



KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

Ieva Barauskaitė

**VARTOTOJO IDENTIFIKAVIMO NAUDOJANT BALSŲ
ATPAŽINIMĄ METODŲ TYRIMAS**

Baigiamasis magistro projektas

Vadovas

Doc. dr. Kastytis Ratkevičius

KAUNAS, 2017

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

**VARTOTOJO IDENTIFIKAVIMO NAUDOJANT BALSŲ
ATPAŽINIMĄ METODŲ TYRIMAS**

Baigiamasis magistro projektas

Infomatika (621I10003)

Vadovas

(parašas) Doc. dr. Kastytis Ratkevičius

(data)

Recenzentas

(parašas) Doc. dr. Vytautas Rudžionis

(data)

Projektą atliko

(parašas) Ieva Barauskaitė

(data)

KAUNAS, 2017



KAUNO TECHNOLOGIJOS UNIVERSITETAS

.....
(Fakultetas)

.....
(Studento vardas, pavardė)

.....
(Studijų programos pavadinimas, kodas)

Baigiamojo projekto „Vartotojo identifikavimo naudojant balso atpažinimą metodų tyrimas“

AKADEMINIO SAŽINGUMO DEKLARACIJA

_____ m. _____ d.
Kaunas

Patvirtinu, kad mano, Ievos Barauskaitės, baigiamasis projektas tema „Vartotojo identifikavimo naudojant balso atpažinimą metodų tyrimas“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka)

(parašas)

TURINYS

LENTELIŲ SĄRAŠAS.....	6
PAVEIKSLŲ SĄRAŠAS.....	7
TERMINŲ IR SANTRUMPŲ ŽODYNAS	9
ĮVADAS.....	10
1 PROBLEMINĖS SRITIES ANALIZĖ	11
1.1.1 Bendrasis modelis.....	11
1.1.2 Identifikavimo sistemų vertinimo charakteristikos	12
1.1.3 Signalo apdorojimas	13
1.1.4 Požymių išskyrimas.....	14
1.1.5 Požymių palyginimas	19
1.2 Esamų sprendimų analizė.....	24
1.3 Programiniai paketai	27
2 PROJEKTAVIMAS	29
3 REALIZACIJA	31
3.1 Paslėptųjų Markovo modelių paketas	31
3.2 MATLAB paketai	36
3.2.1 Požymių išskyrimas.....	36
3.2.2 Dirbtiniai neuronų tinklai	37
3.2.3 Bajeso klasifikatorius	37
4 EKSPERIMENTAS	39
4.1 Eksperimento duomenys	39
4.2 Eksperimento eiga.....	39
4.3 Eksperimento rezultatai.....	40
4.3.1 Rezultatai su pradinėmis reikšmėmis	41
4.3.2 Rezultatai skirtingiems požymių koeficientų skaičiams	49
4.3.3 Paslėptųjų Markovo modelių paketo rezultatai	56
4.3.4 Rezultatai skirtingoms dirbtinio neuronų tinklo architektūroms.....	58
5 IŠVADOS.....	60
6 LITERATŪRA.....	61
7 PRIEDAS. Garso įrašų paskirstymo atvejų lentelė	64
8 PRIEDAS. Rezultatai su pradinėmis reikšmėmis	66
9 PRIEDAS. Rezultatai skirtingiems požymių koeficientams.....	73
10 PRIEDAS. Paslėptųjų Markovo modelių paketo rezultatai	75

Barauskaitė, Ieva. Vartotojo identifikavimo naudojant balso atpažinimą metodų tyrimas. Magistro baigiamasis projektas / vadovas doc. dr. Kastytis Ratkevičius; Kauno technologijos universitetas, Informatikos fakultetas.

Mokslo kryptis ir sritis: Informatika

Reikšminiai žodžiai: *vartotojo identifikavimas, balso atpažinimas.*

Kaunas, 2017. 70 p.

SANTRAUKA

Kalbančiojo atpažinimo uždavinys yra biometrinis uždavinys, kuris kyla iš bendresnės uždavinių grupės – kalbos apdorojimo uždavinių. Lyginant su kitais biometrijos uždaviniais – tokiais kaip identifikavimas pagal akies rainelę ar pirštų antspaudus vartotojo identifikavimo pagal balsą tikslumas yra žemesnis, tačiau techninė įranga pigesnė ir dažniau sutinkama kituose įrenginiuose. Šiame darbe yra lyginami keli vartotojo identifikavimo pagal balsą metodai identiškiems garso įrašams, ir skirtingomis jų paskirstymo konfigūracijoms.

Surname, Name. *Research Of User Identification Using Voice Recognition Methods*. Master's thesis in Informatic's / supervisor assoc. prof. Kastytis Ratkevičius. The Faculty of Informatics, Kaunas University of Technology.

Research area and field: Informatics

Key words: user identification, voice recognition

Kaunas, 2017. 70 p.

SUMMARY

Speaker recognition task depends for more general tasks group – speech processing tasks. In comparison with other biometric tasks such as iris or fingerprint identification – user identification accuracy is lower but hardware is cheaper and more common in other devices. This work is comparison of several user identification by voice methods with the same audio recordings and different configurations of their distribution.

LENTELIŲ SĄRAŠAS

Lentelė 1.1 Diktoriaus atpžinimo sistemų palyginimas	25
Lentelė 3.1 Modelio failo parametrų reikšmės [29].....	32
3.2 „CONFIG“ parametrų aprašymas [29].....	33
Lentelė 3.3 melcepst funkcijos parametrų reikšmės	37
Lentelė 4.1 Maksimalios ir minimalios identifikavimo tikslumo vertės įrašų paskirstymams pagal pirmą ir antrą atvejį.....	42
Lentelė 4.2 Maksimalios ir minimalios identifikavimo tikslumo vertės įrašų paskirstymams pagal trečią ir ketvirtą atvejį	43
Lentelė 4.3 Maksimalios ir minimalios klaidingo priėmimo lygio vertės įrašų paskirstymams pagal pirmą ir antrą atvejį.....	45
Lentelė 4.4 Maksimalios ir minimalios klaidingo priėmimo lygio vertės įrašų paskirstymams pagal trečią ir ketvirtą atvejį	45
Lentelė 4.5 Klaidingo atmetimo lygių maksimalios ir minimalios vertės įrašų paskirstymams pagal pirmą ir antrą atvejį.....	47
Lentelė 4.6 Klaidingo atmetimo lygių maksimalios ir minimalios vertės įrašų paskirstymams pagal trečią ir ketvirtą atvejį	47
Lentelė 4.7 Maksimalios ir minimalios mokymo laikų vertės įrašų paskirstymams pagal pirmą ir antrą atvejį.....	48
Lentelė 4.8 Maksimalios ir miniamlaio mokymo laikų vertės įrašų paskirstymams pagal trečią ir ketvirtą atvejį.....	49
Lentelė 4.9 Maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams su dirbtiniais neuronų tinklais (ANN).....	53
Lentelė 4.10 Maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams (nuo 1 iki 20 arba nuo 1 iki 21) su Bajeso klaidifikatoriumi (NBC).....	55
Lentelė 4.11 Būsenų skaičius skirtingoms frazėms.....	56
Lentelė 7.1 Pirmo ir antro eksperimento atvejų garso įrašų pasiskirstymas	64
Lentelė 7.2 Trečio ir ketvirto eksperimento atvejų garso įrašų paskirstymas	65
Lentelė 8.1 Metodų trumpinių lentelė	66
Lentelė 9.1 Maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams su dirbtiniais neuronų tinklais (ANN) antro atvejo garso įrašų paskirstymui	73
Lentelė 9.2 Maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams (nuo 1 iki 20 arba nuo 1 iki 21) su Bajeso klaidifikatoriumi (NBC) antro atvejo garso įrašų paskirstymui.....	74

