hz

Diktoriai: moterys nuo 001 iki 010
vyrai nuo 001 iki 010

Fonemos fragmentas LC (kairysis)



39 pav. Klasifikavimo rezultatai nuo 4000 iki 5000Hz kairiojo fragmento



40 pav. Neuroninio tinklo apsimokymas nuo 4000 iki 5000Hz kairiojo fragmento

Rezultatas

Paėmus fonemos fragmentą LC nuo 4000 iki 5000Hz, buvo gautas 74.84% atpažinimo tikslumas.

Ekspertas Nr.14

Duomenų paruošimas

Apmokymui:

Dažnis: 4000 Hz – 5000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrai nuo 011 iki 030

Fonemos fragmentas SP

(stacionarioji - vidurinė)

Testavimui:

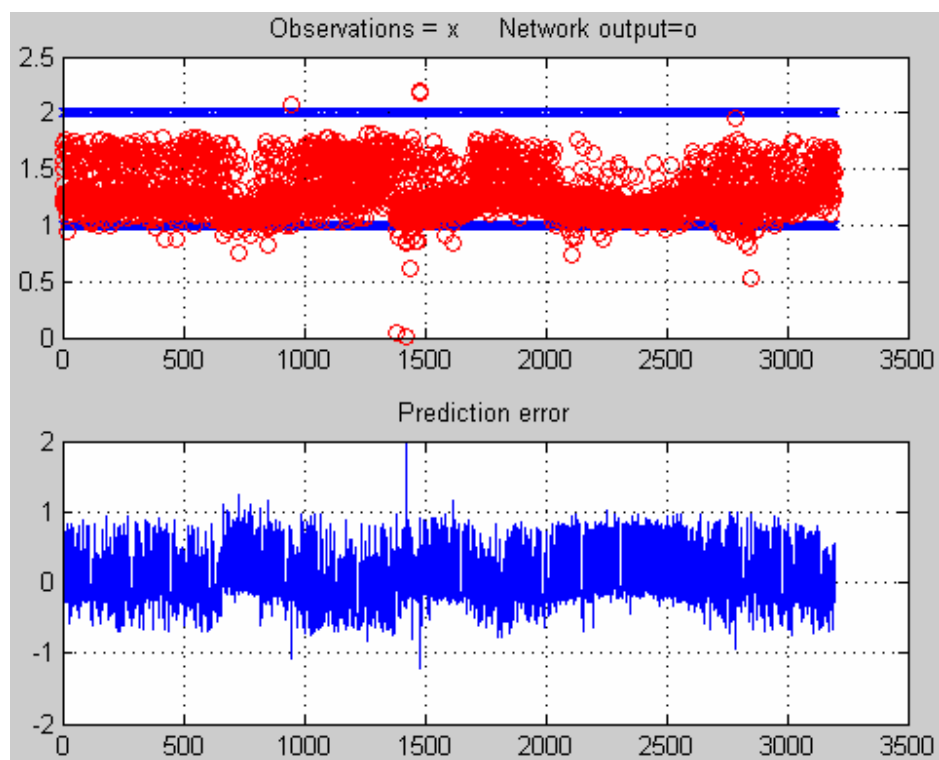
Dažnis: 4000 Hz – 5000 Hz

Diktoriai: moterys nuo 001 iki 010

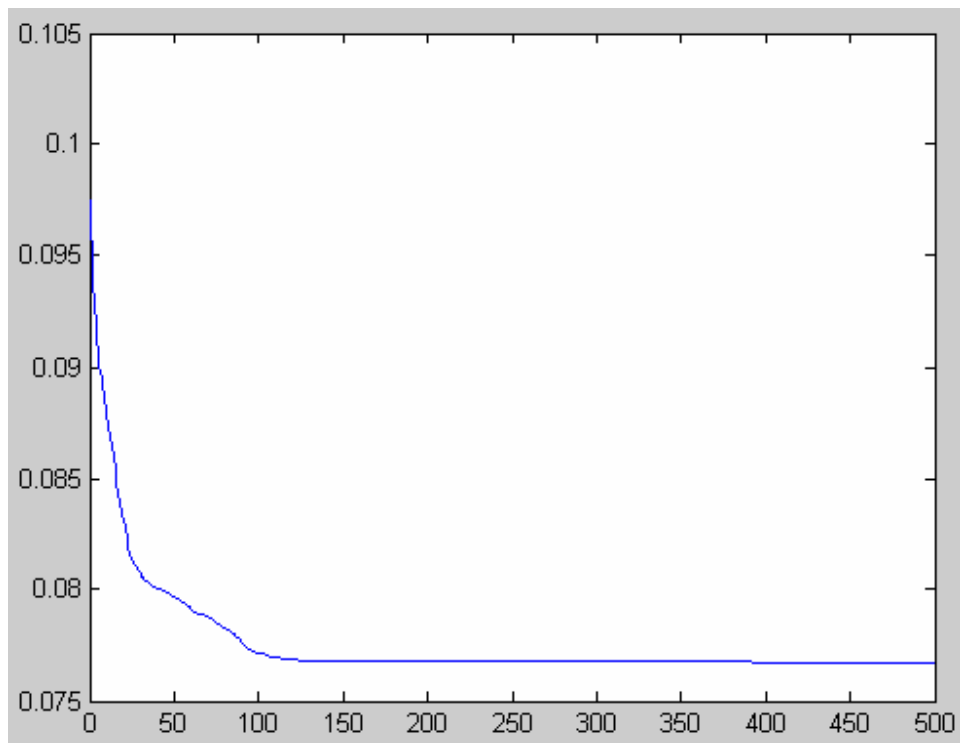
vyrai nuo 001 iki 010

Fonemos fragmentas SP

(stacionarioji - vidurinė kairysis)



41 pav. Klasifikavimo rezultatai nuo 4000 iki 5000Hz stacionaraus fragmento



42 pav. Neuroninio tinklo apsimokymas nuo 4000 iki 5000Hz stacionaraus fragmento

Rezultatas

Paėmus fonemos fragmentą SP nuo 4000 iki 5000Hz, buvo gautas 77.15% atpažinimo tikslumas.

Ekspertas Nr.15

Duomenų paruošimas

Apmokymui:

Dažnis: 4000 Hz – 5000 Hz

Diktoriai: moterys nuo 011 iki 030
vyrai nuo 011 iki 030

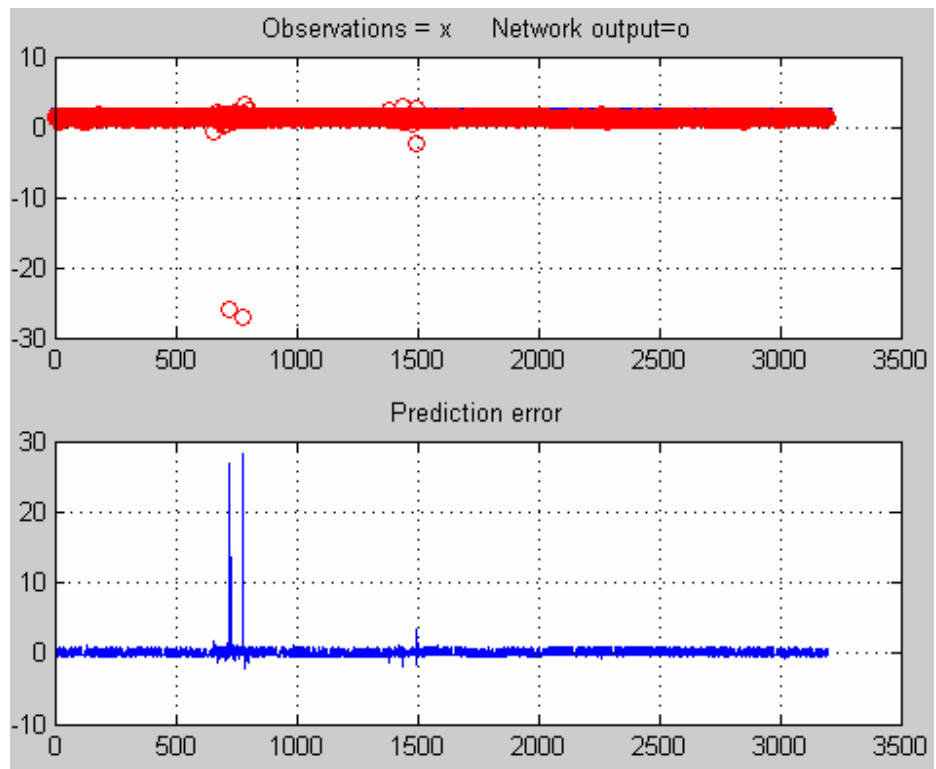
Fonemos fragmentas RC
(dešinysis)

Testavimui:

