

ŠIAULIŲ UNIVERSITETAS
TECHNOLOGIJOS FAKULTETAS
ELEKTRONIKOS KATEDRA

Martynas Dobrovolskis, Kęstutis Zonys

ŠNEKOS ATPAŽINIMAS
Magistro darbas

Darbo vadovas:
doc. dr. G. Daunys

Šiauliai, 2005

SUMMARY

Voice recognition technologies appeared in the period of general device miniaturization, when all technologies were commonly integrated into one lust. There is no space for buttons and displays anymore.

To have a good system of Lithuanian language recognition, a number of throughout researches must be implemented. Only after selecting the most efficient speech recognition scheme, we can proceed to the development of software adapted to the contemporary time. The aim of this paper is to determine, how efficient speech recognition is possible using neuron networks. MFCC and LPC coefficients were chosen as the parameters characterizing the phonemes. The paper attempts at the determination of the coefficients, which lead to the most efficient recognition of phonemes. For testing, programs PRAAT and MatLab were used.

After implementing a number of phoneme recognition experiments in the research work, the results were obtained, which lead to the following conclusions:

1. In case of using neuron network for the recognition of isolated sounds and characterizing the phonemes by MFCC or LPC coefficients, the possibility of recognition does not exceed 90 per cent. It is not enough for quality recognition of Lithuanian speech.
2. In case of using MFCC coefficients, separate phonemes are recognized better than using LPC coefficients. The difference is about 15 per cent.
3. The advantage of LPC coefficients in comparison with MFCC is the curve of recognition possibility, which is more even.
4. To obtain more efficient phoneme recognition, at least several parameters characterizing the phonemes should be used.

TURINYS

ĮVADAS.....	5
1. ŠNEKOS ATPAŽINIMO SISTEM APŽVALGA.....	6
1.1. Programų apžvalga.....	7
1.1.1. Dragon naturally speaking.....	7
1.1.2. IBM ViaVoice.....	8
1.1.3. PHILIPS free speech.....	9
1.1.4. Горинич.....	9
1.2. Dragon naturally speaking programos testavimas.....	10
1.3. IBM Via Voice programos testavimas.....	11
1.4. Paslėptieji Markovo modeliai (PMM).....	12
1.4.1. PMM panaudojimas šnekos atpažinime.....	12
1.5. Neuroniniai tinklai.....	14
2. KALBOS SIGNALŲ POŽYMIAI.....	19
2.1. Kalbos signalo analizavimas.....	19
2.2. Fonemas charakterizuojančių koeficientų gavimo eiga.....	21
2.2.1. Lango funkcijos.....	24
2.2.2. Furjė transformacija.....	25
2.2.3. MFCC koeficientų gavimas.....	27
2.2.4. Pagrindinė LPC metodo idėja.....	28
2.2.5. Formančių įvertis.....	29
2.3. Fonemos samprata.....	30
2.3.1. Lietuvių kalbos balsių fonemos.....	30
2.3.2. Lietuvių kalbos priebalsių fonemos.....	33
2.3.3. Lietuvių kalbos dvibalsių fonemos.....	37
3. ŠNEKOS ATPAŽINIMO METODO, PANAUDOJANT NEURONINĮ TINKLĄ, TESTAVIMAS IR GAUTŲ REZULTATŲ APIBENDRINIMAS.....	41
3.1. Izoliuotų fonemų atpažinimas be signalo normalizavimo.....	41
3.2. Izoliuotų fonemų atpažinimas su signalo normalizavimu.....	46

3.3. Izoliuotų fonemų atpažinimas neuroninį tinklą apmokant dviejų diktorių fonemomis.....	52
3.4 Izoliuotų fonemų, įrašytų triukšmingoje aplinkoje, atpažinimas.....	58
IŠVADOS.....	62
LITERATŪRA.....	63
PRIEDAI.....	65

ĮVADAS

Rimčiau balso atpažinimo technologiją imta tobulinti pastarąjį dešimtmetį. Beje, nuo to laiko spėjo pasikeisti ir mūsų supratimas apie šios technologijos taikymo ribas. Anksčiau būtume manę, kad užtektų kompiuteriui diktuoti tekstą ar įsakyti rinkti norimą telefono numerį, o šiais laikais automatinį kalbos atpažinimą (angl: ASR – automatic speech recognition) pasirodė įmanoma panaudoti ir tokiems dalykams, apie kuriuos net nepagalvodavome: žaislams ir grotuvams, automobiliams ir pasilinksminimo sistemoms, mobiliesiems telefonams ir delninkams. Šiandien turime tokią tobulą skaitmeninę techniką, kad automatinio kalbos atpažinimo nesutrikdytų nei atsitiktiniai triukšmai, nei pasikeitusi balso intonacija. Be to, pasirodė galingesni bei pigesni ir drauge kompaktiškesni procesoriai bei atminties mikroschemos, todėl visa balso atpažinimo sistema gali tilpti į mažą dėžutę.

Balso atpažinimo technologijos pasirodė visuotinės prietaisų miniatiūrizacijos laikotarpiu, kai visą technologiją įprasta integruoti į vieną lustą. Mygtukams ir displejams jau nebelieka vietos. Tarkim, turėdami balso sąsają, galėtume pasiklausyti muzikos vien ištarę dainos pavadinimą - valdymo įtaisais nereiktų jos ieškoti tarp įvairių meniu.

Šiuo metu yra nemažai smulkių ir stambių automatinį kalbos atpažinimą tyrinėjančių įmonių. Tarp jų - garsiosios IBM ir Philips. Abi kompanijos jau yra sukūrusios ne vieną balso atpažinimo įrenginį bei sukaupusios dešimtmečių patirtį. Yra ir smulkesnių įmonių: Sensory, Advanced Recognition Technology ir Voice Signal Technologies ir kt.

1. ŠNEKOS ATPAŽINIMO SISTEMŲ APŽVALGA

Iš esmės balso atpažinimo technologija yra ne naujas dalykas. Pagrindus šiai technologijai aštuntojo dešimtmečio pradžioje padėjo IBM korporacijos ir Carnegie Mellon universiteto mokslininkai. Nuo to laiko šią sritį ėmė plėtoti įvairių kompanijų ir universitetų tyrėjų grupės. Balso apdorojimo technologijos - tai tokia veiklos sritis, kuri apima įvairiausių mokslus: kompiuteriją, taikomąją matematiką, elektrotechniką, lingvistiką ir informatiką.

Kai žmogus šneka, mikrofonas garso bangas verčia analoginiu signalu, o šis keičiamas skaitmeniniu. Iš skaitmeninio signalo kas 10 ar 20 milisekundžių išrenkami informacijos langai. Kiekvienas langas turi būti toks trumpas, kad per jį nekistų informacijos dažninės savybės, ir toks ilgas, kad apimtų bent vieną dažnio periodą. Balso apdorojimo sistema iš kiekvieno lango išrenka tik jai reikalingą spektrinės priklausomybės informaciją, o visa kita - atmeta.

Toliau ASR gautus rezultatus palygina su bibliotekoje saugomų žodžių atitikmenimis. Teoriškai kalbos duomenis būtų galima lyginti su visų bibliotekos žodžių akustine duomenų baze, kurioje atsispindi netgi žodžių tarimo ypatybės - akcentas, tarmė ir t.t. Bet galop, kai bus rasti geriausi ištartų žodžių atitikmenys, pasirodys, kad toks žodžių atpažinimo būdas per lėtas, taigi šiuo būdu užduočių neįmanoma atlikti realiu laiku.

Yra ir kitas atpažinimo principas. Pirmiausia ASR ištartą žodį lygina su kalbos fonemų ir alofonų (pozicinių fonemos variantų) akustine duomenų baze. Fonema yra mažiausias kalbos garsinės sistemos vienetas, skiriantis tos kalbos žodžius. Lietuvių kalboje yra 65 fonemos ir keli tūkstančiai alofonų. Taigi, balso atpažinimo sistemos bibliotekos žodžiai gali būti pateikiami pagal fonemų ir alofonų visumą, t. y., pagal žodžių tarimą.

Balso atpažinimo principas gali remtis ir kalbos modeliais, kurie žodžius jungia į frazes ar sakinius. Paprastesniame gramatikos - kalbos modelyje, sistema žodžius atpažįsta tik iš konteksto. Gramatinis modelis gerai veikia dialogo, įsakinėjimo ar valdymo atvejais. Jei nenorima, kad kalbėtojas būtų varžomas būtinybės kalbėti įsakmiai, kalbos modelis gali remtis statistiniais kalbos ypatumais - juk kai kurių žodžių junginiai yra dažnai vartojami. Balso sistemų veiksmingumas priklauso nuo to, kaip sudarytas atpažinimo algoritmas, ir nuo to, kaip tarpusavyje susiję atpažinimo metodai. Pradėti analizuoti balsą galima įvairiais būdais. Tarkim, sistema atpažino kelias fonemas. Toliau balsas gali būti analizuojamas tik tinkamiausiais artiniais. Kartais sistema identifikuoja žodį dar nepasibaigus tam skirtam

laikui, nes nustatoma aukšta šio žodžio atpažinimo tikimybė. Kartais analizės pabaigoje sistema nusprendžia, kad identifikavimo tikimybė nėra didelė, arba randa kelis išstartų žodžių atitikmenis. Tokiais atvejais sistema gali paprašyti pakartoti sakinį.

Triukšmų įtakos mažinimas. Balso atpažinimo technologijas naudinga įdiegti automobiliuose. Tokią įrangą turintį automobilį vairuoti daug saugiau, nes vairuotojas mobiliajam telefonui gali diktuoti, į jį net nepažiūrėjęs, o be to, patogiau valdyti prietaiso skydelį bei orientuotis kelyje. Deja, automobilyje yra daugybė visokių triukšmo šaltinių: variklio ir vėjo gaudesys, išorės bei radijo imtuvo triukšmas, keleivių kalbos. Triukšmo ir balso dažnius nesunku atskirti. Svarbiausia, nustatyti pašalinių garsų (tarp jų ir atsitiktinių žmogaus išstartų žodžių) ypatumus, pagal kuriuos būtų galima atskirti triukšmą nuo balso garsų ir jį filtruoti. Kai kuriuos triukšmus, pvz. įvairiu greičiu važiuojančio automobilio variklio ūžesį, galima išmatuoti iš anksto, kad vėliau būtų nesunku jų atsikratyti.

1.1. Programų apžvalga

1.1.1. Dragon naturally speaking

Dragon Naturally Speaking programa gerai įveda tekstą ir leidžia lengvai persijungti tarp diktavimo, taisymo ir formatavimo režimų. Dragon kompanijos paketas nemažai pralenkė kitas kompanijas teksto atpažinime. Šis paketas priartėjo prie 95% atpažinimo tikslumo.

Kai Dragon vis dėlto padaro klaidą, galima įeiti į taisymo režimą, pasakius “delete that” (pašalinti) arba “scratch that” (išbraukti), o po to vėl pakartoti žodį ar žodžių junginį teisingai. Formatuoti tekstą šioje programoje yra taip pat labai lengva: reikia išskirti tekstą ir išstarti tarkim tokius žodžius – “set font Arial 24” (Arial 24 dydžio šrifto nustatymas), “center that” (centravimas) arba “bold that” (paryškintas šriftas). Tačiau diktavimo rezultatą greičiau ir patogiau pataisyti su pele ir klaviatūra.

Šios programos dėka, balsu galima valdyti ir tokias programas kaip Microsoft Word 97 ir naujesnes versijas. Galima valdyti komandinį meniu, bet tai daryti nėra labai patogu, nes reikia nemažai pavargti, kad Word’as paklustu. Daug paprasčiau dirbti su pele ar klaviatūra.

Dragon kompanija perspėja, kad kompiuteriuose su lėtais procesoriais atpažinimas gali būti su didelėmis pauzėmis.

Geriausiai ši programa tinka diktuoti paprastus, neformalius tekstus, tipinius elektroninius laiškus.

Excel 97 programoje braižyti lenteles ar įdiktuoti skaičius su Dragon programa yra gana sunku. Tačiau su valdymo meniu dirbti yra labai paprasta. Pelės kliktelėjimą atitinka žodis - click (spragtelėjimas).

Nors Dragon paketas ir nusileidžia kai kuriems konkurentams klaidų taisyme ir teksto formatavime, jis pralenkia visus gebėjimu iš pirmo karto teisingai užrašyti ištartą tekstą.

1.1.2. IBM ViaVoice

Jeigu reikia kursorių valdyti po darbstačią ne su pele, o balso pagalba, tai ViaVoice programa, IBM korporacijos gaminys, bus geriausias pasirinkimas, nes joje ši funkcija realizuota labai sėkmingai. Tačiau ji negali lygintis su Dragon programos kalbos atpažinimu.

Testuojama programa gan gerai susitvarkė su testinio laiško tekstu, bet suklypo prie kai kurių vardų ir sutrumpinimų. Pavyzdžiui, pavardę Bernardo ji parašė kaip Bernad O, o miestelio pavadinimą Westwood – West would it, o Peterborough – kaip Peter burrow. Tai žymiai padidino atpažinimo klaidų procentą.

Kaip ir Dragon, IBM korporacija garantuoja paprastą persijungimą tarp diktavimo režimo, taisymo režimo, diktavimo paraidžiui ir komandų įvedimo. Paketas gerai supranta ką norima tam tikru atveju daryti.

Via Voice viduje galima naudoti tokias programas kaip Word, Excel ir Internet Explorer Mail. Diktavimas Word'e vyksta beveik be užlaikymų, bet kitose programose tenka šiek tiek palaukti, kol padiktuotas tekstas bus apdorotas. Tačiau Via Voice geriau dirba Excel programoje su skaičiais ir lentelėmis.

Dar vienas Via Voice privalumas – nuostabiai organizuotas Windows darbstačio valdymas. Kad, pavyzdžiui, paleisti Excel, užtenka pasakyti "open Excel", norint iškviesti meniu punktą užtenka jį ištarti. Galima išsrinkti mygtukus tariant ant jų parašytus žodžius (tokius, kaip OK arba Cancel).

Tuo atveju jei programa neatpažino komandų, joje numatytas apmokymas, tačiau to griebtis tenka retai. Jeigu reikia sumažinti kasdieninį darbą su klaviatūra, Via Voice bus geras pasirinkimas. Vis dėl to ši programa atpažystant diktuojamą atsitiktinį tekstą padaro nemažai klaidų.

1.1.3. PHILIPS free speech

Philips FreeSpeech programą galima išbandyti nemokamai. Ji palaiko standartinį WordPad diktavimo langą, ir galimybę diktuoti bet kurioje Windows programoje, kur yra teksto įvedimas, ir meniu valdymas.

Ši programa gan gerai atpažįsta bazinę leksiką, bet kiekvienas neįprastas žodis sukelia keblumą. Diktuoti skaičius, kaip ir Dragon Naturally Speaking, kad programa juos atpažintų reikia po vieną. Perkėlimo ir formatavimo komandos labai panašios į kitų paketų. Tačiau FreeSpeech dažnai su užsispyrimu ignoruoja nurodymus išskirti tekstą arba perkelti kursorių. Daug patogiau šioje programoje tai daryti su pele ar klaviatūra.

1.1.4. Горинич

Tai pirmoji rusų šnekos atpažinimo automatinė sistema. Skirta diktavimui ir kompiuterio valdymui rusų kalba.

Programos galimybės:

Teksto įvedimas rusų ir anglų kalbomis, balsinis valdymas atskiromis operacinių sistemų funkcijomis, tekstinių redaktorių funkcijomis, leidžia apiforminti dokumentus, diplominius darbus (įskaitant formules).

Kaip sistemos branduolys panaudota amerikiečių programa "Dragon Dictate" ir rusiškas modulis, kuris užtikrina rusiško teksto įvedimą ir balsinį valdymą rusų kalba (tas pačias funkcijas atlieka ir anglų kalba).

Teksto rinkimo greitis balsu priklauso nuo kompiuterio pajėgumo ir gali siekti 500 – 700 simbolių per minutę.

Baigus dirbti su programa ji visada siūlo išsaugoti vartotojo biblioteką.

Paruoštos komercinės versijos žodyną sudaro iki 10 000 žodžių. Programa reikalauja vartotojo balso ir tarimo nustatymo, kas užima tų patų laiką. O vėliau programa ir vartotojas vis geriau supranta vienas kitą.

Diktuojama tariant atskirus žodžius, tai yra po kiekvieno ištarto žodžio daroma nedidelė pauzė. Visų pirma sistema orientuota į informacijos įvedimą, o ne į koregavimą, tačiau turi priemonių ištaisyti "blogai išgirstą" tekstą, bet tai galima padaryti ir su klaviatūra, kaip ir įvesti žodžius kurių nėra programos žodyne.

1.2. Dragon naturally speaking programos testavimas

Viena iš nesunkiai vartotojui prieinamų programų, kurių veikimo principas pagrįstas kalbos atpažinimu, yra Dragon naturally speaking. Programa skirta angliškai kalbančiam vartotojui. Ši programa leidžia kompiuterį valdyti balsu. Naudojantis šia Dragon naturally speaking programa galima diktuoti tekstą, kurį kompiuteris atpažysta ir užrašo. Norėdami įsitikinti kiek tiksliai yra atpažystamas skaitomas tekstas, atlikome keletą bandymų:

- dešimt kartų iš eilės buvo kartojamas angliškas žodis “ house “;
- dešimt kartų iš eilės buvo kartojama kompiuterio valdymo komanda “ File ”;
- į programos atmintį įrašėme lietuvišką žodį “ spinta “ ir jį kartojome dešimt kartų;
- įrašius antrą panašaus skambesio žodį “ stinta “ tikrinome kaip programa sugeba atskirti panašaus skambesio žodžius;

Atliekant pirmąjį bandymą dešimt kartų kartojome žodį “ house “ . Programa teisingai šį žodį atpažino ir užrašė 5 kartus. Kitais kartais buvo užrašomi panašiai skambantys žodžiai (Pvz.: mouse, warehouse,...). Bandymo metu nustatėme, kad atpažinimo tikimybė yra 0,5.

Kadangi programa leidžia balsu valdyti kompiuterį, tikrinome kokia yra kompiuterio valdymo funkcijų atpažinimo tikimybė. Dešimt kartų kartojant “ File “ kompiuteris šią komandą suprato devynis kartus iš dešimties, tai gi, valdymo komandos atpažinimo tikimybė yra 0,9.

Dragon naturally speaking programa turi galimybę įrašyti ir įgarsinti norimą žodį, ar frazę. Atlikome bandymą su lietuvišku žodžiu “spinta”. Įgarsindami šį žodį mes tuo pat metu balsą įrašėme į kompaktinį diską. Mūsų atliekamas bandymas susideda iš dviejų dalių. Pirmuoju atveju į mikrofoną kalba žmogus, antruoju atveju yra atkuriamas garsas iš kompaktinio disko. Kalbant žmogui, programa atpažino ir užrašė žodį “spinta” septynis kartus iš dešimties. Atkuriant garsą iš kompaktinio disko, žodis buvo atpažintas dešimt kartų iš dešimties.

Įvesto naujo ir programai nežinomo žodžio atpažinimo tikimybė kalbant žmogui yra 0,7 , o atkuriant įrašytą garsą iš kompaktinio disko, tikimybė artima vienetai.

Atliekant ketvirtąjį bandymą norėta išsiaiškinti kaip programa atpažysta panašaus skambesio žodžius. Tuo tikslu įrašėme žodį “ stinta “. Įgarsindami šį žodį mes jį taip pat įrašėme į kompaktinį diską. Bandymas susideda iš trijų etapų, pirmuoju atveju kalba asmuo

įgarsinęs šį žodį, antruoju – kitas žmogus, trečiuoju – paleidžiamas įrašas iš kompaktinio disko. Pirmojo atvejo atpažinimo tikimybe 0,6. Kartais programa neteisingai atpažysta žodį ir užrašo “ spinta “. Kalbant kitam žmogui atpažinimo tikimybė 0,2, o paleidus įrašą iš kompaktinio disko 0,8.

1.3. IBM Via Voice programos testavimas

Kita programa, su kuria atlikome analogiškus bandymus yra “ IBM Via Voice” Jos paskirtis tokia pat kaip ir “ Dragon Naturally speaking recognition” Programos įdiegimo metu yra atliekamas detalesnis balso kalibravimas, todėl procesas vykdomas ilgiau.

Bandydami “IBM Via Voice” programą, kalbos atpažinimo tikimybei nustatyti, naudojome tuos pačius žodžius ir tą pačią metodiką.

- dešimt kartų iš eilės buvo kartojamas angliškasis žodis “ house “;

Žodis “house” iš dešimties kartų buvo teisingai atpažintas aštuonis kartus, atpažinimo tikimybė 0,8

- dešimt kartų iš eilės buvo kartojama kompiuterio valdymo komanda “ file “;

Komanda “ file” buvo suprasta devynis kartus iš dešimties, atpažinimo tikimybė 0,9

Bandant įrašyti lietuviškus žodžius į programos žodynėlį buvo gauti neigiami rezultatai. Programa neatpažino mūsų įvestų žodžių, todėl negalime palyginti su “ Dragon naturally speaking recognition” programa.

Bandymų apibendrinimas

Atlikus eilę bandymų su programomis iš gautų rezultatų matome, kad “Dragon naturally speaking recognition” programa labiau tinkama kompiuterio valdymui balsu, nei diktuojamo teksto užrašymui. Kompiuterio valdymo komandos atpažystamos su tikimybe artima vienetai. Teksto atpažinimo tikimybė yra žymiai mažesnė - 0,7.

“IBM Via Voice” programa labiau tinkama teksto rašymui, nes jos kalbos atpažinimo tikimybė yra didesnė, nei “Dragon naturally speaking recognition” ir siekia 0,8 – 0,9.

Būtina pažymėti, kad bandymui naudojome “Dragon naturally speaking recognition 5” ir “IBM Via voice 7” programas. Naudojant naujasnes programų versijas rezultatai gali skirtis.

1.4. Paslėptieji Markovo modeliai (PMM)

Beveik visos šiuo metu rinkoje siūlomos ASR sistemos remiasi paslėptaisiais Markovo modeliais (Hidden Markov Models – HMM, liet.: PMM). Jais aprašomas fonemų ir alofonų tarimas bei šių tarimų trukmė. Markovo modeliai jau taikomi tris dešimtmečius, ir jų populiarumas tik didėja. PMM artinys turi neginčijamų pranašumų. Matematikos požiūriu jis yra labai patogus, lengvai suprantamas ir įgyvendinamas.

Kad būtų sukurta duomenų bazė, su kuria reiktų lyginti išstartus žodžius, PMM irgi reikia "pamokyti". Kai kurias užduotis atlikti sistemą moko pats kalbėtojas. Tuomet atpažinimo sistema mokės atpažinti būtent to žmogaus kalbą. Kitais atvejais kalbančiųjų gali būti daug. Tuomet sistemos kūrėjai parenka daugelio žmonių kalbos pavyzdžių ir sudaro mokymo duomenų vidurkį.

Jei automatinė kalbos atpažinimo sistema skirta naudoti automobilyje, balso pavyzdžiai turi būti įrašyti automobilyje, kad duomenų bazėje būtų atsižvelgta į šios ribotos erdvės akustines charakteristikas. Ištikrųjų parengti gerus sistemos mokymo duomenis yra taip pat svarbu, kaip ir parašyti veiksmingus žodžių paieškos algoritmus.

1.4.1. PMM panaudojimas šnekos atpažinime

Apibendrinta atpažinimo sistema

1.1 pav. pavaizduota nenutrūkstamos kalbos atpažinimo sistema. Joje signalas apdorojimas taip:

- a. *Savybių analizė.* Čia atliekama spektrinė ir/arba laikinė kalbos signalo analizė, kurios metu gaunami stebėjimo vektoriai, kurie aprašo įvairius kalbos garsus. Gauti vektoriai naudojami PMM mokymui.
- b. *Elemento atitikimo sistema.* Čia, pirmiausiai, turi būti pasirinktas kalbos atpažinimo elementas. Tai gali būti įvairios lingvistinės žodžio dalys tokios, kaip fonemos, dvibalsiai, skiemenys ir pan., arba visas žodis ar net keli žodžiai. Bendru atveju, kuo paprastesnis yra pasirinktas elementas, tuo jų mažiau yra kalboje ir tuo sudėtingesnė yra jų struktūra nenutrūkstamame kalbos atpažinime. Didelio žodyno (daugiau kaip 1000 žodžių) kalbos atpažinime paprastai yra privalomas paprastų elementų (žodžio dalių) naudojimas, nes kitaip būtų sunku sudaryti tinkamą žodžių rinkinį PMM mokymui. Tačiau tam tikrais specializuotais

atvejais (kai žodyne yra mažai žodžių) abu variantai yra tinkami. Nepriklausomai nuo pasirinkto varianto, kalbos elemento požymiai randami apmokant modelį.

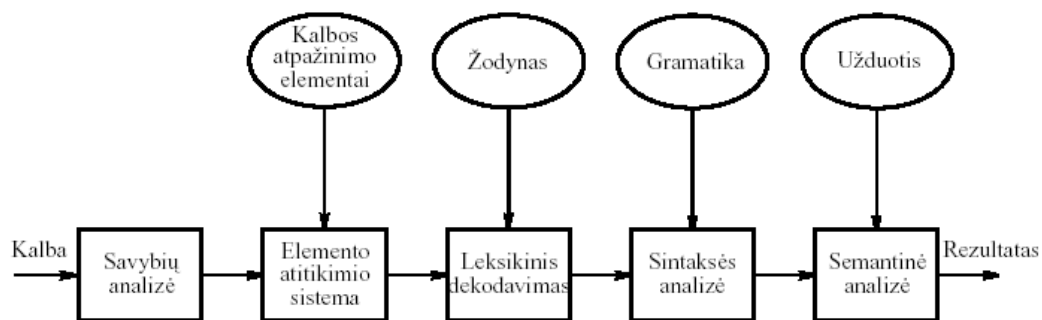
Paprastai kiekvienas elementas yra charakterizuojamas kažkoku PMM, kurio parametrai randami pagal mokymo duomenis. Elemento atitikimo sistema pateikia atitikimo tikėtinumą, visų turimų kalbos atpažinimo elementų, lyginant su nežinomu įvesties signalu.

c. *Leksikinis dekodavimas*. Šis procesas iškelia reikalavimus elemento atpažinimo sistemai, kad joje ištirti keliai atitiktų kalbos elementų sekai, kuri yra žodžių žodyne (leksikone). Šioje procedūroje atpažystamų žodžių žodynas turi būti sudarytas pagal pasirinktus atpažinimo elementus.

Jei elementai yra atskiri žodžiai ar žodžių junginiai, tai tada šis leksikinio dekodavimo žingsnis yra nereikalingas ir atpažinimo sistema supaprastėja.

d. *Sintaksės analizė*. Šis procesas taip pat iškelia dar kelis reikalavimus elemento atitikimo sistemai, kad joje ištirti keliai ne tik atitiktų kalbos elementų sekai, kuri sudaro žodžius, bet ir šie žodžiai turi būti išsidėstę tokia tvarka, kuri atitinka gramatikos reikalavimus.

e. *Semantinė analizė*. Šis procesas, taip kaip ir du prieš jį sekę procesai, toliau didina reikalavimus parenkant atpažinimo paieškos kelius. Vienas iš būdų kaip patenkinti semantinius reikalavimus yra naudoti dinaminį atpažinimo sistemos būsenos modelį. Priklausomai nuo atpažinimo sistemos būsenos tam tikri sintaksiškai teisingi žodžių junginiai yra atmetami. Tai palengvina atpažinimo uždavinį ir didina sistemos efektyvumą.



1.1 pav. Nenutrūkstamos kalbos atpažinimo sistemos panaudojant PMM blokinė diagrama.

Yra dar vienas faktorius, kuris turi svarbią įtaką realizuojant kalbos atpažinimo sistemą. Tai yra tylos periodų atskyrimas kalbos signalu. Yra trys priimtini būdai tai padaryti:

1. Detektuoti kalbos buvimą pagal signalo energiją ir trukmę.

2. Sudaryti tylos periodo statistinį modelį ir įėjimo signalą vaizduoti kaip seką (*tyla*)-kalba-(*tyla*), kurioje tylos dalis nėra būtina, nes jos po kalbos gali nebūti.

3. Išplėsti kalbos vieneto modelį prijungiant prie jo pradžios ar pabaigos tylos periodą, tuo būdu tyła bus modeliuojama visuose kalbos vienetų modeliuose.

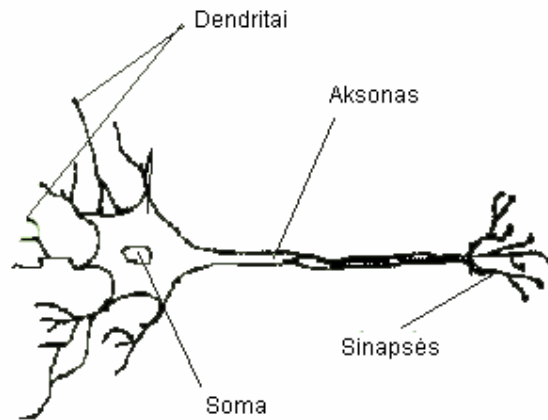
1.5. Neuroniniai tinklai

Vienas iš idomesnių klasifikatorių, kurie gali būti apsimokyti, yra neuroniniai tinklai. Jau pats pavadinimas byloja, kad šis duomenų atpažinimo, klasifikavimo ir prognozavimo principas turi analogijų su žmogaus mąstymu. Taip ir yra - šis principas buvo kuriamas imituojant žmogaus smegenų veikimą - informacijos talpinimą bei paiešką neuronų tinkluose. Todėl šis principas dažnai dar yra vadinamas ir kitu vardu - daugiasluoksnis perceptronas. Jo pagrindas - funkcija, įgijanti reikšmes iš binarinės aibės $\{0, 1\}$. Kadangi veiksmai kompiuteryje pagrįsti binarine algebra, šio principo panaudojimas pritaikant skaičiavimo mašinas yra paprastesnis nei kitų, nes gana efektyviai galima išnaudoti kompiuterio architektos ypatumus. Nagrinėję objektų atpažinimo ir klasifikacijos problemas buvo padaryta daug išvadų, kurios buvo sėkmingai pritaikytos tiriant žmogaus smegenų veiklą. Kaip pavyzdys gali būti pradinio žingsnio parinkimas pirmajame klasifikacijos etape, kai atsitiktiniu būdu bandoma parinkti funkcija labiausiai aproksimuojanti turimus duomenis.

Dirbtiniai neuroniniai tinklai vis plačiau naudojami dėl keleto pagrindinių priežasčių:

- Neuroninis tinklas yra galingas modeliavimo aparatas, įgalinantis modeliuoti ypač sudėtingas funkcijas. Bendru atveju neuroninis tinklas yra netiesinė struktūra. Neuroniniai tinklai padeda išvengti dimensiškumo problemos, išskylančios modeliuojant daugelio kintamųjų funkcijas tiesiniais metodais.
- Neuroniniai tinklai apmokomi iš pavyzdžių. Turint duomenų pavyzdžių ir naudojant mokymo algoritmus, neuroninis tinklas pritaikomas prie duomenų struktūros. Nors neuroninio tinklo vartotojas turi turėti tam tikrų įgudžių, kaip parinkti ir paruošti duomenis, parinkti neuroninį tinklą ir jo struktūrą, tačiau norint sėkmingai naudoti neuroninius tinklus pakanka mažiau žinių nei naudojant, pavyzdžiui, tradicinius statistikos metodus.

Vienasluoksnis perceptronas (VSP). Šis klasifikatorius – vienas pirmųjų neurono modelių. Trumpai priminsime biologinio neurono struktūrą, kurią matome 1.2 pav :



1.2 pav. Biologinio neurono struktūra.

„Įėjimo sluoksnyje“ esantys dendritai priima informaciją ir perduoda ją somai (branduoliui), kuriame vyksta tam tikras procesas (gautos informacijos transformavimas). Šio proceso rezultatai aksonais perduodami į sinapses, kurios turi ryšį su kitais neuronais.

Nagrinėjame kategorijas K_1, K_2 bei N -matį vektorių X . Tarkime, jog žinome, kad kategorijos yra tiesiškai atskiriamos. Tegu, kad $X=(x_1, x_2, \dots, x_N)$.

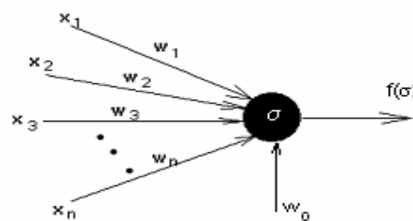
Klasifikatorių konstruojame taip:

1. Imame $N+1$ skaičių $W=(w_0, w_1, w_2, \dots, w_N)$ ir sudarome svorinę sumą:

$$\sigma = \sum_{k=1}^N x_k w_k \quad (1.1)$$

2. Apibrėžiame aktyvacijos funkciją $f(\sigma) = 1$, jei $\sigma > w_0$ ir 0 priešingu atveju.

Gautą klasifikatorių galime pavaizduoti sekančia schema matoma 1.3 pav:



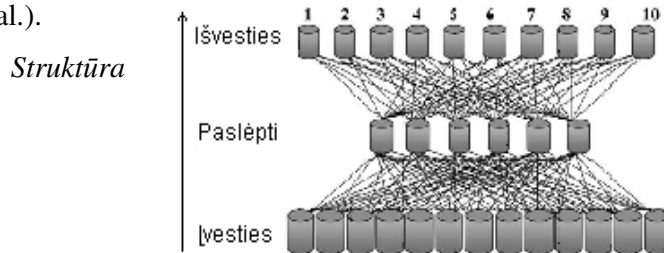
1.3 pav. Biologinio neurono modelis.

kur X - “dendritai”, per kuriuos gaunama informacija, svoriai W - „vidinė informacija“ (neurono parametrai), sumos σ skaičiavimas - „branduolyje vykstantis procesas“

, o $f(\sigma)$ (aktyvacijos funkcija) žymi neurono būseną (suaktyvėjęs ar ne). Svoris w_0 nusako kokį skaičių σ turi pasiekti, kad neuronas „suaktyvėtų“ t.y. $f(\sigma) = 1$. Kitaip tariant:

$$f(\sigma) = 1 \Leftrightarrow \sigma > w_0 \Leftrightarrow \sum_{k=1}^N x_k w_k > w_0 \quad (1.2)$$

Daugiasluoksnis perceptronas. Šis klasifikatorius buvo sukurtas 1986 metais (Rumelhart et al.).