PAVEIKSLŲ SĄRAŠAS

1.1 pav. Bendra vartotojo identifikavimo sistemos principinė schema.....	11
1.2 pav. Vartotojo identifikavimo sistemos principinė schema, identifikavimo procesui	11
1.3 pav. Požymių apibendrinimas [9].....	15
1.4 pav. Paprastas kalbos trakto vamzdinis modelis [10].....	16
1.5 pav. Vienodo girdimumo kreivės [13]	17
1.6 pav. Dviejų funkcijų sąsūkos pavyzdys	18
1.7 pav. Kepstrinė analizė [11].....	19
1.8 pav. Mel kepstro koeficientų radimas	19
1.9 pav. Dinaminis laiko skalės kraipymas dviems signalams [14].....	21
1.10 pav. Tipinio biologinio neurono supaprastintas atvaizdas [15].....	23
2.1 pav. Paspėtųjų Markovo modelių paketo (HTK) duomenų srauto diagrama.....	29
2.2 pav. MFCC-NBC, MFCC39-NBC, LPCC-NBC, PLPC-NBC, MFCC-ANN, MFCC39-ANN, LPCC-ANN ir PLPC-ANN duomenų srauto diagrama.	30
3.1 pav. „gram.dict“ failo pavyzdys	31
3.2 pav. „failai.scp“ failo pavyzdys.....	31
3.3 pav. „trans_FAGNGRA.scp“ failo pavyzdys.....	31
3.4 pav. Modelio failo pavyzdys	32
3.5 „CONFIG“ failo turinys	33
3.6 pav. „test_mfc.scp“ failo pavyzdys	33
3.7 pav. „hmm1ist“ failo turinio pavyzdys	34
3.8 pav. „dict“ failo turinio pavyzdys.....	34
3.9 pav. „testref.mlf“ failo turinio pavyzdys	35
3.10 pav. HTK terminalo komandų pavyzdžiai	35
3.11 pav. Rezultatų failo pavyzdys	36
3.12 pav. melcepst funkcijos kvietimo pavyzdys.....	36
3.13 pav. lpcauto funkcijos kvietimo pavyzdys	37
3.14 pav. rastapl funkcijos kvietimo pavyzdys	37
4.1 pav. Vieno metodo eksperimento schema	40
4.2 pav. Metodų identifikavimo tikslumai ketvirto atvejo garso įrašo paskirstymui	41
4.3 pav. Metodų klaidingo priėmimo lygiai ketvirto atvejo garso įrašo paskirstymui.....	43
4.4 pav. Metodų klaidingo priėmimo lygiai ketvirto atvejo garso įrašo paskirstymui.....	44
4.5 pav. Klaidingo atmetimo lygiai ketvirto atvejo garso įrašų paskirstymui.....	46
4.6 pav. Mokymo laikai ketvirto atvejo garso įrašų paskirstymui	48
4.7 pav. Vieno įrašo testavimo laikas (sekundėmis) garso įrašų paskirstymui pagal ketvirtą atvejį	49
4.8 pav. Tiesinės prognozės kepstro koeficientų ir dirbtinių neuronų tinklų (LPCC-ANN) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį.....	50
4.9 pav. Mel skalės kepstro koeficientų ir dirbtinių neuronų tinklų (MFCC-ANN) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį.....	51
4.10 pav. Suvokimo tiesinės prognozės ir dirbtinių neuronų tinklų (PLPC-ANN) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį.....	52
4.11 pav. Tiesinės prognozės kepstro koeficientų ir Bajeso klasifikatoriaus (LPCC-NBC) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį.....	53
4.12 pav. Mel skalės kepstro koeficientų ir Bajeso klasifikatoriaus (MFCC-NBC) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį.....	54
4.13 pav. Suvokimo tiesinės prognozės ir Bajeso klasifikatoriaus (PLPC-NBC) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį.....	54
4.14 pav. Tiesinės prognozės kepstro koeficientų ir Bajeso klasifikatoriaus (LPCC-NBC) identifikavimo tikslumas koeficientų skaičiui nuo 22 iki 40, garso įrašų paskirstymui pagal ketvirtą atvejį.....	56
4.15 pav. Ketvirto atvejo rezultatai skirtingoms būsenoms ir mišiniams	57

4.16 pav. Ketvirto atvejo identifikavimo tikslumas skirtingoms būsenoms	57
4.17 pav. Ketvirto atvejo mokymo laikas (sekundėmis) skirtingam būsenų skaičiui	58
4.18 pav. Įrašo testavimo laikai (sekundėmis) skirtingiems būsenų skaičiams įrašų paskirstymui pagal ketvirtą atvejį	58
4.19 pav. Identifikavimo tikslumas su 19 Mel skalės kepstro koeficientų, skirtingam neuronų skaičiui paslėptajame sluoksnyje, garso įrašų paskirstymui pagal ketvirtą atvejį	59
8.1 pav. Metodų identifikavimo tikslumai pirmo atvejo garso įrašo paskirstymui	67
8.2 pav. Metodų identifikavimo tikslumai antro atvejo garso įrašo paskirstymui	67
8.3 pav. Metodų identifikavimo tikslumai trečio atvejo garso įrašo paskirstymui	68
8.4 pav. Metodų klaidingo priėmimo lygiai pirmo atvejo garso įrašų paskirstymui	68
8.5 pav. Metodų klaidingo priėmimo lygiai antro atvejo garso įrašo paskirstymui	69
8.6 pav. Metodų klaidingo priėmimo lygiai trečio atvejo garso įrašų paskirstymui	69
10.1 pav. Pirmo atvejo identifikavimo tikslumas skirtingam būsenų skaičiui	75

TERMINŲ IR SANTRUMPŲ ŽODYNAS

LPCC	(ang. <i>linear predictive ceptral coefficients</i>) tiesinės prognozės kepstro koeficientai
PLPC	(ang. <i>perceptual linear prediction coefficients</i>) percepcinės tiesinės prognozės koeficientai
MFCC	(ang. <i>Mel frequency cepstral coefficients</i>) Mel kepstro skalės koeficientai
NBC	(ang. <i>Naive Bayes classifier</i>) Naivus Bajeso klasifikatorius
ANN	(ang. <i>Artificial Neural Network</i>) dirbtinis neuronų tinklas
HTK	(ang. <i>Hidden Markov Toolkit</i>) paslėptųjų Markovo modelių paketas