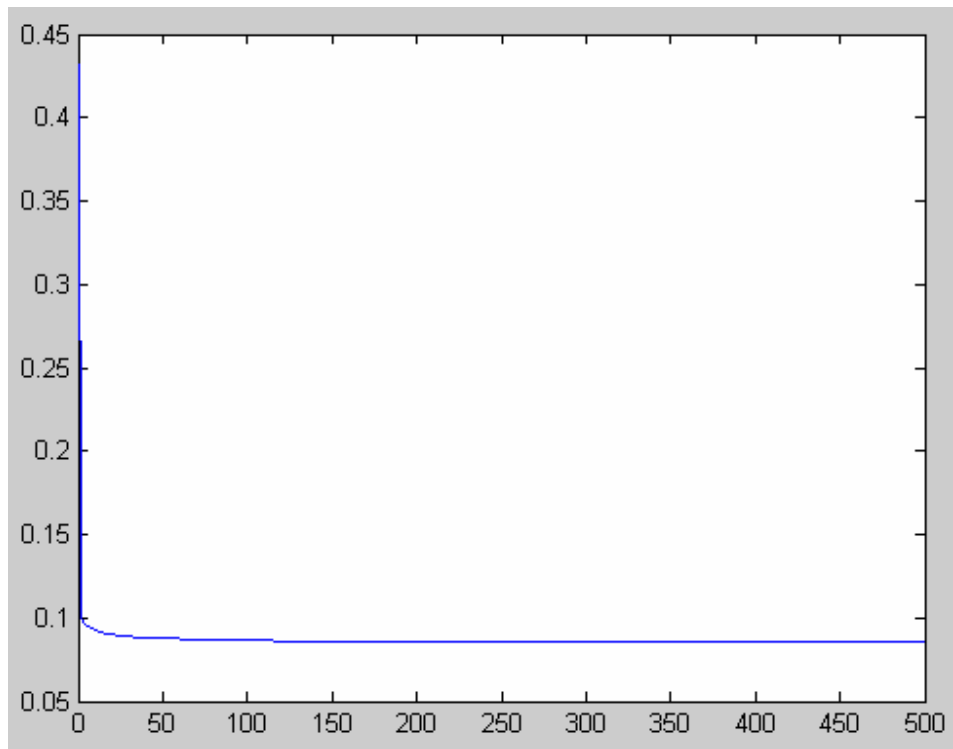
Dažnis: 4000 Hz – 5000 Hz

Diktoriai: moterys nuo 001 iki 010
vyrai nuo 001 iki 010

Fonemos fragmentas RC
(dešinysis)



43 pav. Klasifikavimo rezultatai nuo 4000 iki 5000Hz dešinio fragmento



44 pav. Neuroninio tinklo apsimokymas nuo 4000 iki 5000Hz dešinio fragmento

Rezultatas

Paėmus fonemos fragmentą RC nuo 4000 iki 5000Hz, buvo gautas 74.68% atpažinimo tikslumas.

Ekspertas Nr.16

Duomenų paruošimas

Apmokymui:

Dažnis: 5000 Hz – 6000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrai nuo 011 iki 030

Fonemos fragmentas LC (kairysis)

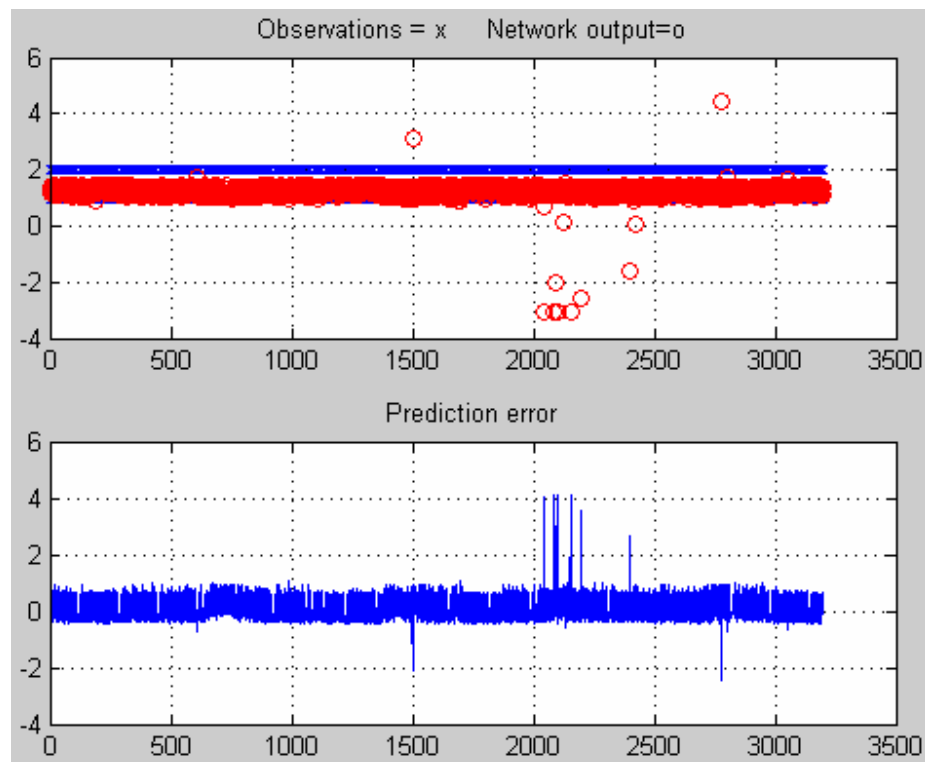
Testavimui:

Dažnis: 5000 Hz – 6000 Hz

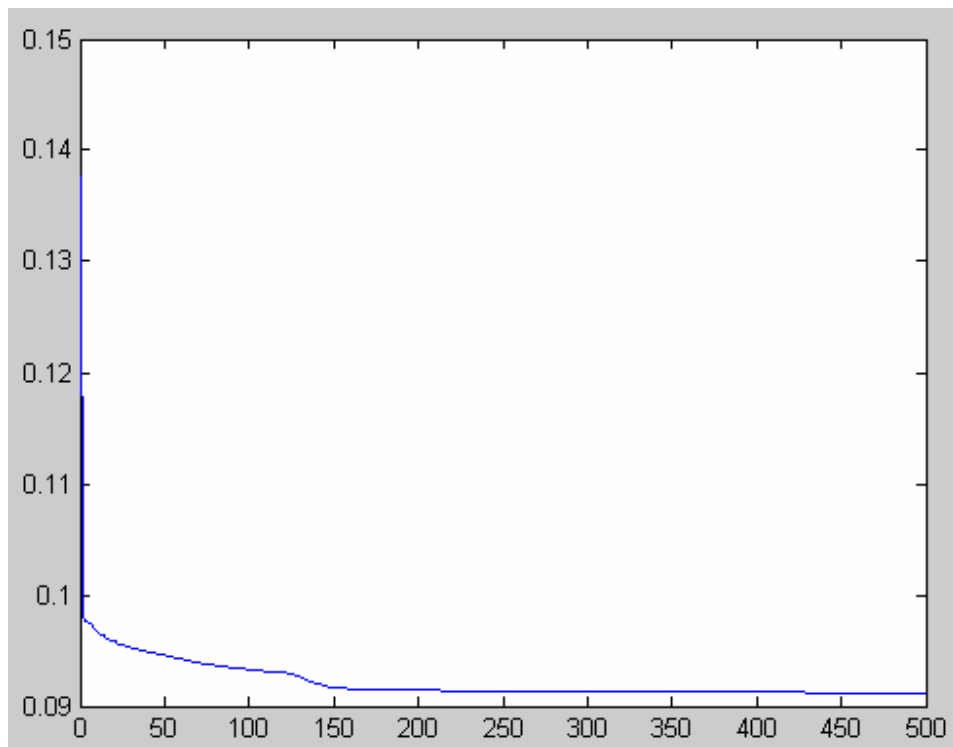
Diktoriai: moterys nuo 001 iki 010

vyrai nuo 001 iki 010

Fonemos fragmentas LC (kairysis)



45 pav. Klasifikavimo rezultatai nuo 5000 iki 6000Hz kairiojo fragmento



46 pav. Neuroninio tinklo apsimokymas nuo 5000 iki 6000Hz kairiojo fragmento

Rezultatas

Paėmus fonemos fragmentą LC nuo 5000 iki 6000Hz, buvo gautas 71.71% atpažinimo tikslumas.