1.4 pav. Daugiasluoksnio perceptrono struktūra.

Daugiasluoksnis perceptronas (DSP) gaunamas jungiant VSP į sluoksnius, kaip parodyta 1.4 pav. Neurono modelis išlieka toks pats, kaip ir VSP, tik dažniausiai naudojama diferencijuojama aktyvacijos funkcija. Tai įgalina naudoti metodus, kuriuose dirbama su diferencijuojamomis funkcijomis (pvz. gradientinis ekstremumų paieškos metodas).

DSP sluoksniai skirstomi į tris grupes:

1. Įvesties sluoksnis, kurio neuronai perduoda gautus duomenis kiekvienam paslėptojam sluoksnio neuronui (žr. 1.2 pav.) jų netransformuodami.
2. Paslėptasis sluoksnis. Paslėpti sluoksniai gali būti keli.
3. Išvesties sluoksnis.

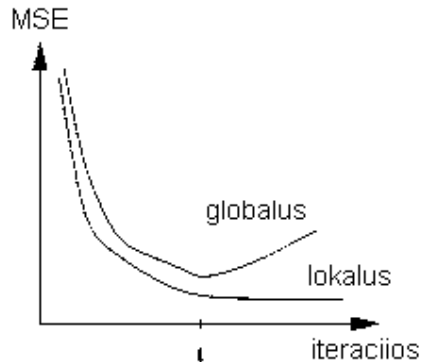
DSP galima išivaizduoti kaip m -matę sudėtinę funkciją nuo n -mačių įėjimų (kur n – įėjimo sluoksnio neuronų skaičius, o m – išėjimo sluoksnio neuronų skaičius).

DSP tinklo mokymo ypatybės:

1. *Pradiniai svoriai.* Jeigu naudojame logistinę f -ją ar hiperbolinį tangentą naudinga parinkti pradinius svorius iš pakankamai siauros 0 aplinkos, kadangi priešingu atveju aktyvacijos funkcijos išvestinė jau pačiose pirmose iteracijose bus maža, o iš to seka, jog tada algoritmo konvergavimas sulėtėja. Rekomenduojama pradinius svorius imti su vidurkiu = 0 ir dispersija = 1.

2. *Mokymo iteracijų skaičius.* Jeigu turėdami fiksuotą mokymo aibę paimsime labai mažą leistiną klaidą, taip automatiškai padidindami iteracijų skaičių, per kurias tinklas turi išmokyti duomenis, gali pasireikšti „žirklių efektas“ – kada tinklas prisitaiko prie turimų

duomenų (lokalus MSE mažėja), o tuo tarpu globalių duomenų MSE – pradeda didėti. (1.5 pav.). MSE - Klasifikatoriaus apsimokymo lygį charakterizuojanti vidurtinė kvadratinė klaida (angl.: MSE - Mean Squared Error).



1.5 pav. Ryšys tarp MSE ir iteracijų skaičiaus.

Todėl vietoj iteracijų didinimo dažnai naudingiau bandyti rasti tašką, kuriame globalusis MSE pradeda didėti (vadinamoji *early stopping* strategija).

Žirklių efekto atsiradimą galima traktuoti taip: DSP pradeda naudoti ne tik informaciją, charakteringą analizuojamų duomenų imčiai, bet ir informaciją iš konkrečios jos realizacijos (apmokymo aibė su kuria dirbama).

3. *Adaptyvus mokymo greitis*. Nors tinklo skaičiavimuose mokymo greitis imamas kaip konstanta, bet praktikoje naudinga jį mokymo eigoje keisti. Pavyzdžiui naudojant tokią strategiją:

pradinis greitis $\eta = \text{const}$

toliau kas 1% iteracijų η :

jeigu MSE mažėja - $\eta := \eta * A$ (kur $A = \text{const} > 1$)

jeigu MSE didėja - $\eta := \eta * B$ (kur $0 < B = \text{const} < 1$).

Analogiškai galima kontroliuoti ir aktyvacijos funkcijos prisiplojimą.

4. *Tinklo išėjimo ribos*: Praktiniuose uždaviniuose dažnai nepakanka atsakymo iš (0,1) (logistinės f-jos atveju) arba (-1,1) (hiperbolinio tangento atveju).

Norėdami išplėsti tinklo atsakymą galime modifikuoti aktyvacijos funkciją, padaugindami ją iš atitinkamos konstantos (pvz. $N * \tanh(x)$, $x \in \mathfrak{R}$) atvaizduos realiųjų skaičių ašį į intervalą (-N,+N). Suprantama, tada figūruojančias funkcijos išvestines reiks padauginti iš N.

Daugiasluoksniams perceptronams egzistuoja efektyvus tikslumo atžvilgiu mokymo algoritmas – tai vadinamasis klaidos atbulinio sklidimo algoritmas. Jis grindžiamas gradientinio nusileidimo metodu.

Neuroninių tinklų privalumai. Apsimokymo galimybė leidžia žymiai sutrumpinti skaičiavimo laiką ir naudoti mažesnius resursus - tuo būdu tokių skaičiavimų savikaina sumažėja, o darbo efektyvumas išauga. Dar vienas neuroninių tinklų metodo privalumas - tai galimybė pasiekti norimai didelį tikslumą (tai yra sąlygota to, kad rezultatas yra 'auginamas' ir artinamas prie norimos reikšmės norimai ilgai), ko dažnai neleidžia kiti metodai.

Neuroninių tinklų trūkumai. Didelio tikslumo pasiekimas gana ilgai 'auginant' rezultatą sąlygoja rezultato gremėzdiškumą, o tai gali būti viena iš priežasčių, neleidžiančių naudotis gautu rezultatu. Kitas trūkumas - gana didelių (pavyzdžiui, kompiuterinių, jei skaičiavimai atliekami naudojant kompiuterį) resursų poreikis, ypač siekiant labai tikslių rezultatų. Kadangi šis metodas paremtas ir atsitiktine paieška, blogas pirminis konstantų bei intervalų parinkimas gali nulemti ilgą skaičiavimą bei rezultato gremėzdiškumą, pasiekiant tą patį tikslumą, kurio siekus naudojant kitas pradines konstantas ir intervalus, rezultatas būtų gautas greičiau.

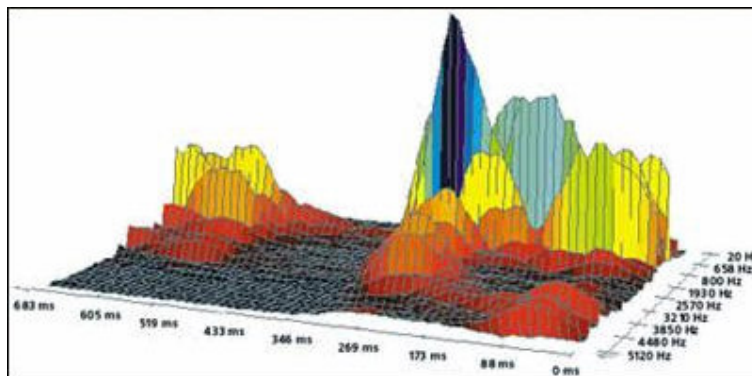
Neuroninių tinklų naudojimo rekomendacijos, atsižvelgiant į jų teigiamas ir neigiamas sąvybes. Šis metodas yra pakankamai galingas ir produktyvus, kai reikia pasiekti didelį analizės tikslumą. Tačiau didelių tikslumų siekimas reikalauja didelių resursų, ir ši priklausomybė panaši į geometrinę progresiją. Todėl šis metodas rekomenduotinas, kai reikalinga ypač tiksli analizė, o nauda iš rezultatų gavimo leidžia investuoti į brangesnius resursus skaičiavimams.

2. KALBOS SIGNALŲ POŽYMIAI

2.1. Kalbos signalo analizavimas

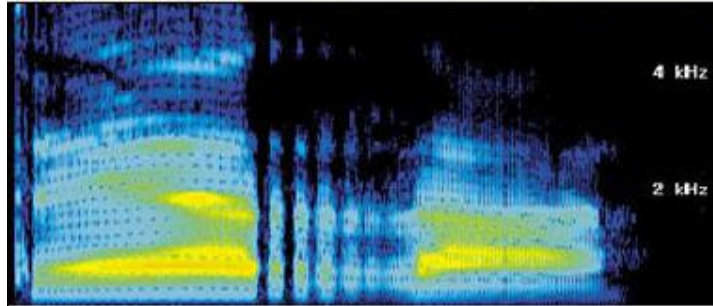
Šiuolaikiniais personaliniais kompiuteriais galima atlikti kalbos signalo analizę laikinėje ar spektrinėje dalyje. Kalbos signalo oscilograma gaunama panaudojant mikrofoną ir kompiuterinę garso plokštę, apdorojant signalą paprastu garso redaktoriumi. Daugiau informacijos gali suteikti spektrinis kalbos signalo pavidalas, tada garsas atvaizduojamas daugeliu sinusoidžiu. Signalu atvaizdavimas spektru, atliekamas panaudojant greitąją Furjė transformaciją. Greitoji Furjė transformacija yra naudojama daugelyje programų, kurių paskirtis kalbos signalo atpažinimas realiaame laike.

Kalbininkai ir lingvistai savo tyrinėjimuose dažniausiai naudoja kalbos signalą, kuris yra atvaizduotas trimatėmis ir dvimatėmis sonogramomis. Naudojant trimatį vaizdavimą koordinatinių ašys atitinka laiką, dažnį ir spektrą (energiją). Dvimačiame atvaizdavime koordinatinių ašis, kuri skirta energijai, pakeičiama spalvos intensyvumu. Kuo ryškesnė spalva tuo didesnę garso energiją ji žymi. 2.1 pav. pavaizduota trimatė sonograma, kurioje atvaizduotas ištartas žodis. Gerai matyti spektrinių dedamųjų kitimas laike.



2.1 pav. Trimatė žodžio sonograma.

2.2 pav. pavaizduota to paties žodžio dvimatė sonograma.



2.2 pav. Dvimatė žodžio sonograma.

Pagal kalbos signalo sonogramas galima analizuoti kalbos signalo toną ir formantines trajektorijas. Žmogaus balsą charakterizuoja keletas svarbių parametrų. Pagal juos galima skirti moterišką balsą nuo vyriško, aukštą balsą nuo žemo ir t.t. Vienas iš tokių parametrų yra galingumo spektras (vidutinis, intensyvus spektras, spektrinių komponentų koreliacija). Formantinės charakteristikos (formančių dažnis, jų amplitudė ir plotis, kitimo dažnis). Kepstrinės charakteristikos, balso tono parametrai (vidutinė reikšmė, dispersija, statiniai momentai, mikro ir makro variacijos). Statinės ritmikos charakteristikos ir kalbos tempas (garsų per sekundę skaičius, kalbos segmentų ilgis ir kalbinės pauzės). Čia išvardintos tik pačios pagrindinės kalbos signalo charakteristikos, į jas visas būtina atsižvelgti kuriant kalbos atpažinimo sistemą, jei norime jog ji būtų universali, t.y. tiktų plačiam vartotojų ratui.

Akustikos požiūriu kalbinis signalas gali būti skaidomas į tris tipus

- Kalbos garsas
- Triukšmas
- Kalbinės pauzės

Kalbos garsas susideda iš sinusoidžių spektro, kurio reišmės kinta priklausomai nuo dažninės signalo reišmės. Laiko atžvilgiu toks signalas gali būti laikomas periodišku. Triukšmo spektras panašus į balto triukšmo. 2.3 pav. pavaizduota raidės “a” oscilograma:

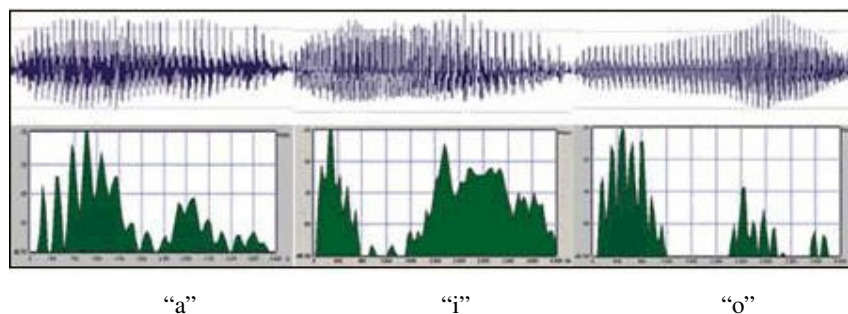


2.3 pav. Tariant garsą “a” gauta oscilograma.

Analizuojant formantes svarbus jų kitimas laike. Formantės žymi ne tik artikuliacines žmogaus balso savybes, bet yra ir savotiškos diferencijuotos fonemos. Paprastai “švariame” garso signalo galime išskirti 4 - 5 formantes. Aukštos formantės mažiau pastebimos spektre,

tačiau butent jos turi svarbiausią informaciją apie žmogaus individuališkumą. Tai plačiai pritaikoma kriminalistikoje. Formančių trajektorijos vyriškų ir moteriškų balsų yra skirtingos, jau nekalbant apie vaikų ir paauglių balsus, kurių formantės tiesiog “gyvena” savo gyvenimą ir sulyginimas tokių formančių su esančiomis duomenų bazėje pasidaro gana keblus.

2.4 pav., viršutinėje grafiko dalyje pateikiamos trijų balsių: “a”, “i” ir “o” tipinės signalo formos, žemiau jų tuo pačiu eiliškumu pavaizduoti balsių spektrai, kur matosi harmonikų pikai ir kalbos formantės. Horizontalioje ašyje pavaizduotas dažnis, vertikaloje signalo amplitudė.



2.4 pav. Trijų skirtingų balsių spektras.

2.4 pav. aiškiai matyti kaip išsiskiria garsų formantės. Pvz. analizuojant balsę “a” išryškėja trys formantės: 300 – 1500Hz pirmoji, 2200 – 3100Hz antroji, 3300 – 3900 trečioji. Vos tik pažvelgus į spektrus matyti požymiai skiriantis vieną balsę nuo kitos. Pavyzdžiui balsė “i” visada turi gana plačią aukštų formančių dalį.

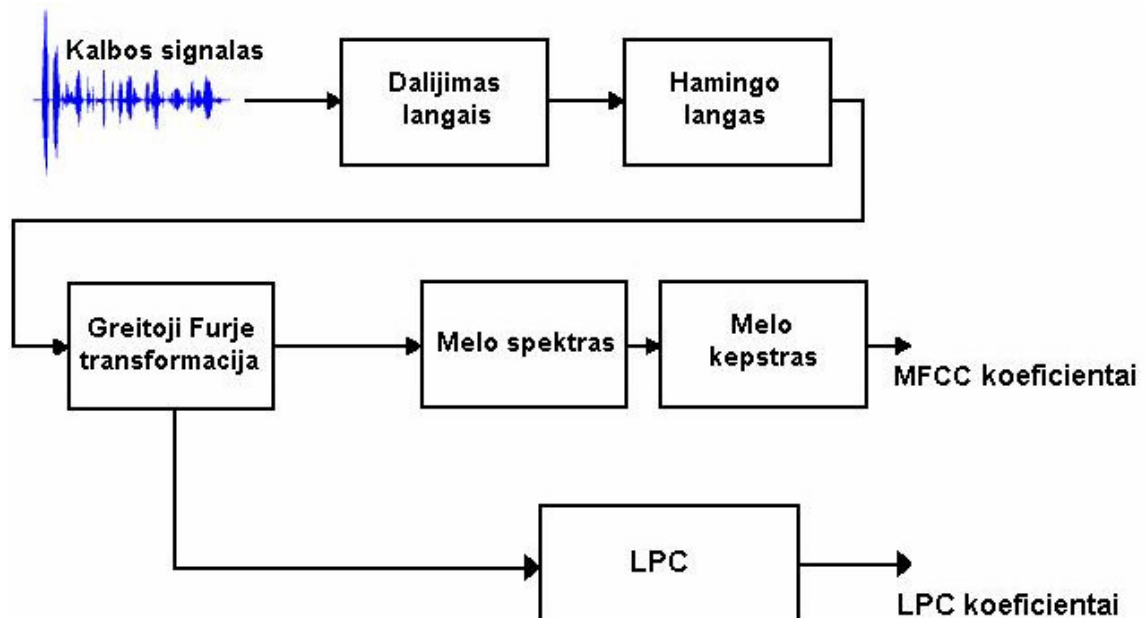
2.2. Fonemas charakterizuojančių koeficientų gavimo eiga

Žmogaus balsas keičiasi laikui bėgant, priklauso nuo emocinės būsenos, fonogramos įrašymo patalpos ir aibės kitų faktorių. Dėl šių priežasčių stengiamasi surasti tokius kalbos signalo arba jo trumpų fragmentų (fonemų) požymius, kurie mažai priklausytų nuo išvardintų kalbančiojo ir įrašymo sąlygų kintamų faktorių.

Garsų atpažinimui naudojamas ne garso akustinis vektorius, o kompaktiškesnis požymių rinkinys, išskiriamas iš akustinio vektoriaus. Nuo to kaip gerai požymių rinkinys charakterizuoja garsą labai stipriai priklauso atpažinimo tikslumas. Galimi įvairūs požymių rinkiniai. Dažniausiai naudojami – LPC (tiesinės prognozės) ir MFCC (Melo dažnių skalės kepriniai koeficientai) koeficientai. Tačiau galimi įvairūs kitokių požymių rinkiniai bei mišrūs rinkiniai, kombinuojantys įvairius požymius ir jų gerąsias savybes.

Savo atliekamame šnekos atpažinimo eksperimente naudosime lietuvių kalbos fonemas, kurių viso yra 65. Fonemomis vadinami pirminiai šnekamosios kalbos vienetai. Šnekamosios kalbos fonemų rinkinys yra rašytinės kalbos raidžių vienetų rinkinio, vadinamo abėcėle, analogas. Rašytinį tekstą, naudojant pasirinktų (sukurtų) taisyklių rinkinį, galima transformuoti į fonetiniais vienetais užrašomą tekstą. Rašytinės kalbos transformavimas į fonetiniais vienetais užrašytą kalbą vadinamas transkribavimu. Idealiu atveju atpažįstant šnekamąją kalbą reikia kalbos signalą transformuoti į fonetinių vienetų įrašą, kurį galutiniame kalbos atpažinimo etape reikia transformuoti į rašytinį tekstą. Mes savo darbe kaip fonemas charakterizuojantį parametą pasirinkome Melo dažnių skalės koeficientus (MFCC) ir LPC koeficientus.

Fonemas charakterizuojančių koeficientų gavimo eiga, siekiant iš kalbos signalo gauti MFCC ir LPC koeficientus pateikiama 2.5 pav.



2.5 pav. Struktūrinė schema skirta MFCC ir LPC koeficientams iš kalbos signalo gauti.

Kalbos signalas iš mikrofono patenka į kompiuterio garso plokštę. Šiame įrenginyje analoginis signalas paverčiamas skaitmeniniu. Natūralios žmogaus kalbos dažnis išsitenka diapazone apytikriai 100–8000 Hz, o pagal *Nyquist–Kotelnikov* teoremą analoginis signalas pilnai atstatomas iš diskretizuoto įrašo, jeigu diskretizacijos dažnis nemažesnis nei $2 \times F_{\max}$, kur F_{\max} – didžiausias analoginio signalo dažnis. Pagal šią teoremą reikėtų naudoti ~16kHz

diskretizavimą, bet visuotinai priimta, kad 11 kHz diskretizavimas pakankamas kalbos atpažinimui.

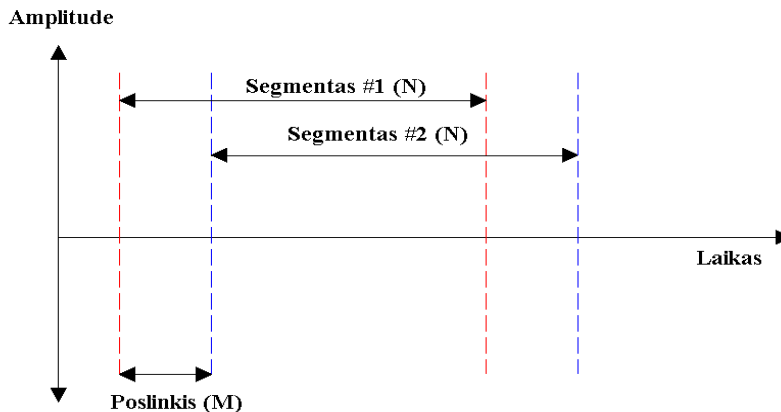
Tam, kad atpažinimo procesui neturėtų įtakos kalbėtojo kalbos garsumas, garso signalą prieš dalijant į atskirus langus reikia normalizuoti. Garso normalizavimas atliekamas pagal maksimalią amplitudę

Kalbos signalas paprastai yra gana dinamiškas, o daugelis požymių išskyrimo funkcijų skirtos sąlyginai statiniam, t.y. stipriai nekeičiančiam savybių signalui, todėl garsas yra segmentuojamas nedidelės trukmės atkarpomis, taip vadinamais langais (N diskrečių reikšmių), kuriose signalas yra sąlyginai statiškas, o segmentams išskirti požymiai apjungiami į bendrą požymių vektorių. Kad neprarasti garso savybių kitimo požymių kitas segmentas skaičiuojamas ne nuo praeito segmento galo, o nuo praeito segmento pradžios + poslinkis, mažesnis už segmento ilgį ($M < N$). 2.6 pav. pavaizduota kaip vykdomas signalo dalijimas langais.

Visus kalbos garsus, kuriuos galime tarti norimai ilgai (pvz. *a*, *s*, *m*, *ž*) atitinka stacionarūs fonogramos intervalai. Tariant tokius garsus trumpam yra užfiksuojama kalbos trakto organų padėtis, dėl ko yra galimybė naudojantis fonograma “išmatuoti” kalbos trakto parametrus ir pagal gautus parametrus atpažinti šį fragmentą atitinkantį kalbos fonetinį vienetą. Nestacionarius fonetinius vienetus (pvz. *k*, *p*, *b*) taip pat atitinka fiksuota kalbos trakto padėtis, todėl jiems atpažinti, nors ir ne taip sėkmingai, galima taikyti tą pačią metodiką (t.y. nestacionarių garsų fonogramos fragmentus laikysime stacionariais ir jų parametrus/požymius išskirsime pagal tą pačią metodiką taikytą lokalizuotų stacionarių garsų atveju).

Stacionarių intervalų suradimui pritaikysime dažnai kalbos atpažinimo uždaviniuose naudojamą metodą. Pagal jį sulyginami du greta esantys kalbos signalo kadrų (segmentai), paskaičiuojant tikėtino santykio atstumą tarp jų. Kai atstumas viršija eksperimentiškai iš anksto pasirinktą slenkstį, pranešama apie stacionaraus intervalo pabaigą.

Kai mes norime atpažinti fonemą, darome prielaidą, kad turime nežinomo fonetinio vieneto fonogramą ir n žinomų fonetinių vienetų fonogramų. Fonemų atpažinime svarbu ar atpažystama fonema ir žinomų fonemų įrašai yra vieno asmens ar ne. Paprastumo dėlei laikysime, kad ir viena ir kita yra vieno asmens, kitaip tariant yra atpažystamos vieno diktoriaus fonemos. Atpažinimo tikslas yra išrinkti iš lyginamųjų fonemų įrašų egzempliorių, kuris yra kokias nors prasme “artimiausias” tiriamajam.



2.6 pav. Signalų skaidymo langais pavyzdys.

Atlikus kalbos signalo dalijimą į atskirus langus, signalas praleidžiamas pro Hammingo langą. Po to kiekvienam langui atliekama greitoji Furje transformacija. Jos principas bus aprašomas vėliau.

2.2.1. Lango funkcijos

Fonemų skaitmeninius signalus stengiamasi glaustai aprašyti nedideliu kiekiu skaitinių parametrų, kurie apskaičiuojami iš signalo. Gauti parametrai vadinami požymiais, kurie ir yra pateikiamia atpažinimui.

Prieš pereidami prie fonemų skaitmeninių signalų konkrečių požymių aprašymo, susipažinsime su klasikiniu pirminiu signalo apdorojimu, vadinamu dauginimu iš lango funkcijos. Šio veiksmo esmė labai paprasta – kiekviena skaitmeninio signalo reikšmė yra dauginama iš fiksuotos lango funkcijos daugiklio ("svorio"), kurio reikšmė priklauso nuo taško indekso. Lango funkcija parenkama taip, kad svoriai būtų maži neneigiami skaičiai arti signalo kraštų ir arti vieneto indekso reikšmėms, kurios yra arti signalo centro. Tokiu būdu yra sprendžiama taip vadinama "krašto" problema, nes fonemos skaitmeninis signalas "išpjaunamas" iš kalbos signalo ir todėl kraštinės reikšmės yra nelabai natūralios fizikine prasme. Pavyzdžiui mažiausią indeksą atitinkanti reikšmė gali būti didelė, nors natūralūs signalai kaip taisyklė prasideda nuo artimomis nuliui reikšmėmis. Taip pat dažnai fonemų kraštinės reikšmės yra įtakojamos gretimų fonemų reikšmėmis, dėl ko padidėja tikimybė fonemą atpažinti neteisingai.

Tarkime skaitmeninės fonemos reikšmės yra:

$$X_0, X_1, X_2, \dots, X_{N-1},$$

o lango "svorius" (angl. *weights*) žymėsime:

$w_0, w_1, w_2, \dots, w_{N-1}$. Tuomet trikampis, Hanningo ir Hamingo langai bus apibrėžti tokiomis formulėmis:

Trikampis (kai N nelyginis):

$$w_n = \min(n, N-1-n) / ((N-1)/2)$$

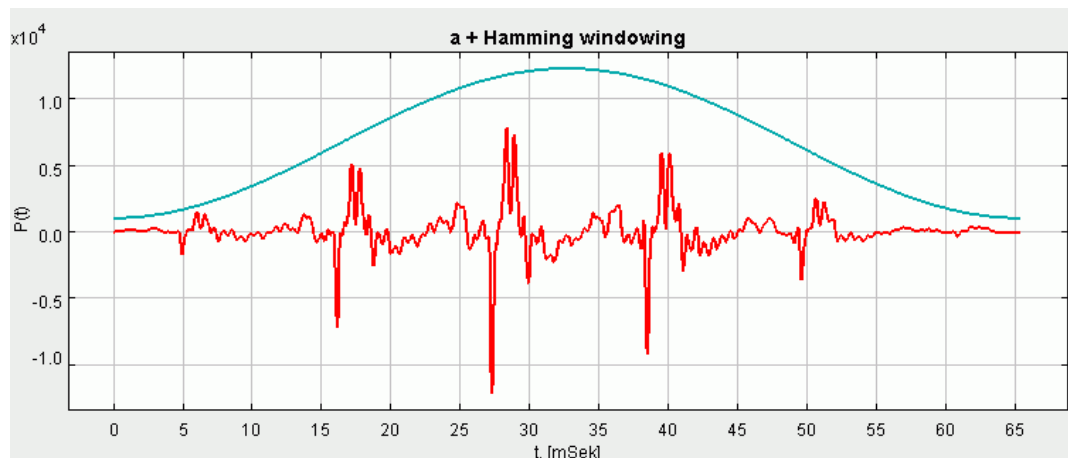
Hanningo:

$$w_n = 0.5(1 - \cos(2\pi n/(N-1)))$$

Hamingo:

$$w_n = 0.54 - 0.46\cos(2\pi n/(N-1))$$

2.7 pav. iliustruoja skaitmeninį fonemos "a" apdorojimą Hamingo langu. Hamingo langas dažniau taikomas trumpiems signalams (iki 0.1 sek), o Hanningo - ilgesnės trukmės signalams.

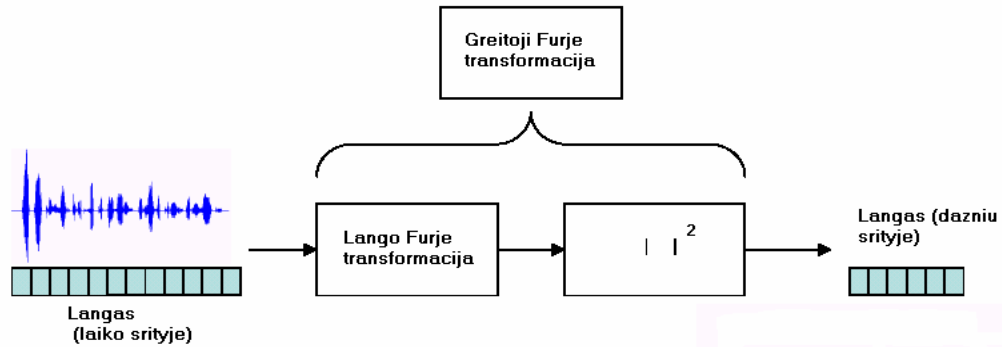


2.7 pav. Balsio "a" signalas apdorotas Hamingo langu.

2.2.2. Furjė transformacija

Tam, kad signalą išskaidyti į jo dažnines dedamasias naudojama diskretinė Furjė transformacija. Bendru atveju transformacija perverčia N atitinkamų reikšmių amplitudės signalą į $N/2+1$ koeficientų poras $Re(n)$, $Im(n)$. Pagal formules galima transformuoti koeficientų poras – amplitudės \sin ir \cos – į kitą labiau naudingą mums porą – amplitudę, \sin $A(n)$ ir šios sinusoidės fazę $Ph(n)$.

2.8 pav. matoma greitosios Furje transformacijos eiga.



2.8 pav. Greitosios Furje transformacijos principas.

Tokiu būdu, Furje transformacija perveda N amplitudės reikšmių į $N/2$ sinusoidės (atskirus dažnius) su amplitudėmis $A(n)$ ir fazėmis $Ph(n)$, plus dar nulinių koeficientų porą – paprastą konstantą (sin ir cos su pastoviu nuliniu argumentu).

Paprastas pavyzdys, išreikštas skaičiais. Tarkime mes turime audio signalą, kurio juosta 44,1 kHz, paėmę aštuonias reikšmes (atskirų atskaitų), norime gauti dažninį padalijimą. Rezultate gausis 5 poros koeficientų – 4 sinusoidės su dažniais 22,05 kHz, 14,7kHz, 11,025 kHz, 8,82 kHz, ir sinusoidė su dažniu 0 kHz, - tai yra konstanta.

Diskretinės Furje transformacijos sprendimą galima įgyvendinti dviem būdais. Pirmo standartinio būdo esmė ta, kad reikia padauginti pradinę funkciją iš sinuso ar kosinuso, sumuoti visą diskretinę funkciją į vieną skaičių ir rezultatą užrašyti koeficientais. Dar šį metodą vadina koreliacijos metodu, jeigu funkcija koreliuoja su ieškomais sinusais arba kosinusais, tai suma visų diskretinių atskaitų mums sako, kiek bendro tarp jų buvo. Šis metodas gana lėtas ir naudojamas norint išspręsti nedideles transformacijas apie 10 – 20 taškų.

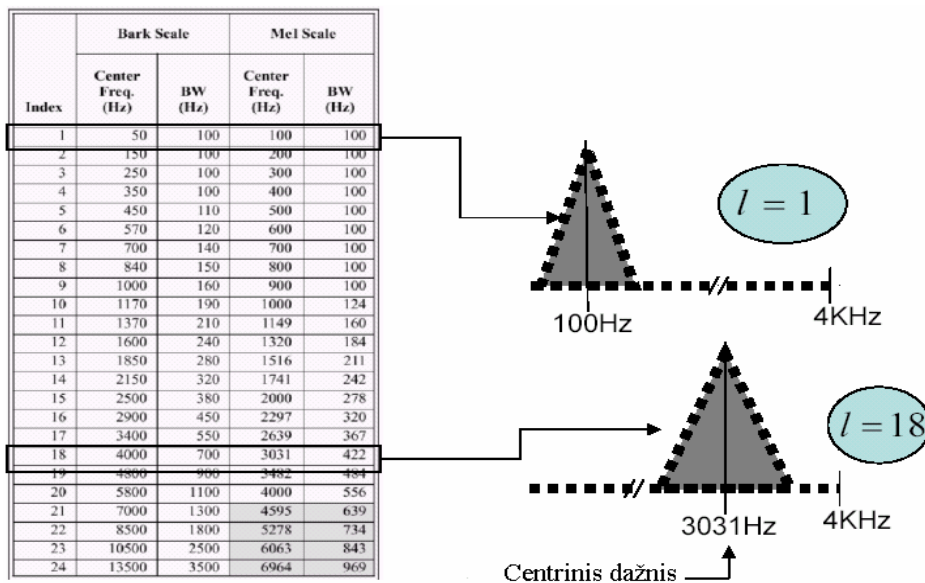
Dėl proceso pagreitinimo egzistuoja gudrus algoritmas pavadintas greitąja Furje transformacija, angl. – FFT. FFT dirba su kompleksiniais skaičiais ir transformacijų dydžiais, susidarančiais iš dvejeta laipsniu (128,...,1024, 2048 ir t.t.). Nereikia galvoti, kad FFT kažkas kita negu Furje padalijimas. Tai visiškai tas pats tik šimtus kartų greičiau. Kompleksinai koeficientai – tai ne kas kita, kaip koeficientai prieš $\cos(\text{Im}[n])$, realieji prieš sin. Daugumoje šiuolaikinių algoritmų naudojama FFT, dėl to šitas pavadinimas tvirtai įsitvirtino už visų algoritmų, kurie išskaido signalą į dažnius.