ĮVADAS

Kalbančiojo atpažinimo uždavinys yra biometrinis uždavinys, kuris kyla iš bendresnės uždavinių grupės – kalbos apdorojimo uždavinių. Daug įvairių kalbančiojo atpažinimo uždavinių buvo apibrėžtų atsakant į didėjančius technologiškai orientuoto gyvenimo būdo poreikius. Visi šie uždaviniai gali būti suskirstyti į kelių klasių ir dviejų klasių. Dviejų klasių uždaviniams yra priskiriamas diktoriaus verifikavimas. Diktoriaus verifikavimo procesas prasideda nuo tapatybės pateikimo ir turi du galimus atsakymus – tapatybė patvirtinta arba tapatybė nepatvirtinta. Kelių klasių uždaviniams yra priskiriamas diktoriaus identifikavimas. Diktoriaus identifikavimas – tai automatizuotas procesas, kurio metu, remiantis informacija, gaunama iš kalbančio asmens ar kalbančių asmenų skleidžiamos garso bangos, nustatoma kalbančio asmens tapatybė. Asmens identifikavimo pagal balsą metodai naudojami arba yra planuojami naudoti skambučių centruose – klientams identifikuoti, kriminalistikoje, kai turimas garso įrašas su keliais kalbančiais asmenimis, sistemose, turinčiose prieigos valdymą, saugumo lygiui padidinti, balsu valdomose programose – patogesniai prieigos valdymui. Pagrindinė šių uždavinių idėja yra prielaida, kad žmogaus balsas yra unikalus kiekvienam individui ir gali būti naudojamas tų individų atskyrimui. Tačiau, iki šiol nėra sukurtos universalios požymių sistemos ar teorijos kaip žmogus atskiria vieną balsą nuo kito, leidžiančios patikimai atskirti skirtingus balsus.

Tyrimo sritis. Vartotojo identifikavimas pagal balsą susideda iš įrašo sudarymo, įrašo nuskaitymo, signalo apdorojimo, požymių išskyrimo ir požymių klasifikavimo. Į šio darbo sritį įeina požymių išskyrimas ir požymių klasifikavimas.

Tyrimo objektas. Vartotojo identifikavimas pagal balsą.

Tikslas. Vartotojo identifikavimo pagal balsą metodų palyginimas.

Uždaviniai :

1. Ištirti skirtingų požymių tipų : Mel skalės kepstro koeficientų (MFCC), suvokimo tiesinės prognozės koeficientų (PLPC) ir tiesinės prognozės kepstro koeficientų (LPCC) įtaką vartotojo identifikavimo sistemos tikslumui naudojant Bajeso klasifikatorių ir dirbtinius neuronų tinklus

2. Ištirti skirtingų klasifikatorių (Bajeso ir dirbtinių neuronų tinklų) įtaką vartotojo identifikavimo sistemos tikslumui ir greitaveikai

3. Palyginti pirmuose dviejuose uždaviniuose minėtų metodų rezultatus su paslėptųjų Markovo modelių paketo rezultatais

1 PROBLEMINĖS SRITIES ANALIZĖ

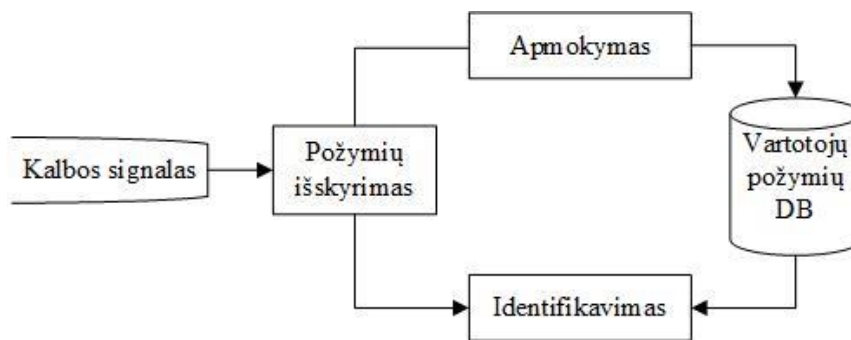
Analizės tikslas yra aprašyti suformuluotus vartotojo identifikavimo pagal balsą požymių išskyrimo ir klasifikavimo metodus.

1.1.1 Bendrasis modelis

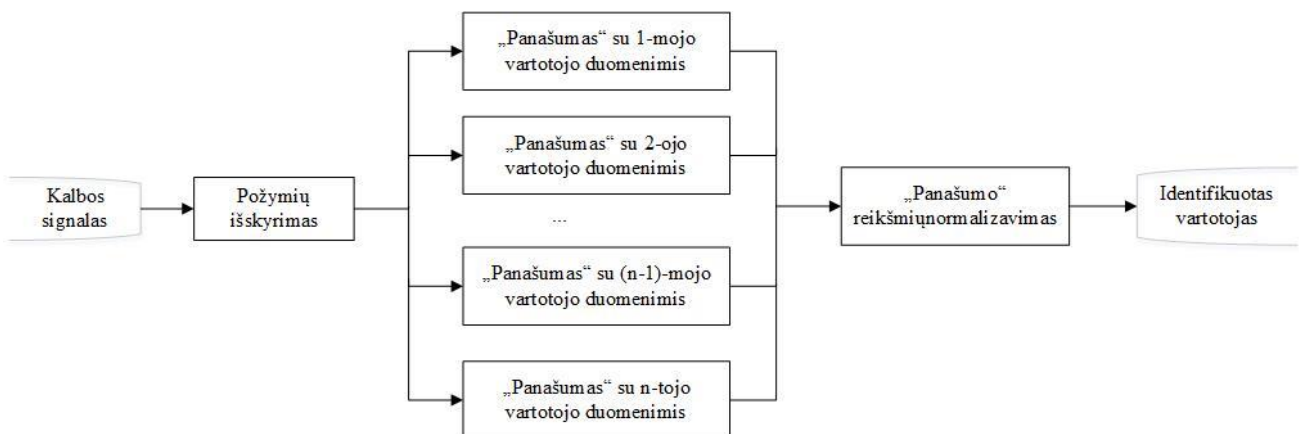
Mokymo metu (1.1. pav) atitinkamiems vartotojams priskirti požymiai įrašomi į duomenų bazę. Identifikuojant asmenį (1.2. pav) pagal balsą yra apskaičiuojami „panašumai“ tarp požymių išskirtų iš pateikto garso įrašo ir požymių, atitinkančių vartotojus duomenų bazėje. Panašumu yra laikomas atstumas tarp požymių vektorių arba pateikto garso įrašo priklausymo tam tikram vartotojui tikimybė. Identifikuotas vartotojas yra tas, kuriam tarp jį atitinkančių požymių vektorių ir pateikto garso įrašo požymių vektorių atstumas yra mažiausias arba tikimybė, kad iš garso įrašo išskirti požymiai priklauso vartotojo požymių klasei yra didžiausia.