Ekspertas Nr.17

Duomenų paruošimas

Apmokymui:

Dažnis: 5000 Hz – 6000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrų nuo 011 iki 030

Fonemos fragmentas SP

(stacionarioji - vidurinė)

Testavimui:

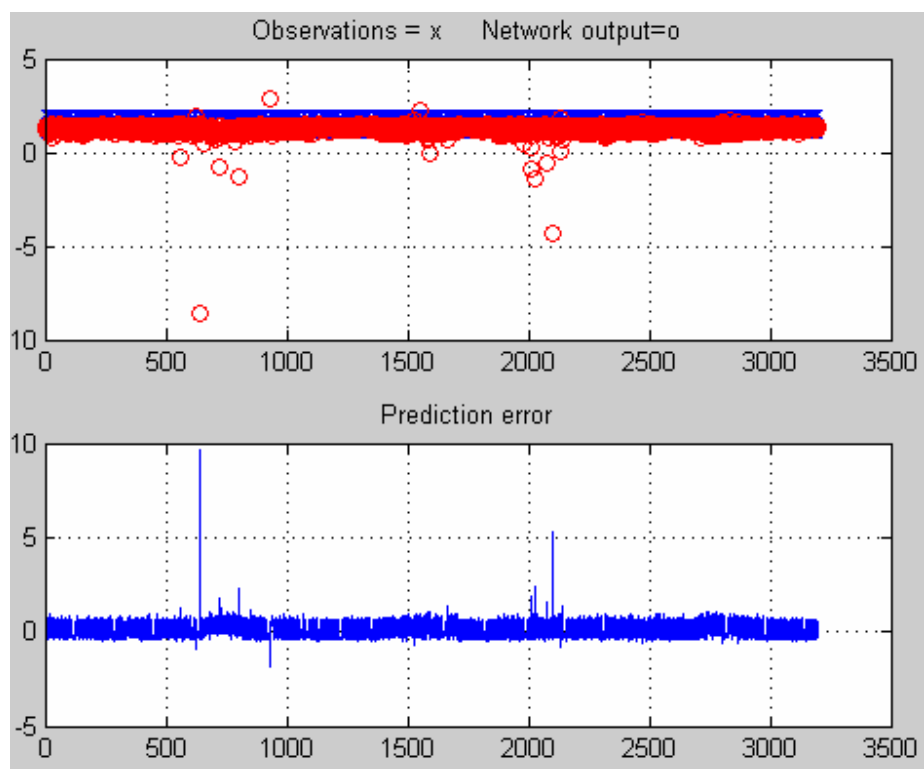
Dažnis: 5000 Hz – 6000 Hz

Diktoriai: moterys nuo 001 iki 010

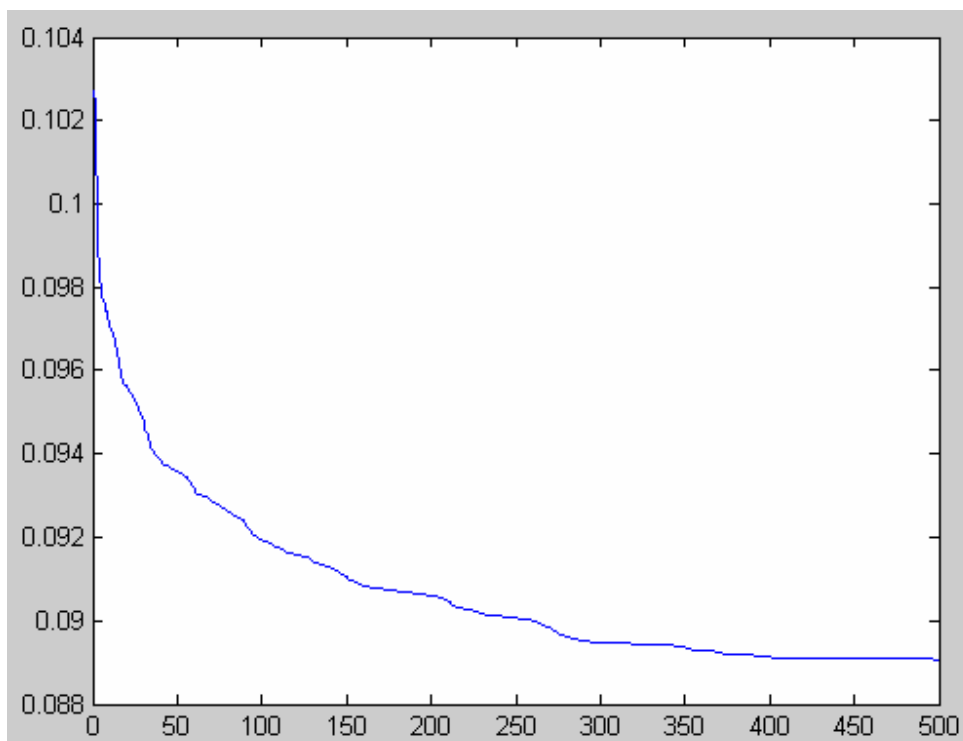
vyrų nuo 001 iki 010

Fonemos fragmentas SP

(stacionarioji - vidurinė kairysis)



47 pav. Klasifikavimo rezultatai nuo 5000 iki 6000Hz stacionaraus fragmento



48 pav. Neuroninio tinklo apsimokymas nuo 5000 iki 6000Hz stacionaraus fragmento

Rezultatas

Paėmus fonemos fragmentą SP nuo 5000 iki 6000Hz, buvo gautas 71.46% atpažinimo tikslumas.

Ekspertas Nr.18

Duomenų paruošimas

Apmokymui:

Dažnis: 5000 Hz – 6000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrai nuo 011 iki 030

Fonemos fragmentas RC

(dešinysis)

Testavimui:

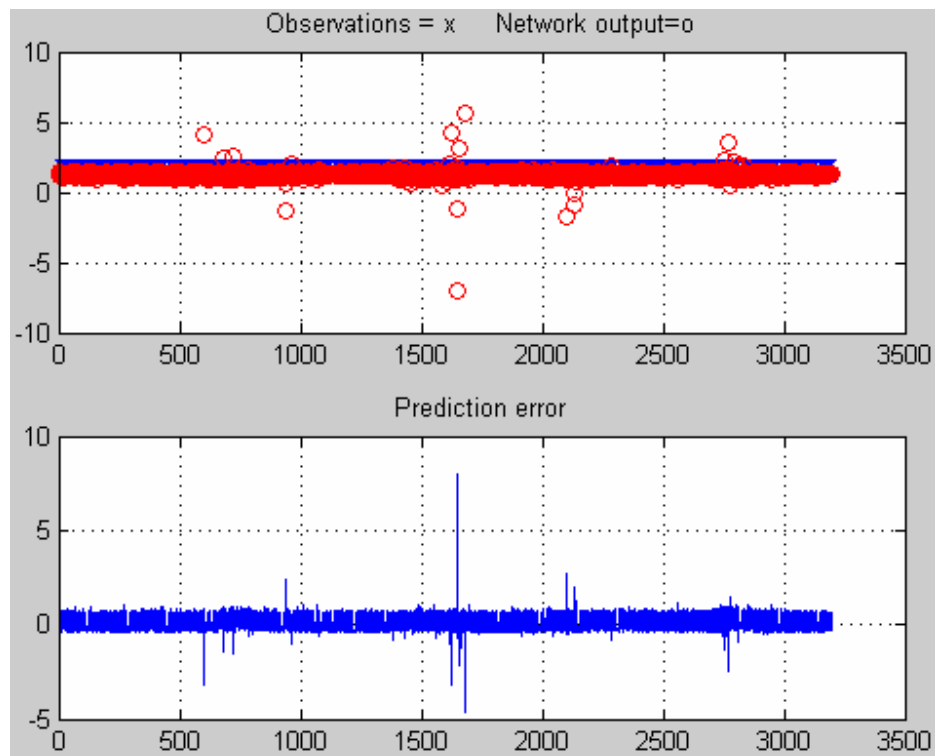
Dažnis: 5000 Hz – 6000 Hz

Diktoriai: moterys nuo 001 iki 010

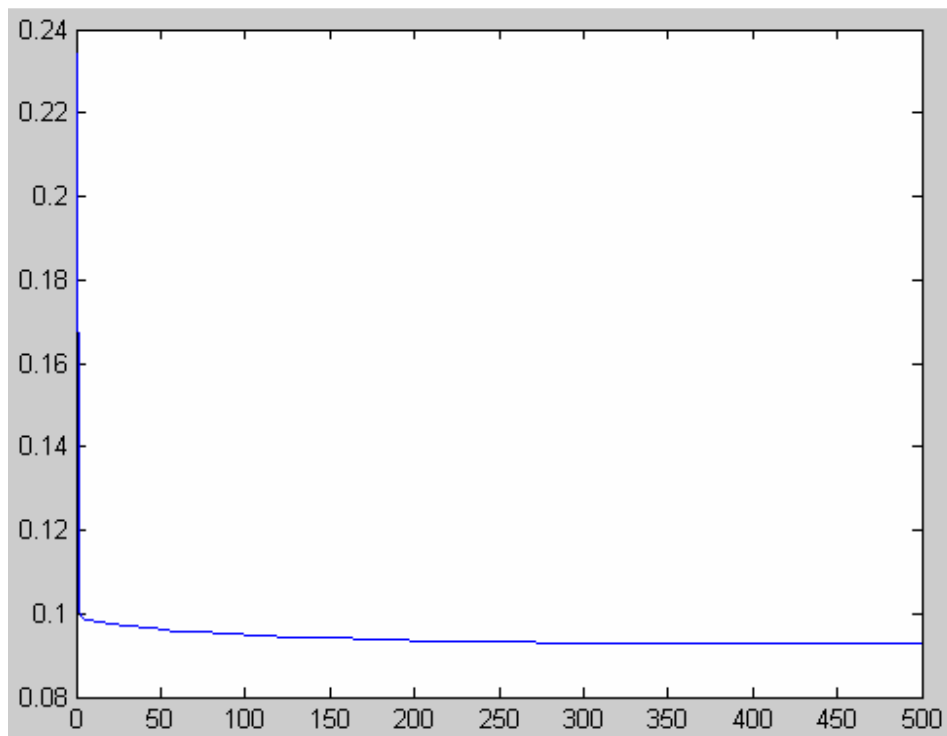
vyrai nuo 001 iki 010

Fonemos fragmentas RC

(dešinysis)



49 pav. Klasifikavimo rezultatai nuo 5000 iki 6000Hz dešinio fragmento



50 pav. Neuroninio tinklo apsimokymas nuo 5000 iki 6000Hz dešinio fragmento

Rezultatas

Paėmus fonemos fragmentą RC nuo 5000 iki 6000Hz, buvo gautas 71.24% atpažinimo tikslumas.