Svarbios Furje transformacijos savybės:

- Realiai išskaidžius ir atgal sudėjus signalą, niekada nieko neprarandame. Tai labai svarbi savybė.
- Dažninis išsprendimas priklauso nuo transformacijos dydžio ir sudaro pusę nuo šito dydžio. Jei $FFT = 512$, rezultate mes gauname amplitudes ir fazes 256 – ių tolygiai išsidėsčiusių dažnių.
- Signalo pervedimas į dažninį pavidalą galimas tiktais blokais (langais).

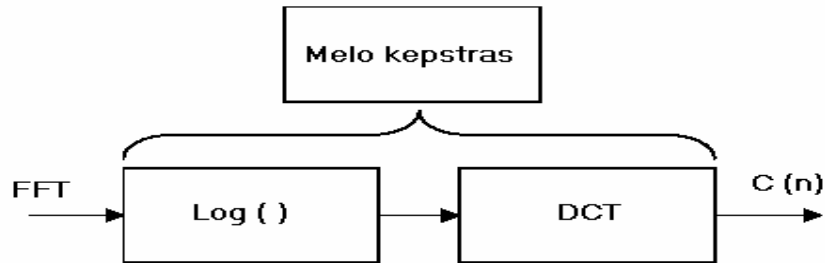
2.2.3. MFCC koeficientų gavimas

Bet kuris kalbos signalas turi informacijos perteklių. Atlikus Furje transformaciją informacijos perteklius nėra pašalinamas. Tam, kad signalą koduoti mažesniu bitų skaičiumi yra naudojami filtrų bankai. Naudojant filtrų bankus visa dažnių juosta yra dalijama į mažesnes juostas ir surandama kiekvienai juostai tenkanti energija. Buvo nustatyta, kad daugiausia energijos tenka žemiems dažniams, dėl to pasiūlyta naudoti Melo dažnių skalę. Konstruojant filtrų banką Melo skalėje, ji yra dalijama į aštuonias ar daugiau dalių. Užsienio literatūroje filtrų skaičių siūloma pasirinkti pagal naudojamą dažnių juostą (2.9 pav.), tačiau praktikoje užtenka naudoti ir mažesni filtrų banką. Pavyzdžiui 4000 Hz juostai sudaryti aštuonių filtrų banką.



2.9 pav. Lentelė Melo dažnių skalės filtrų banko sudarymui.

Signalui praėjus pro filtrų banką yra panaikinas informacijos perteklius ir jau galima iš Melo spektro gauti Melo kepstrą, o po to ir MFCC koeficientus. Tam atliekami veiksmai pavaizduoti 2.10 pav.



2.10 pav. Melo kepstro struktūrinė schema.

Vietoj atvirkštinės Furje transformacijos paskutiniame lange parodyta DCT. **DCT: Discrete Cosine Transform** (Diskretinė kosinuso transformacija). Šis metodas turi tam tikrų pranašumų prieš atvirkštinę Furje transformaciją:

- Įgyvendinti atvirkštinę Furje transformaciją reikalinga sudėtinga aritmetika
- DCT aritmetika paprasta
- DCT apibrėžia ta pačią funkcija kaip ir FFT tik tiksliau lyginant su pradiniu signalu
- DCT labiau efektinga MFCC skaičiavime.

Iš kalbos signalo gavus MFCC koeficientus iš jų galima paskaičiuoti požymių vektorių x_k , po L požymių kiekvienam kadrai. Taip gaunami požymių vektoriai $x_k = [x_{k1}, x_{k2}, \dots, x_{kL}]$, kur $k = 1, 2, \dots, K$, K – langų skaičius. Požymių vektorių x_k negalima tiesiogiai panaudoti neuroninio tinklo mokymui, nes jų skaičius kiekvienam signalui yra skirtingas. Buvo panaudotas k -vidurkių klasterizavimo algoritmas. Vektoriai klasterizuoti į C klasterių ir kaip fonemos požymiai naudojam gautų klasterių centrai $c_i = [c_{i1}, c_{i2}, \dots, c_{iL}]$, kur $i = 1, 2, \dots, C$. Tokiu būdu radus panašiausius klasterių atitikmenis vykdomas kalbos atpažinimas ir konkrečią fonemą atitinkanti raidė išvedama į kompiuterio ekraną.

2.2.4. Pagrindinė LPC metodo idėja

LPC metode daroma prielaida, kad kalbos signalas yra inicijuojamas garso šaltinio kintamo skerspjūvio vamzdelio viename iš galų. Pas žmogų garso šaltinis yra virpančios balso

stygos. Šis šaltinis charakterizuojamas intensyvumu (garsumu) ir dažniu (pagrindiniu tonu). Gerklės ir burnos kalbos traktas formuoja kintamo skersmens vamzdelį. Kalbos traktas charakterizuojamas rezonansiniais dažniais, vadinamais formantėmis.

Tiesinės prognozės metodu yra įvertinamos formantės, atskiriant jas nuo kalbą generuojančio šaltinio, lemiančio kalbos garsumą ir toną. Formančių eliminavimas iš kalbos signalo kartais vadinamas atvirkštiniu filtravimu, ir likęs signalas vadinamas liekamuoju.

Skaičiai aprašantys formantes ir liekamasis signalas gali būti saugojami ir siunčiami atskirai. Kalbos sintezavimas tiesinės prognozės metodu gaunamas apgręžiant metodą: liekamasis signalas tampa kalbos signalą generuojančiu šaltiniu, o iš formančių gaunamas filtras, atspindintis kalbos trakto geometriją. Filtruojant liekamąjį signalą ir gaunamas sintezuotas garsas.

2.2.5. Formančių įvertis

Pagrindinė LPC metodo problema yra įvertinti formantes remiantis įrašytu kalbos signalu. Metodu tikslas yra gauti tam tikrą diferencialinę lygtį, kurios esmė yra išreikšti kuo tiksliau eilinę garso slėgio imtį remiantis keliom prieš tai žinomom kalbos signalo reikšmėm. Kadangi naudojama tiesinė aproksimacija, tai gaunama skirtuminė diferencialinė lygtis yra vadinama tiesine prognoze, o pats signalo kodavimo metodas vadinamas tiesinės prognozės metodu.

Diferencialinės išraiškos koeficientai vadinami prognozės (tiesinės) koeficientais. Šie koeficientai aprašo formantes, taigi metodo esmė yra apskaičiuoti tiesinės prognozės koeficientus. Formulės apskaičiuoti LPC koeficientus yra gaunamos minimizuojant liekamojo signalo vidutinę kvadratinę paklaidą.

Rezultate yra gaunama tiesinė algebrinių lygčių sistema prognozės koeficientų atžvilgiu. Praktiškai tenka spręsti tokias problemas:

- Apskaičiuoti pagal kalbos signalo fragmento duomenis tiesinių lygčių sistemos matricos koeficientus ir laisvuosius narius.
- Išspręsti gautą tiesinę lygčių sistemą.

Naudojami keli matricos koeficientų gavimo metodai (autokoreliacijos, kovariacijos, rekursyvinės gardelės), kurie garantuoja sprendinio vienatį ir skaičiavimų efektyvumą.

Niuansai: vamzdelis nelygus vamzdeliui. Gali atrodyti keista, kad kalbos signalas gali būti aprašytas tokiu paprastu vamzdelio modeliu. Tam, kad toks paprastas modelis galėtų, vamzdelis turi neturėti nei vieno išsišakojimo. Kiekviena šaka aprašoma antirezonansiniais dažniais, kurie pasižymi tam tikro ilgio bangų slopinimu. Matematinė prasme vamzdelį su išsišakojimais geriau aprašinėti tokiu modeliu, kuris leistų reguliuoti signalo nulius spektro srityje.

Įprasti balsiai gerai aprašomi paprastu kintamo skerspjūvio vamzdeliu. Tačiau nosiniai garsai, kurių suformavime aktyviai dalyvauja šoninis nosies ertmės kanalas, reikalauja kiek sudėtingesnio modelio. Tačiau praktiškai į tai dažnai neatsižvelgiama ir problemos sprendimas nukeliamas į liekamojo signalo valdymą.

2.3. Fonemos samprata

Kalbant apie fonemų atpažinimą pirma reikėtų išsiaiškinti fonemos termino prasmę. Mes nuo mokyklos laikų žinome *raidžių* (*a, q, b, c, ... ,ž*) prasmę. Raidės naudojamos rašytine forma užrašyti kalbą, mintis ir panašiai. Fizinė prasme kalba susidaro kalbos padargais generuojant atmosferinio slėgio mikropokyčius, kurie plinta oru ir yra priimami bei suvokiami žmogaus klausos aparatu.

Fonema yra vadinama neskaidoma elementari kalbą sudarančio garsinio signalo dalis. Panašiai kaip ir raidžių atveju, kiekviena kalba turi jai būdingų fonemų (*garsų*) baigtinį sąrašą, kurį galima vadinti "fonemų abėcėle". Lietuvių kalbai fonemų abėcėlė kol kas nėra standartizuota, tačiau kuriant kalbos sintezavimo ir atpažinimo sistemas autoriai sudaro tokių fonemų sąrašus.

2.3.1. Lietuvių kalbos balsių fonemos

Balsiai sudaro maždaug trečdalį visų kalbos sraute vartojamų garsų. Kitaip sakant, vidutiniškai kas trečias lietuvių bendrinės kalbos garsas yra vienabalsis (neskaitant dvigarsių dėmenų). Ypač dažni trumpieji – neįtemptieji balsiai [i], [a], taip pat ilgasis – įtemptasis [o']. Šie trys balsiai savo dažnumu pralenkia visus kitus draugėn sudėtus devynis balsius. Retai kalboje pasitaiko ilgasis – įtemptasis [e'] ir ypač retai trumpieji – neįtemptieji [o], [e].

Akustiškai balsiai klasifikuojami pagal tembro aukštį, spektro sklaidą, bemoliškumą, trukmę ir įtempimą.

Tembro aukštis. Net ir be prietaisų, vien klausia galime nustatyti, kad balsis [u'] yra daug žemesnio tembro negu balsis [i']. Tariant užpakalinės eilės balsius liežuvis būna atsitraukęs atgal, burnos ertmė pasidaro palyginti didelė, plati ir rezonuoja žemu tembru. Todėl visi užpakalinės eilės balsiai [u'], [u], [o'], [o], [a'], [a] yra žemo tembro. Priešakinės eilės balsiai [e'], [e],[i'], [i], yra aukšto tembro.

Balsių tembro aukštį rodo antrosios formantės F_2 padėtis spektre. Žemo tembro balsių F_2 yra žemųjų dažnių (500 – 1500 Hz), o aukšto tembro balsių F_2 – aukštųjų dažnių (1500 – 2500 Hz) srityje. Pats žemiausias balsis yra [u'], o aukščiausias [i']. Tarp šių balsių pagal tebrą išsidėsto visi kiti balsiai. Žemo ir aukšto tembro balsių skiriamoji riba yra tarp [a'] ir [e'], [a] ir [e].

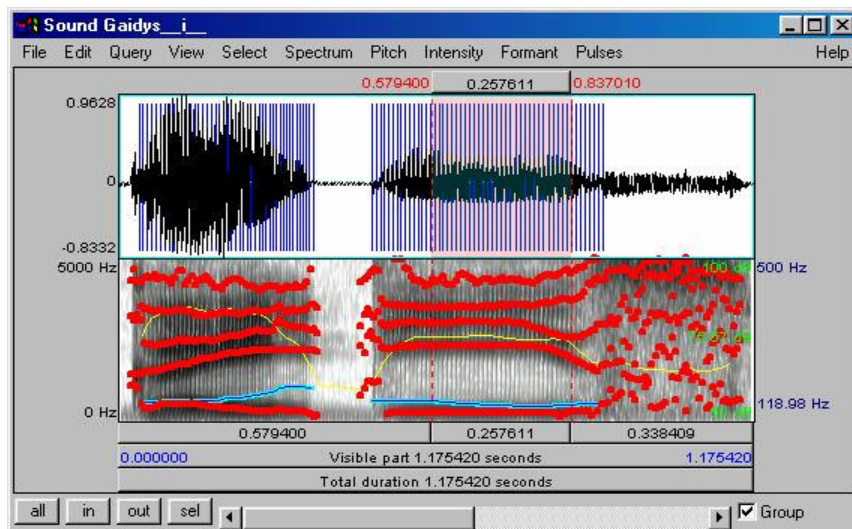
Spektro sklaida. Antrosios formantės kitimo diapazonas yra platus. Aukštutinio pakilimo balsių F_2 yra viename spektro pakraštyje, o žemutinio pakilimo balsių F_2 – priartėjusi prie spektro vidurio. Vidutinio pakilimo balsių F_2 užima tarpinę padėtį. Pirmoji formantė F_1 priklauso nuo žmogaus ryklės ertmės. Tariant aukštutinio pakilimo balsius liežuvis būna pakilęs aukštyn, todėl ne tik balsių [u'], [u], bet ir [i'], [i] pirmoji formantė yra žemųjų dažnių (200 – 300 Hz) srityje. Artikuliuojant žemutinio pakilimo balsius [a'], [a], [e'], [e] liežuvis nusileidžia žemyn. Tokia ryklės ertmė lemia palyginti aukštą šių balsių pirmąją formantę, tačiau ji nebūna aukštesnė kaip 900 Hz.

Atsižvelgiant į abiejų formančių padėtį galima nustatyti balsių spektro sklaidą. Aukštutinio pakilimo balsių F_1 ir F_2 yra pačiuose spektro pakraščiuose (abi formantės pačių žemųjų dažnių srityje arba pirmoji formantė pačių žemųjų, o antroji formantė aukštųjų dažnių srityje).

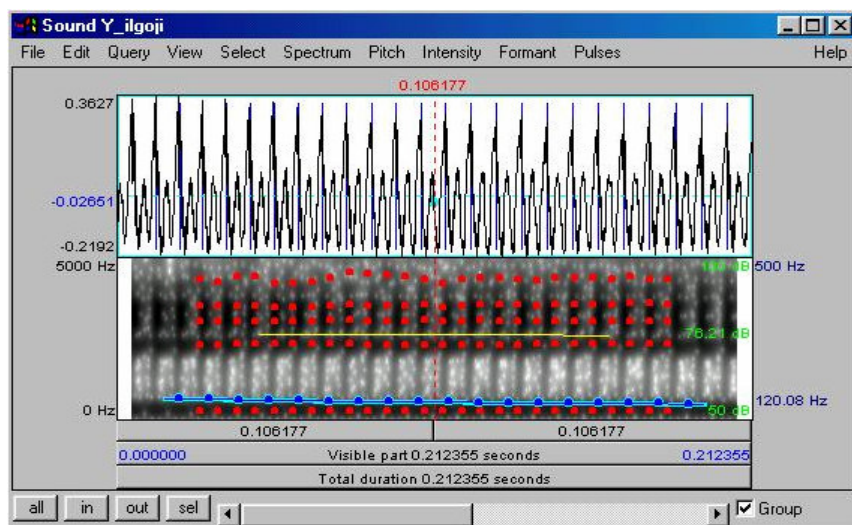
Trukmė ir įtempimas. Balsių trukmės skirtumus galima aiškinti paprastai: panašiomis sąlygomis ilgieji balsiai skamba maždaug 2 kartus ilgiau negu atitinkami trumpieji balsiai.

Lietuvių kalboje yra 12 balsių garsų. Bendrinėje kalboje yra 12 raidžių balsių garsams žymėti. Tačiau šis skaičių sutapimas nereiškia idealaus garsų ir raidžių atitikimo. Kai kurie balsiai žymimi dviem skirtingomis raidėmis, ir, kita vertus kartais ta pati raidė reiškia skirtingus balsius.

2.11 pav. Matomas žodžio “gaidys” signalas Praat lange. Iš šio žodžio buvo iškirpta “y ilgoji” fonema. Jos signalas matomas 2.12 pav. Kitoms balsių fonemoms gauti taip pat buvo įdiktuoti žodžiai, kuriuose balsinių fonemų sąvybės geriausiai atsispindi.



2.11 pav. Lietuvių kalbos žodžio “gaidys” signalas PRAAT lange.



2.12 pav. Fonemos “Y ilgoji” signalas PRAAT lange.

12 balsių garsų, lietuvių bendrinėje kalboje funkcionuoja kaip savarankiškos fonemos.

Kiekviena fonema yra fonologiškai savarankiška viena nuo kitos.

- 1). [a] – fonemą atitinkanti raidė: a; (**a**bu);
- 2). [aʹ] – fonemą atitinkančios raidės: a, a; (**ra**mūs, **dra**šūs);
- 3). [e] – fonemą atitinkanti raidė: e; (**ge**resni);
- 4). [eʹ] – fonemą atitinkančios raidės: e, e; (**te**sė, **ge**šta);
- 5). [eʹ] – fonemą atitinkanti raidė: e; (**sė**dėti);
- 6). [i] – fonemą atitinkanti raidė: i; (**i**ma);
- 7). [iʹ] – fonemą atitinkančios raidės: y, i; (**i**monė, **gai**dys);

8).[o] – fonemą atitinkanti raidė: o; (dėsnīgumas galioja tik tarptautiniuose žodžiuose. Pvz: motoras);

9). [o'] – fonemą atitinkanti raidė: o; (kovoti);

10).[ê] – fonemą atitinkanti raidė: e; (dėsnīgumas galioja tik tarptautiniuose žodžiuose. Pvz: epas, foje);

11). [u] – fonemą atitinkanti raidė: u; (upė);

12). [u'] – fonemą atitinkančios raidės: ū,ų; (skųsti, lųžo);

2.3.2.Lietuvių kalbos priebalsių fonemos

Priebalsių akustinį pagrindą sudaro šlamesiai, kurie kyla, stipriai oro srovei veržiantis pro bet kurioje kalbos trakto dalyje susidariusią uždarumą ar ankštumą.

Visų priebalsių artikuliacijai yra būdinga:

a) kalbos padargų uždaruma ar ankštuma, pro kurią eidama oro srovė sukelia šlamesį;

b) labai stipri (palyginti su balsiais) iš plaučių iškvepiama oro srovė, tariant dusliuosius, ir kiek silpnesnė, tariant skardžiuosius priebalsius;

c) kalbos padargų įtempimas, lokalizuotas toje vietoje, kur susidaro uždaruma ar ankštuma

Lietuvių literatūrinė kalba turi 45 priebalsines fonemas: *b, b', c, c', ch, ch', č, č', d, d', dz, dz', dž, dž', f, f', g, g', h, h', j, k, k', l, l', m, m', n, n', p, p', r, r', s, s', š, š', t, t', v, v', z, z', ž, ž'*. Ženklu ' minkštoji fonema atskiriama nuo kietosios. Paprastai priebalsių fonemos sudaro kietųjų ir atitinkamų minkštųjų poras: *p:p', b:b', t:t', d:d', k:k', g:g' c:c', dz:dz', č:č', dž:dž', m:m', n:n', s:s', š:š',z:z', ž:ž', f:f', v:v', ch:ch', h:h', l:l', r:r'*. Fonema *j* neturi kietosios poros.

Kietumas ir minkštumas yra vienas pagrindinių lietuvių literatūrinės kalbos priebalsių sistemos fonologinių požymių. Kietosios ir minkštosios priebalsinės fonemos, einančios prieš užpakalinės eilės balsius ir dvibalsius yra fonologiškai priešpastatomos ir šiais atvejais skiria žodžius, jų formas ir morfemas, pvz.: kumpo:kumpio, lobo:lobio ir t.t.

Priebalsiai klasifikuojami pagal šlamesio židinio sudarymo būdą ir aktyviųjų kalbos padargų (balso stygų, lūpų ir liežuvio) veikimą.

Šlamesio židinį gali sudaryti kalbos padargų uždaruma, ankštuma arba kalbos padargų virpėjimas. Pagal šlamesio židinio sudarymo būdą priebalsiai skirstomi į 1) uždarumas, 2) ankštumas, 3) virpamuosius priebalsius.

Uždarumos priebalsiai yra šie : *p, p', b, b', t, t', d, d', k, k', g, g', c, c', dz, dz', č, č', dž, dž', m, m', n, n'*.

Ankštumos priebalsiai yra: *v, v', f, f, s, s', š, š', z, z', ž, ž', l, l', ch, ch', h, h', j*

Virpamieji yra sonantai: *r, r'*.

Uždaromos priebalsiai skirstomi į tris grupes: a) sprogstamuosius (eksplozyvinius), b) afrikatas ir c) nosinius sonantus.

Ankštumos priebalsiai skirstomi į dvi grupes: a) pučiamuosius (frikatyvinius) ir b) šoninius sonantus

Pagal kitų aktyviųjų kalbos padargų veikimą lietuvių literatūrinės kalbos priebalsiai skirstomi į šias grupes: 1) lūpinius, 2) liežuvio priešakinius, 3) liežuvio vidurinius, 4) liežuvio užpakalinius.

Pagal artikuliacijos metu susidarančių šlamesio židinių skaičių lietuvių literatūrinės kalbos afrikatos ir pučiamieji priebalsiai skirstomi i 1) vienžidinius ir 2) dvižidinius.

Pagal balso stygų veikimą visi lietuvių kalbos priebalsiai skirstomi į skardžiuosius *b, b', d, d', g, g', dz, dz', dž, dž', m, m', n, n', z, z', ž, ž', l, l' r, r', v, v', j, h, h'*, tariamus suartėjusiomis, įtemptomis ir virpančiomis balso stygomis, ir dusliuosius *p, p', t, t', k, k', c, c', č, č', s, s', š, š', f, f, ch, ch'*, tariamus nesuartėjusiomis, neįtemptomis ir nevirpančiomis balso stygomis.

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

Su priebalsiais juos sieja šie bruožai:

a) sonantai, kaip ir kiti priebalsiai, tariami įtemptiant, nors žymiai silpniau, tik tą kalbos padargų dalį, kurioje susidaro uždaruma arba ankštuma, todėl sonantų būdingieji tonai nėra tokie aiškūs, kaip balsių;

b) sonantai, kaip ir visi kiti priebalsiai, turi aiškiai lokalizuotą uždarumą arba ankštumą.

Į balsius jie panašūs šiais bruožais:

a) daug silpnesne, negu kitų priebalsių, iškvepiama oro srove ir nedideliu šlamesių kiekiu;

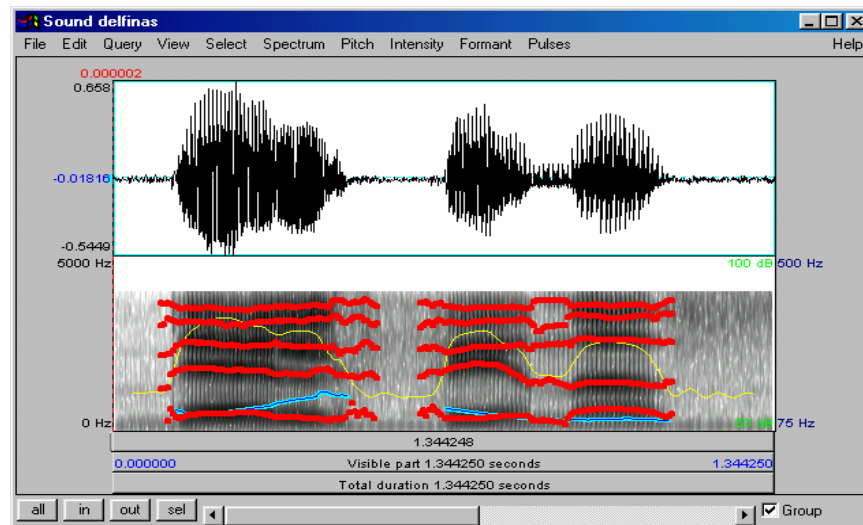
b) nepaprastu toningumu - lietuvių literatūrinėje kalboje jie visi yra skardūs ir neturi atitinkamų dusliųjų porų;

c) kaip ir balsiai, jie niekada neduslėja, o priebalsiai prieš juos niekada neskardėja, pvz.: *dvelkti, išrovė (ne išrovė)*;

d) eidami mišriųjų dvigarsių antraisiais dėmenimis, sonantai tariami su tono aukštumo moduliacijomis ir įvairiose fonetinėse pozicijose esti nevienodo ilgumo.

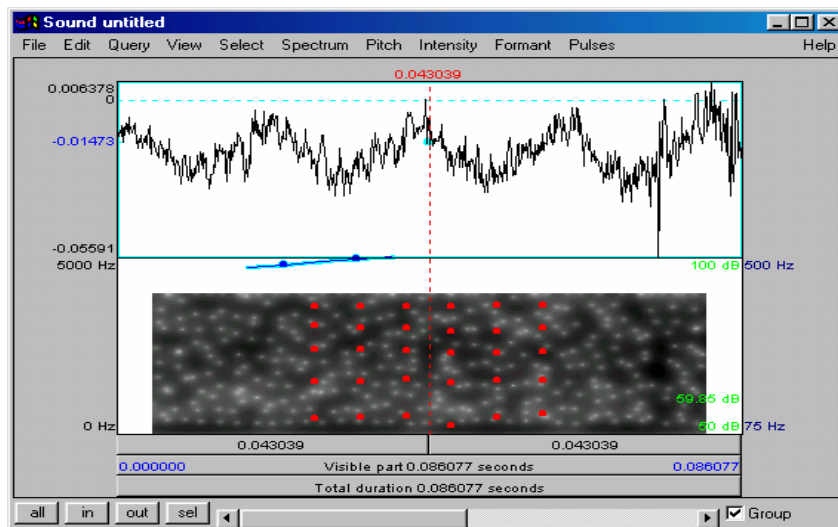
Tam, kad išskirti priebalsių fonemas buvo įrašyti žodžiai, kuriuose ,anot kalbininkų, labiausiai atsispindi fonemų sąvybės. Įrašyti žodžiai ir iš jų iškirpta priebalsių fonemą:

Ūpas – p, upė – p', troba – b, grobis – b', matas – t, vertimas – t', radau – d, didelis – d', sakau – k, sakymas – k', auga – g, daugybė – g', caras – c, procesas – c', dzandzaliuoti – dz, dzindzėti – dz', ginčas – č, svečias – č', džandža – dž, medžio – dž', namas – m, raminti – m', vienas – n, kišenė – n', elfas – f, delfinas – f', gyvas – v, ravėti –v', kasa – s, klausyti – s', lazda – z, zyzia – z', viršus – š, rašiau – š', dažau – ž, vežimas – ž', jauja – j, mechanizmas – ch, psichika – ch', buhalteris – h, farmaldehidas – h', aulas – l, galingas – l', skara – r, daryti – r'.

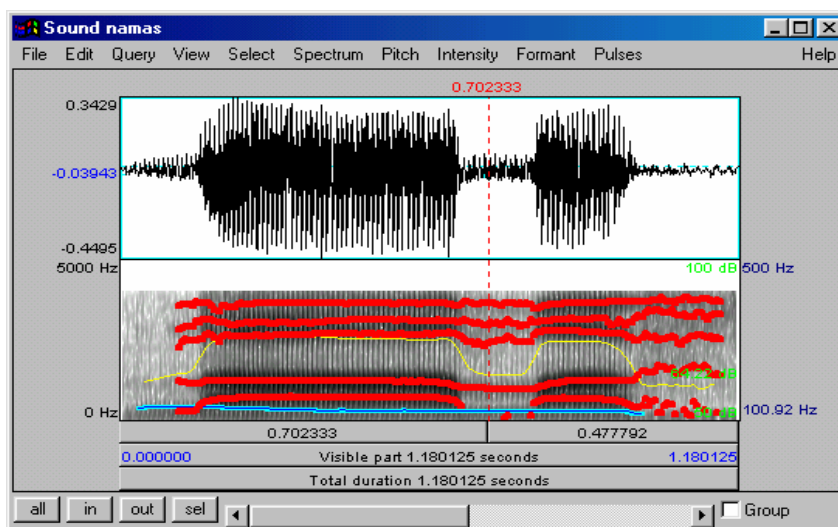


2.13 pav. Lietuvių kalbos žodžio “delfinas” signalas PRAAT lange.

Fonemų išskyrimą ir analizavimą atlikome Praat programa, paveikslėliuose pateikiama keletas darbo pavyzdžių. 2.13 pav. matome žodžio delfinas analizės langą. Viduryje esantis “trūkis” apytiksliai apibrėžia fonemos f ‘ ribas. F yra duslusis priebalsis, jo susidaryme nedalyvauja balso stygos todėl jis neturi pagrindinio dažnio.

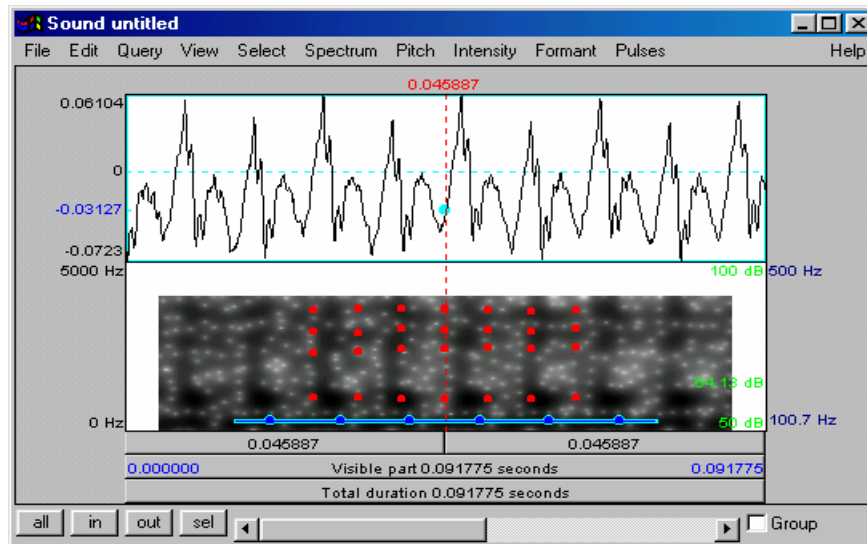


2.14 pav. Iš žodžio delfinas iškirptos fonemos f' signalas PRAAT lange.



2.15 pav. Lietuvių kalbos žodžio "namas" signalas PRAAT lange.

2.15 pav. pavaizduotas žodžio "namas" signalas. Tarp aiškiai pastebimų balsių "a" esantis tarpas apibrėžia fonemos "m" ribas. Matome, kad fonema turi pagrindinį dažnį, vadinasi jos susidarime dalyvauja ir balso stygos.



2.16 pav. Fonemos “m” signalas PRAAT lange.

Analogiškai buvo analizuojami ir kiti ankščiau paminėti žodžiai ir iškerpamos reikalingos priebalsių fonemos.

2.3.3. Lietuvių kalbos dvibalsių fonemos

Dvibalsiu, arba diftongu, fonetinį pagrindą sudaro dviejų balsinių elementų junginys tame pačiame žodžio skiemenyje.

Dvibalsių artikuliacija yra sudėtinga: jie prasideda vienu balsiniu elementu ir baigiasi kitu, sudarydami neperskiriamą junginį, būtinai tariamą viename skiemenyje. Artikuliuojant dvibalsį, kalbos padargai slenka nuo pirmojo balsinio elemento pozicijos, pakeliui sudarydami keletą pereinamųjų balsinių elementų, kol pasiekia antrojo balsinio elemento poziciją.

Lietuvių literatūrinė kalba turi 8 dvibalsines fonemas: *ai, au, ei, eu, oi, ui, ie, uo*. Dvibalsiai *ie, uo* skiriasi iš visų dvibalsių sistemos ir savo artikuliacija priartėja prie diftongoidų, t. y. prie balsių, panašių į dvibalsius. Todėl dvibalsiai *ie, uo* laikomi sutaptiniais dvibalsiais, o kiti dvibalsiai - sudėtiniais.

Sudėtiniai dvibalsiai *ai, ei, au, oi, eu, ui* laikomi savarankiškais fonemomis dėl šių priežasčių:

1) jų abu dėmenys priklauso tam pačiam skiemeniui ir vienas bet kurio dvibalsio dėmuo niekada negali būti savarankiškas morfolginis vienetas, pavyzdžiui, žodžiuose *vaikas, vaikų, vaikams* ir t. t. *ai* yra šaknies dvibalsis, žodžių *eilė, eiti, eiga* šaknyse yra dvibalsis *ei*;

2) skiemens atžvilgiu jie yra nedalūs, t. y. tarp abiejų dvibalsio dėmenų negalima išvesti skiemens ribos;

3) dvibalsiai yra panašaus ilgumo, kaip ilgieji balsiai.

Dvibalsių morfologinio ir skiemeninio nedalumo negalima painioti su visada skirtingose morfemose ir skirtinguose skiemenyse esanciais balsiais a ir i, a ir u ir kt., kurie niekada nesudaro dvibalsių, pavyzdžiui, žodžiuose ne-išeik, pa-imk, pa-upys, su-ošė.