Egzistuoja du identifikavimo uždavinio porūšiai:

1. Uždaros aibės. Turi fiksuotą skaičių N vartotojų;
2. Atviros aibės. Gali turėti $N+1$ skaičių vartotojų, iš kurių $(N+1)$ -mojo duomenų bazėje gali nebūti.



1.1 pav. Bendra vartotojo identifikavimo sistemos principinė schema



1.2 pav. Vartotojo identifikavimo sistemos principinė schema, identifikavimo procesui

1.1.2 Identifikavimo sistemų vertinimo charakteristikos

Asmens identifikavimo algoritmų kokybė vertinama greitaveika ir tikslumu. Biometrinis tikslumas – terminas, nusakantis, kaip tiksliai veikia biometrinė sistema, kuris gali būti nusakomas keliais kriterijais [1], [2]:

- Identifikavimo santykiu (ang. *identification rate*) – teisingai identifikuotų ir visų užklausų santykis;
- Klaidingu teigiamu identifikavimo rodikliu (ang. *false positive identification rate*) arba KPL (klaidingo priėmimo lygiu) – dalis užklausų pavyzdžių, atitinkančių neįtrauktus į sistemą vartotojus, kurių tapatybė gražinama kaip žinoma. Parodo neteisingų atitikimų tikimybę-

$$\text{KPL} = \frac{P}{R} \quad (1.1)$$

P – klaidingų priėmimų skaičius, R – visų užklausų skaičius;

- Klaidingu neigiamu identifikavimo rodikliu (ang. *false negative identification rate*) arba klaidingo priėmimo lygiu (KAL) – dalis užklausų pavyzdžių, atitinkančių vartotojus, įtrauktus į sistemą, kurių tapatybė yra gražinama, tačiau yra klaidinga. Parodo neteisingų neatitikimų tikimybę-

$$\text{KAL} = \frac{Q}{W} \quad (1.2)$$

Kur Q – klaidingų atmetimų skaičius, W – visų užklausų skaičius;

- Bendru klaidų lygiu (BKL) (ang. *total error rate*) – bendras KPL ir KAL klaidų lygis-

$$\text{BKL} = \frac{P+Q}{R+W} \quad (1.3)$$

- Teisingu teigiamu identifikavimo rodikliu (ang. *true positive identification rate*) – dalis užklausų pavyzdžių, atitinkančių vartotojus, įtrauktus į sistemą, kurių tapatybė atpažįstama teisingai;
- Lygių klaidų lygiu (LKL) apibūdinančiu tašką, kuriame KPL ir KAL yra arčiausiai nulio arba lygios. Naudojamas kaip pagalbinis pirmos eilės efektyvumo indikatorius 1:1 verifikacijos sistemose [1]. Kuo mažesnė lygių klaidų lygių vertė, tuo didesnis biometrinės sistemos tikslumas;
- Registravimo sutrikimo lygiu FTE (ang. *failure to enroll*) – dalis duomenų, kurie laikomi netinkamais ir į sistemą nepriimami (gauti duomenys laikomi netinkamais ar per prastos kokybės);
- Išskyrimo sutrikimų lygiu FTC (ang. *failure to capture*) – dalis duomenų, kurie laikomi tinkamais, tačiau nėra tinkami požymių išskyrimui;
- Eiliškumo klaidų rodikliu (ang. *ranking error rate*) – dalis užklausų pavyzdžių, atitinkančių vartotojus, įtrauktus į sistemą, kurių tapatybė atpažįstama teisingai. Gautas didesnis nei likusiems panašumas;
- Etalonų talpa – duomenų rinkinių, kurie gali būti įvesti į sistemą maksimalus skaičius.
- Duomenų atvaizdavimo kreivėmis [1]:

ROC kreive (ang. *receiver operating characteristic curve*) – klaidingų teigiamų ir teisingų teigiamų rodiklių reikšmės, skirtingiems slenksčiams. Naudojama sprendimui priimti esant tam tikriems parametrams, pvz. slenksčiui. ROC grafikai braižomi skirtingose ašyse atidedant KPL ir KAL reikšmes arba vienoje ašyje atidedant KPL vertes, kitoje – teisingų atpažinimų tikimybės vertes. Grafikas, kurio abscisių ašyje yra KPL vertės, o ordinačių KAL vertės dar vadinamas DET (ang. *detect-error trade-off*).

DET kreive (ang. *detection error trade of curve*) – klaidingų teigiamų ir klaidingų neigiamų rodiklių reikšmės, skirtingiems slenksčiams.

CMC kreive (ang. *cumulative match characteristic curve*) – atitinka dalį testavimo užklausų pavyzdžių, atitinkančių teisingą tapatybę k-toje vietoje. Kur k – teisingos tapatybės vieta išrikiuotuose panašumuose [3].

Šių kriterijų vertes pagal [4] lemia signalų kokybės, kalbos signalo ilgis, kalbos signalo turinys ir vartotojų skaičius.

Pagal įvesties signalo turinį, asmens identifikavimo pagal balsą sistemos yra skirstomos į priklausančias ir nepriklausančias nuo teksto, kurios pagal veikimo principą taip pat yra klasifikuojamos. Pagrindiniai asmens identifikavimo pagal balsą sistemų tipai [5]:

- Nuo teksto priklausomos, su iš anksto numatytu slaptažodžiu.
- Nuo teksto priklausomos, su iš anksto numatytais skirtingais slaptažodžiais kiekvienam vartotojui.
- Nuo teksto nepriklausomos. Tekstas parenkamas sistemos.
- Nuo teksto nepriklausomos. Tekstas parenkamas vartotojo.

1.1.3 Signalų apdorojimas

Pagrindinis komponentas triukšmui kalbos signalė sumažinti yra VAD (ang. *voice activity detection*), kurio paskirtis - atskirti signalo segmentus, kuriuose dominuoja kalba, nuo tų, kuriuose dominuoja triukšmas (arba tyła). VAD algoritmai gali būti dalinami į du pagrindinius blokų požymių išskyrimą ir sprendimo modulį. Vieni iš pagrindinių parametrų pagal kuriuos nustatomas skirtumas tarp triukšmo ir kalbos segmentų [6]: energijos lygmuo (ang. *energy level*), žadinimo signalo pagrindinis dažnis (ang. *pitch*), nulių kirtimų dažnis (ang. *zero crossing rates*), parododantis kiek kartų signalas kerta laiko ašį per sekundę, kepstro požymiai (ang. *cepstral features*). Pagrindiniai sprendimo modulių tipai - fiksuoto slenksčio, euklidinių atstumų, SVM (ang. *support vector machines*) ir statistiniai.

VAD gali būti naudojamas tylos periodams iš garso įrašų iškirpti. Suformavus triukšmo spektrą iš tylos signalo kadru, kuriuose kalba nėra aktyvi naudojami, triukšmo šalinimo metodai, kuriems reikalingas triukšmo spektras, pvz. spektrinės atimties metodas.