Ekspertas Nr.19

Duomenų paruošimas

Apmokymui:

Dažnis: 6000 Hz – 7000 Hz

Diktoriai: moterys nuo 011 iki 030
vyrai nuo 011 iki 030

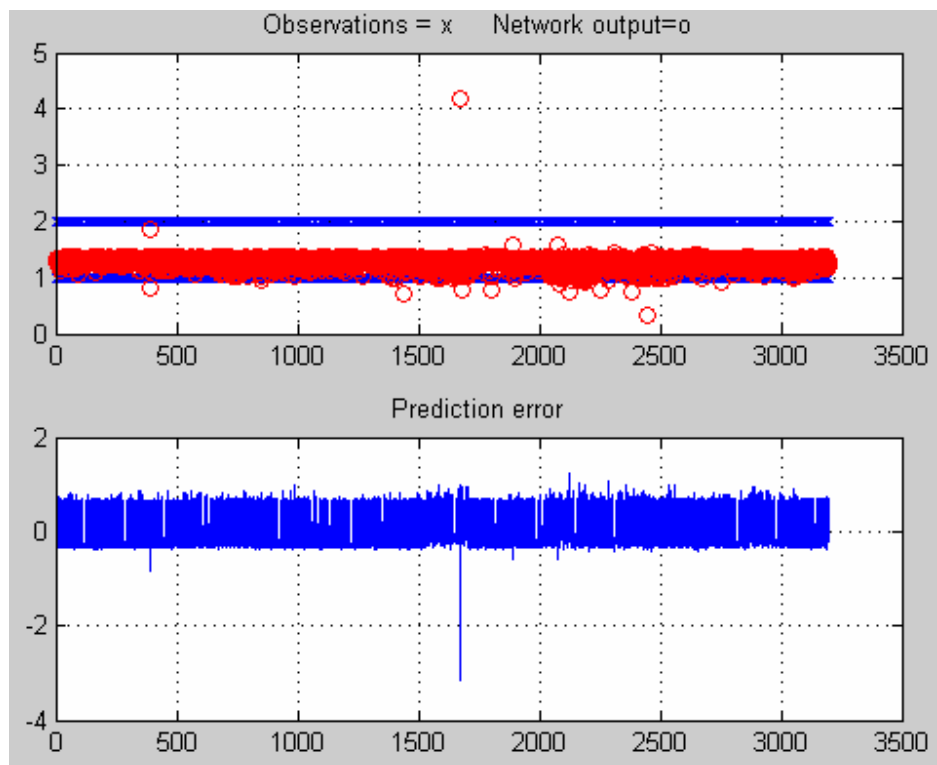
Fonemos fragmentas LC (kairysis)

Testavimui:

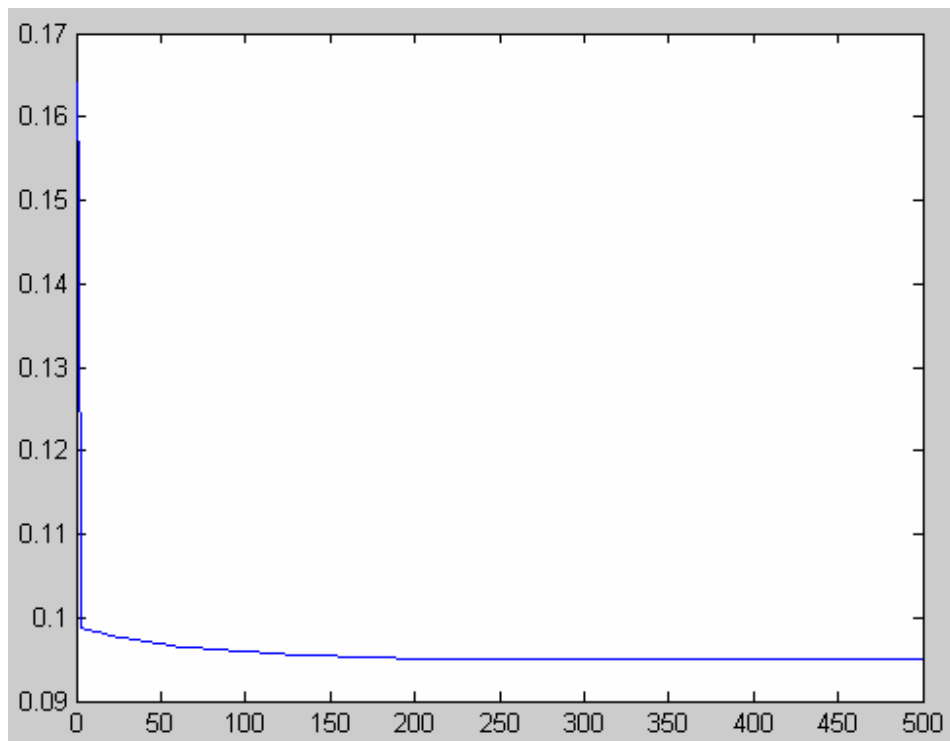
Dažnis: 6000 Hz – 7000 Hz

Diktoriai: moterys nuo 001 iki 010
vyrai nuo 001 iki 010

Fonemos fragmentas LC (kairysis)



51 pav. Klasifikavimo rezultatai nuo 6000 iki 7000Hz kairiojo fragmento



52 pav. Neuroninio tinklo apsimokymas nuo 6000 iki 7000Hz kairiojo fragmento

Rezultatas

Paėmus fonemos fragmentą LC nuo 6000 iki 7000Hz, buvo gautas 71.99% atpažinimo tikslumas.

Ekperimentas Nr.20

Duomenų paruošimas

Apmokymui:

Dažnis: 6000 Hz – 7000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrai nuo 011 iki 030

Fonemos fragmentas SP

(stacionarioji - vidurinė)

Testavimui:

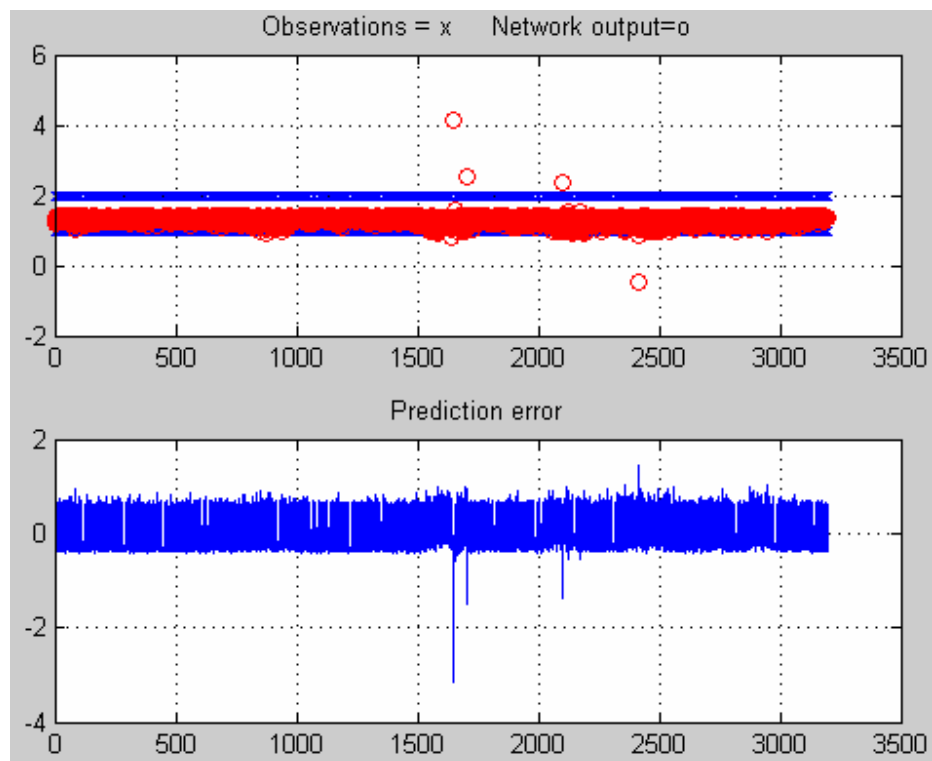
Dažnis: 6000 Hz – 7000 Hz

Diktoriai: moterys nuo 001 iki 010

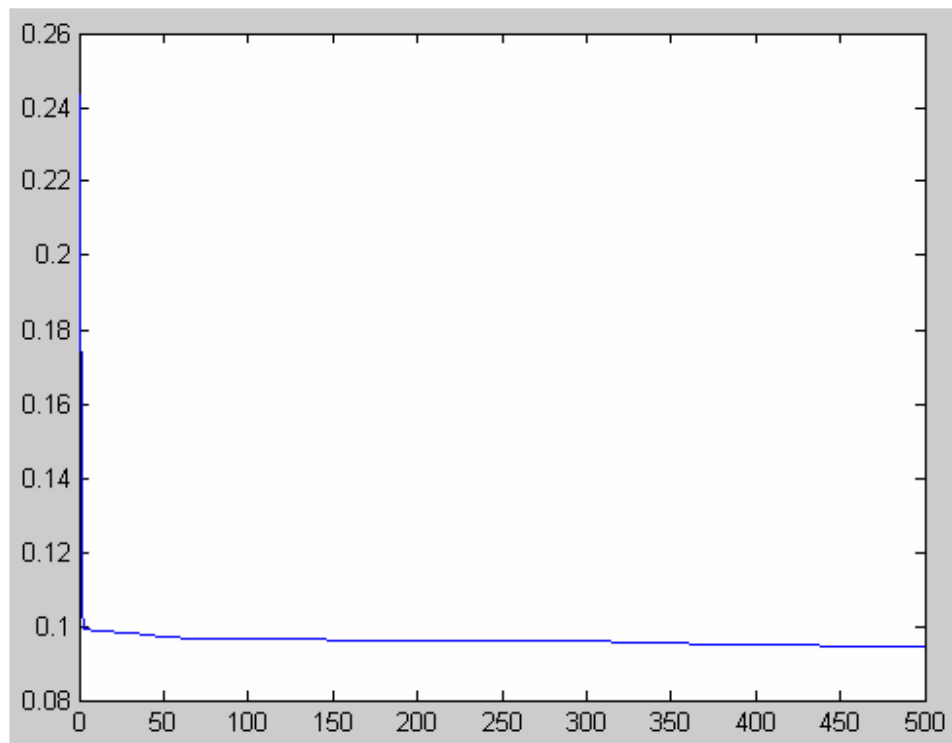
vyrai nuo 001 iki 010

Fonemos fragmentas SP

(stacionarioji - vidurinė kairysis)



53 pav. Klasifikavimo rezultatai nuo 6000 iki 7000Hz stacionaraus fragmento



54 pav. Neuroninio tinklo apsimokymas nuo 6000 iki 7000Hz stacionaraus fragmento

Rezultatas

Paėmus fonemos fragmentą SP nuo 6000 iki 7000Hz, buvo gautas 71.93% atpažinimo tikslumas.

Ekspertas Nr.21

Duomenų paruošimas

Apmokymui:

Dažnis: 6000 Hz – 7000 Hz

Diktoriai: moterys nuo 011 iki 030
vyrai nuo 011 iki 030

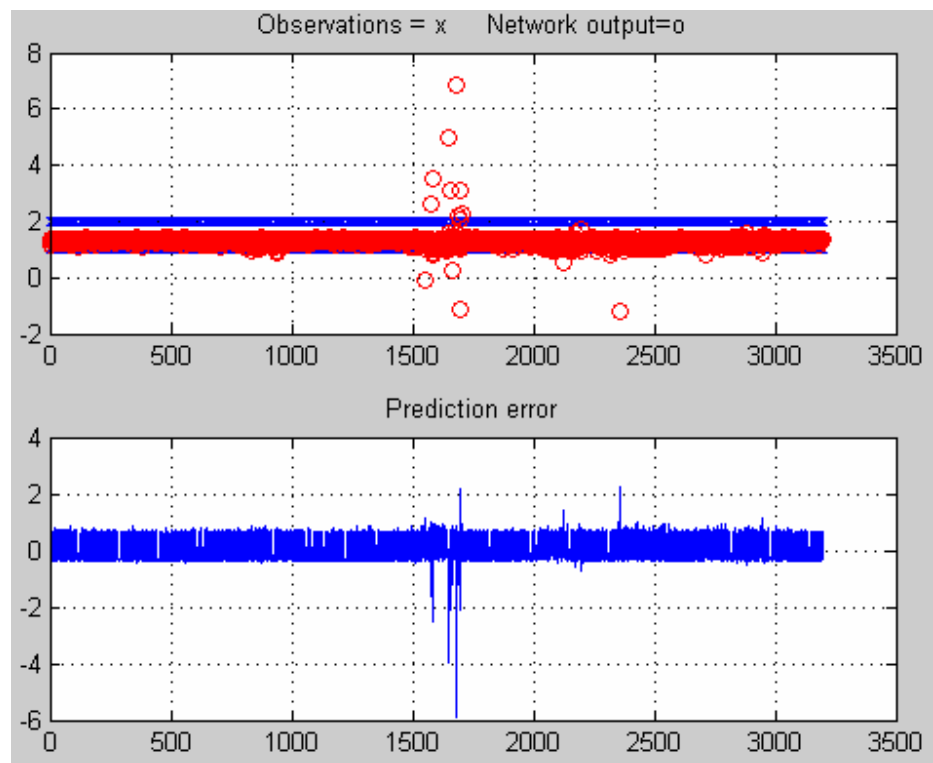
Fonemos fragmentas RC
(dešinysis)

Testavimui:

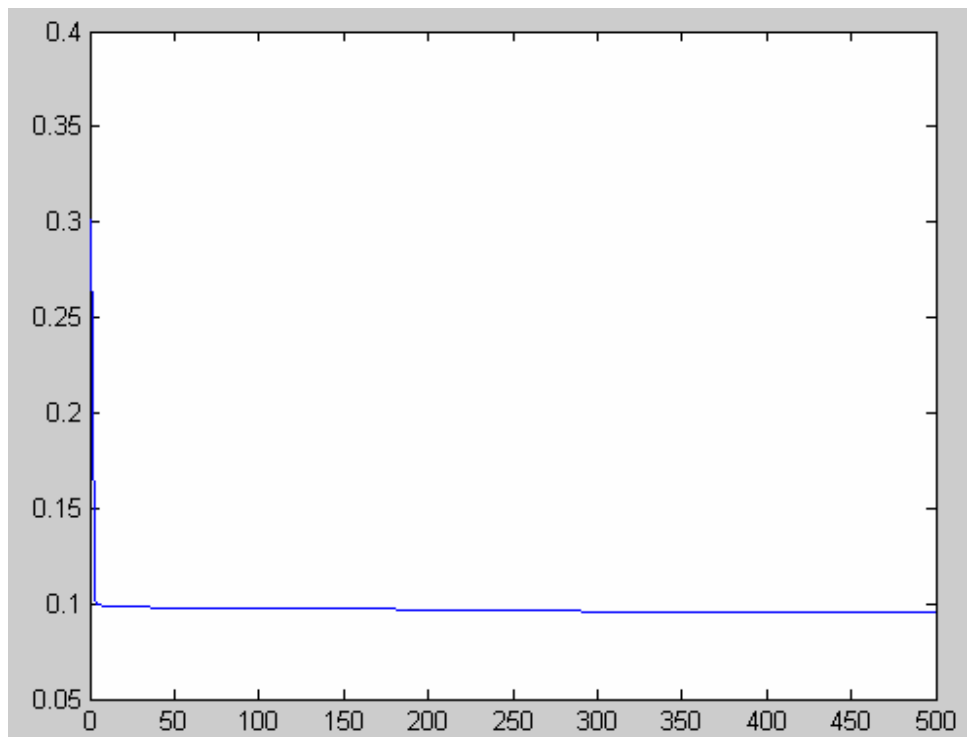
Dažnis: 6000 Hz – 7000 Hz

Diktoriai: moterys nuo 001 iki 010
vyrai nuo 001 iki 010

Fonemos fragmentas RC
(dešinysis)



55 pav. Klasifikavimo rezultatai nuo 6000 iki 7000Hz dešinio fragmento



56 pav. Neuroninio tinklo apsimokymas nuo 6000 iki 7000Hz dešinio fragmento

Rezultatas

Paėmus fonemos fragmentą RC nuo 6000 iki 7000Hz, buvo gautas 71.93% atpažinimo tikslumas.