Tie patys dvibalsiai, kaip ir balsiai, su skirtingomis priegaidėmis laikomi ne savarankiškoms fonemomis, o tik vienos fonemos poziciniais variantais.

Kiekvienai dvibalsinei fonemai būdingi šie poziciniai variantai:

1. Tvirtapradis, labai intensyvus, tariamas įtemptais kalbos padargais.

2. Tvirtagalis, tariamas ne taip intensyviai, kaip tvirtapradis, mažiau įtemptais kalbos padargais.

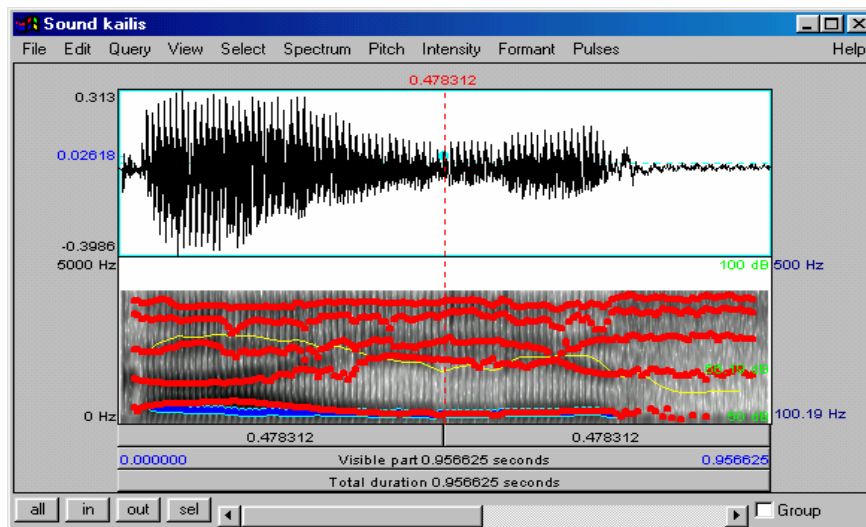
3. Nekirčiuotas, tariamas dar mažiau intensyviai, negu tvirtagaliai variantai, neįtemptais kalbos padargais.

Tam, kad išskirti dvibalsių fonemas buvo kompiuteriu įrašyti žodžiai, kuriuose yra dvibalsinės fonemos. Atlikus žodžių dažninę ir spektrinę analizę iš jų iškirptos dvibalsių fonemos.

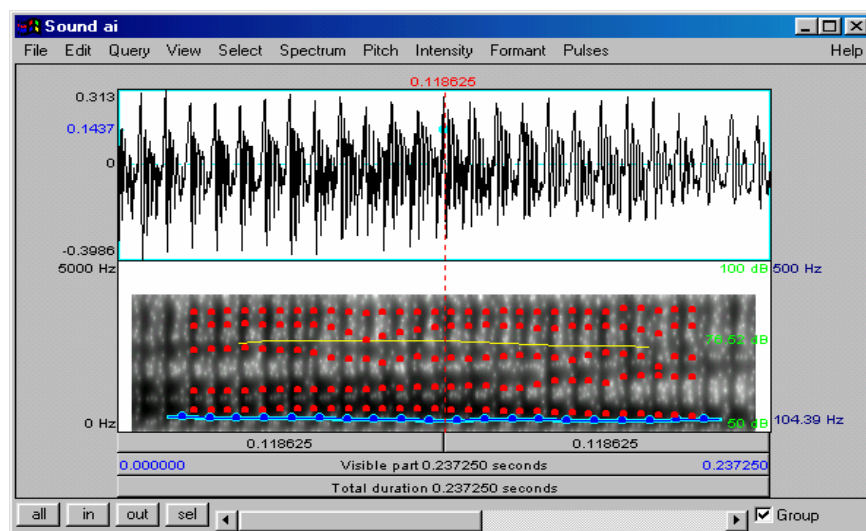
Buvo atrinkti ir įrašyti tokie žodžiai (šalia užrašyta žodyje esanti dvibalsio fonema):

Kailis – ai, kaulas – au, leidimas – ei, leukocitai – eu, boikotas – oi, muilas – ui, kietis – ie, kuodas – uo.

Kaip ir priebalsių analizavime, taip ir dvibalsių fonemų išskyrimo buvo naudojama Praat programa. Žemiau esančiuose paveikslėliuose pateikiama keletas darbo pavyzdžių.

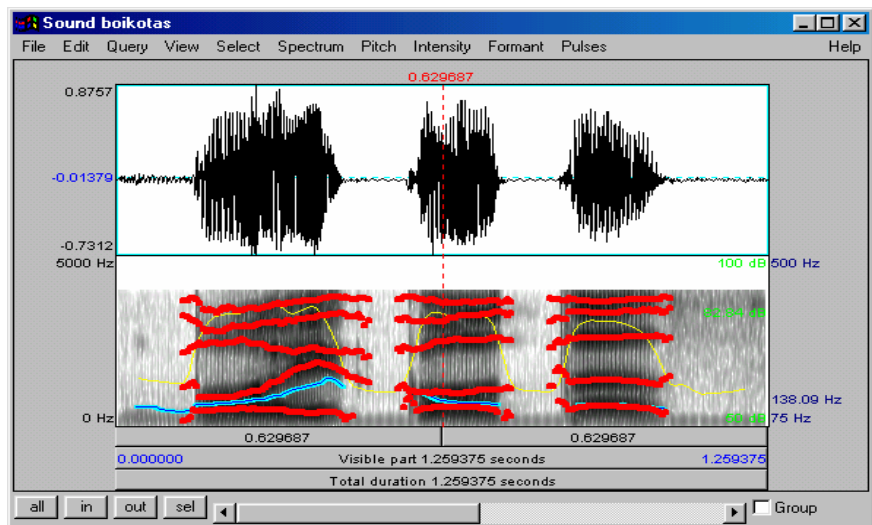


2.17 pav. Lietuvių kalbos žodžio “kailis” signalas PRAAT lange.

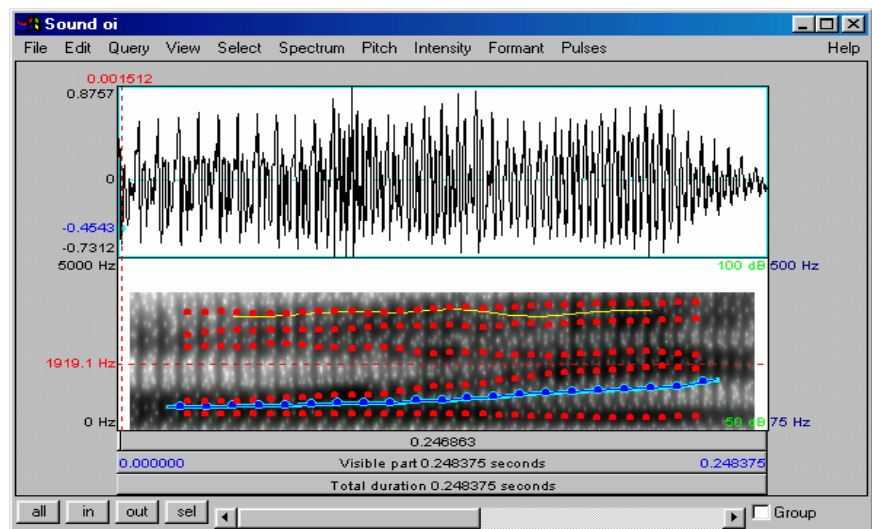


2.18 pav. Fonemos “ai” signalas PRAAT lange.

2.18 pav. gerai matomas perėjimas iš garso a prie garso i. Fonemos ai trukmė panaši kaip ir ilgųjų balsių.



2.19 pav. Lietuvių kalbos žodžio “boikotas” signalas PRAAT lange.



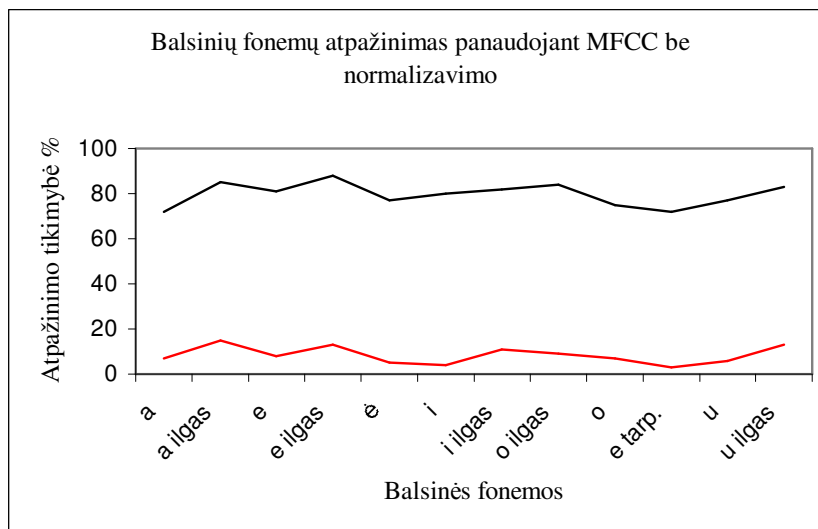
2.20 pav. Fonemos “oi” signalas PRAAT lange.

2.20 pav. pavaizduota fonemos oi analizė, kurioje matyti kaip kinta fonemos forma laike. Fonemos oi viduryje matomas perėjimas nuo o prie i garso. Perėjimo trukmė 0,045s, tuo tarpu visos fonemos trukmė 0,24s. Keičianti kalbos tempui, keistusi ir fonemos ilgis pagal laiką, tačiau perėjimas nuo vieno garso prie kito išliktų bet koku atveju, pasikeistų tik jo trukmė.

3. ŠNEKOS ATPAŽINIMO METODO, PANAUDOJANT NEURONINĮ TINKLĄ, TESTAVIMAS IR GAUTŲ REZULTATŲ APIBENDRINIMAS

3.1. Izoliuotų fonemų atpažinimas be signalo normalizavimo

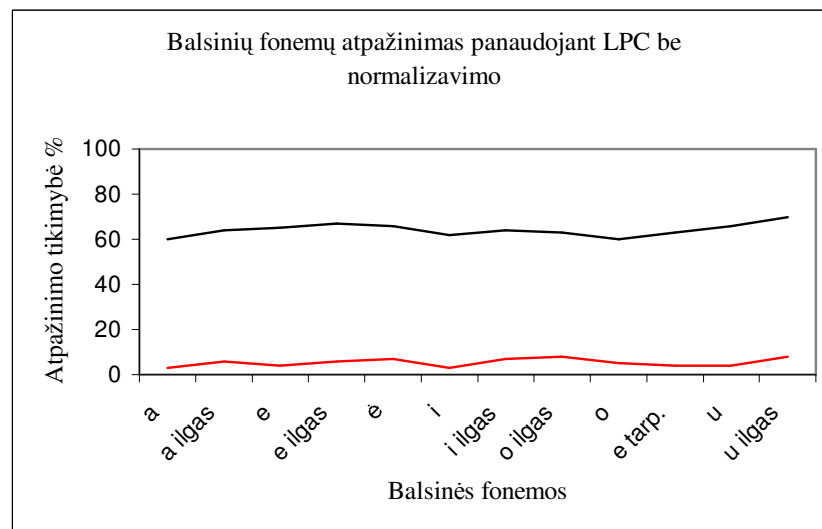
Balsių atpažinimo eksperimentas. Pirmiausia buvo atliekamas atpažinimo eksperimentas su balsinėmis fonemomis. Neuroninis tinklas buvo apmokytas vieno iš autorių balsinėmis fonemomis, kaip fonemas charakterizuojantys koeficientai panaudoti MFCC. Tada iš atsitiktinai pasirinktų žodžių iškirptos balsinės fonemos. Viso iškirpta po 300 kiekvienos fonemos pavyzdžių. Iškirpus fonemas buvo atliktas atpažinimo eksperimentas, kurio rezultatai matomi 3.1 pav.



3.1. Balsinių fonemų atpažinimo tikimybė panaudojant MFCC koeficientus.

3.1 pav. matome, kaip eksperimento metu buvo atpažystamos balsinės fonemos. Matosi, kad ilgųjų balsių atpažinimo tikimybė yra didesnė nei trumpųjų. Skirtumas apytiksliai 8 procentai. Geriausiai buvo atpažinta e ilgoji fonema (88%), o blogiausiai a trumpoji ir e tarptautiniuose žodžiuose (72%). Balsinių fonemų atpažinimo ekperimento metu panaudotos to paties diktoriaus fonemos kaip ir neuroninio tinklo apmokyme. Padarytas ir atvirkščias variantas: vieno diktoriaus šabloninėmis fonemomis apmokytas neuroninis tinklas, o atpažinti duodamos kito diktoriaus įdiktotos fonemos. Rezultatai pažymėti raudona linija, taip pat matomi 3.1 pav.. Matosi, jog atpažinimas ženkliai pablogėjo ir geriausiu atveju tesiekia 15 procentų (a ilgoji), prasčiausiai atpažinta e fonema esanti tarptautiniuose žodžiuose (3%).

Sekančiame etape neuroninis tinklas apmokytas fonemas charakterizuojančiu parametru panaudojant LPC koeficientus. Pirmiausiai atliktas balsių atpažinimo eksperimentas su to paties diktorius fonemomis. Rezultatai pateikiami 3.2 pav.



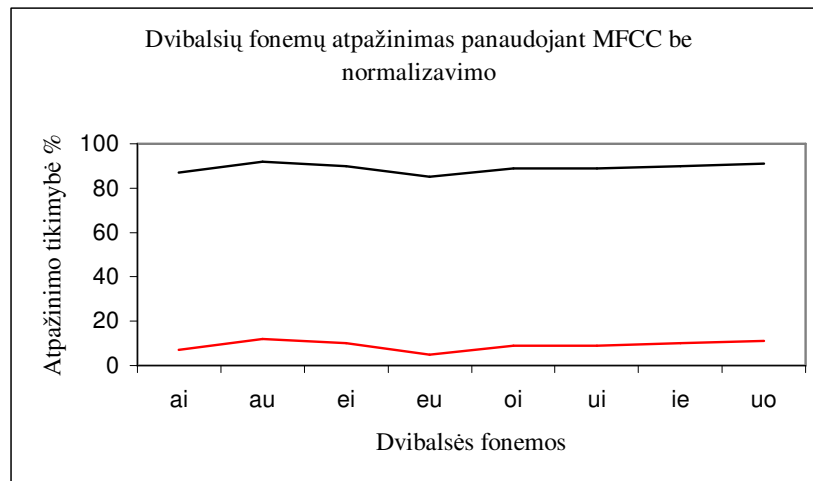
3.2 pav. Balsinių fonemų atpažinimo tikimybė panaudojant LPC koeficientus.

3.2 pav. matyti, kad jei fonemas charakterizuojančiais koeficientais naudojame LPC tai atpažinimo tikimybė yra mažesnė nei naudojant MFCC koeficientus. Geriausiai atpažinta buvo u ilgoji fonema (70%), o prasčiausiai a ir o trumposios (60%). Tačiau atpažinimo tikimybės kreivė gavosi tolygesnė, visos balsinės fonemos atpažintos su mažiau išsibarsčiusia tikimybe. Vėliau atliktas bandymas su skirtingais diktoriais, kurio rezultatai pavaizduoti 3.2 pav. apatinėje dalyje raudona linija. Matosi, jog atpažinimui ir tinklo apmokymui panaudojant skirtingų diktorių fonemas ir jas charakterizuojant LPC koeficientais, atpažinimo tikimybė labai sumažėja. Tačiau kreivė vistiek išlieka tolygesnė nei analogiškame bandyme panaudojus MFCC koeficientus.

Dvibalsių atpažinimo eksperimentas

Tokie pat fonemų atpažinimo eksperimentai, kaip ir su balsinėmis fonemomis, buvo atliekami ir su izoliuotomis dvibalsių fonemomis. Iškirptos šabloninės dvibalsių fonemos ir jos panaudotos neuroninio tinklo apmokymui. Pirmiausia atliktas eksperimentas su MFCC koeficientais, neuroninio tinklo apmokimui ir atpažinimui panaudotos to paties diktorius

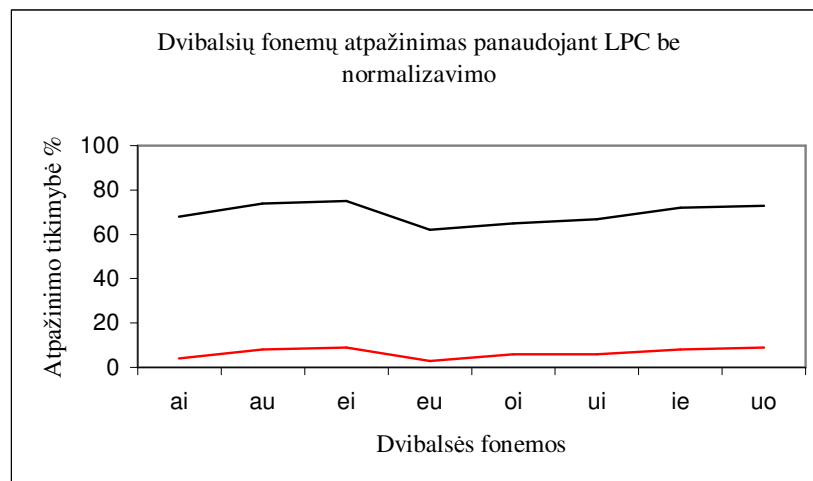
fonemos. Vėliau pakartotas bandymas tinklo apmokymui naudojant vieno diktoriaus fonemas, o atpažinti bandoma antrojo diktoriaus dvibalsių fonemas. Rezultatai pateikiami 3.3 pav.. Matosi, kad dvibalsės fonemos su vienu diktoriumi atpažystamos kiek geriau nei balsinės. Eksperimento su dviem diktoriais atpažinimo tikimybė, pavaizduota 3.3 pav. apatinėje dalyje, gana žema ir panaši į analogišką bandymą su balsinėmis fonemomis.



3.3 pav. Dvibalsių fonemų atpažinimo tikimybė panaudojant MFCC koeficientus.

Geriausiai atpažystama fonema au (92%), blogiausiai eu (82%). Neuroninio tinklo apmokymui naudojant vieno diktoriaus fonemas. Atpažinti bandant kitą diktorių geriausiai atpažintosa taip pat au fonama(12%), blogiausiai eu fonema (5%) fonemos.

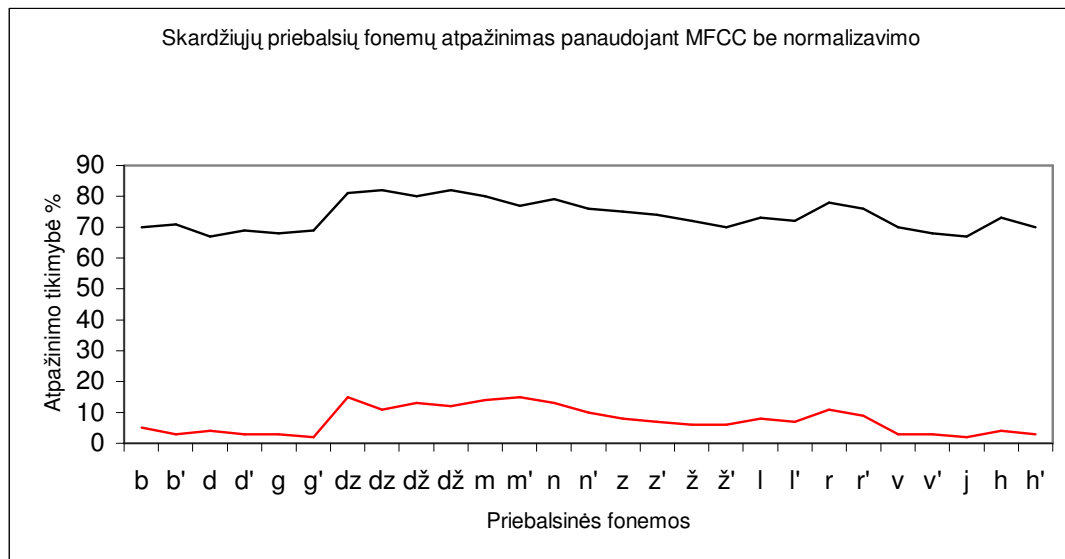
Dvibalsių fonemų atpažinimo tikimybė ištirta ir panaudojant LPC koeficientus. Rezultatai pateikiami 3.4 pav..



3.4 pav. Dvibalsių fonemų atpažinimo tikimybė panaudojant LPC koeficientus.

Dvibalsių fonemų atpažinime naudojant MFCC koeficientus, atpažinimo tikimybės kreivė gavosi tolygesnė, nei su LPC koeficientais. 3.4 pav. viršutinė grafiko kreivė rodo atpažinimo tikimybę eksperimentą atliekant su vieno diktorius fonemomis, o apatinė – atpažinimo tikimybę naudojant dviejų diktorių fonemas. Pirmu atveju geriausiai atpažinta ei fonema (75%), o blogiausiai eu fonema (62%), antru atveju geriausiai atpažintos fonemos ei ir uo (9%), blogiausiai – eu (3%).

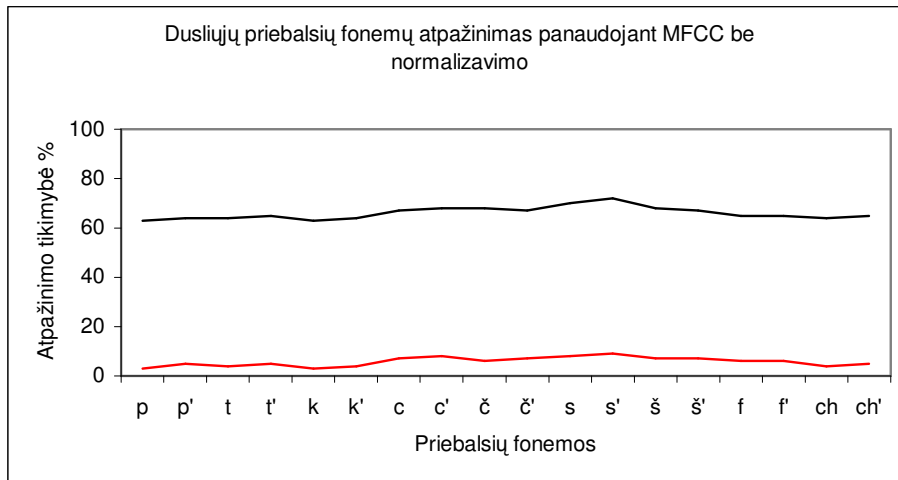
Priebalsių atpažinimo eksperimentas. Prieš atlikdami izoliuotų priebalsių fonemų atpažinimo eksperimentą, priebalsius padalinome į dvi grupes: dusliuosius ir skardžiuosius. Pirmiausia atliktas atpažinimo bandymas fonemas charakterizuojant MFCC koeficientais. Rezultatai pateikiami 3.5 ir 3.6 pav.



3.5 pav. Skardžiųjų priebalsių fonemų atpažinimo tikimybė panaudojant MFCC.

3.5 pav. pavaizduota skardžiųjų priebalsių fonemų atpažinimo tikimybė panaudojant MFCC koeficientus. Viršutinė grafiko kreivė rodo atpažinimo tikimybę dirbant su vieno diktorius fonemomis, apatinė kreivė – rezultatai gauti tinklą apmokant vieno diktorius fonemomis, o atpažinti bandant kitą diktorių. Eksperimente su vienu diktoriumi geriausiai atpažintos dz' ir dž' fonemos (82%), o blogiausiai j fonema (67%). Antru atveju geriausiai atpažintos dz ir m' fonemos (15%), blogiausiai g' ir j fonemos (2%).

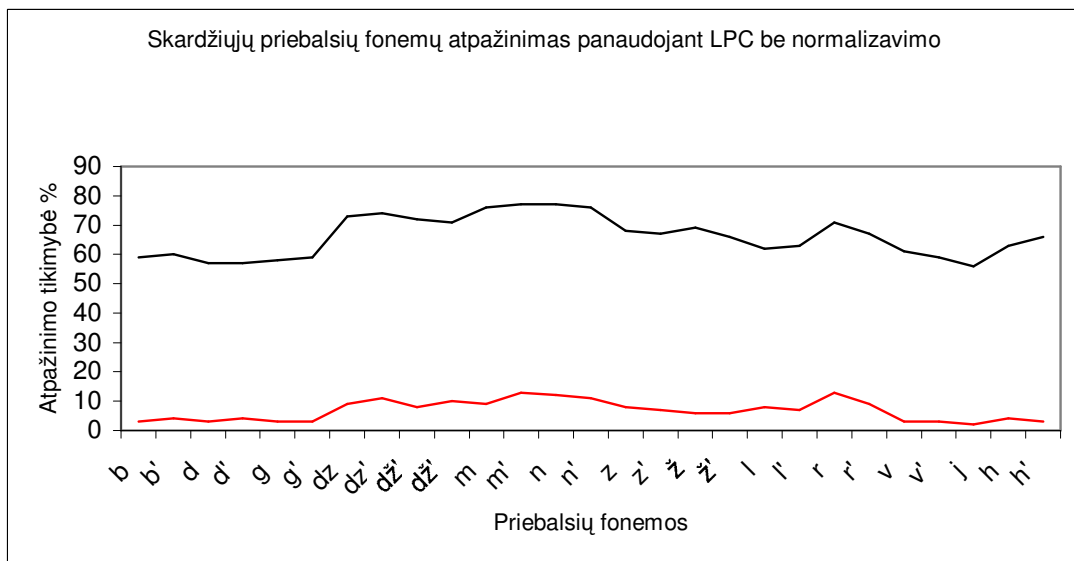
Toks pats bandymas atliktas ir su dusliosiomis priebalsių fonemomis. Rezultatai matomi 3.6 pav..



3.6 pav. Dusliųjų priebalsių fonemų atpažinimo tikimybė panaudojant MFCC.

3.6 pav. pavaizduotos kreivės atspindi dusliųjų priebalsių atpažinimo rezultatus. Viršutinė kreivė – atpažinimo bandymas su vieno diktoriaus fonemomis, geriausiai atpažinta s' fonema (72%), blogiausiai p ir k fonemos (63%). Eksperimente su dviem diktoriais, kurio rezultatai pavaizduoti apatinėje kreivėje, geriausiai atpažinta s' fonema (9%), blogiausiai – k ir p fonemos (3%). Iš gautų rezultatų matome, kad dusliosios fonemos atpažystamos šiek tiek prasčiau už skardžiasias.

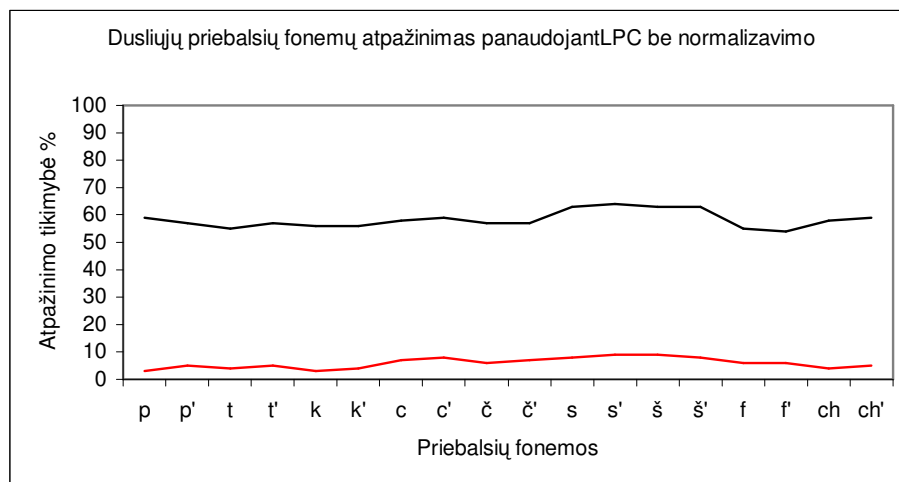
Priebalsių atpažinimo eksperimentas taip pat atliktas fonemas charakterizuojant LPC koeficientais.



3.7 pav. Skardžiujų priebalsių fonemų atpažinimo tikimybė panaudojant LPC.

Pirmiausia atliktas eksperimentas norint ištirti skardžiųjų priebalsių atpažinimo tikimybę. Rezultatai matomi 3.7 pav. Viršutinė grafiko kreivė rodo skardžiųjų priebalsių fonemų atpažinimo tikimybę dirbant su vieno diktorius fonemomis, apatinė (raudona) atpažinimo tikimybė eksperimentui panaudojant dviejų diktorių fonemas. Kai tinklas apmokomas ir atpažinti bandoma to paties diktorius fonemas, geriausiai atpažystamos m' ir n fonemos (77%), blogiausiai – j fonema (56%). Tinklą apmokant vieno diktorius fonemomis, o atpažinti bandant antrąjį diktorių, rezultatai žymiai suprastėja. Geriausiai atpažystamos m' ir r fonemos (13%), o blogiausiai – j fonema (2%).

Analogiškas atpažinimo eksperimentas atliktas ir su izoliuotų dusliųjų priebalsių fonemomis. Rezultatai matomi 3.8 pav.



3.8 pav. Dusliųjų priebalsių fonemų atpažinimo tikimybė panaudojant LPC.

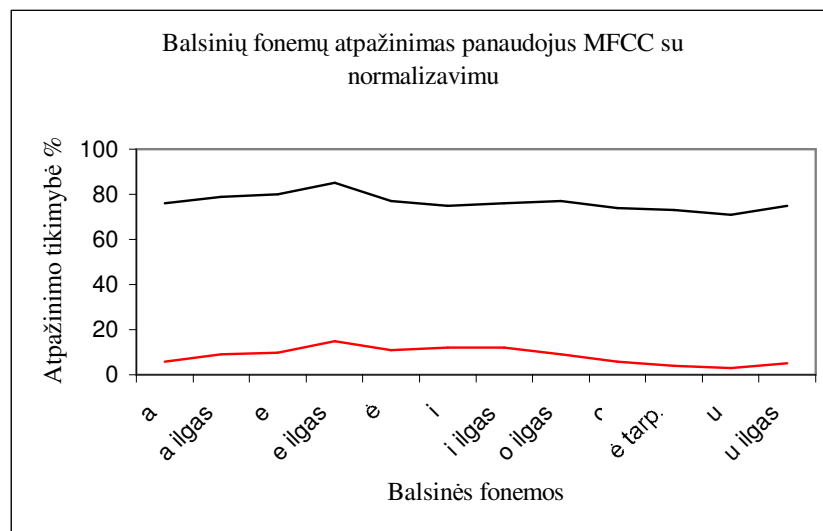
Viršutinė grafiko, esančio 3.8 pav., kreivė atspindi rezultatus gautus atpažinimui naudojant vieno diktorius fonemas. Matome, kad geriausiai atpažystama s' fonema (64%), blogiausiai – f' fonema (54%). Apatinė grafiko kreivė vaizduoja rezultatus, eksperimento metu, tinklą apmokant vieno diktorius fonemomis, o atpažinti bandant antrą diktorių. Geriausiai atpažystamos s' ir š fonemos (9%), o blogiausiai k ir p fonemos (3%).

3.2. Izoliuotų fonemų atpažinimas su signalo normalizavimu

Visi prieš tai buvę mūsų tiriamojo darbo eksperimentai buvo atlikti nenormalizavus fonemų amplitudės pagal maksimalią reikšmę. Todėl atlikome izoliuotų garsų atpažinimo bandymą ir su normalizavimu. Garso normalizavimas pagal maksimalią amplitudę reikalingas

tam, kad atpažinimui neturėtų įtakos atpažįstamo kalbos garso ištarimo garsumas (amplitudės dydis). Apačioje esančiuose grafikuose pateikiami balsinių fonemų atpažinimo rezultatai, prieš tai fonemas normalizavus. Normalizavimas atliktas pasinaudojant Matlab programoje esančia funkcija: $\text{wave} = \text{wave} / \max(\text{abs}(\text{wave}))$.

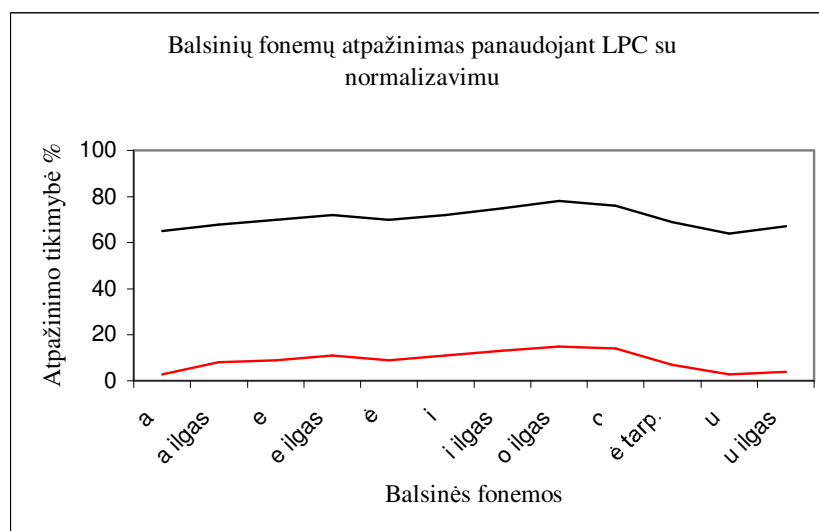
Balsių fonemų atpažinimo eksperimentas atlikus normalizavimą. 3.9 pav. parodytas balsinių fonemų atpažinimas naudojant MFCC koeficientus. Eksperimentas atliktas pagal tokia pat eiga, kaip ir prieš tai esantys. Pirmiausia bandoma atpažinti vieno diktoriaus fonemas, po to vieno diktoriaus fonemomis apmokomas neuroninis tinklas ir bandoma atpažinti antrojo diktoriaus fonemas.