Spektrinė atimtis (ang. *spectral subtraction*). Tinkamas stacionariems arba lėtai kintantiems laiko atžvilgiu triukšmams. Triukšmas $N(f)$ įvertinamas kalbos neaktyvumo periodais ir atimamas iš viso signalo $X(f)$. Gaunamas „švarus“ kalbos signalas [7] :

$$|S(f)| = |X(f)| - |N(f)| \quad (1.4)$$

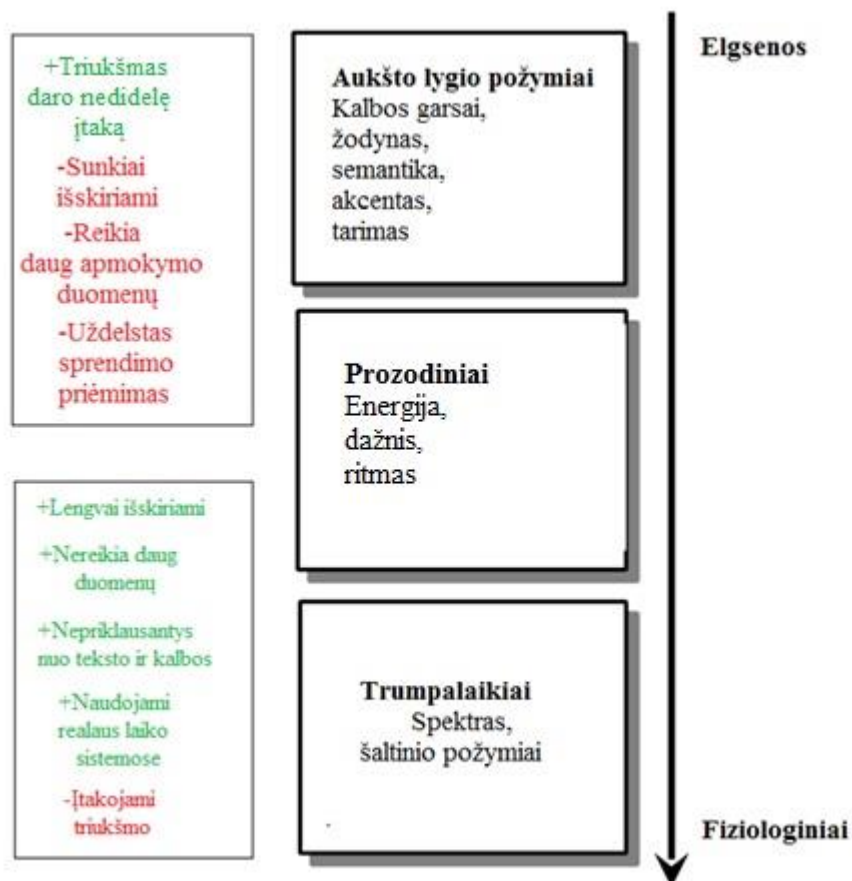
1.1.4 Požymių išskyrimas

Daugelis kalbančiojo atpažinimo sistemų naudoja akustinę informaciją, kuri yra laikoma fiziniiais bruožais, gaunamais iš kalbos signalo charakteristikų. Tokios charakteristikos gali būti gaunamos iš daugelio požymių išskyrimo technikų ir turėtų būti [8]:

- lengvai „išmatuojamos“,
- atsparios imitacijai,
- rodyti nedideles fliktuacijas tarp skirtingų aplinkų,
- stabilios laike,
- dažnai sutinkamos natūralioje kalboje.

Požymius naudojamus asmeniui atpažinti pagal balsą galima suskirstyti į penkias klases [8]:

- Spektrinius (ang. *spectral features*). Atvaizduoja 10 - 30ms kalbos signalo kadro spektrą. Daugiau atspindi fizines balso trakto charakteristikas.
- Dinaminius (ang. *dynamic features*). Parodo spektrinių ar kitų požymių kitimą laike.
- Šaltinio (ang. *source features*). Susiję su žadinimo signalu.
- Suprasegmentinius (ang. *suprasegmental features*). Apima ilgesnius laiko intervalus.
- Aukšto lygio (ang. *high-level features*). Pvz.: tam tikro žodžio vartojimas.



1.3 pav. Požymių apibendrinimas [9]

Tiesinės prognozės kepstro koeficientai (LPCC), percepcinės tiesinės prognozės koeficientai (PLPC) ir Mel skalės kepstro koeficientai (MFCC) yra dažniausiai naudojami požymių išskyrimo metodai asmeniui atpažinti pagal balsą. Visi jie priskiriami trumpalaikiams spektro požymiams.

1.1.4.1 Tiesinės prognozės kepstro koeficientai

Tiesinės prognozės (ang. *linear prediction coding*) modelis paremtas paprastu kalbos generavimo modeliu. Balso traktas yra modeliuojamas kaip keletas sujungtų tokio pat ilgio ir skirtingų diametrų vamzdžių (1.5 pav.) Daroma prielaida, kad balsaskylė generuoja kalbą arba triukšmą. Su tam tikrom sąlygom (pvz.: jokių energijos nuostolių balso trakte, jokių netiesinių efektų ir t.t.) perdavimo funkcija gali būti aprašoma pagal :

$$A(z) = \frac{1}{1 - \sum_{i=1}^P a_i z^{-i}} \quad (1.5)$$

kur P yra vamzdžio segmentų skaičius, z yra z - transformacija (naudojama diskretaus laiko signalo konvertavimui į kompleksinių dažnių skalę). Koeficientai $a_1 \dots a_P$ yra tiesiogiai susiję su rezonansiniais dažniais balso trakte, vadinamais formantėmis. Ir neša informaciją apie balso trakto formą.

Tariant atskirus garsus, pvz. „A“ ir „L“ skiriasi balso trakto konfigūracija bei žadinimo signalas. Vokalizuotų garsų energija yra stipriausia. Formantėmis yra vadinami balsių signalo spektro maksimumai. Antiformantėmis - vadinami balsių signalo spektro minimumai. Teigiama, kad dažnių juostoje 200-5000 Hz galima išskirti 3-5 formantes. Kiekvieną ištartą garsą atitinka tam tikra balso trakto konfigūracija, formančių padėtis dažnių srityje bei amplitudės dydis. Nustatyta, kad asmuo sąmoningai gali keisti pirmų dviejų formančių padėtį, tuo tarpu aukštesnių formančių padėtis priklauso nuo asmens anatominių ypatumų.

Vidutinį atstumą tarp formančių apsprendžia vidutinis balso trakto ilgis L . I -tosios formantės dažnį galima išreikšti :

$$F(i) = \frac{c}{2L} \left(i - \frac{1}{2}\right) \quad (1.6)$$

kur C – garso bangos sklidimo greitis, L – vidutinis balso trakto ilgis.