Ekspirimentas Nr.22

Duomenų paruošimas

Apmokymui:

Dažnis: 7000 Hz – 8000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrai nuo 011 iki 030

Fonemos fragmentas LC (kairysis)

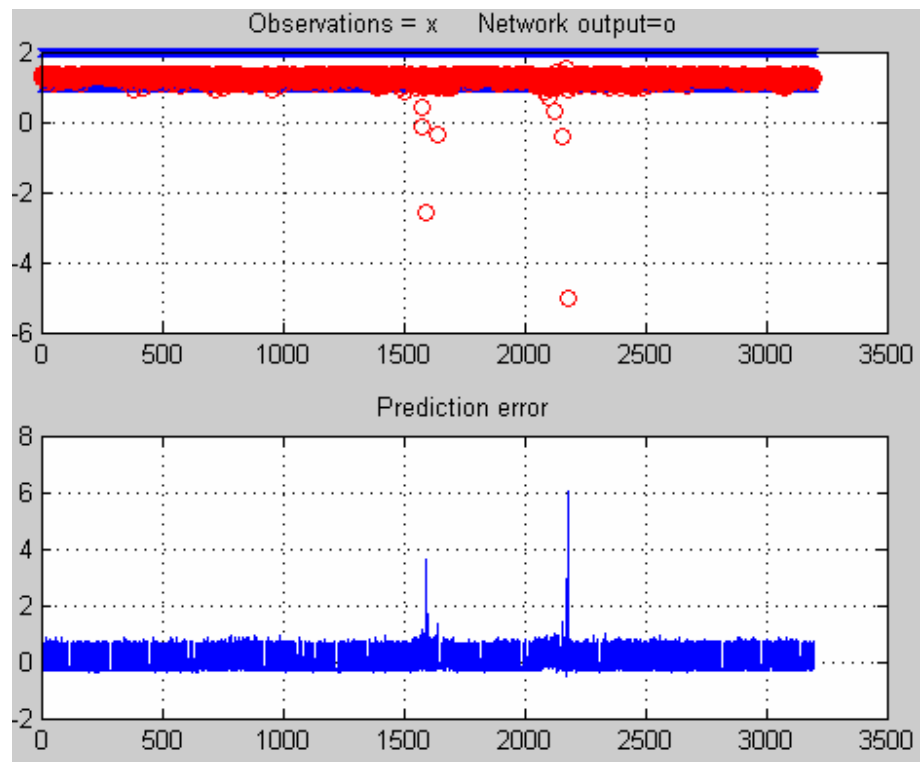
Testavimui:

Dažnis: 7000 Hz – 8000 Hz

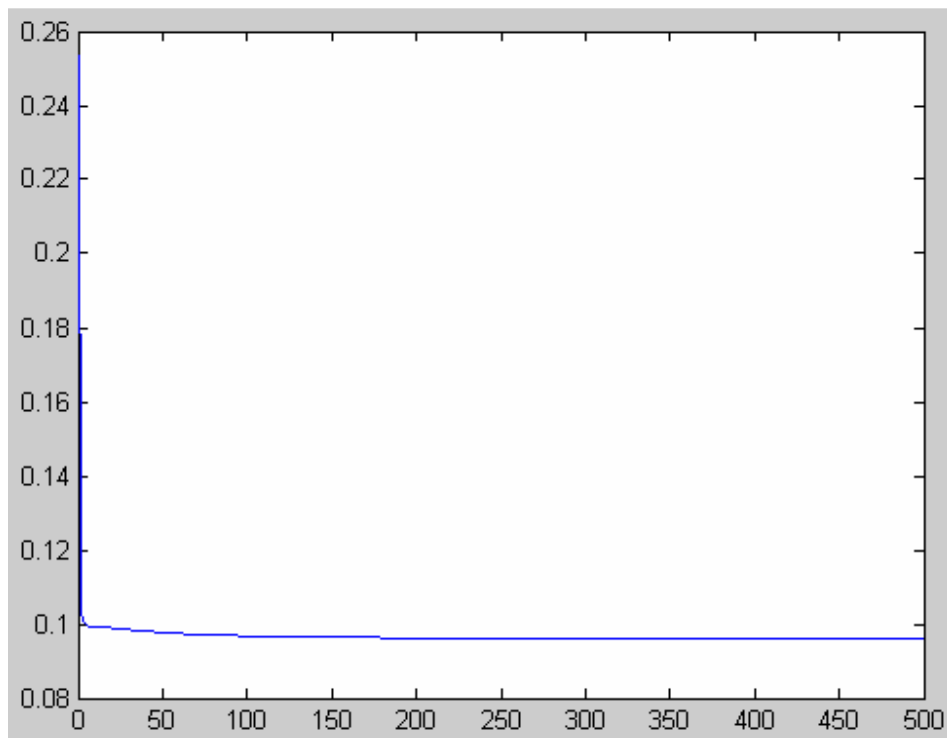
Diktoriai: moterys nuo 001 iki 010

vyrai nuo 001 iki 010

Fonemos fragmentas LC (kairysis)



57 pav. Klasifikavimo rezultatai nuo 7000 iki 8000Hz kairiojo fragmento



58 pav. Neuroninio tinklo apsimokymas nuo 7000 iki 8000Hz kairiojo fragmento

Rezultatas

Paėmus fonemos fragmentą LC nuo 7000 iki 8000Hz, buvo gautas 71.89% atpažinimo tikslumas.

Ekspertas Nr.23

Duomenų paruošimas

Apmokymui:

Dažnis: 7000 Hz – 8000 Hz

Diktoriai: moterys nuo 011 iki 030
vyrai nuo 011 iki 030

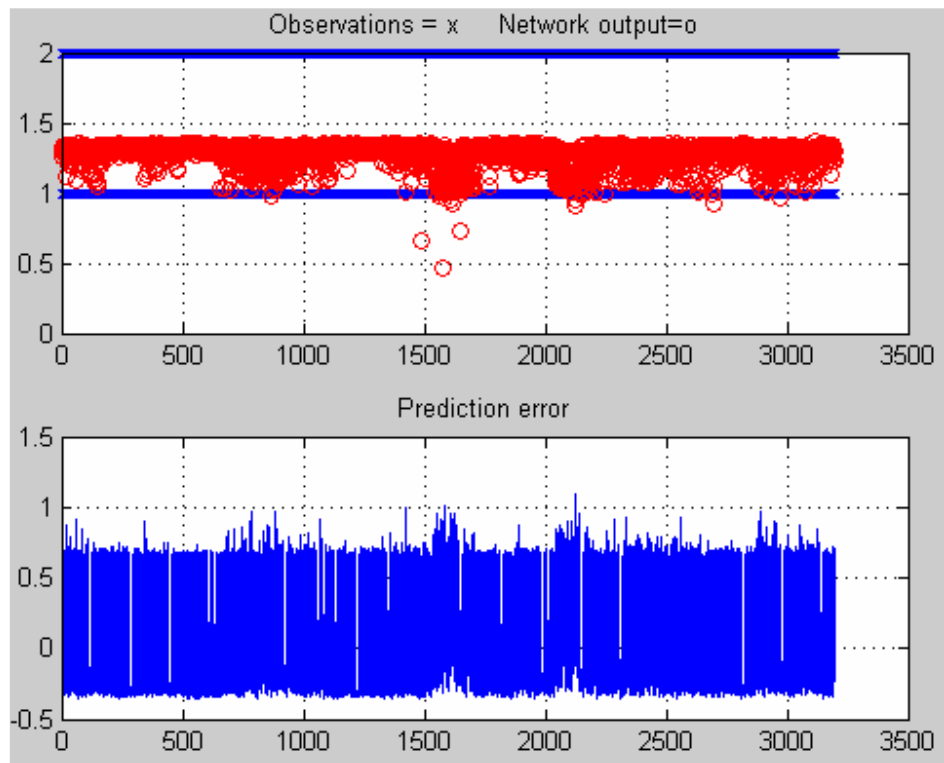
Fonemos fragmentas SP
(stacionarioji - vidurinė)

Testavimui:

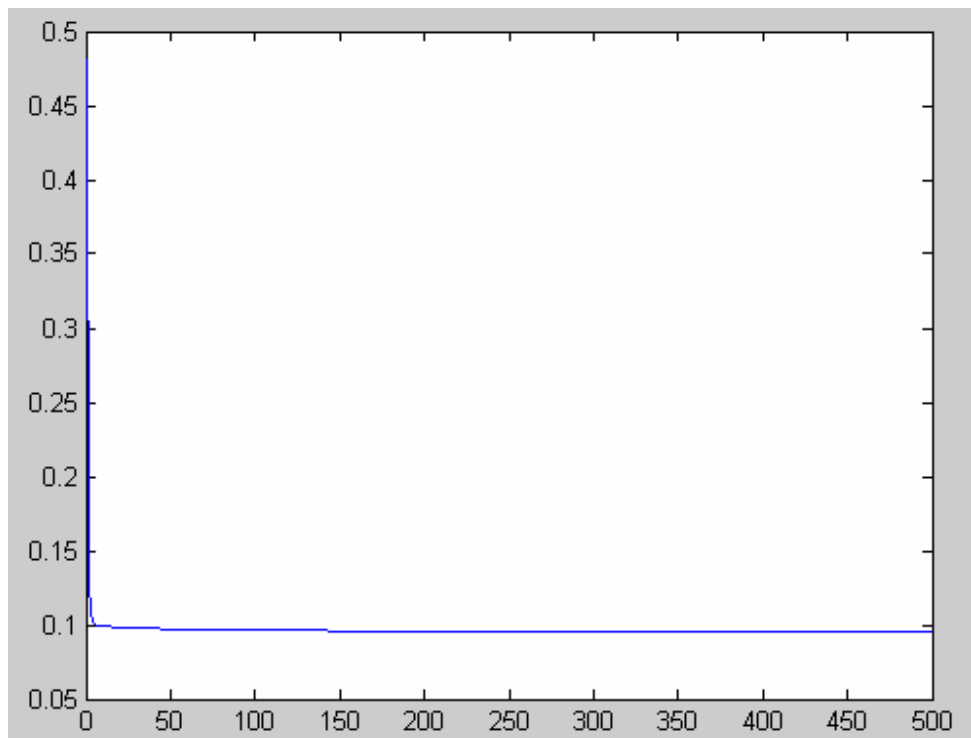
Dažnis: 7000 Hz – 8000 Hz

Diktoriai: moterys nuo 001 iki 010
vyrai nuo 001 iki 010

Fonemos fragmentas SP
(stacionarioji - vidurinė kairysis)