3.9 pav. Normalizuotų balsinių fonemų atpažinimo tikimybė panaudojus MFCC koeficientus

Kaip matome 3.9 pav. eksperimento metu geriausiai atpažinta e ilgoji fonema (85%), o blogiausiai u trumpoji (71%). Panaudojant dviejų diktorių fonemas taip pat geriausiai atpažinta e ilgoji (15%), o blogiausiai u trumpoji (3%). Bandant atpažinti izoliuotas balsinias fonemas ir nenaudojant normalizavimo funkcijos geriausiai buvo atpažinta taip pat e ilgoji fonema (88%). Vadinasi atlikus normalizavimą atpažinimas sumažėjo 3 %.

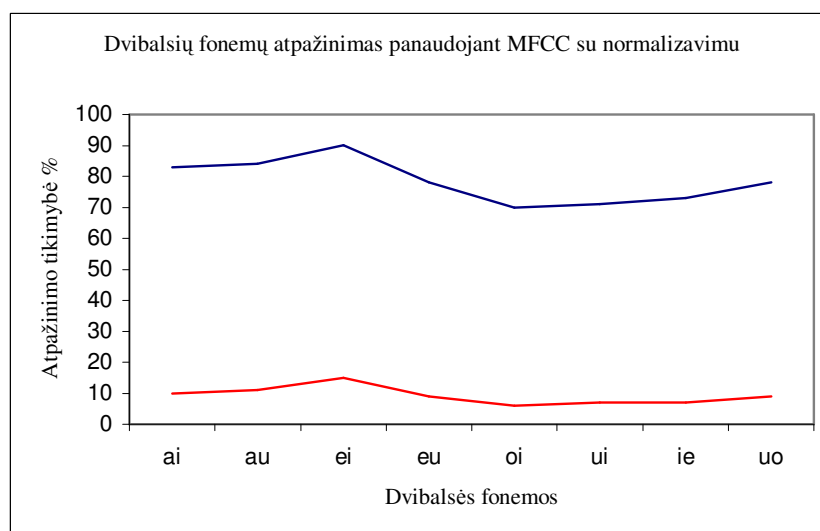
Analogiškas balsinių fonemų atpažinimo eksperimentas, normalizavus fonemas pagal maksimalią amplitudę, buvo atliktas ir su LPC koeficientais. Rezultatai pateikiami 3.10 pav. Jame matosi, kad geriausiai atpažystama fonema o ilgoji (78%), o blogiausiai u trumpoji (64%). Eksperimente su dviejų diktorių fonemomis geriausiai buvo atpažinta o ilgoji fonema (15%), o blogiausiai a ir u trumposios fonemos (3%).



3.10 pav. Normalizuotų balsinių fonemų atpažinimo tikimybė panaudojus LPC koeficientus

Atlikus eksperimentą su LPC koeficientais matyti, kad po amplitudės normalizavimo atpažinimo tikimybė padidėjo. Kaip pavyzdį galime imti o ilgąją fonemą: be normalizavimo ji atpažinta su 63 % tikimybe, o atlikus normalizavimą atpažinimas padidėjo iki 78%.

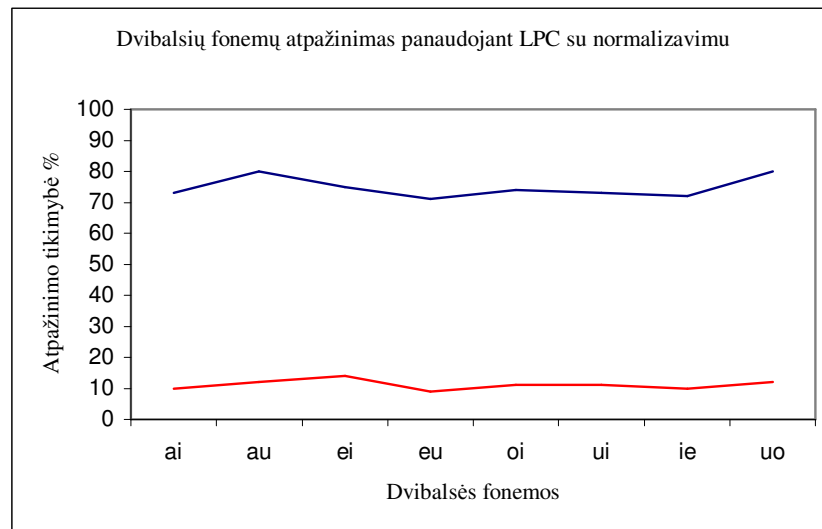
Dvibalsių fonemų atpažinimo eksperimentas atlikus normalizavimą. Normalizavę dvibalsių fonemas pagal maksimalią jų amplitudę, atlikome jų atpažinimo bandymą. Pirmiausia atpažinimo eksperimentas atliktas dvibalsių fonemas charakterizuojant MFCC koeficientais.



3.11 pav. Normalizuotų balsinių fonemų atpažinimo tikimybė panaudojus MFCC koeficientus

Rezultatai pateikiami 3.11 pav. Viršutinė grafiko kreivė rodo eksperimento rezultatus kai neuroninio tinklo apmokymui ir atpažinimui panaudojamos vieno ir to paties diktoriaus fonemos. Geriausiai atpažystama fonema ei (90%), blogiausiai oi (70%). Atliekant atpažinimo eksperimentą su dviem diktoriais geriausiai atpažinta taip pat ei fonema (15%), blogiausiai oi fonema (6%), rezultatai pažymėti raudona kreive grafiko apačioje.

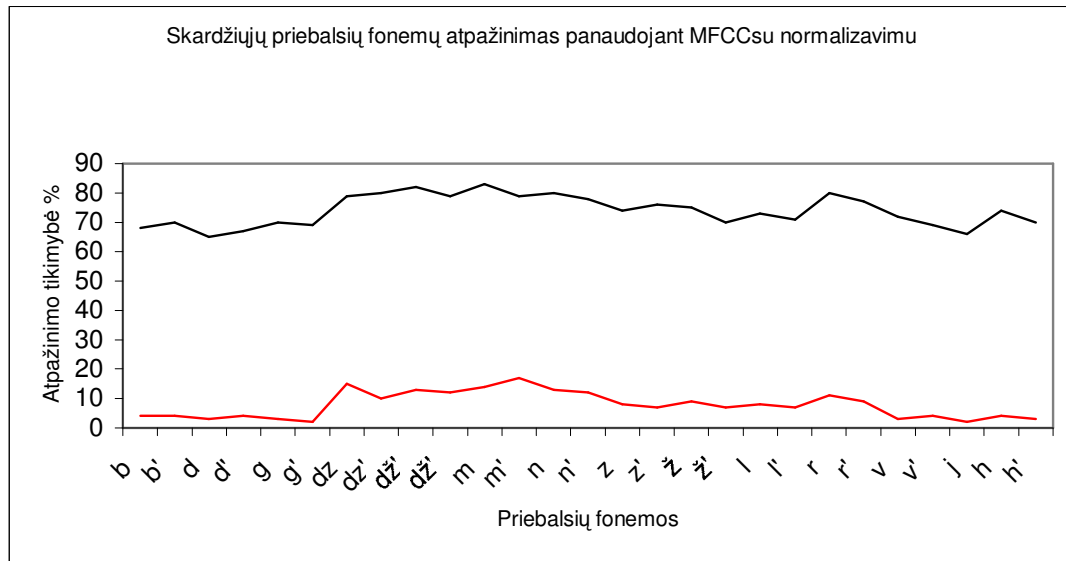
Analogišką atpažinimo bandymą atlikome dvibalses fonemas charakterizuodami LPC koeficientais. Eksperimento rezultatai matomi 3.12 pav.



3.12 pav. Normalizuotų balsinių fonemų atpažinimo tikimybė panaudojus LPC koeficientus

Viršutinė grafiko kreivė atspindi eksperimento rezultatus kai neuroninio tinklo apmokymui ir atpažinimui panaudojamos vieno ir to paties diktoriaus fonemos. Geriausiai atpažystamos fonemos au ir uo (80%), blogiausiai eu (71%). Atliekant atpažinimo eksperimentą su dviem diktoriais geriausiai atpažinta ei fonema (14%), blogiausiai eu fonema (9%), rezultatai pažymėti raudona kreive grafiko apačioje. Kaip ir bandyme su normalizuotomis balsių fonemomis, normalizavus dvibalsių fonemas ir jas charakterizuojant LPC koeficientais atpažinimo tikimybė padidėjo.

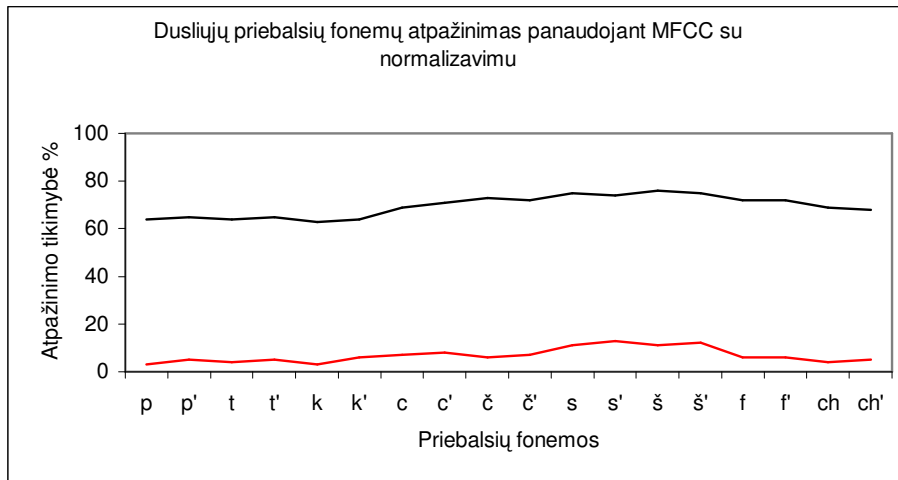
Priebalsių fonemų atpažinimo eksperimentas atlikus normalizavimą. Analogiški bandymai atlikus signalo normalizavimą atlikti su priebalsių fonemomis. Pirmiausia fonemas charakterizavome MFCC koeficientais. 3.13 pav. matoma skardžiųjų priebalsių atpažinimo tikimybės kreives.



3.13 pav. Normalizuotų skardžiųjų priebalsių fonemų atpažinimo tikimybė panaudojus MFCC koeficientus

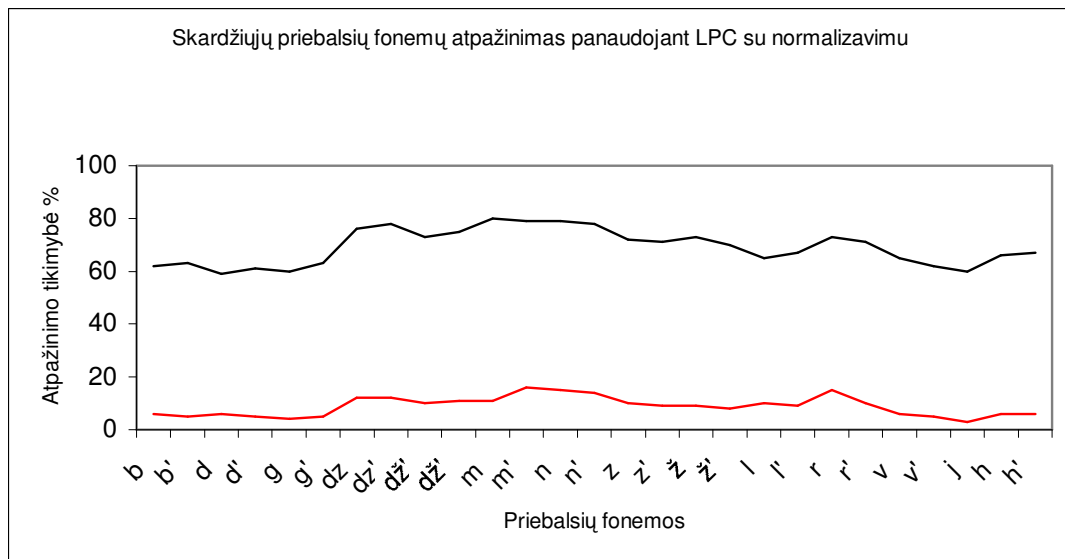
3.13 pav. pavaizduota skardžiųjų priebalsių fonemų, atlikus normalizavimą, atpažinimo tikimybė panaudojant MFCC koeficientus. Viršutinė grafiko kreivė rodo atpažinimo tikimybę dirbant su vieno diktoriaus fonemomis, apatinė kreivė – rezultatai gauti tinklą apmokant vieno diktoriaus fonemomis, o atpažinti bandant antrą diktorių. Eksperimente su vienu diktoriumi geriausiai atpažinta m fonema (83%), o blogiausiai j fonema (66%). Antru atveju geriausiai atpažinta m' fonema (17%), blogiausiai g' ir j fonemos (2%).

Toks pat atpažinimo eksperimentas atliktas ir su normalizuotomis dusliųjų priebalsių fonemomis. Rezultatai matomi 3.14 pav. Jame matosi, jog geriausiai atpažystama fonema š (76%), o blogiausiai k fonema (63%). Eksperimente su dviem diktoriais, geriausiai buvo atpažinta s' fonema (13%), o blogiausiai p ir k fonemos (3%). Iš gautų rezultatų matosi, kad kai kurių fonemų atpažinimas po normalizavimo pagerėjo 2 – 3 procentais.



3.14 pav. Normalizuotų duslių priebalsių fonemų atpažinimo tikimybė panaudojus MFCC koeficientus

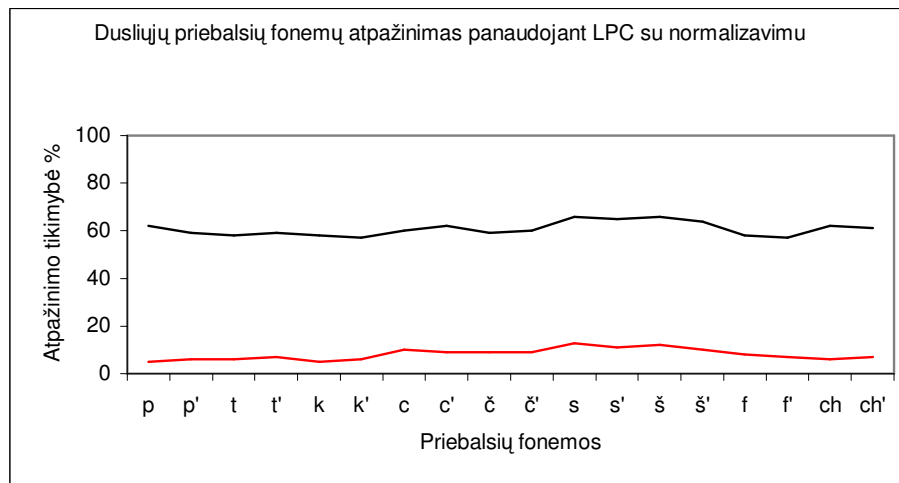
Atlikę normalizavimą pagal maksimalią signalo amplitudės vertę, priebalsių fonemas taip pat charakterizavome ir LPC koeficientais, bei atlikome jų atpažinimo bandymą. Pirmiausia atliktas bandymas atpažinimo eksperimentui panaudojant skardžiasias priebalsių fonemas. Eksperimento eiga lygiai tokia pati, kaip ir prieš tai buvusių: pirma neuroninis tinklas apmokomas vieno diktorius fonemų koeficientais ir atpažinti bandoma tą patį diktorių, antru atveju tinklas apmokomas pirmo diktorius fonemomis, o atpažinti bandoma kito diktorius priebalsių fonemas. Normalizuotų skardžiųjų priebalsių atpažinimo rezultatai pateikiami 3.15 pav. esančiame grafike.



3.15 pav. Normalizuotų skardžiųjų priebalsių fonemų atpažinimo tikimybė panaudojus LPC koeficientus

Viršutinė grafiko, esančio 3.15 pav., kreivė rodo eksperimento rezultatus kai neuroninio tinklo apmokymui ir atpažinimui panaudojamos vieno ir to paties diktoriaus fonemos. Geriausiai atpažystama fonema m (80%), blogiausiai j ir g fonemos (60%). Atliekant atpažinimo eksperimentą su dviem diktoriais geriausiai atpažinta m' fonema (16%), blogiausiai j fonema (3%), rezultatai pažymėti raudona kreive grafiko apačioje.

Lygiai toks pat izoliuotų garsų atpažinimo bandymas atliktas su dusliosiomis priebalsių fonemomis. Rezultatai matomi 3.16 pav.



3.16 pav. Normalizuotų duslių priebalsių fonemų atpažinimo tikimybė panaudojus LPC koeficientus

Kaip matome iš 3.16 pav. esančio grafiko viršutinės kreivės geriausiai buvo atpažintos s ir š fonemos (66%), o prasčiausiai f' ir k' fonemos (57%). Eksperimento su dviejų diktorių fonemomis rezultatus vaizduoja apatinė raudona grafiko kreivė. Matome, kad didžiausia atpažinimo tikimybė yra s fonemos (13%), mažiausia – p ir k fonemų (5%).

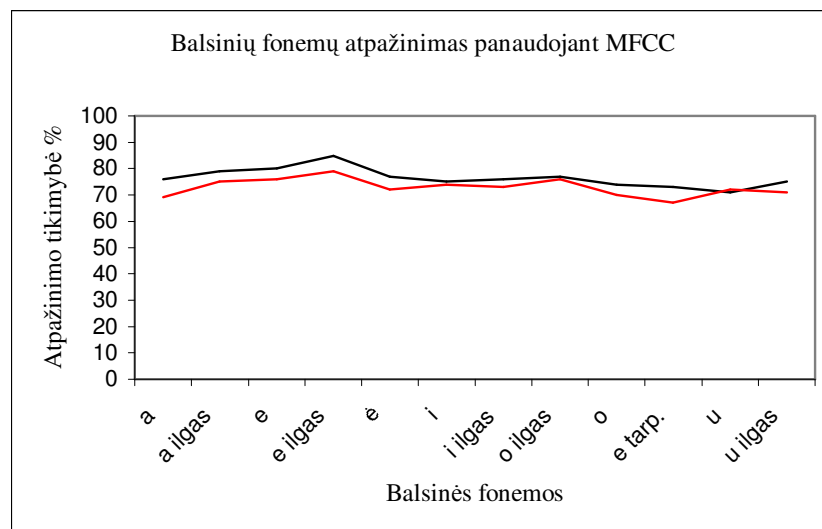
Kaip ir bandyme su normalizuotomis balsių ir dvibalsių fonemomis, normalizavus priebalsių fonemas ir jas charakterizuojant LPC koeficientais atpažinimo tikimybė padidėjo.

3.3 Izoliuotų fonemų atpažinimas neuroninį tinklą apmokant dviejų diktorių fonemomis

Trečiame tiriamojo darbo etape atlikome izoliuotų fonemų atpažinimo eksperimentą, neuroninį tinklą apmokydami abiejų diktorių fonemomis. Atpažinti bandėme tik vieno iš diktorių įdiktuoatas fonemas. Eksperimentai buvo atliekami tam, kad patikrinti kiek sumažėjo

atpažinimo tikimybė, kai neuroninis tinklas apmokomas didesniu kiekiu panašių viena į kitą fonemų. Fonemos suklasifikavome sekančiu principu: prie pirmo diktoriaus fonemų pridėjome indeksą 1, prie antro pridėtas indeksas 2, tam, kad galėtume atskirti katro diktoriaus fonema neuroninis tinklas atpažino. Kaip pavyzdį imant balsines fonemas, neuroninis tinklas buvo apmokytas 24 balsių fonemomis. Kadangi atpažinimo eksperimentuose, su signalo normalizavimu pagal maksimalią amplitudę, daugeliu atvejų rezultatai buvo geresni nei be normalizavimo, šiam bandymui taip pat pasirinkome normalizuotas fonemas.

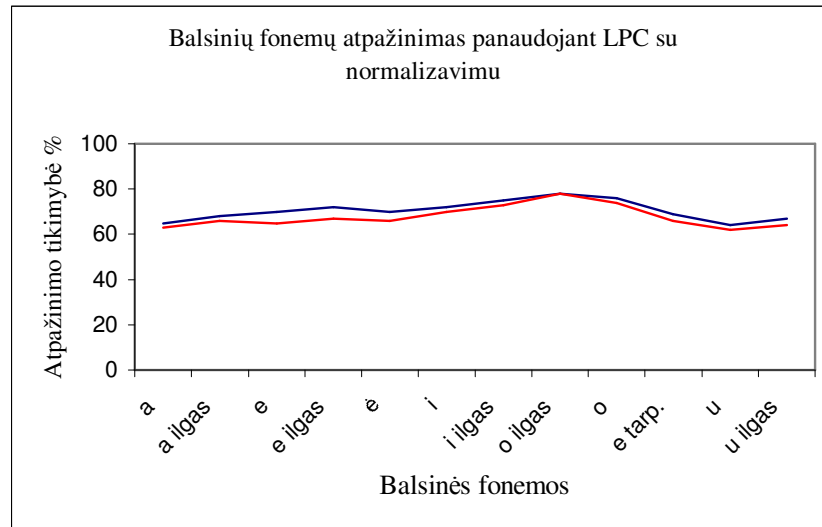
Balsių atpažinimo eksperimentas. Eksperimento rezultatai, kai fonemos charekterizuotos MFCC koeficientais pateikiami 3.17 pav.



3.17 pav. Balsių atpažinimo tikimybė panaudojus MFCC koeficientus ir tinklą apmokant dviejų diktorių fonemomis

3.17 pav. esančiame grafike viršutinė juoda kreivė vaizduoja prieš tai buvusio balsių atpažinimo eksperimento rezultatus. Apatinė, raudona kreivė, rodo bandymo rezultatus kai neuroninis tinklas turėjo atpažinti vieno diktoriaus balsines fonemas iš 24 galimų variantų. Matome, kad atpažinimo tikimybė šiek tiek sumažėjo. Geriausiai atpažinta e ilgoji fonema (79%, buvo 85%), blogiausiai atpažinta e tarptautiniuose žodžiuose fonema (67%, o buvo 73%).

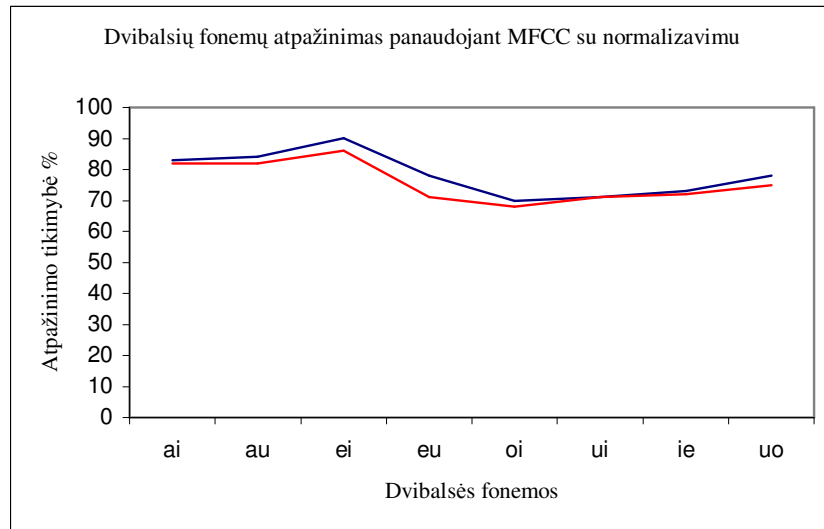
Balsių atpažinimo bandymas, tinklą apmokant abiejų diktorių fonemomis atliktas ir panaudojant LPC koeficientus. Bandymas atliktas tokia pat tvarka kaip ir su MFCC koeficientais. Rezultatai matomi 3.18 pav.



3.18 pav. Balsių atpažinimo tikimybė panaudojus LPC koeficientus ir tinklą apmokant dviejų diktorių fonemomis

3.18 pav. raudona grafiko kreivė vaizduoja bandymo rezultatus. Juoda grafiko kreivė – ankstesni rezultatai kai neuroninis tinklas buvo apmokytas vieno diktoriaus fonemomis. Grafike akivaizdžiai matosi, kad izoliuotų balsių atpažinimo tikimybė neuroninį tinklą apmokius dviejų diktorių balsinėmis fonemomis šiek tiek sumažėjo, skirtumas iki 5 procentų. Geriausiai buvo atpažinta o ilgoji fonema (78%, buvo 78%), blogiausiai atpažinta u fonema (62%, o buvo 64%). Labiausiai sumažėjo e (65%, o buvo 70%) ir e ilgosios fonemos (67%, o buvo 72%).

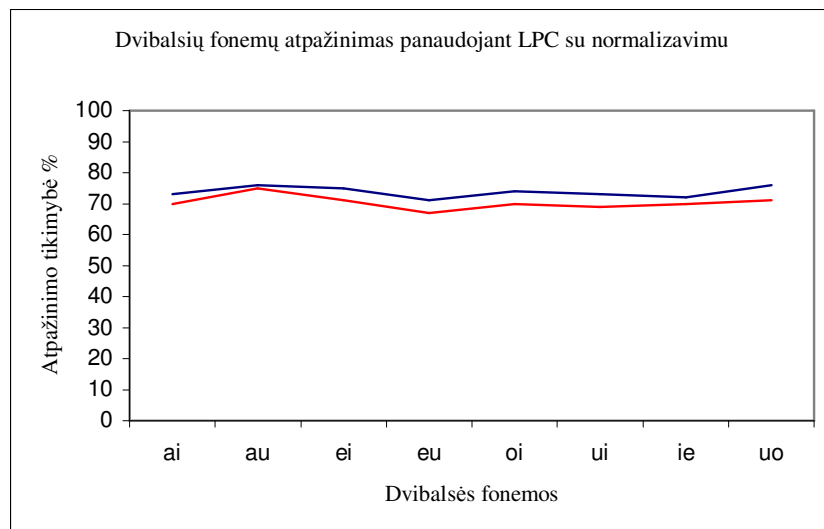
Dvibalsių atpažinimo eksperimentas. Analogiški bandymai atlikus signalo normalizavimą ir tinklą apmokant abiejų diktorių fonemomis atlikti ir su izoliuotais dvibalsiais. Pirmiausia atliktas bandymas fonemas charakterizavus MFCC koeficientais. 3.19 pav. matome dvibalsių fonemų atpažinimo tikimybės kreives.



3.19 pav. Dvibalsių atpažinimo tikimybė panaudojus MFCC koeficientus ir tinklą apmokant dviejų diktorių fonemomis

3.19 pav. esančiame grafike pavaizduoti lietuvių kalbos dvibalsių atpažinimo eksperimento rezultatai (raudona kreivė). Palyginimui išbrėžta juoda kreivė, kurioje ankstesnio eksperimento rezultatai kai neuroninis tinklas buvo apmokytas vieno diktoriaus fonemomis. Geriausiai atpažinta ei fonema (86%, o buvo 90%), blogiausiai fonema oi (68%, o buvo 70%).

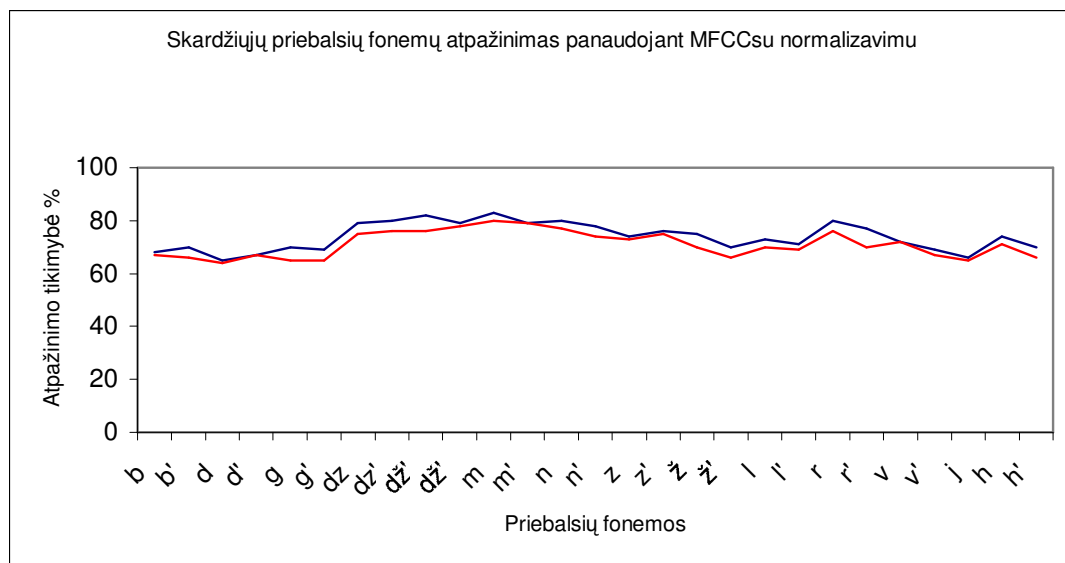
3.20 pav. matome eksperimento rezultatus kai neuroninis tinklas buvo apmokytas dviejų diktorių fonemomis. Jos charakterizuotos LPC koeficientais.



3.20 pav. Dvibalsių atpažinimo tikimybė panaudojus LPC koeficientus ir tinklą apmokant dviejų diktorių fonemomis

Eksperimente geriausiai buvo atpažinta au dvibalsė fonema (75%, buvo 76%), blogiausiai atpažinta eu fonema (67%, o buvo 71%). Labiausiai sumažėjo uo fonemos atpažinimas (71%, o buvo 76%).

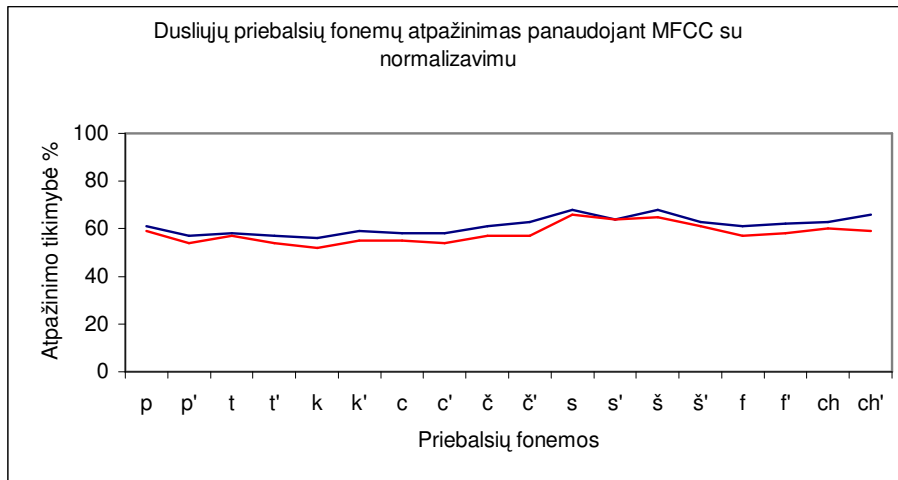
Priebalsių atpažinimo eksperimentas. Prieš atlikdami izoliuotų priebalsių fonemų atpažinimo eksperimentą, tinklą apmokant dviejų diktorių priebalsių fonemomis, lietuvių kalbos priebalsius padalinome į dvi grupes: dusliuosius ir skardžiuosius. Pirmiausia atliktas atpažinimo bandymas fonemas charakterizuojant MFCC koeficientais. Rezultatai matomi 3.21 pav. ir 3.33 pav.



3.21 pav. Skardžiųjų priebalsių atpažinimo tikimybė panaudojus MFCC koeficientus ir tinklą apmokant dviejų diktorių fonemomis

3.21 pav. esančiame grafike viršutinė juoda kreivė vaizduoja prieš tai buvusio skardžiųjų priebalsių atpažinimo eksperimento rezultatus. Apatinė, raudona kreivė, rodo bandymo rezultatus kai neuroninis tinklas turėjo atpažinti vieno diktoriaus skardžiųjų priebalsių fonemas iš 54 galimų variantų. Matome, kad atpažinimo tikimybė šiek tiek sumažėjo. Geriausiai atpažinta m fonema (80%, buvo 83%), blogiausiai atpažinta d fonema (64%, o buvo 65%). Labiausiai sumažėjo r' fonemos atpažinimo tikimybė (70%, buvo 77%).