Formantės charakterizuojamos dažniu ir amplitude. Formantės pločiu vadinama vieta kur rezonansinio piko plotis pasiekia $\frac{1}{2}$ maksimumo. Kalbant keičiasi formančių padėtis, tačiau jų skaičius lieka pastovus. Tariant vokalizuotus garsus balso traktas kinta pakankamai lėtai, nevokalizuotų garsų trukmė yra žymiai mažesnė. Kai kurie spektro maksimumai spektro gaubiamojoje išnyksta, todėl juos apskaičiuoti iš spektro gaubiamosios yra sudėtinga, o kartais ir neįmanoma [2].

Perdavimo funkcijos koeficientai gali būti skaičiuojami tiesiai iš signalo minimizuojant tiesnės prognozės paklaidą :

$$e(n) = s_n - \sum_{i=1}^p a_i s_{n-i} \quad (1.7)$$



1.4 pav. Paprastas kalbos trakto vamzdinis modelis [10]

Bendru atveju prognozavimo paklaida $e(n)$ yra atsiktinis dydis, kurio vidurkis lygus 0, o dispersija σ^2 . Sekantis kalbos signalo vektorius elementas gali būti aproksimuotas pagal kelis (p) prieš tai buvusius elementus [11] :

$$s(n) = - \sum_{i=1}^p a_i s_{n-i} \quad (1.8)$$

$s(n)$ – dabartinė išvestis, p – tiesinės prognozės modelio eilė (ang. *prediction order*), a_k – tiesinės prognozės modelio koeficientai (ang. *prediction coefficients*), s_{n-i} – signalo vektoriaus elementai. LPC išvestis yra vektorius su p koeficientų (žinomų kaip autoregresiniai koeficientai).

Yra keli skirtingi kriterijai tiesinės prognozės paklaidos minimizavimui. Vienas iš jų *Yule-Walker* lygtis. Pasirinkus naudoti matematinę kvadratinę viltį (ang. *squared expectation*) naudojamas autokoreliacinis metodas. Kiti metodai, pvz. kovariacijos veda prie skirtingų lygčių. Abiem atvejais daroma prielaida, kad balso traktas ir signalas nekinta kadre [10].

1.1.4.2 Suvokimo tiesinės prognozės koeficientai

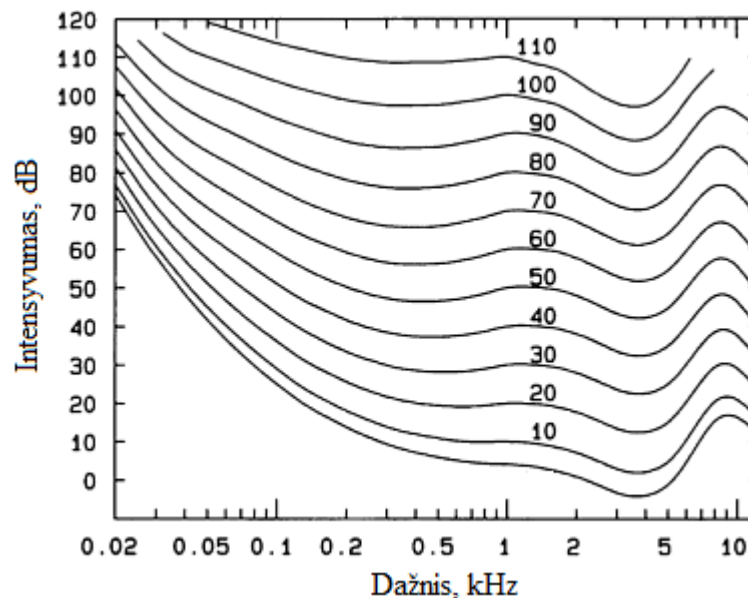
Suvokimo arba percepcinės tiesinės prognozės koeficientai (ang. *perceptual linear prediction*) modeliuoja žmogaus klausos sistemos savybes. PLP algoritmą sudaro šie etapai [12]:

- kalbos signalo pjūvio amplitudinio spektro apskaičiavimas;
- dažnių skalės hercais į dažnių skalę barkais transformavimas. Herco skalė gali būti transformuojama į Barkų pagal –

$$B = 13 \arctan(0.76f) + 3.5 \left(\frac{f}{7.5}\right)^2 \quad (1.9)$$

kur f – dažnis hercais.

- atlikti transformuoto į barkų skalę amplitudinio spektro sąsūką su vienodo girdimumo kreive (angl. *equal loudness curve*). Vienodo girdimumo kreivė (kartais vadinama Fletčerio - Munsono kreive) – tai empiriškai nustatyta giliai netiesinė kreivė, kurioje fiksuojama žemutinė garso lygio riba, kada žmogus pradeda girdėti tam tikro dažnio garsus.



1.5 pav. Vienodo girdimumo kreivės [13]

Vienodo girdimumo kreivės parodo kokio garsumo turėtų būti kiti dažniai, lyginant su garsumu 1kHz dažnyje, pagal žmogaus klausą. Pvz.: 0.5kHz signalo 55dB garsumas žmogaus bus suvokiamas kaip 60dB garsumas 1kHz signalui – t.y. 0.5kHz 60dB signalas žmogaus bus suvokiamas kaip garsesnis už 1kHz 60dB signalą.

Spektrinė sąsūka – matematinė operacija tarp dviejų funkcijų f ir g , apibūdinanti funkcijos g persiklojimą kai ji yra perkeliama per funkciją f –

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau = \int_{-\infty}^{\infty} g(\tau)f(t - \tau)d\tau \quad (1.10)$$

t – šiam atvejui reiškia laiką, o τ – atsistiktinį kintamąjį, funkcijų argumentų reikšmių perkėlimui. Atvejui kai, funkcijų argumentų reikšmės nėra mažesnės už nulį tinka tik -

$$(f * g)(t) = \int_0^t f(\tau)g(t - \tau)d\tau \quad (1.11)$$

- atlikti gauto vienodo girdimumo spektro pirminį stiprinimą pirmos eilės rekursiniu filtru;
- atlikti spektro ištiesinimą įvertinant intensyvumo – girdimumo kreivę;
- apskaičiuoti gauto spektro tiesinės prognozės koeficientus.

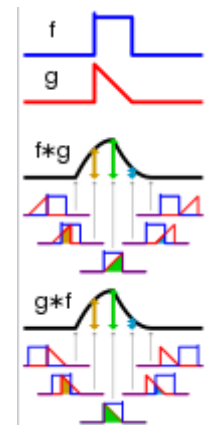
RASTA-PLP algoritmas kilęs nuo anglišių žodžių Relative Spectral (RASTA) ir preceptual linear prediction (PLP). Chronologine tvarka pirmiausia buvo pasiūlytas PLP algoritmas, o vėliau pastarasis buvo modifikuotas ir įvestas RASTA-PLP algoritmas.