59 pav. Klasifikavimo rezultatai nuo 7000 iki 8000Hz stacionaraus fragmento



60 pav. Neuroninio tinklo apsimokymas nuo 7000 iki 8000Hz stacionaraus fragmento

Rezultatas

Paėmus fonemos fragmentą SP nuo 7000 iki 8000Hz, buvo gautas 72.11% atpažinimo tikslumas.

Ekperimentas Nr.24

Duomenų paruošimas

Apmokymui:

Dažnis: 7000 Hz – 8000 Hz

Diktoriai: moterys nuo 011 iki 030

vyrų nuo 011 iki 030

Fonemos fragmentas RC

(dešinysis)

Testavimui:

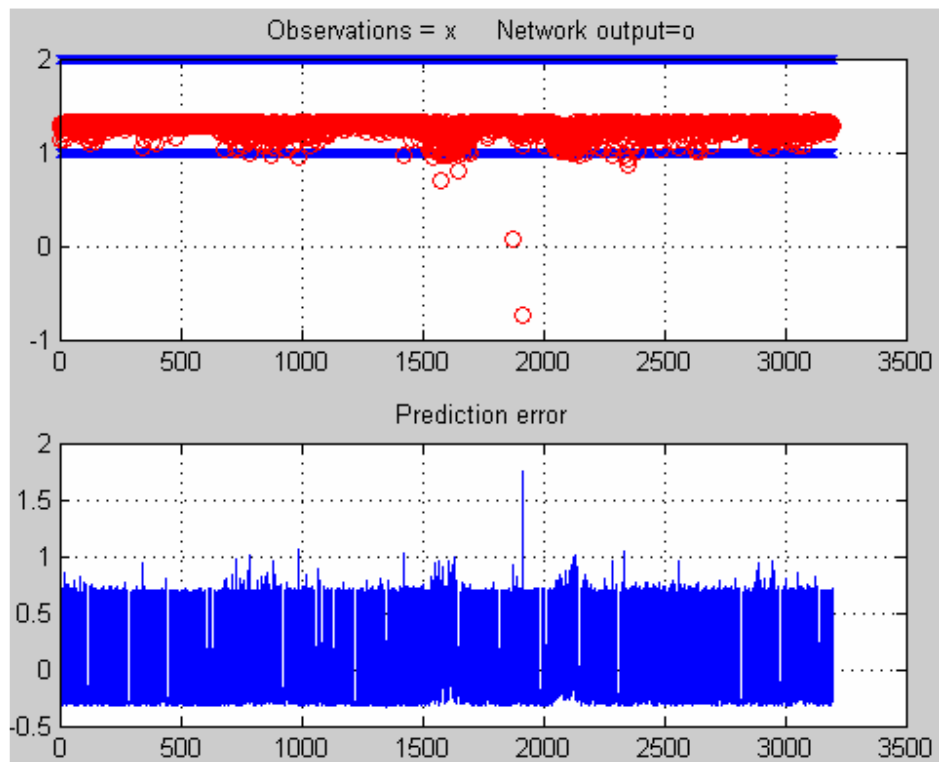
Dažnis: 7000 Hz – 8000 Hz

Diktoriai: moterys nuo 001 iki 010

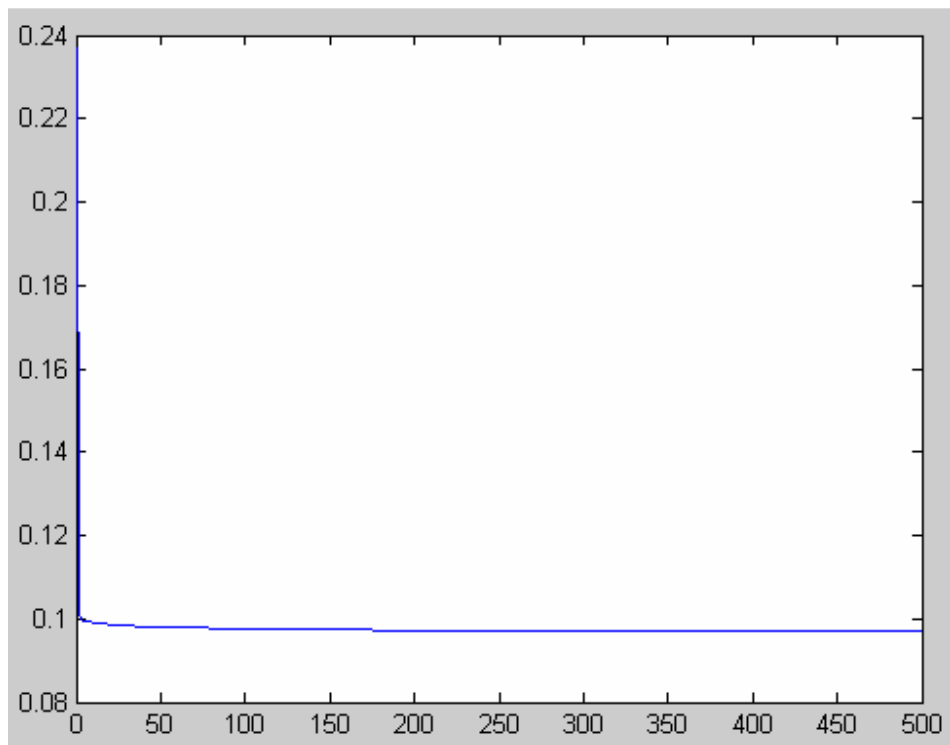
vyrų nuo 001 iki 010

Fonemos fragmentas RC

(dešinysis)



61 pav. Klasifikavimo rezultatai nuo 7000 iki 8000Hz dešinio fragmento



62 pav. Neuroninio tinklo apsimokymas nuo 7000 iki 8000Hz dešinio fragmento

Rezultatas

Paėmus fonemos fragmentą RC nuo 7000 iki 8000Hz, buvo gautas 72.08% atpažinimo tikslumas.

Eksperimento išvados

Kaip ir tikėtasi, atlikus eksperimentą geriausi rezultatai buvo gauti žemiausioje ir vidurinėse dažnių juostose, tai yra matoma pateiktoje rezultatų lentelėje.

1 lentelė

Eksperimentų rezultatų lentelė (proc.)

	LC	SP	RC
7-8 Khz	71,89	72,11	72,08
6-7 Khz	71,99	71,93	71,77
5-6 Khz	71,71	71,46	71,24
4-5 Khz	74,84	77.15	74.68
3-4 Khz	73.43	75.37	74.74
2-3 Khz	71.74	74.90	72.24
1-2 Khz	71.99	72.08	71.49
0-1 Khz	73.08	75.56	75.31

Iš lentelės matoma, kad geriausi rezultatai buvo gauti nuo 0Khz iki 1Khz dažnių juostoje stacionarioje dalyje (SP) 75.56 % ir dešiniajame kontekste (RC) 75.31 % tikslumu. Taip pat buvo gautas geras rezultatas tarp 3Khz ir 4Khz dažnių juostose stacionarioje dalyje (SP)75.37 % tikslumu.

Geriausias rezultatas buvo gautas nuo 4Khz iki 5Khz dažnių juostose stacionarioje dalyje (SP) 77.15 % tikslumu.

IŠVADOS

Pagrindiniai kalbos atpažinimo principai yra: kalbos nuskaitymas, kalbos požymių išskyrimas ir kalbos atpažinimas.

Dabartines kalbos atpažinimo sistemas galima klasifikuoti į: priklausomas nuo diktoriaus, nepriklausomas nuo diktoriaus, atskirai tariamų žodžių atpažinimas ir nepertraukiamas kalbos atpažinimas, viso žodžio ir žodžio smulkesnių dalių atpažinimas bei viso žodžio ar fonemos atpažinimas pagal žodyno dydį

Fonemų atpažinimas – tai šnekos atpažinimas, kurio metu priima sprendimą neuroninis tinklas, kokiai fonemų klasei (šiuo atveju balsis ar pusbalsis) labiausiai atitinka atpažinta fonema.