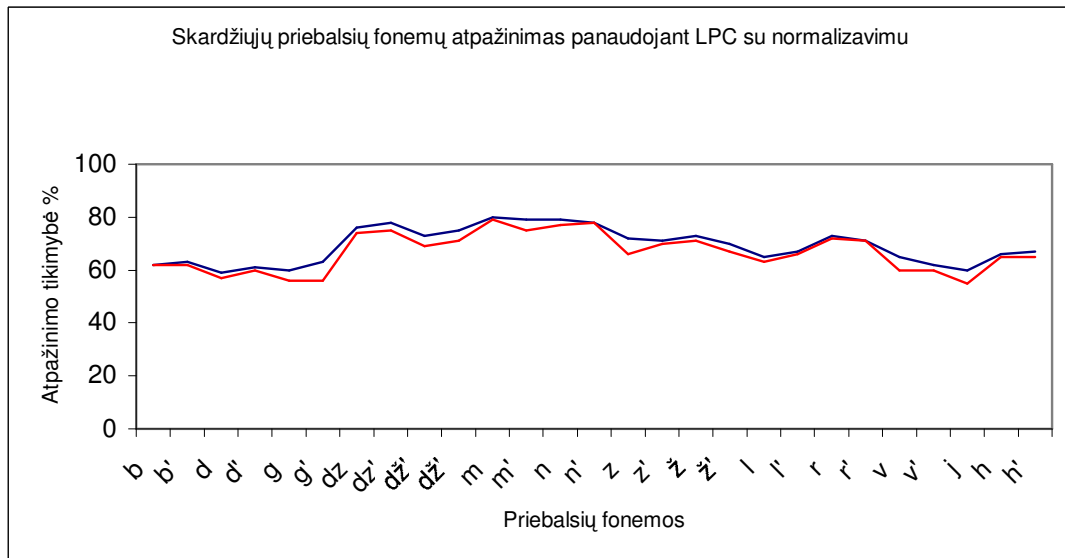
Tokiu pat metodu patikrinta duslių priebalsių fonemų atpažinimo tikimybė. Rezultatai pateikiami 3.22 pav.



3.22 pav. Dusliųjų priebalsių atpažinimo tikimybė panaudojus MFCC koeficientus ir tinklą apmokant dviejų diktorių fonemomis

3.22 paveikslėlyje esančiame grafike matosi, kad dusliųjų priebalsių atpažinimo tikimybė taip pat sumažėjo. Geriausiai atpažinta s fonema (66%, buvo 68%), blogiausiai atpažinta k fonema (52%, o buvo 56%). Labiausiai sumažėjo ch' fonemos atpažinimo tikimybė (59%, buvo 66%).

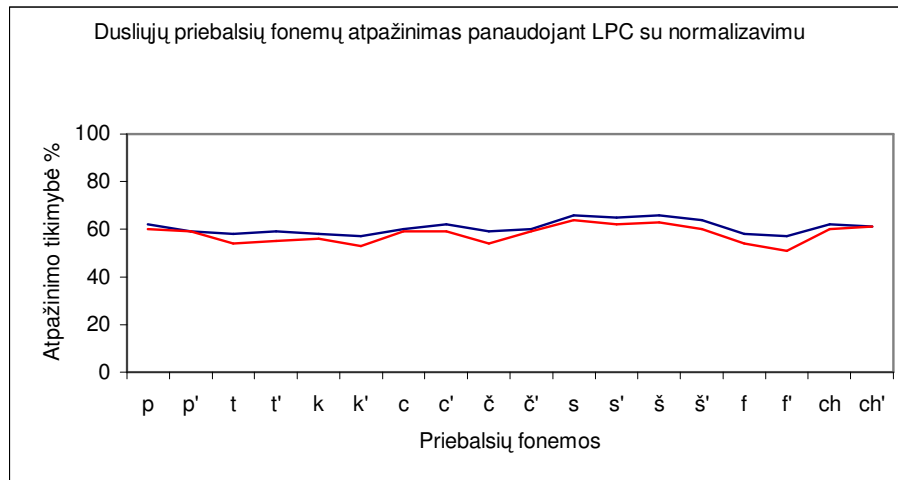
Priebalsių atpažinimo tikimybė, kai tinklas apmokytas dviejų diktorių fonemomis, taip pat ištirta priebalsius charakterizuojant LPC koeficientais. Skardžiųjų priebalsių eksperimento atpažinimo rezultatai pavaizduoti 3.23 pav.



3.23 pav. Skardžiųjų priebalsių atpažinimo tikimybė panaudojus LPC koeficientus ir tinklą apmokant dviejų diktorių fonemomis

Eksperimente su LPC koeficientais geriausiai buvo atpažinta m skardžioji fonema (79%, buvo 80%), blogiausiai atpažinta j fonema (55%, o buvo 60%). Labiausiai sumažėjo g' fonemos atpažinimas (56%, o buvo 63%).

Bandymo su dusliosiomis priebalsių fonemomis rezultatai pateikiami 3.24 pav.

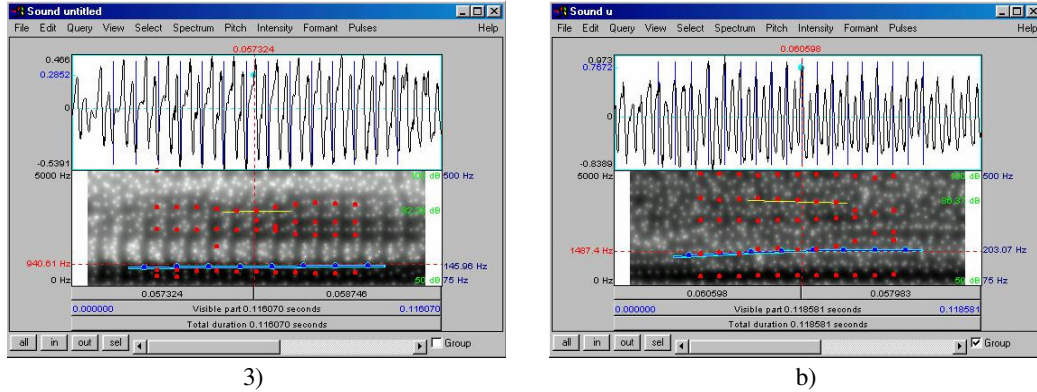


3.24 pav. Dusliųjų priebalsių atpažinimo tikimybė panaudojus LPC koeficientus ir tinklą apmokant dviejų diktorių fonemomis

3.24 paveikslėlyje esančiame grafike viršutinė juoda kreivė vaizduoja prieš tai buvusio dusliųjų priebalsių atpažinimo eksperimento rezultatus. Apatinė, raudona kreivė, rodo bandymo rezultatus kai neuroninis tinklas turėjo atpažinti vieno diktorius priebalsių fonemas iš 36 galimų variantų. Matome, kad atpažinimo tikimybė šiek tiek sumažėjo Eksperimente panaudojant LPC koeficientus geriausiai buvo atpažinta s duslioji fonema (64%, buvo 66%), blogiausiai atpažinta f' fonema (51%, o buvo 57%). Labiausiai sumažėjo f' fonemos atpažinimas (51%, o buvo 57%).

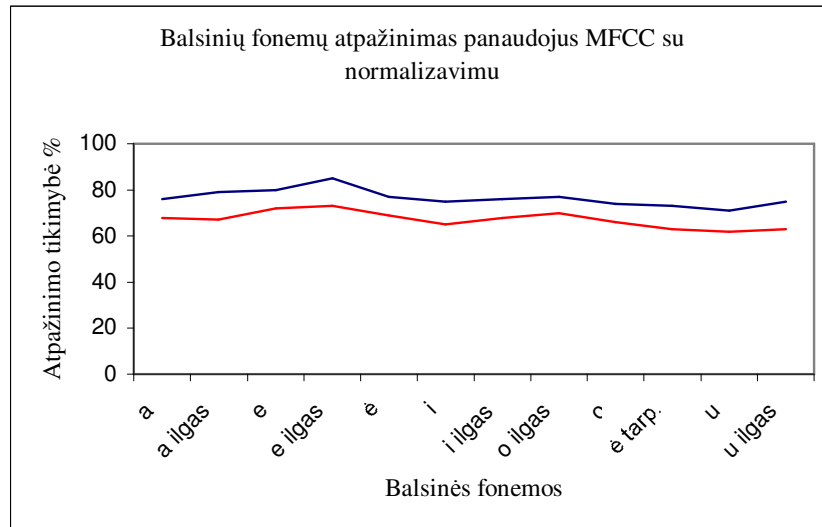
3.4 Izoliuotų fonemų, įrašytų triukšmingoje aplinkoje, atpažinimas

Paskutinis tiriamojo magistro darbo bandymas atliktas su izoliuotomis fonemomis, jų įrašimui pasirinkus triukšmingą aplinką. Eksperimentu norėta nustatyti triukšmo įtaka fonemų atpažinimui. Tam darant įrašą buvo įjungtas radijo imtuvas, garsas - vidutinės žmogaus kalbos lygio. Kaip skiriasi fonemos u signalas, kuri įrašyta triukšmingoje aplinkoje nuo tylaus įrašo matyti 3.25 pav. (a dalis - įrašas tylioje aplinkoje, b – įrašas triukšmingoje aplinkoje). Atpažinimui naudojome garsus, kurių amplitudė normalizuota pagal max reikšmę.



3.25 pav. Balsinės fonemos u signalas PRAAT lange (a) – tyli aplinka, b) – triukšminga).

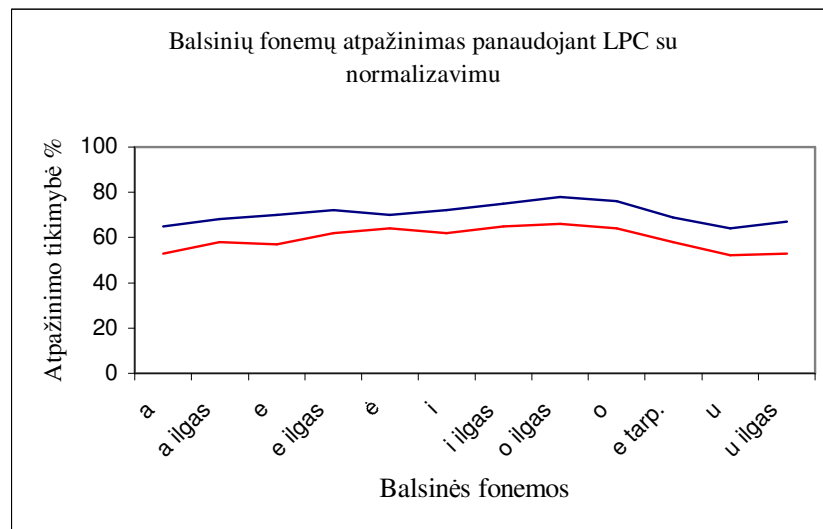
Balsių atpažinimo eksperimentas. Atpažįstant fonemas tinklas buvo apmokytas tyloje aplinkoje įrašytomis fonemomis. Atpažinimui naudotos fonemos, įrašus atlikus triukšmingoje aplinkoje ir jas charakterizavus MFCC koeficientais. Eksperimento rezultatai matomi 3.26 pa.



3.26 pav. Balsinių fonemų, įrašytų triukšmingoje aplinkoje, atpažinimo tikimybė panaudojus MFCC koeficientus.

Viršutinėje grafiko, esančio 3.26 pav. dalyje, dalyje esanti kreivė rodo atpažinimo ekperimento rezultatus, kai buvo dirbama su tyloje aplinkoje įrašytomis fonemomis. Apatinė, raudona kreivė, atspindi rezultatus kurie gauti šio izoliuotų balsinių fonemų atpažinimo eksperimento metu. Balsių atpažinimo tikimybė sumažėjo iki 15% Geriausiai atpažinta e ilgoji fonema (73%, buvo 85%), blogiausiai u fonema (62%, buvo 71%).

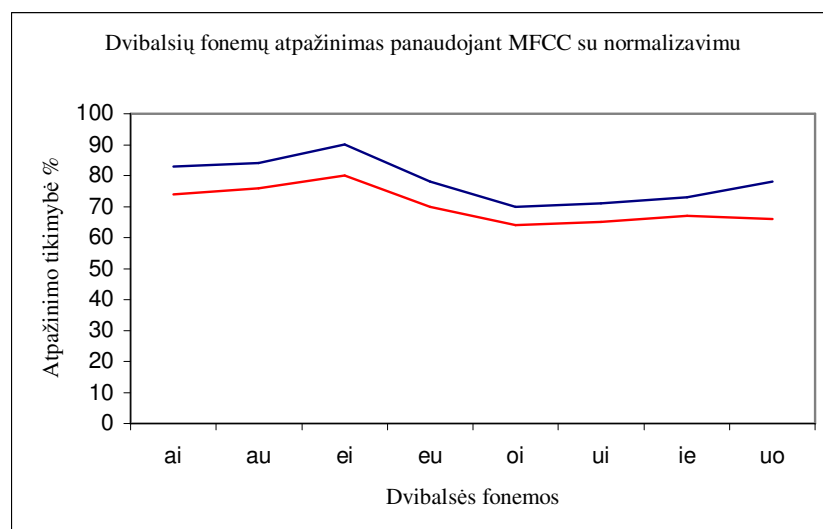
Balsines fonemas charakterizavus LPC koeficientais, gauti kitokie bandymo rezultatai, kurie pavaizduoti 3.27 pav.



3.27 pav. Balsinių fonemų, įrašytų triukšmingoje aplinkoje, atpažinimo tikimybė panaudojus LPC koeficientus.

Apatinė raudona linija žymi šio ekperimento su LPC koeficientais rezultatus. Geriausiai atpažinta o ilgoji fonema (66%, buvo 78%), blogiausiai u fonema (52%, buvo 64%).

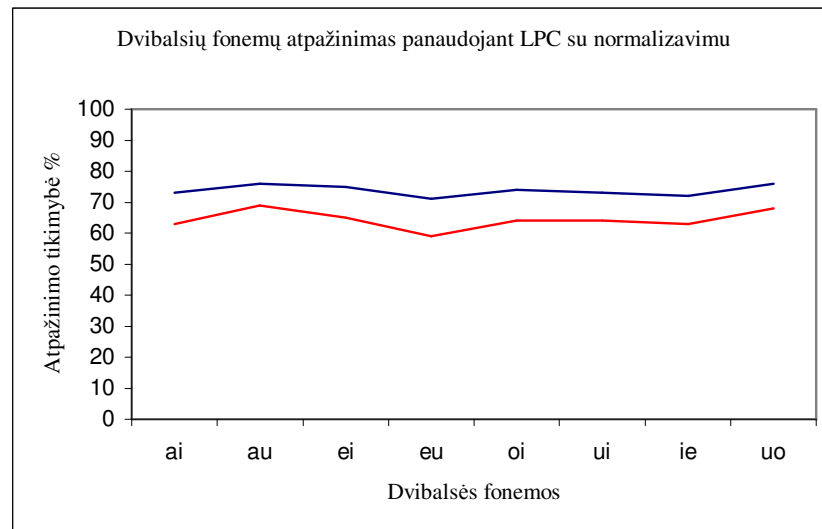
Dvibalsių atpažinimo eksperimentas. Pirmiausia atliktas eksperimentas dvibalses fonemas charakterizavus MFCC koeficientais. Rezultatai pateikiami 2.28 pav.



3.28 pav. Dvibalsių fonemų, įrašytų triukšmingoje aplinkoje, atpažinimo tikimybė panaudojus MFCC.

Gauti rezultatai taip pat prastesni už analogiško bandymo, kai atpažinimui buvo naudojamos tylioje aplinkoje įrašytos fonemos. Geriausiai atpažinta ei fonema (80%, buvo 90%), blogiausiai oi fonema (64%, buvo 70%).

Paskutinis svibalsių atpažinimo eksperimentas atliktas su dvibalsių fonemomis, jų charakterizavimui panaudojus LPC koeficientus. Rezultatai matomi 3.29 pav.



3.29 pav. Dvibalsių fonemų, įrašytų triukšmingoje aplinkoje, atpažinimo tikimybė panaudojus LPC koeficientus.

Geriausiai buvo atpažinta au fonema (69%, buvo 76%), blogiausiai - eu fonema (59%, buvo 71%).

IŠVADOS

Tiriamąjį magistrą darbą metu apžvelgtos šiuolaikinės šnekos atpažinimo sistemos. Pateikiama populiariausių programų vartotojiška apžvalga, bei testai. Buvo išanalizuoti MFCC ir LPC koeficientų gavimo principai. Po to atliktas abiejų darbo autorių įrašytų lietuvių kalbos fonemų atpažinimo bandymas. Fonemų atpažinimui buvo įrašyti lietuvių kalbos žodžiai iš kurių iškirpta po 300 kiekvienos fonemos pavyzdžių.

Tiriamąjį magistrą darbą metu atlikus eilę izoliuotų fonemų atpažinimo eksperimentų, buvo gauti rezultatai iš kurių galima padaryti sekančias išvadas:

1. Atpažinimui naudojant neuroninį tinklą (daugiasluoksnį perceptroną) ir fonemas charakterizuojant MFCC ar LPC koeficientais, atpažinimo tikimybė neviršija 90 procentų. Geriausiai atpažystamos dvibalsių fonemos. To nepakanka kokybiškam lietuvių šnekos atpažinimui.
2. Naudojant MFCC koeficientus pavienės fonemos atpažystamos geriau nei panaudojus LPC koeficientus. Skirtumas tarp vienu ir kitu apytiksliai 15 procentų.
3. LPC koeficientų pranašumas prieš MFCC yra toks, kad atpažinimo tikimybės kreivė yra tolygesnė.
4. Atlikus signalo normalizavimą pagal maksimalią amplitudę ir atpažinimui naudojant LPC koeficientus atpažinimo tikimybė padidėja.
5. Atpažinimui naudojant izoliuotas fonemas įrašytas triukšmingoje aplinkoje atpažinimo tikimybė sumažėja.
6. Norint gauti efektyvesnį fonemų atpažinimą, vertėtų naudoti bent keletą fonemas charakterizuojančių parametrų.

LITERATŪRA

1. A. Adrid, J. P. Barjaktarevic, O. Ozun, M. Smith, P. Steurer. Automatic Speech Recognition for Isolated Words. 2003, [žiūrėta 2004-02-11]. Prieiga per Internetą: <<http://www.ee.ucla.edu/~psteurer/projects/ee214bprojectreport.pdf>>.
2. A.Salomon. Detection of speech landmarks: Use of temporal information/ A.Salomon, Carol Y. Espy - Wilson, O. Deshmukh//Acoustical Society of America. 2004, p.1296-1305.
3. Aleksas Girdenis. Fonologija. Vilnius, 1981. 216 p.
4. Aleksas Girdenis. Teoriniai lietuvių fonologijos pagrindai. Vilnius, 2003. 388 p.
5. Antanas Pakerys. Lietuvių bendrinės kalbos fonetika. Vilnius, 1995. 316p.
6. Bojana Gajich. Speech parameterization for automatic speech recognition in noisy conditions. [žiūrėta 2004-12-03]. Prieiga per Internetą: <http://maxwell.me.gu.edu.au/spl/publications/papers/norsig01_bojana.pdf>
7. Filterbank Analysis. [žiūrėta 2004-10-15]. Prieiga per Internetą: <<http://labrosa.ee.columbia.edu/doc/HTKBook21/node54.html>>
8. H.G. Hirsch, K. Hellwig, S. Dobler. Speech Recognition at Multiple Sampling. 2001, vasaris [žiūrėta 2004-02-10]. Prieiga per Internetą: <<http://dnt.kr.hs-niederrhein.de/papers/eurospeech01.pdf>>.
9. L. Rabiner, B. H. Juang. Fundamentals of Speech Recognition. Prentice Hall, 1993. Lietuvių kalbos gramatika. I tomas/A. Laigonaitė, V. Mažiulis, K. Ulvydas, Vilnius, 1965. 731p.
10. Objektu atpažinimas ir klasifikavimas. Neuroniniai tinklai. [žiūrėta 2005-01-13]. Prieiga per Internetą: <http://vejas.pit.ktu.lt/~bronius/intelektika_old/52_tema/52_tema.html%25>
11. Pokalbis su mašina. [žiūrėta 2004-10-15]. Prieiga per Internetą: <<http://www.rtn.lt/rtn/0204/pokalbis.html>>
12. R. Duda, P. Hart. Pattern Classification and Scene Analysis. Wiley, 1973.
13. S. Haykin. Neural Networks: A Comprehensive Foundation. Prentice Hall, 1999.
14. T. Thrasyvoulou and S. Benton. Speech parameterization using the Mel scale Part II. 2004, sausis [žiūrėta 2005-01-12]. Prieiga per Internetą:

http://hitchcock.dlt.asu.edu/media5/a_spanias/speech-recognition/real-lectures/PDF/SP-4.pdf>.

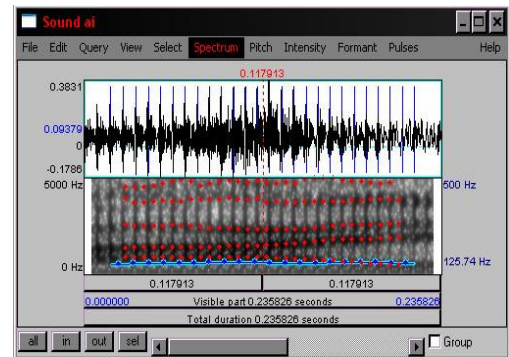
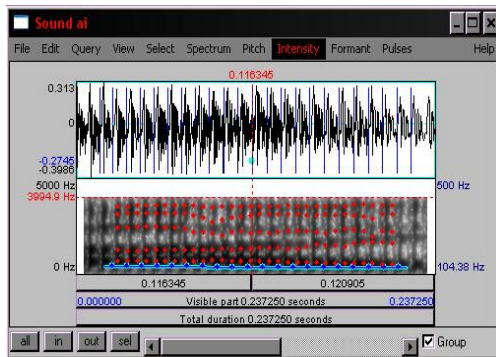
15. Дмитрий Михайлов. FFT Анализ. 2003, vasaris [žiūrēta 2005-02-27]. Prieiga per Internetą: <<http://ts.kmc.ru>>.

16. Четвериков С.В. Нейронные сети и их устройство. 1998, balandis [žiūrēta 2004-04-17]. Prieiga per Internetą: <<http://p6.ru/referats/inf/37/hai-0392.zip.htm>>.

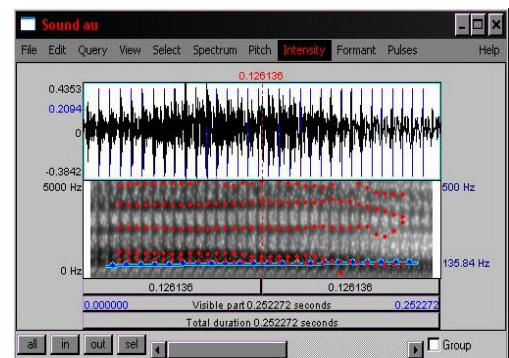
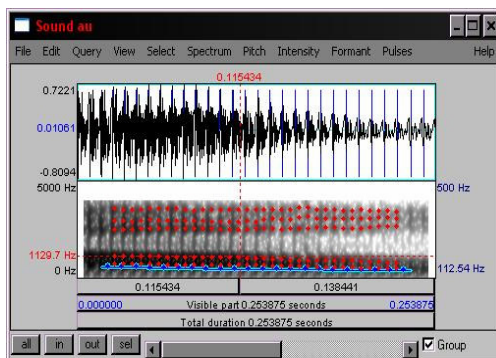
1. PRIEDAS

Lietuvių kalbos fonemų signalai PRAAT lange

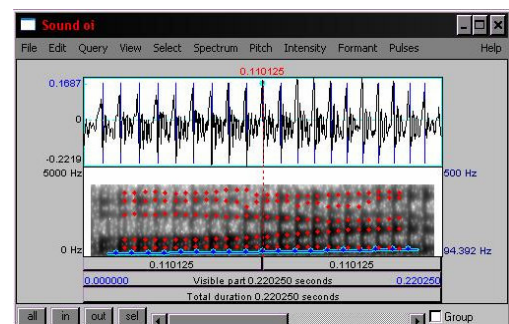
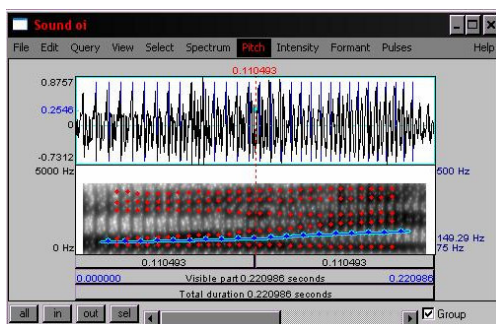
Lietuvių kalbos dvibalsių fonemos



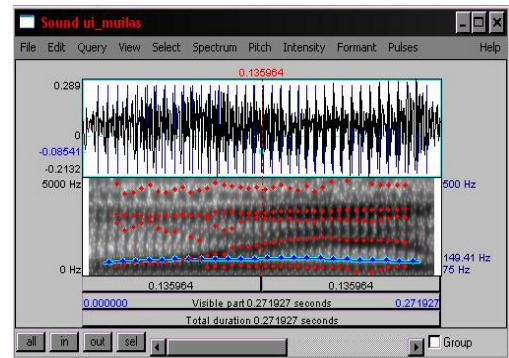
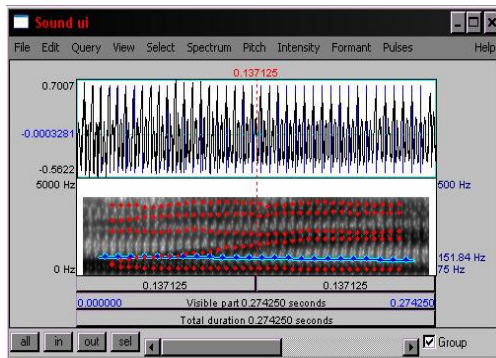
1 pav. Fonemos “ai” signalas PRAAT lange



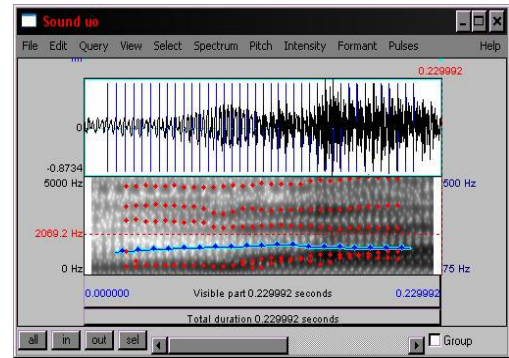
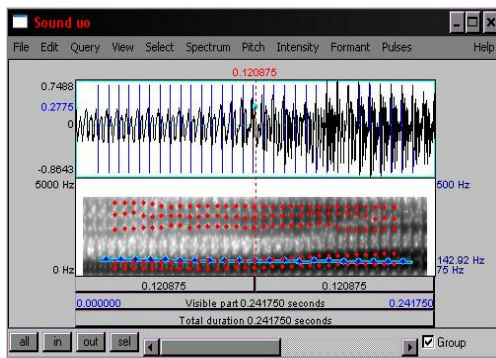
2 pav. Fonemos “au” signalas PRAAT lange



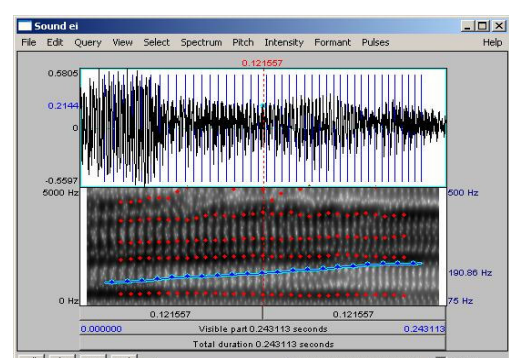
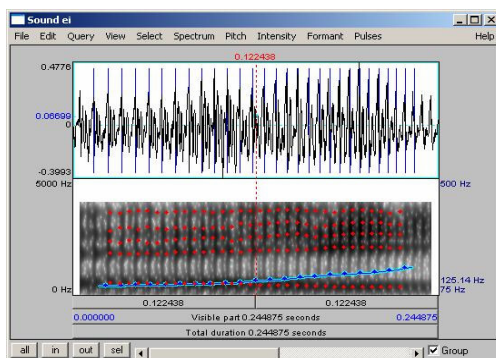
3 pav. Fonemos “oi” signalas PRAAT lange



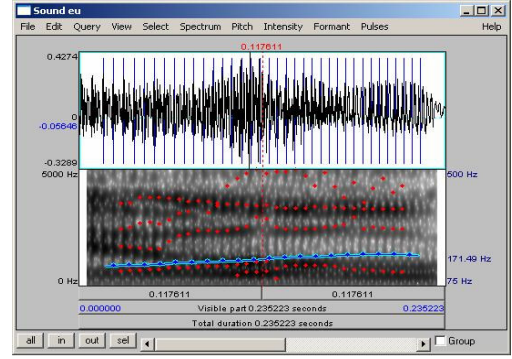
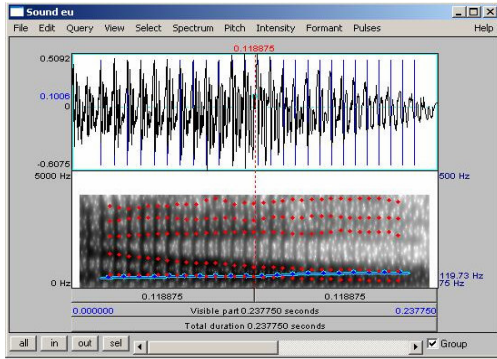
4 pav. Fonemos “ui” signalas PRAAT lange



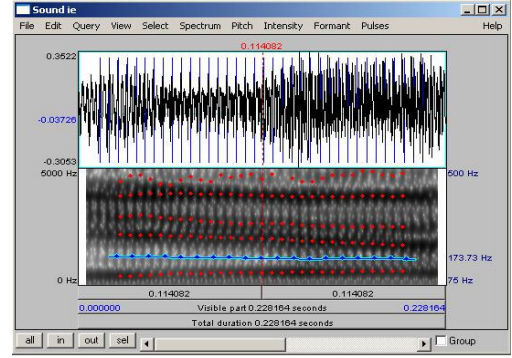
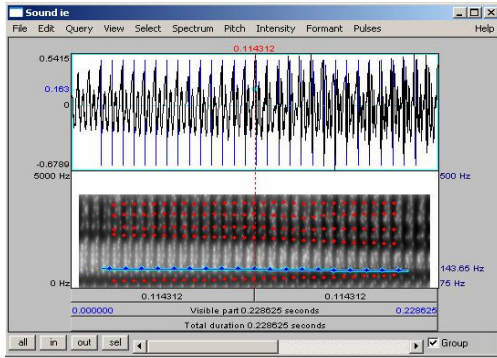
5 pav. Fonemos “uo” signalas PRAAT lange



6 pav. Fonemos “ei” signalas PRAAT lange

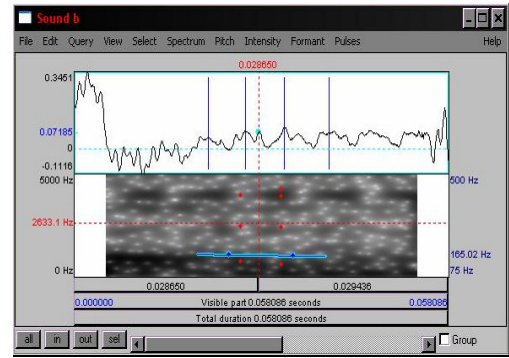
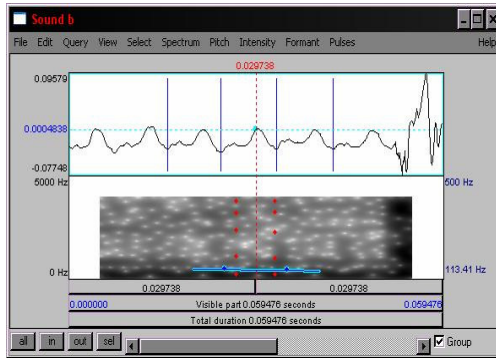


7 pav. Fonemos “eu” signalas PRAAT lange

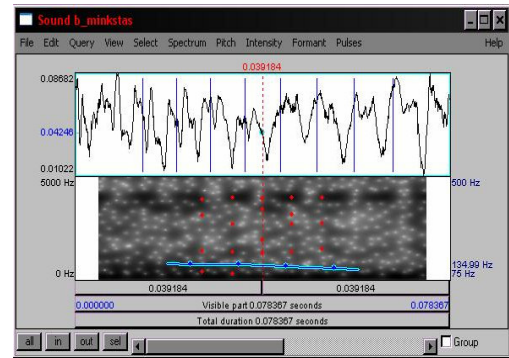
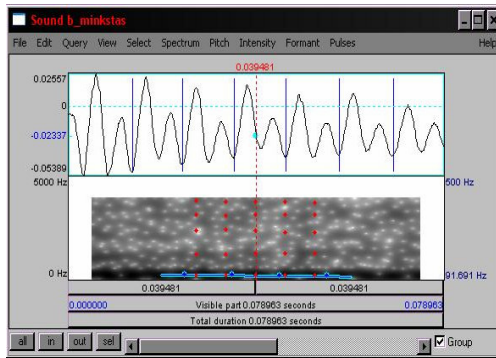


8 pav. Fonemos “ie” signalas PRAAT lange

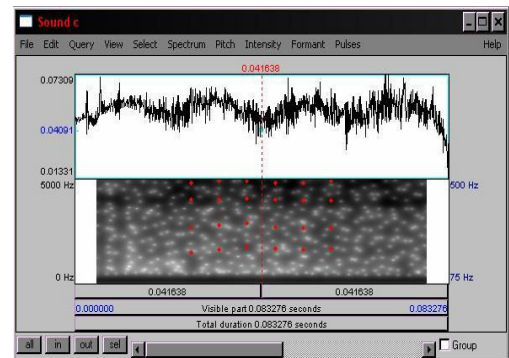
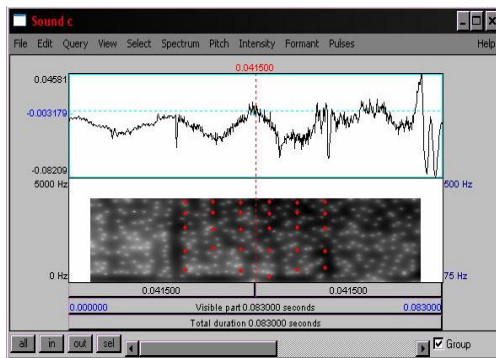
Lietuvių kalbos priebalsių fonemos