RASTA-PLP algoritmą sudaro žemiau pateikti etapai [12]:

- spektro barkų skalėje apskaičiavimas ir logaritnavimas;
- logaritminio spektro kritinėse juostose kitimo laike įvertinimas, naudojant penkių gretimų pjūvių spektro reikšmių regresiją;
- gautoms reikšmėms atliekamas arba ne netiesinis apdorojimas;
- gautų logaritminių laiko išvestinių integravimas naudojant pirmos eilės rekursinį filtrą. Filtrų koeficiento reikšmė turi būti parinkta priklausomai nuo analizės pjūvio dydžio. Rekursinio filtro perdavimo funkcija :

$$H(z) = 0.1 * \frac{2+z^{-1}-z^{-3}-2z^{-4}}{z^{-4}*(1-0.98z^{-1})} \quad (1.12)$$

- panašiai kaip paprastame PLP algoritme pridedama vienodo girdimumo kreivė ir gautas atstojamasis spektras padauginamas iš 0,33;
- gautam logaritminiam spektrui atliekamas atvirkštinis logaritnavimas (eksponentinė funkcija);
- apskaičiuojami šio spektro tiesinės prognozės kepsro koeficientai.



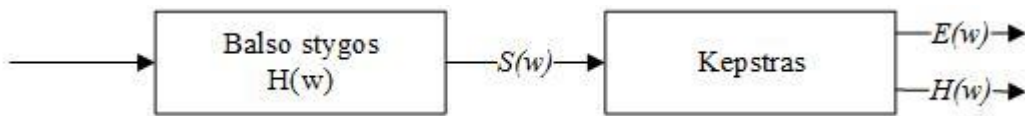
1.6 pav. Dviejų funkcijų sąsūkos pavyzdys

1.1.4.3 Mel skalės kepstro koeficientai

Kalbos signalas $S(w)$ gali būti laikomas balsaskylės sužadavimo (ang. *excitation*) $E(w)$ ir balso trakto signalų $H(w)$ sąsūka (ang. *convolution*) $E(w)*H(w)$. Kepstras gaunamas taikant atvirkštinę diskrečiąją Furjė transformaciją kalbos signalo kadro absoliutinės vertės (ang. *magnitude*) logaritmui. Realiają kepstro dalį galima apskaičiuoti pagal [11]:

$$RC = FFT^{-1}(\log(FFT(s(n)))) \quad (1.13)$$

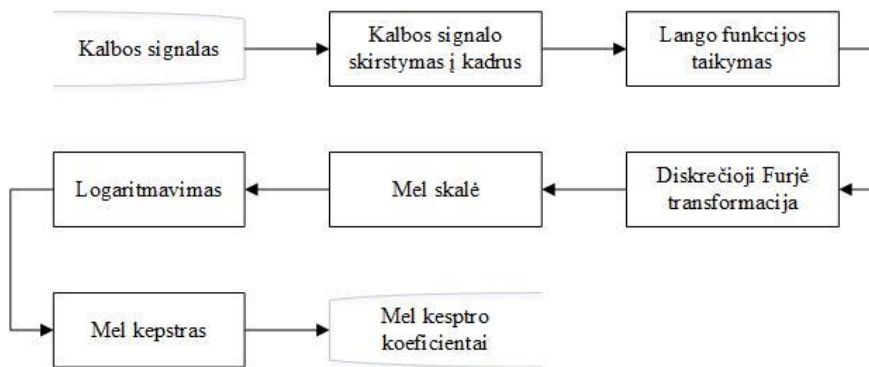
$E(w)$ – laikomas papildomu požymiu kintančiu laike, priklausomai nuo individo emocinės būsenos ir amžiaus. $H(w)$ – laikomas unikaliu kiekvienam vartotojui. Žemesni dažniai labiau atvaizduoja $H(w)$, aukštesni – $E(w)$. Pastebėta, kad žemų dažnių filtrų naudojimas duoda patikimesnes balso trakto charakteristikas.



1.7 pav. Kepstrinė analizė [11]

Laikoma, kad MFCC (ang. *Mel frequency cepstral coefficients*) turėtų atspindėti balso trakto charakteristikas. MFCC paremti teiginiu, kad informacija iš žemesnių kalbos signalų dažnių komponentų yra fonetiškai svarbesnė, nei iš aukštesnių. Mel kepstro skalės koeficientais vadinamos vertės sudarančios Mel skalės kepstą. MFCC gavimas panašus į kepstro skaičiavimą – kalbos signalo kadro absoliutinei vertei prieš atvirkštinę Furjė transformaciją pritaikant Mel filtrą [8]:

$$Mel(f) = 2595 * \log_{10}\left(1 + \frac{f}{700}\right) \quad (1.14)$$



1.8 pav. Mel kepstro koeficientų radimas

1.1.5 Požymių palyginimas

Kalbos signalo požymiais, kai požymiai išskiriami iš garso įrašo turinčio vieną kanalą, yra laikomas N ilgio požymių vektorius. Segmentuotam kalbos signalui, kuris padalintas į M segmentų požymių matricos dydis $M \times N$. $M \times N$ matrica sukuriama kiekvienam vartotojui apmokymo metu. Diktoriaus modeliu vadinamas požymių rinkinys sudarytas iš visų apmokymo įrašų ($M \times N$ požymių matricų). Testavimo metu gaunama $M \times N$ matrica iš kalbos signalo yra lyginama su diktoriaus

modeliu. Diktorių modeliai skirstomi į du tipus šabloninius (ang. *template*) ir stochastinius modelius [14].

Šabloniniuose modeliuose pavyzdžių palyginimas yra deterministinis. Garso įrašas, lyginamas su diktoriaus modeliu, yra laikomas netobula diktoriaus modelio kopija. Diktoriaus modelio ir lyginamo įrašo kadrai yra išrenkami ir išrikiuojami minimizuojant atstumą. Paprasčiausias etaloninis modelis turi vieną etaloną \bar{x} vienam segmentui. Panašumo įvertis tarp x ir tikrinamo pavyzdžio požymių x_i žymimas $d(x_i, \bar{x})$. Diktoriaus modelis yra N apmokymo vektorių centroidas arba vidurkis:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (1.15)$$

Atstumai tarp x_i ir x yra aprašomi:

$$d(x_i, \bar{x}) = (x_i - \bar{x})^T W (x_i - \bar{x}) \quad (1.16)$$

W – svorinė (ang. *weighting*) matrica. Jei W vienetinė matrica atstumai Euklidiniai. Tikėtumas etaloniniams modeliams aproksimuojamas pagal [15]:

$$L = e^{-\alpha d} \quad (1.17)$$

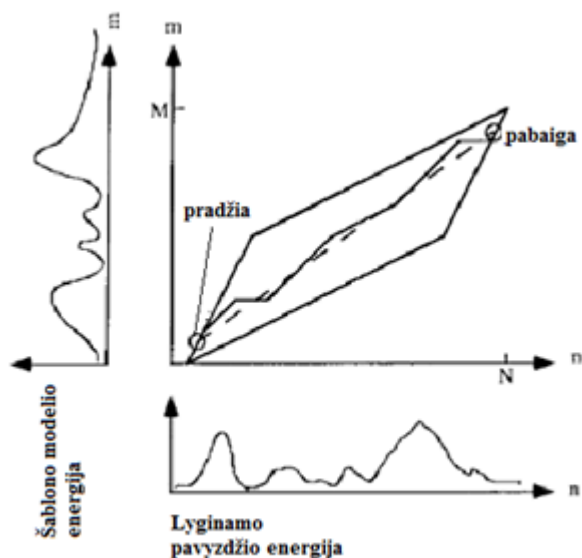
α – teigiama konstanta (normalizavimui). (1.17) išraiška laikoma proporcinga tikėtinumo logaritminėms išraiškoms.