Viena iš pagrindinių plataus spektrinės analizės metodo paplitimo priežasčių yra tai, kad atliekant šią analizę galima išgauti svarbias akustines kalbos signalo charakteristikas įvairiose dažnių juostose.

Sukurtas naujas metodas balsių ir pusbalsių klasifikavimui naudojant spektrinės analizės požymius ir dirbtinį neuroninį tinklą.

Išanalizavus teorines fonetikos taisykles ir apžvelgus panašius tyrimus pastebėta, kad sunkiai pavyksta atskirti balsius nuo pusbalsių.

Išanalizavus balsių ir pusbalsių teorinius fonetikos akustinius ir artikuliacinius požymius nebuvo galima suformuoti vienareikšmiškų kriterijų balsių ir pusbalsių diskriminavimui

Išanalizavus spektrogramas buvo pastebėta, kad balsių ir pusbalsių garsai išsiskiria vidutinėse dažnių juostose.

Šiam tyrimui buvo sukurtas fonemų klasifikavimo modelis, kai atitinkamai balsiai ir pusbalsiai yra išskaidomi į stacionarias dalis.

Pasirinkto modelio eksperimentiniai tyrimai parodė, kad geriausiai balsiai ir pusbalsiai atskiriami nuo 0Khz iki 1Khz dažnių juostoje stacionarioje dalyje 75.56 % ir dešiniajame kontekste 75.31 % tikslumu. Taip pat buvo gautas geras rezultatas tarp 3Khz ir 4Khz dažnių juostose stacionarioje dalyje 75.37 % tikslumu.

Geriausias rezultatas buvo gautas nuo 4Khz iki 5Khz dažnių juostose stacionarioje dalyje 77.15 % tikslumu.

Pasiūlymai:

Nagrinėti tas filtrų juostas ir jų stacionariausias dalis, kuriose buvo gauti geriausi rezultatai tam, kad būtų gaunami dar geresni rezultatai.

LITERATŪROS SĄRAŠAS

1. Acoustic–Phonetic Features for the Automatic Classification of Stop Consonants
Ahmed M. Abdelatty Ali, *Member, IEEE*, Jan Van der Spiegel, *Senior Member, IEEE*, and Paul Mueller [žiūrėta 2006 m. sausis 16 d.]. Prieiga per internetą: http://repository.upenn.edu/cgi/viewcontent.cgi?article=1018&context=ese_papers
2. An acoustic-phonetic feature-based system for automatic phoneme recognition in continuous speech. Ahmed M. Abdelatty Ali, Jan Van der Spiegel, Paul Mueller, Gavin Haentjens and Jeffrey Berman Dept. of Electrical Engineering, University of Pennsylvania, 200 south 33rd St., Philadelphia, PA 19104-6390, USA, (2)Corticon, Inc., 155 Hughes Rd., King of Prussia, PA 19406, USA, and (3)Dept. of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213-3890, USA. [žiūrėta 2006 m. sausio 16d.]. Prieiga per internetą: <http://www.ee.upenn.edu/~jan/Files/Iscas99Speech.pdf>
3. Acoustic-phonetic feature-based signal processing for automatic speech recognition. Ahmed M. Abdelatty Ali, Dept. of Electrical Engineering, University of Pennsylvania. [žiūrėta 2006 m. sausis 16 d.]. Prieiga per internetą: <http://www.seas.upenn.edu/~ahm/results.pdf>
4. Speech recognition via phonetically-featured syllables. Simon King, Paul Taylor, Joe Frankel and Korin Richmond Centre for Speech Technology Research University of Edinburgh, UK. [žiūrėta 2006 m. sausis 16 d.]. Prieiga per internetą: http://www.cstr.ed.ac.uk/downloads/publications/2000/King_et_al_Phonus2000.pdf
5. Rudžionis A., Rudžionis V., Žvinys P. Lietuvių kalbos garsynas// Kompiuterininkų dienos–99, I dalis/ Devintosios mokslinės–praktinės kompiuterininkų konferencijos ir ketvirtosios mokyklinės informatikos mokslo darbai.– Vilnius: Žara, 1999, 86-96p.
6. Žvinys P., Rudžionis A., Rudžionis V. Lietuvių šnekamosios kalbos duomenų bazių sudarymas// Informacinės technologijos’99/ Konferencijos pranešimų medžiaga.– Kaunas: Technologija, 1999, 410-416 p.

7. Rudžionis A., Rudžionis V., Ratkevičius K., Kasparaitis P., Šalna B. Balso technologijų taikymo lietuvių kalbai analizė ir perspektyvinių veiklos krypčių pagrindimas; Kauno Technologijos Universitetas, Vilniaus Universitetas, Teismo Ekspertizės Centras; Kaunas, Vilnius, 2001.
8. INFORMACINĖS TECHNOLOGIJOS' 2005 (mokslinė-teminė konferencija) KAUNO TECHNOLOGIJOS UNIVERSITETAS 2005 m. sausio 26-27 d. [žiūrėta 2006 m. sausis 17 d.]. Prieiga per internetą: http://www.ktu.lt/lt/mokslas/konf05/konf_02/IT2005/Sekc05.pdf
9. Fonemų atpažinimas; Parengta Virginijaus Vapsvos kursinio darbo ir Antano Lipeikos straipsnių pagrindu; [žiūrėta 2006 m. sausio 17 d.]. Prieiga per Internetą: http://www.maf.vu.lt/~bastys/academic/ATE/LPC/fonemu_atp.htm
10. Balso technologijų pasiekimai pasaulyje [žiūrėta 2006 m. sausis 17 d.]. Prieiga per internetą: http://www.likit.lt/all/balso_tech/03_pasiekimai.htm
11. Tan Keng Yan, Colin; Speaker Adaptive Phoneme Recognition Using Time Delay Neural Networks; National University Of Singapore; 2000 [žiūrėta 2007 m. sausio 12 d.]. Prieiga per internetą: http://www.comp.nus.edu.sg/~ctank/msc_thesis.pdf
12. Lietuvių kalba ir literatūros istorija; *www.ualgiman.dtiltas.lt* [žiūrėta 2007 m. sausio 15 d.] . Prieiga per internetą: http://ualgiman.dtiltas.lt/fonetikos_savoka.html
13. Pakerys A. (1995) *Lietuvių bendrinės kalbos fonetika*. Vilnius. Žara.
14. *Lietuvių kalba I*. (1995) Vilnius. Mokslo ir enciklopedijų leidykla.
15. Gintautas Daunys, KALBOS SIGNALŲ APDOROJIMAS; <http://techno.su.lt/~daunys/KSA/> [žiūrėta 2007 m. sausio 16 d.] . Prieiga per internetą: http://techno.su.lt/~daunys/KSA/p1_4.htm
16. Rudžionis V. (1998) *Speech Recognition by phonetic units*. Doctoral thesis. Kaunas.

17. Gintautas Daunys, KALBOS SIGNALŲ APDOROJIMAS;
http://techno.su.lt/~daunys/KSA/ [žiūrėta 2007 m. sausio 16 d.] . Prieiga per internetą:
<http://techno.su.lt/~daunys/KSA/>

18. Driaunys K. (2006) Lietuvių šnekamosios kalbos segmentavimo ir fonetinio atpažinimo tyrimas naudojant LTDIGITS garsyno įrašus. Daktaro disertacija. Vilnius

19. Tebelskis J; Speech Recognition using Neural Networks; School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania; May 1995. [žiūrėta 2007 m. sausio 16 d.]. Prieiga per internetą: <http://isl.ira.uka.de/fileadmin/publication-files/thesis.joe.ps.gz>

20. Mark Filipovič, Atskirai pasakytų žodžių atpažinimo, naudojant neuroninius tinklus, tyrimas; Matematikos ir informatikos institutas 2003. [žiūrėta 2007 m. sausio 16 d.]. Prieiga per internetą: http://www.ktu.lt/lt/apie_renginius/konferencijos/2006/k6_02/IT2003/Sekcija09.pdf

PAPLONSKIS, Julius. (2007) analysis of vowels and semivowels recognition. MBA*Master Graduation Paper. Kaunas: Vilnius University, Kaunas Faculty of Humanities, Department of Informatics. 66 p., it is given 1 table and 62 pictures

SUMMARY

Work purpose. To perform discrimination analysis of vowels and semivowels (using spectral analysis), exclude features of vowel's and semivowel's signals. Identify in which frequency zones differences of vowels and semivowels differentiates the best

Tasks to do:

- Perform articulation features analysis of theoretic Lithuanian language phonetics: vowels and semivowels.
- Accomplish the reviews of discrimination or identification methods, which are used for vowels and semivowels.
- Design the theoretical model that could be applied for discrimination of vowel's and semivowel's using the features of spectral analysis and operating neural networks for classifications.
- Operating the created method execute the experimental test of vowels and semivowels classifications.

To perform the spectral analysis of vowels and semivowels there was suggested a zone filter with framing and phonemes acoustic modeling. There was reached the 77,15 % classification accuracy using neural networks.