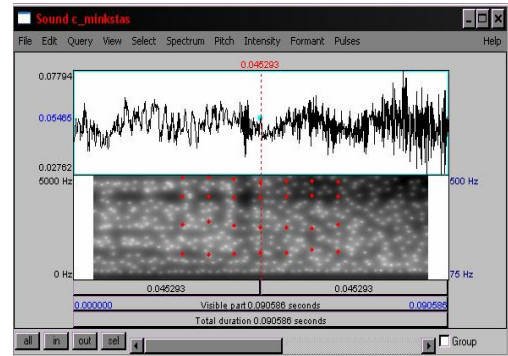
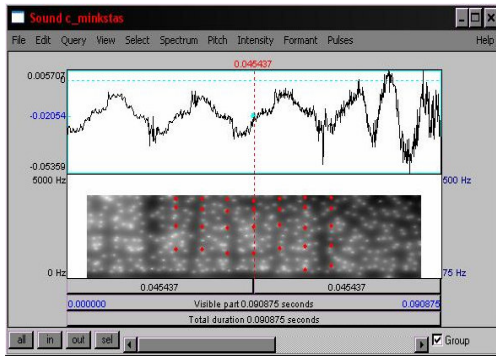
9 pav. Fonemos “b” signalas PRAAT lange



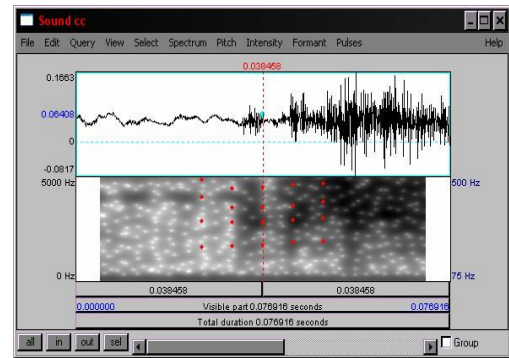
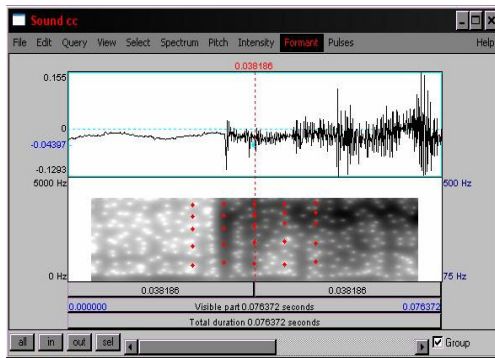
10 pav. Fonemos “b” signalas PRAAT lange



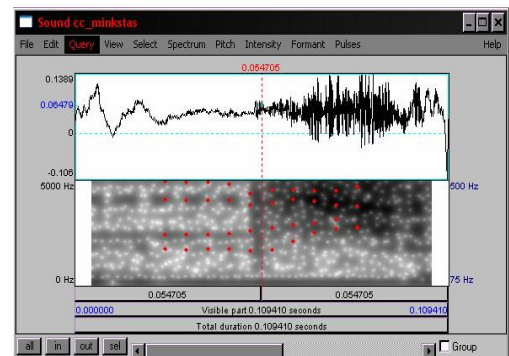
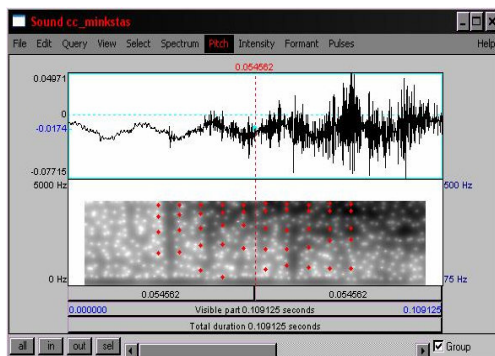
11 pav. Fonemos “c” signalas PRAAT lange



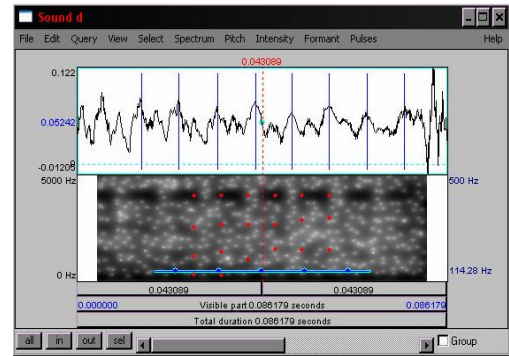
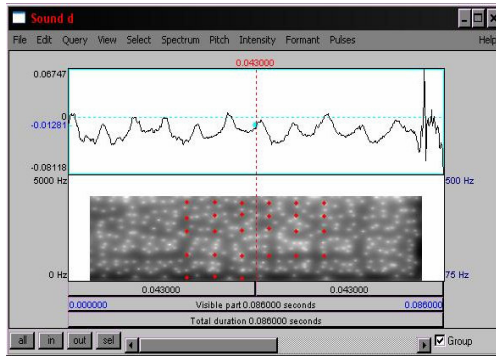
12 pav. Fonemos “c” signalas PRAAT lange



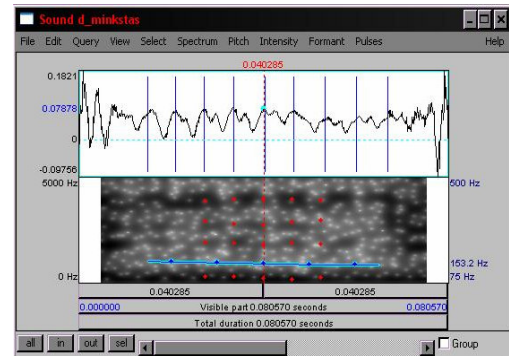
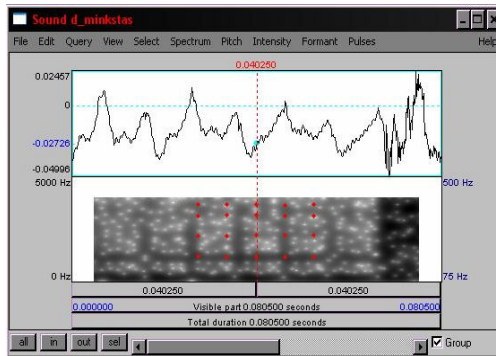
13 pav. Fonemos “č” signalas PRAAT lange



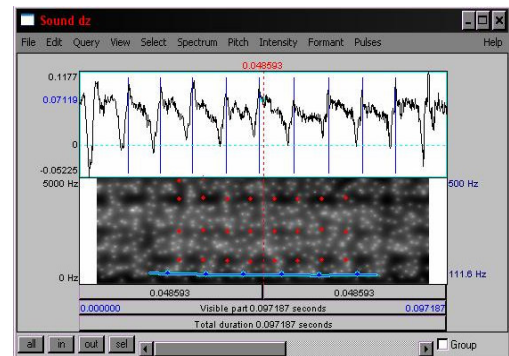
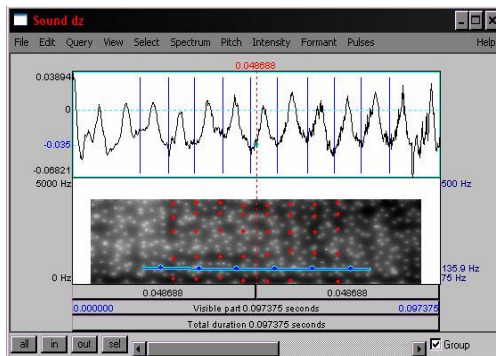
14 pav. Fonemos “č” signalas PRAAT lange



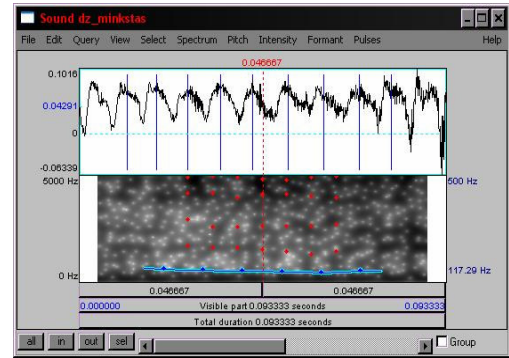
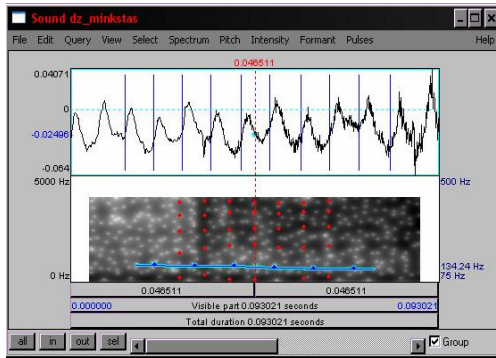
15 pav. Fonemos “d” signalas PRAAT lange



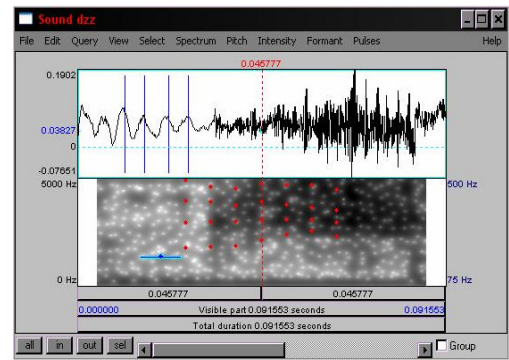
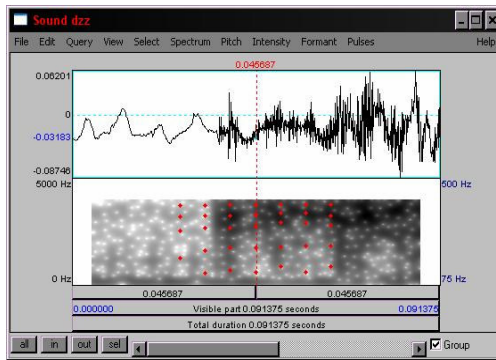
16 pav. Fonemos “d” signalas PRAAT lange



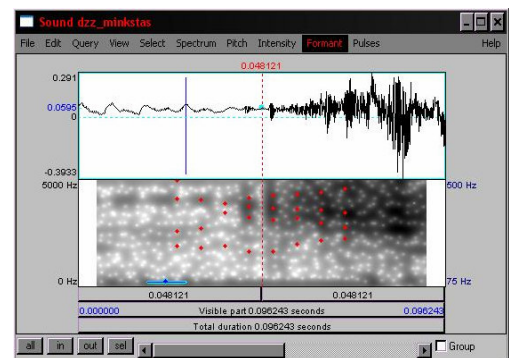
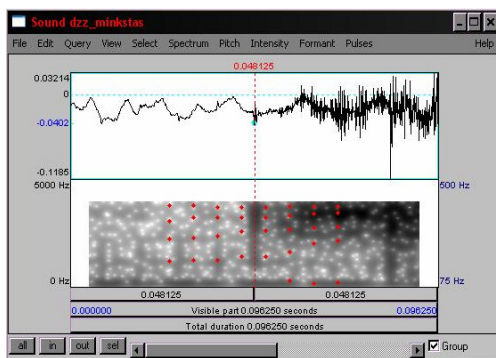
17 pav. Fonemos “dz” signalas PRAAT lange



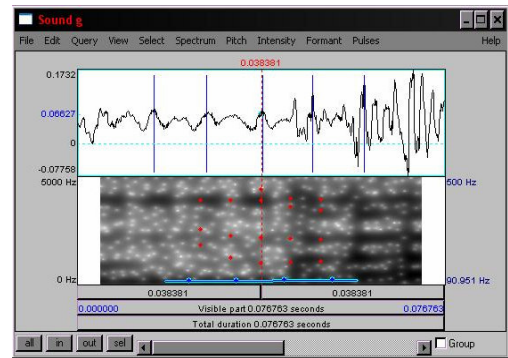
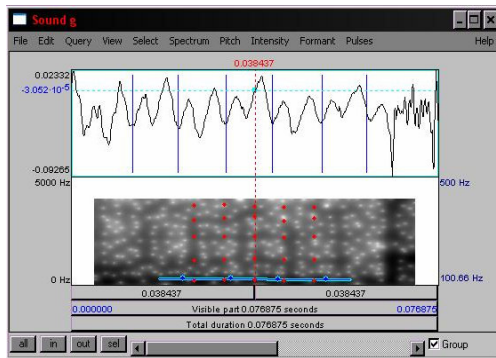
18 pav. Fonemos “dz” signalas PRAAT lange



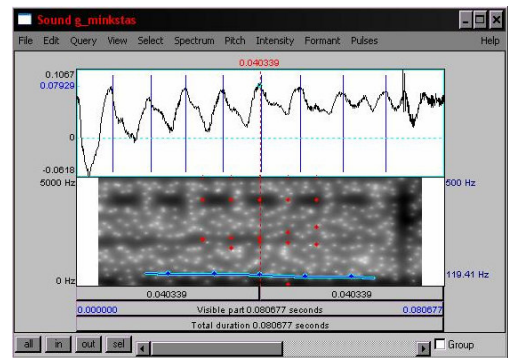
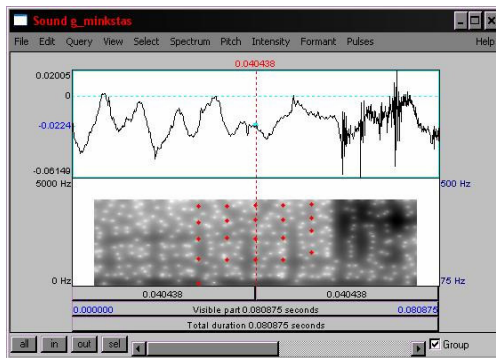
19 pav. Fonemos “dž” signalas PRAAT lange



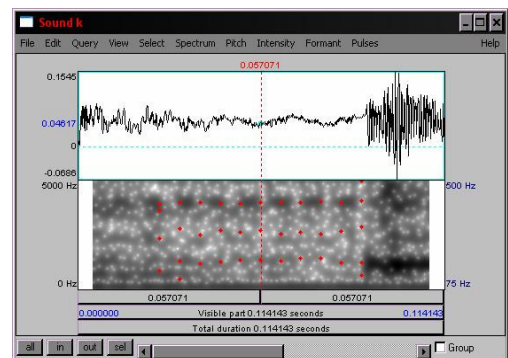
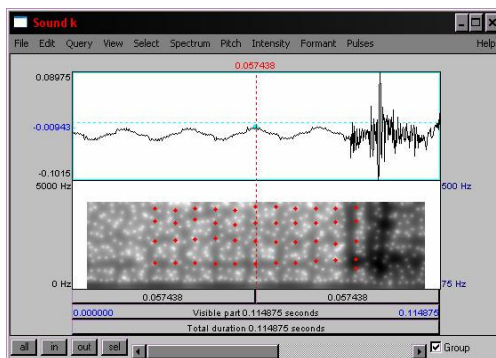
20 pav. Fonemos “dž” signalas PRAAT lange



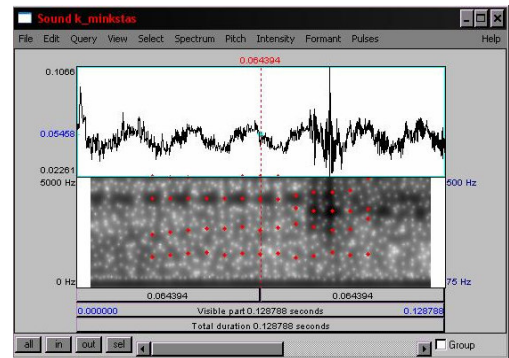
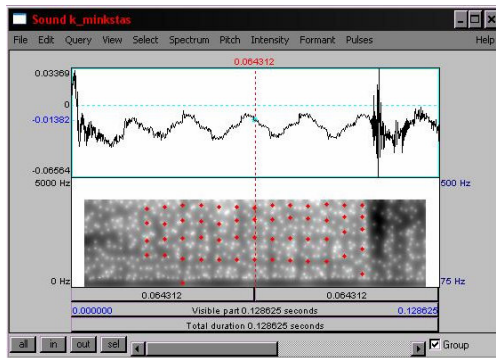
21 pav. Fonemos “g” signalas PRAAT lange



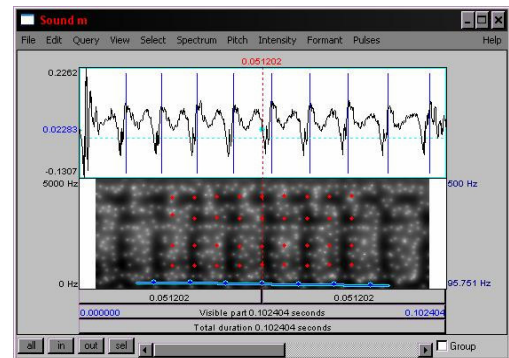
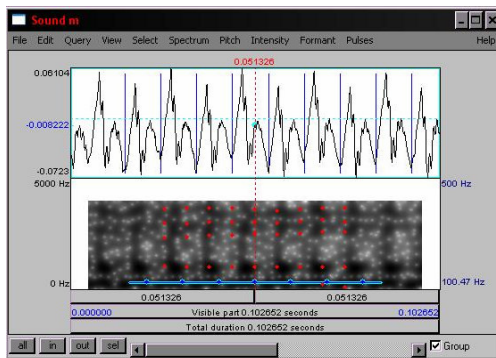
22 pav. Fonemos “g” signalas PRAAT lange



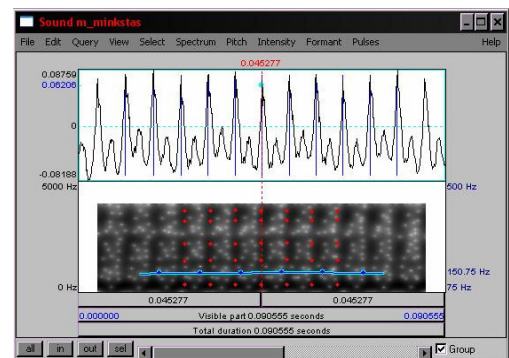
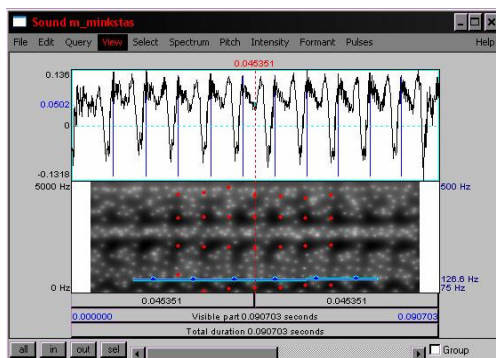
23 pav. Fonemos “k” signalas PRAAT lange



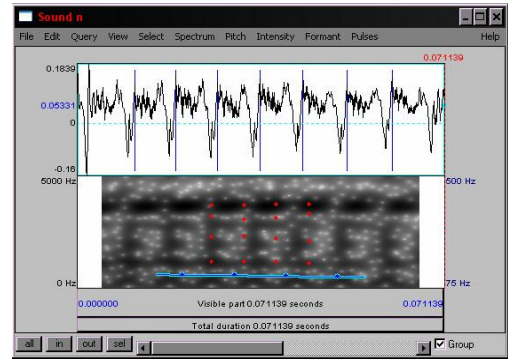
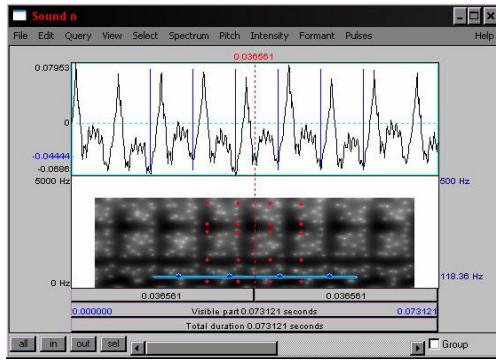
24 pav. Fonemos “k” signalas PRAAT lange



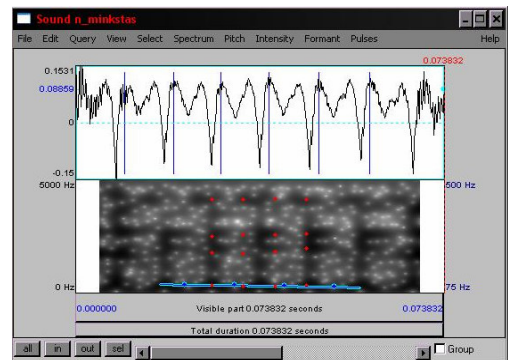
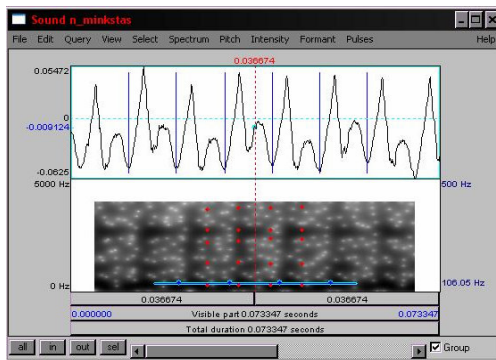
25 pav. Fonemos “m” signalas PRAAT lange



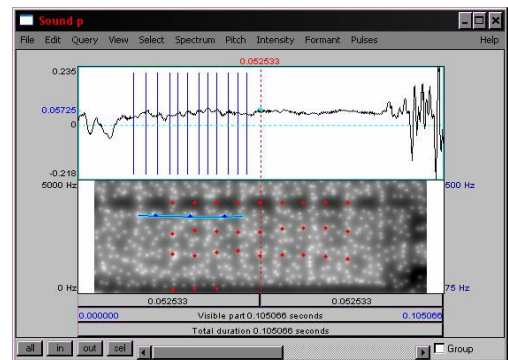
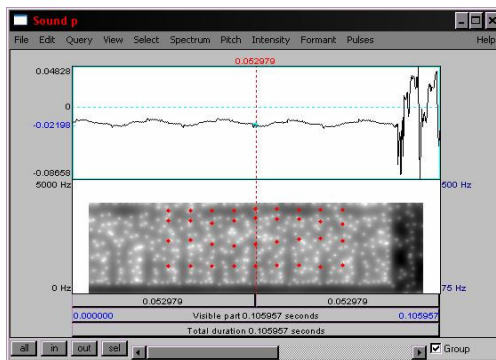
26 pav. Fonemos “m” signalas PRAAT lange



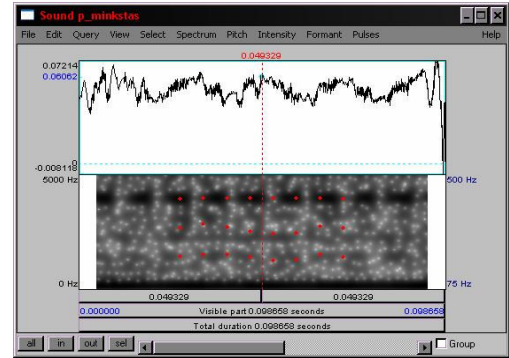
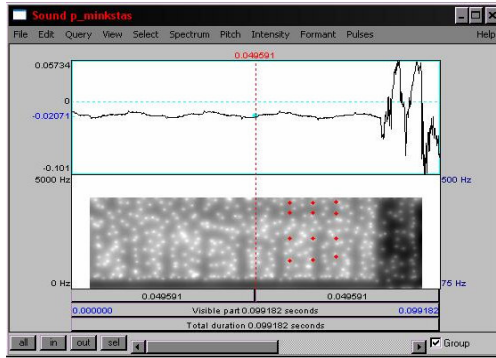
27 pav. Fonemos “n” signalas PRAAT lange



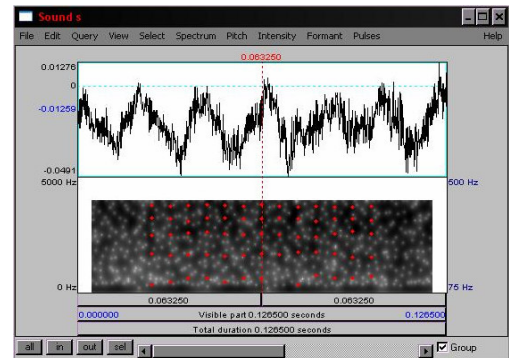
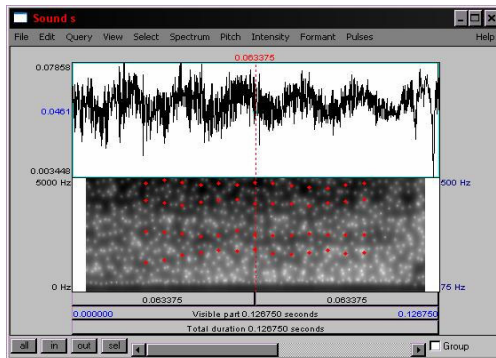
28 pav. Fonemos “n” signalas PRAAT lange



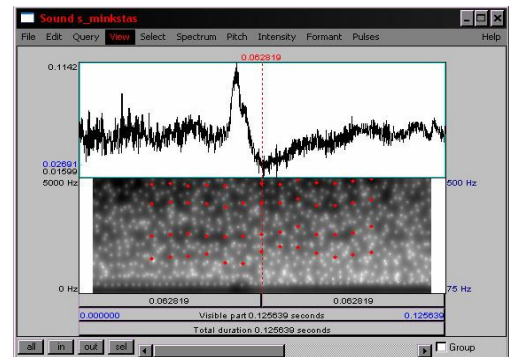
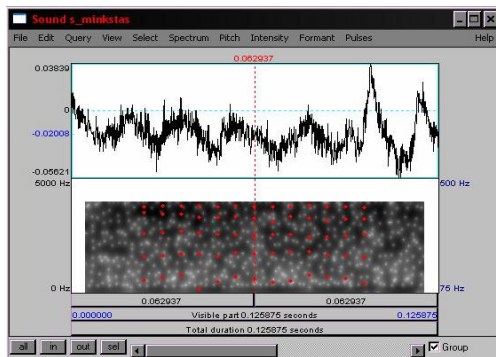
29 pav. Fonemos “p” signalas PRAAT lange



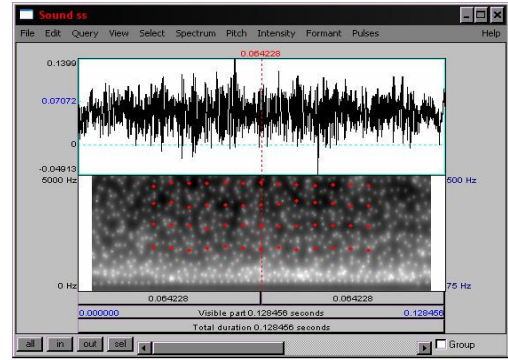
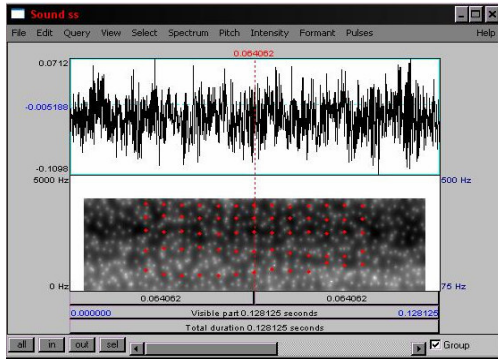
30 pav. Fonemos “p” signalas PRAAT lange



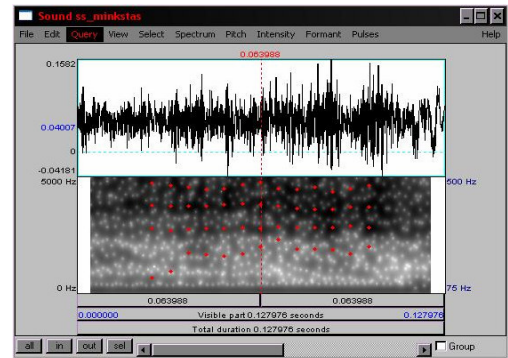
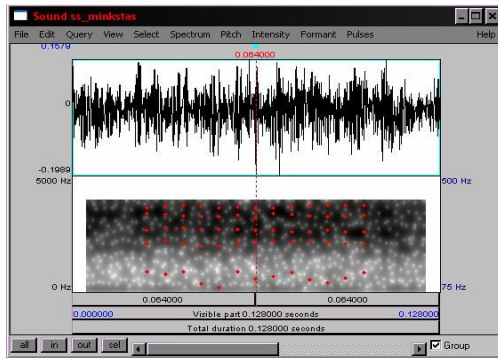
31 pav. Fonemos “s” signalas PRAAT lange



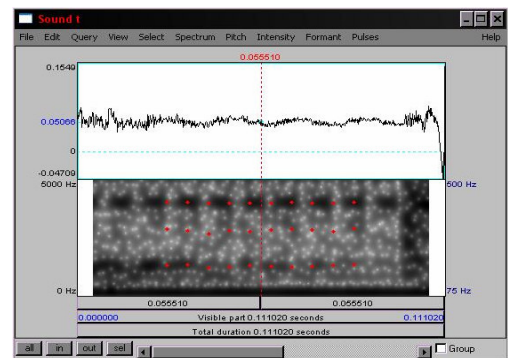
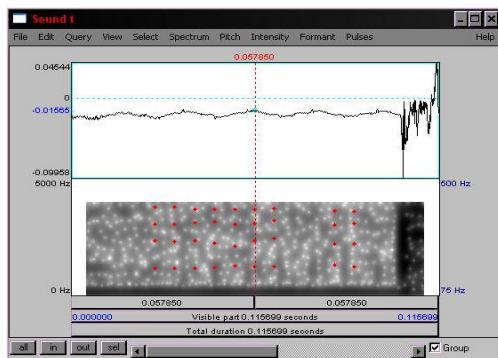
32 pav. Fonemos “s” signalas PRAAT lange



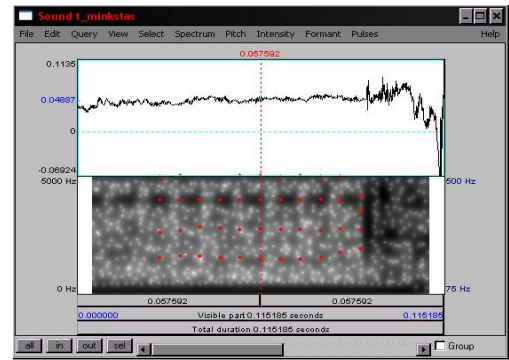
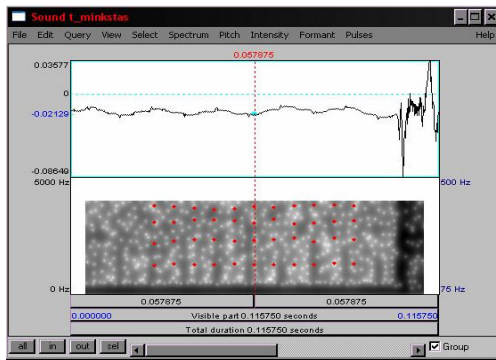
33 pav. Fonemos “s” signalas PRAAT lange



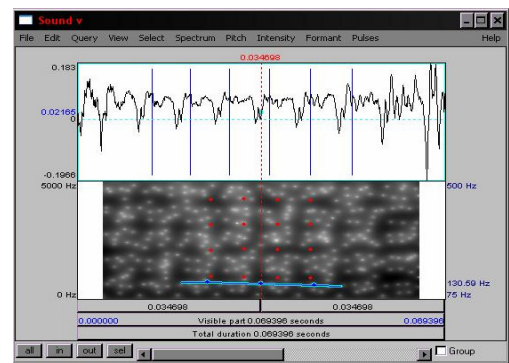
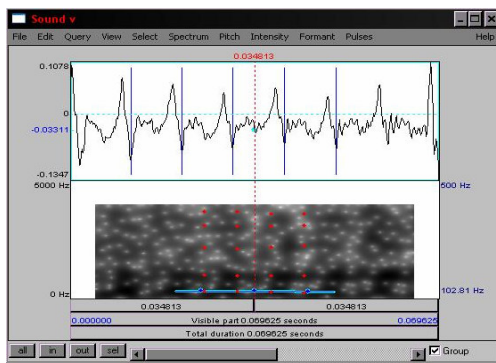
34 pav. Fonemos “s” signalas PRAAT lange



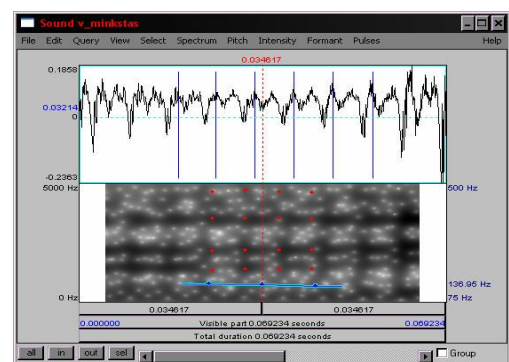
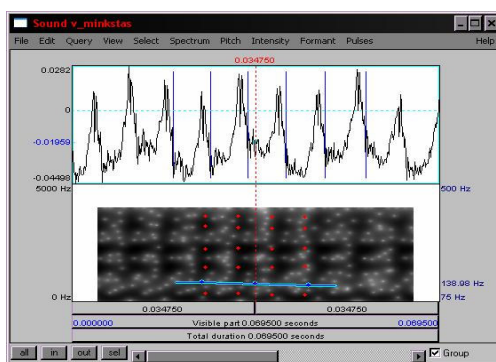
35 pav. Fonemos “t” signalas PRAAT lange



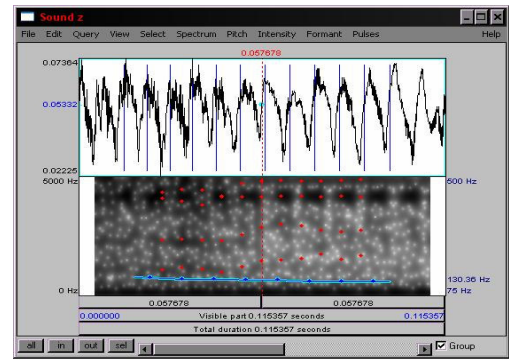
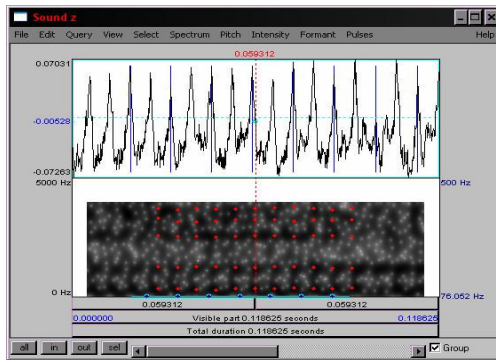
36 pav. Fonemos “t” signalas PRAAT lange



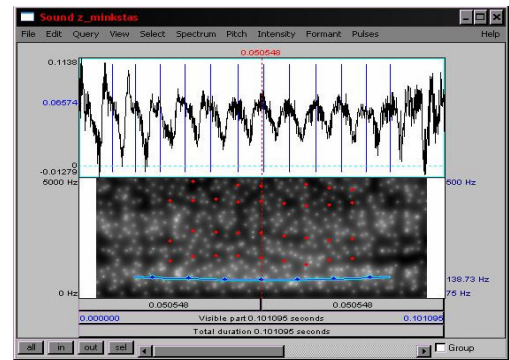
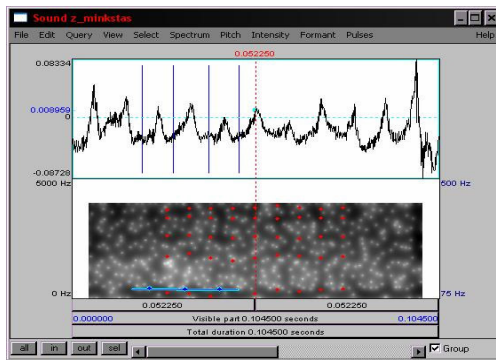
37 pav. Fonemos “v” signalas PRAAT lange



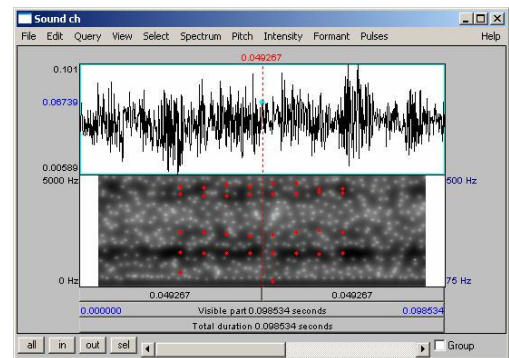
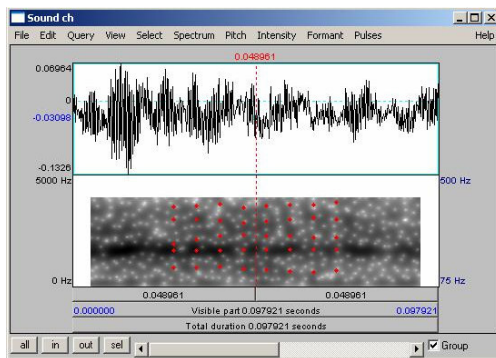
38 pav. Fonemos “v” signalas PRAAT lange



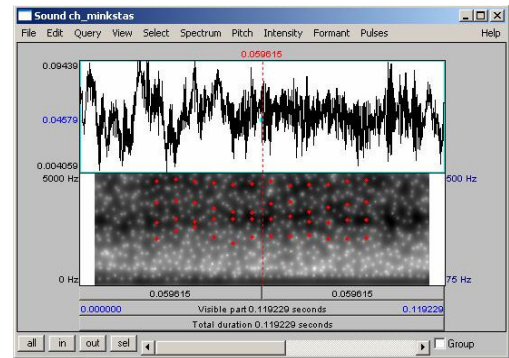
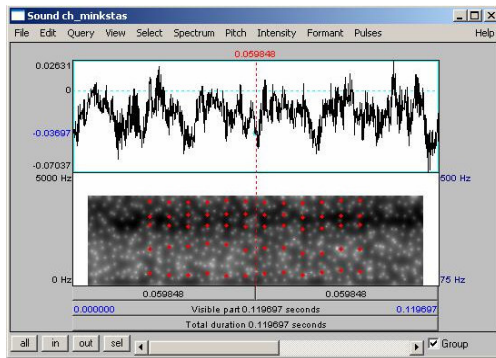
39 pav. Fonemos “z” signalas PRAAT lange



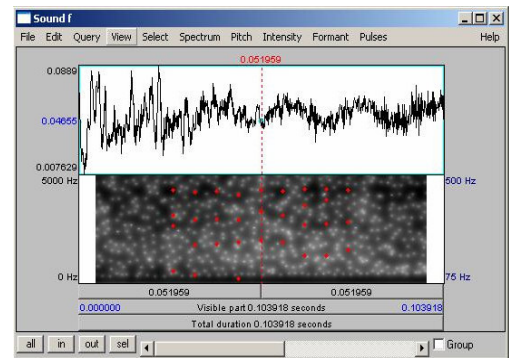
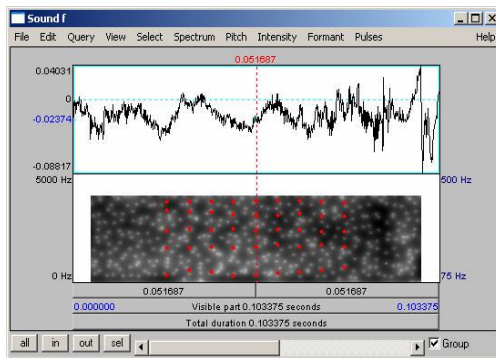
40 pav. Fonemos “z” signalas PRAAT lange



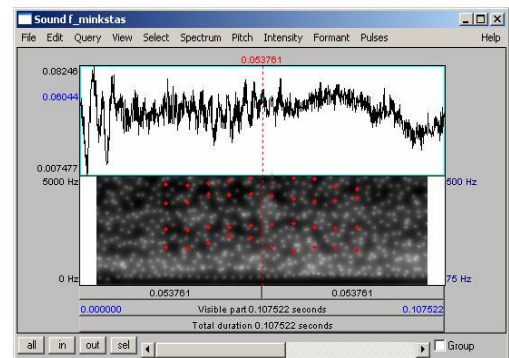
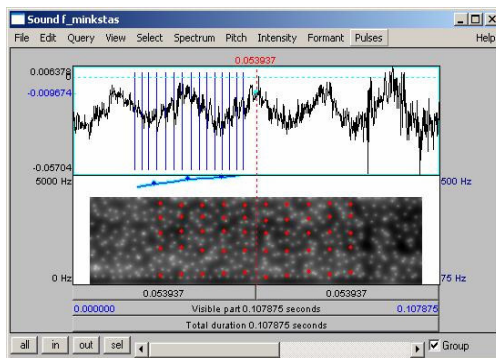
41 pav. Fonemos “ch” signalas PRAAT lange



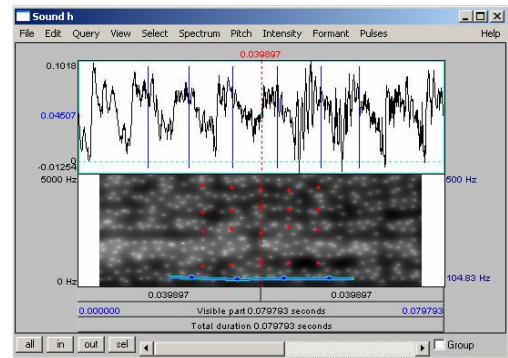
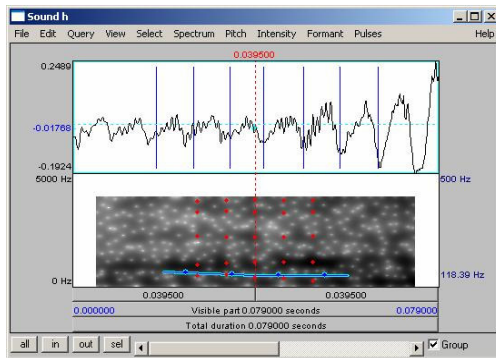
42 pav. Fonemos “ch” signalas PRAAT lange



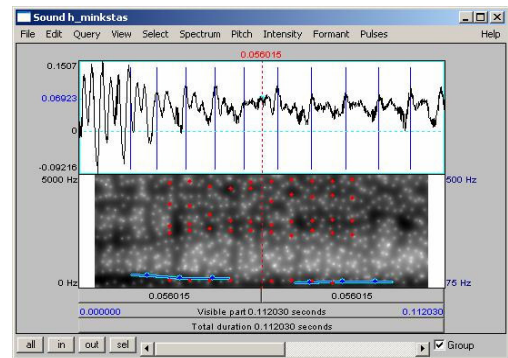
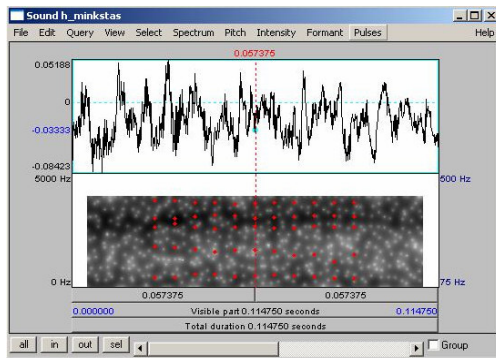
43 pav. Fonemos “f” signalas PRAAT lange



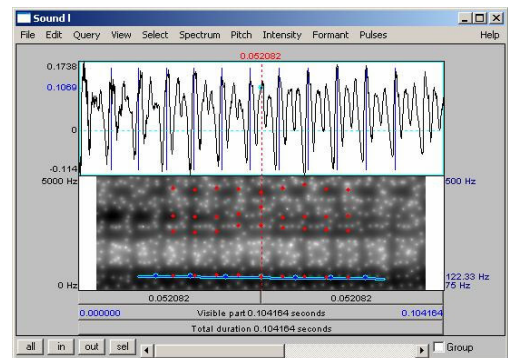
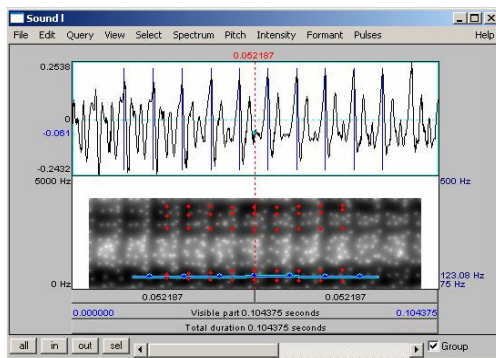
44 pav. Fonemos “f” signalas PRAAT lange



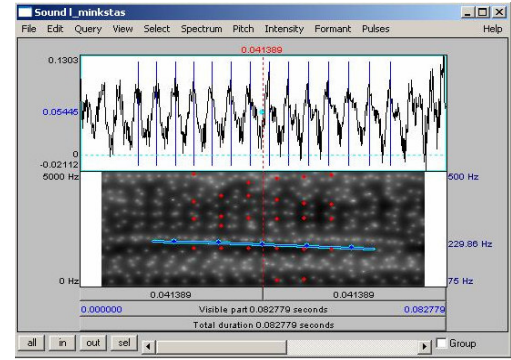
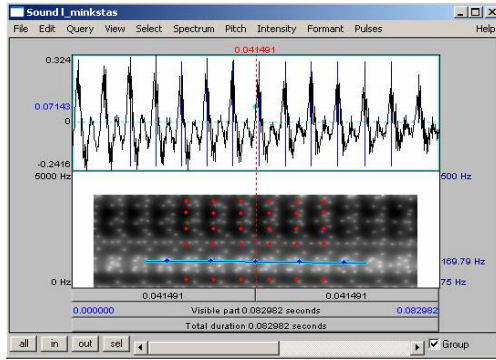
45 pav. Fonemos “h” signalas PRAAT lange



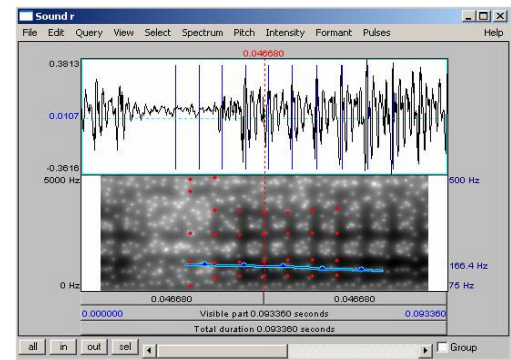
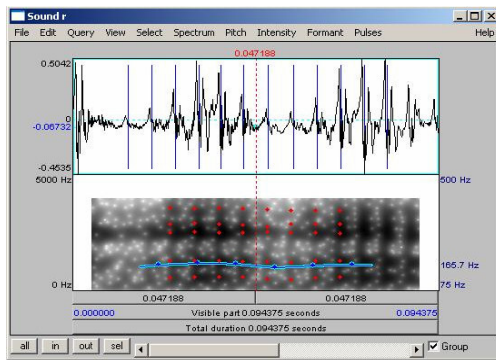
46 pav. Fonemos “h” signalas PRAAT lange



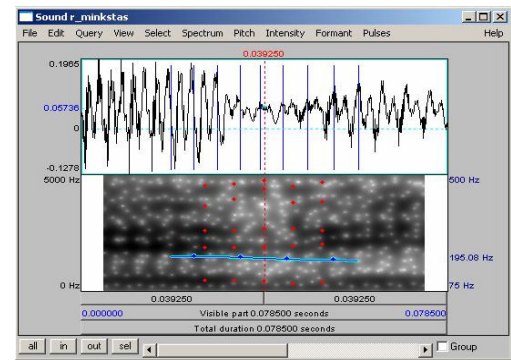
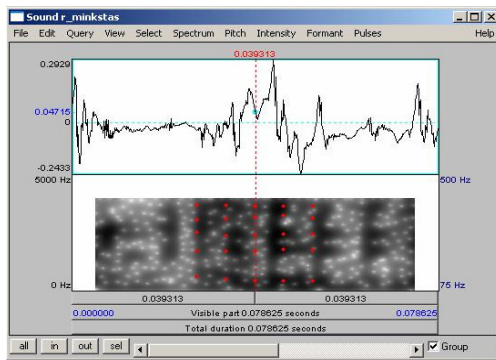
47 pav. Fonemos “l” signalas PRAAT lange



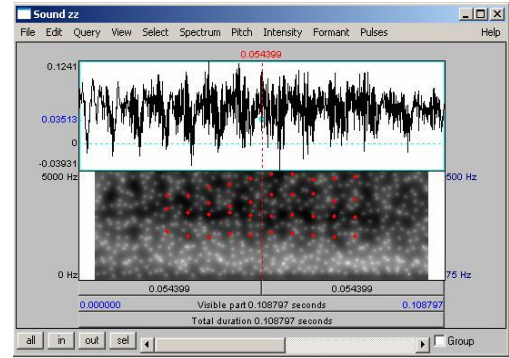
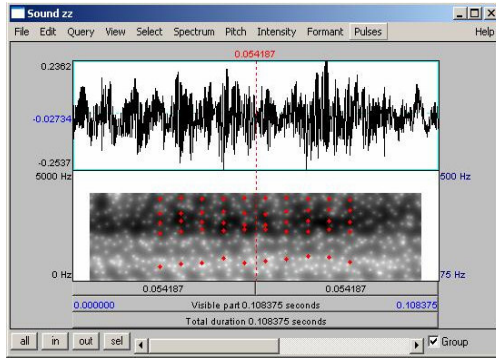
48 pav. Fonemos “l” signalas PRAAT lange



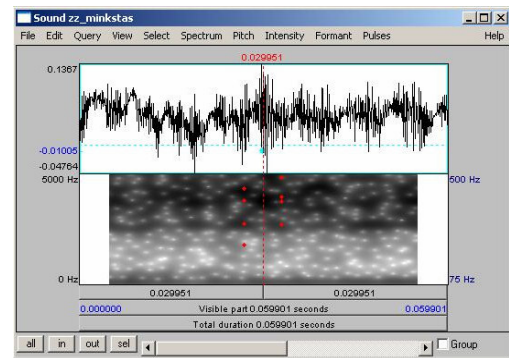
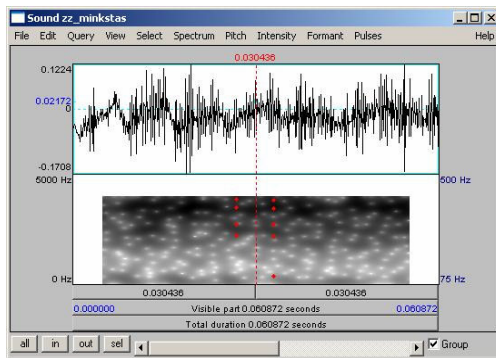
49 pav. Fonemos “r” signalas PRAAT lange



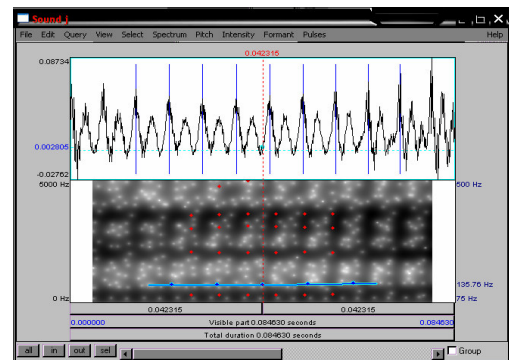
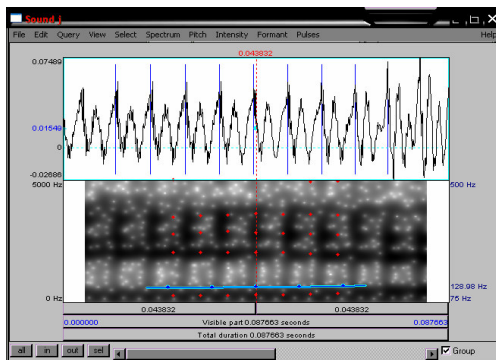
50 pav. Fonemos “r” signalas PRAAT lange



51 pav. Fonemos “ž” signalas PRAAT langoje

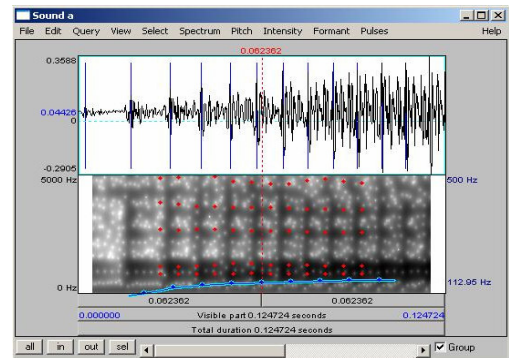
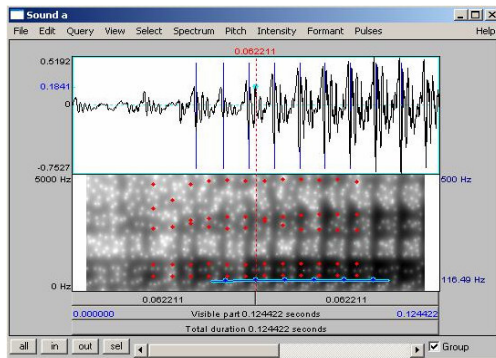


52 pav. Fonemos “ž” signalas PRAAT langoje

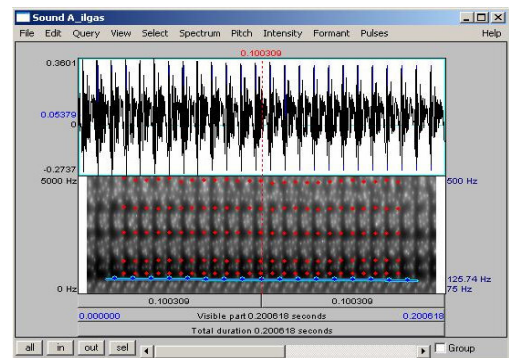
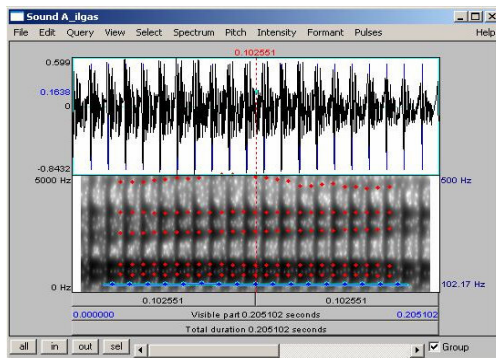


53 pav. Fonemos “j” signalas PRAAT langoje

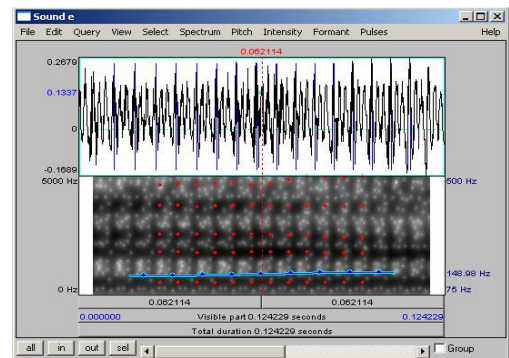
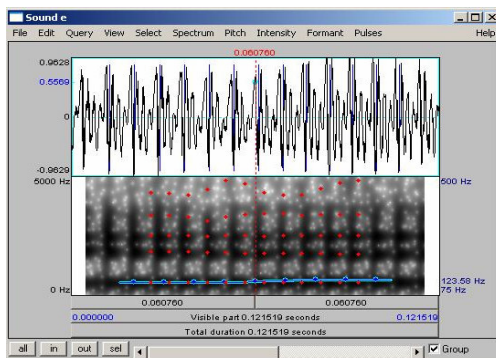
Lietuvių kalbos balsių fonemos



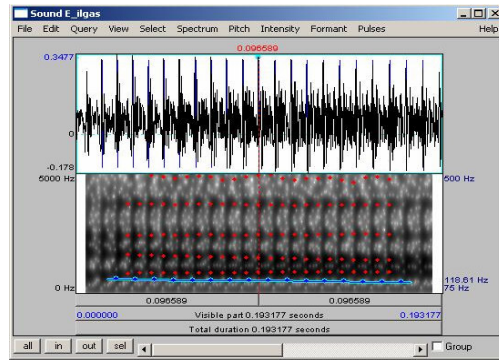
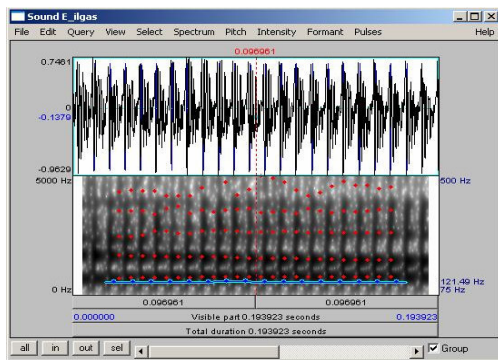
54 pav. Fonemos “a” signalas PRAAT lange



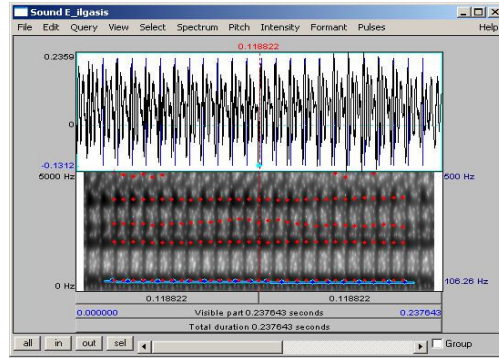
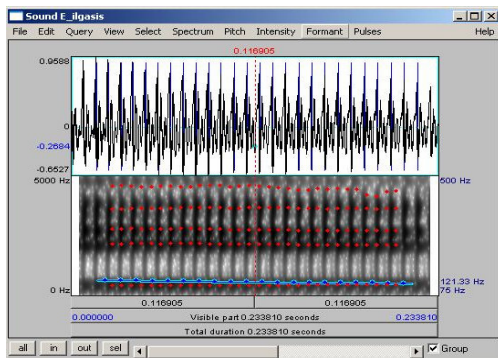
55 pav. Fonemos “a ilgoji” signalas PRAAT lange



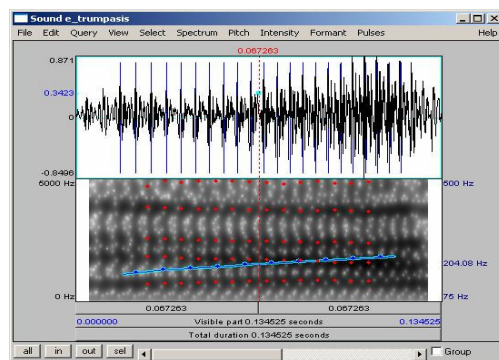
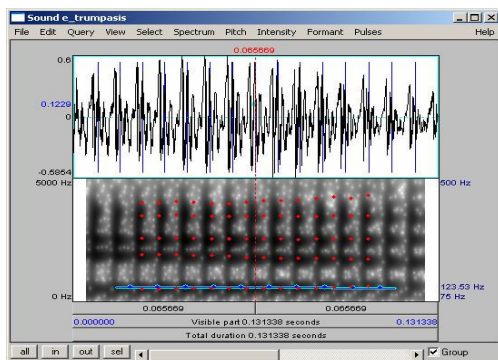
56 pav. Fonemos “e” signalas PRAAT lange



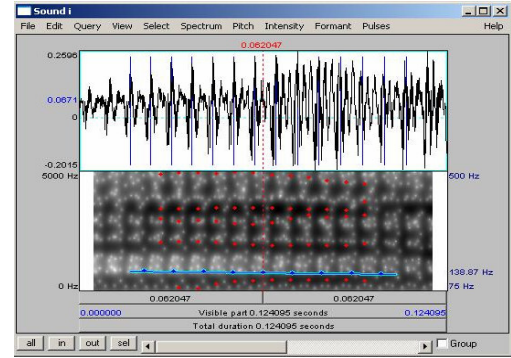
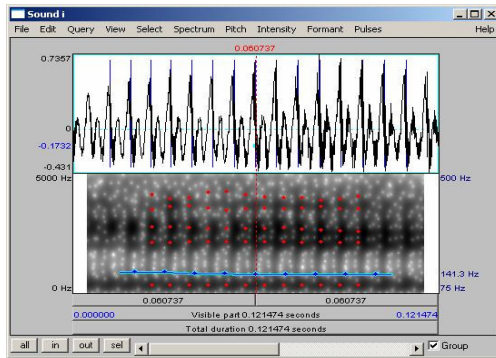
57 pav. Fonemos “e ilgosios” signalas PRAAT lange



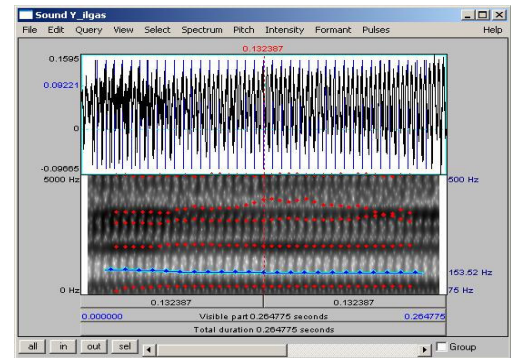
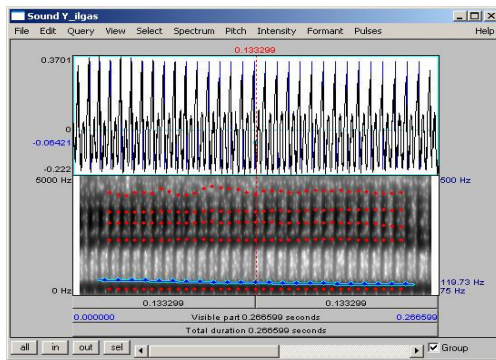
58 pav. Fonemos “è” signalas PRAAT lange



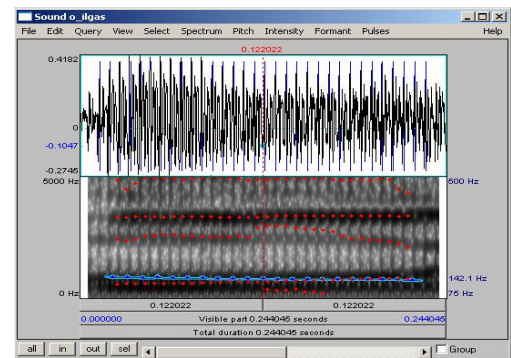
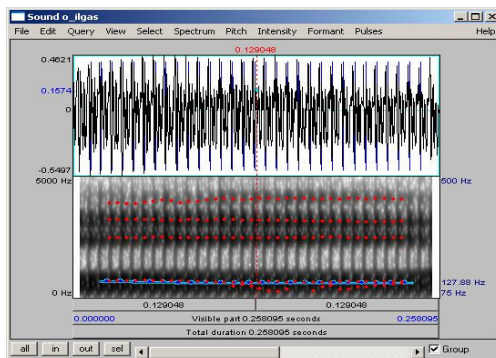
59 pav. Fonemos “è tarptautiniuose žodžiuose signalas” signalas PRAAT lange



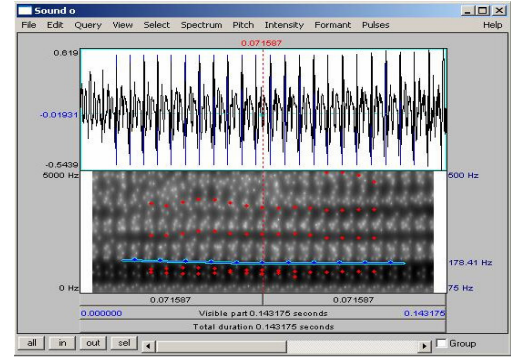
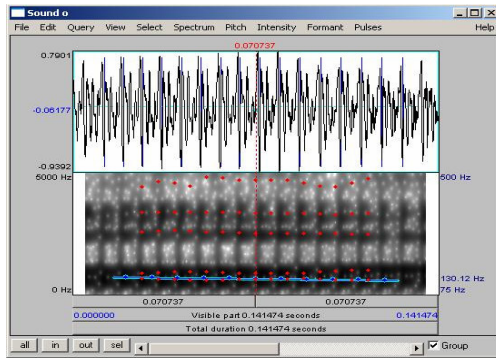
60 pav. Fonemos “i” signalas PRAAT lange



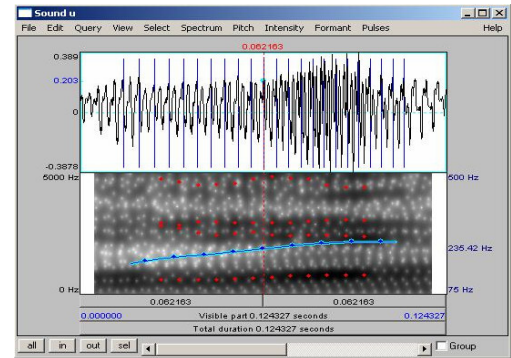
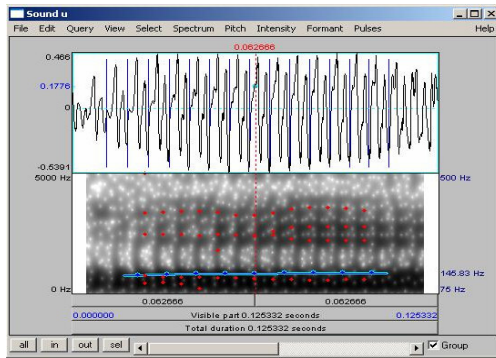
61 pav. Fonemos “i ilgosios” signalas PRAAT lange



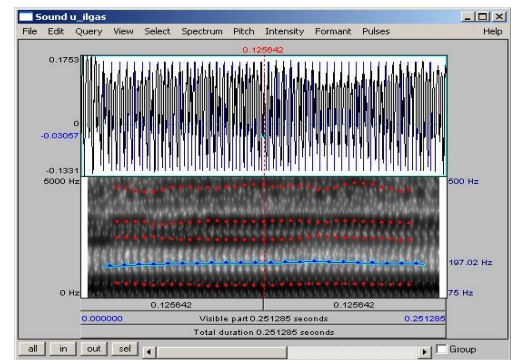
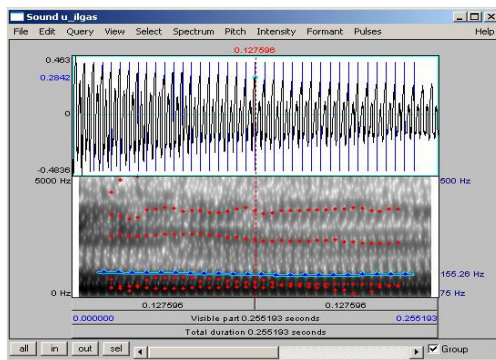
62 pav. Fonemos “o ilgosios” signalas PRAAT lange



63 pav. Fonemos “o trumposios” signalas PRAAT lange



64 pav. Fonemos “u” signalas PRAAT lange



65 pav. Fonemos “u ilgiosios” signalas PRAAT lange