Šabloniniai metodai gali priklausyti arba nepriklausyti nuo laiko. Nuo laiko nepriklausantis šabloninio modeliavimo pavyzdys vektorinio kvantavimo modeliavimas (ang. *vector quantization*). Visos laikinos variacijos yra ignoruojamos ir naudojami tik globalūs vidurkiai (centroidai). Nuo laiko priklausančiuose šabloniniuose modeliuose įvertinamas kintantis kalbėjimo tempas [14].

Populiariausias metodas, kompensuojantis kalbėjimo tempo kintamumą šabloninėse sistemose yra dinaminis laiko skalės kraipymas (ang. *dynamic time warping*). Nuo teksto priklausantis diktoriaus modelis yra seka šablonų $(\bar{x}_1, \dots, \bar{x}_N)$, kuri yra lyginama su sistemai nežinoma seka (x_1, \dots, x_M) . Bendru atveju M nėra lygus N . Asimetrinis sutapimo rezultatas (ang. *match score*) gaunamas iš:

$$z = \sum_{i=1}^M d(x_i, \bar{x}_{j(i)}) \quad (1.18)$$

kur j – yra diktorių modelių indeksai. 1.9 pav. Vaizduojamas kraipymo kelias (ang. *warp path*), kai energija yra naudojama kaip požymis. Iš galimų kelių išrenkamas kelias turintis mažiausią suminį atsrtumą tarp šabloninio modelio ir lyginamo pavyzdžio.



1.9 pav. Dinaminis laiko skalės kraipymas dviems signalams [14].

Stochastiniu signalu vadinamas signalas, kuris netenkina determinuoto signalo sąlygų. Stochastinius signalus generuoja sistemos, kurios parametrai kinta atsitiktiniu dėsnio. Bet kuriuo laiko momentu stochastinio signalo užregistruota vertė yra atsitiktinis dydis, kurio tiksli vertė yra nenuspėjama, tačiau matematinė viltis ar dispersija gali būti žinomos. Stochastiniuose modeliuose (ang. *stochastic models*) diktoriaus modelio ir garso įrašo palyginimas yra tikimybinis – gaunamas tikėtumas arba sąlyginė tikimybė. Lyginama sistemai nežinomų vektorių seka aprašoma tikimybės tankio pasiskirstymo funkcija, kuri sudaroma iš mokymo duomenų. Stochastiniuose modeliuose tikimybė, kad sistemai nežinoma vektorių seka bus skirta gaunama kiekvienam signalo kadru arba vektoriaus parametrų vidurkiui, gautam iš visų kadru. Ši tikimybė yra vadinama sutapimo rezultatu (ang. *match score*) [14]. Stochastiniams modeliams gali būti priskirti Bajeso klasifikatorius ir paslėptosios Markovo grandinės.

1.1.5.1 Naivus Bajeso klasifikatorius

Bajeso klasifikatorius, taip pat, gali būti priskirtas statistiniams ir apmokymo su mokytoju metodams. Jeigu naudojamas normalusis skirstinys, apmokymo metu nustatomi požymių koeficientų vidurkiai ir dispersijos. Kitų skirstinių atvejais nustatomi kiti parametrai. Identifikavimo proceso metu randamas tikėtumas. Identifikuotas diktorius yra tas, kuriam šis tikėtumas yra didžiausias.

1.1.5.2 Paslėptieji Markovo modeliai

Paslėptieji Markovo modeliai diskrečių įvykių stebėjimui aprašomi : stebimų įvykių seka $O = \{O_1, O_2 \dots O_T\}$, kur T yra stebimos vektorių sekos dydis, paslėptų būsenų skaičiumi modelyje $Q = \{q_1, q_2 \dots q_N\}$, klasių skaičiumi M , perėjimų tarp būsenų tikimybėmis $A = \{a_{ij}\}, a_{ij} = P(q_j | \text{laiko_momentu_}t+1 | q_i | \text{laiko_momentu_}t)$, išėjimo klasių tikimybių pasiskirstymu būsenoje j $B =$

$\{b_j(k)\}, b_j(k) = P(v_k_laiko_mometu_t|q_i_laiko_momentu_t)$, pradinių būsenų tikimybių pasiskirstymu $\pi = \{\pi_j\}, \pi_i = P(q_i_laiko_mometu_t = 1)$.

Paslėptiesiems Markovo modeliams charakterizuoti naudojamas parametras $\lambda = \{A, B, \pi\}$, kuris nustatomas iš mokymo duomenų rinkinio ir apibrėžia tikimybinį matą $O = \{O_1, O_2 \dots O_T\}$ – tikimybę $P(O|\lambda)$, kur kiekvienas stebėjimas O_t yra viena iš galimų klasių arba vektorinio kvantavimo knygos V simbolis.

Būsenos aprašomos tikimybės pasiskirstymo funkcija, $p(\mathbf{x}_t|s)$ yra tikimybė kadro požymių vektoriui būsenoje s . Perėjimas tarp būsenų (pvz. iš $t-1$ į t) aprašomas perėjimo tikimybe $p(s_t|s_{t-1})$.

Tikėtinumai T kadru skaičiui, λ , hipotezei [10]:

$$p(X|\lambda) = \sum_{\substack{\text{visoms} \\ \text{būsenoms}}} \prod_{t=1}^T p(s_t|s_{t-1}) p(\mathbf{x}_t|s) \quad (1.19)$$

Stebėjimų sekos O gavimas :

- 1) Laiko momentui suteikiama pradinė reikšmė lygi vienetui $t=1$;
- 2) Parenkama pradinė būsena i atsižvelgiant į pradinės būsenos pasiskirstymo funkciją π ;
- 3) Parenkama O_t atsižvelgus į $b_i(k)$, duotos klasės arba vektorinės knygos simbolio tikybinį pasiskirstymą būsenai i ;
- 4) Parenkama j pagal perėjimų tarp būsenų pasiskirstymo funkciją $\{a_{ij}\}$ ($j=1,2,\dots,N$) duotai būsenai i ;
- 5) Pereinama prie sekančio laiko momento $t \leftarrow t+1$ ir grįžtama į žingsnį 3 jeigu $t < T$. Priešingu atveju iteracinis procesas nutraukiamas.

Apmokymo proceso metu reikia rasti parametrų rinkinį λ^* , kuris tenkintų sąlygą:

$$\lambda^* = \arg \max \prod_{n=1}^N P(O^{(n)}|\lambda) \quad (1.20)$$

kur N – apmokymo imties dydis, o kiekvienas apmokymo imties narys aprašomas stebėjimų seka,

$O^{(n)} = \{O_t^{(n)}\}_t^T$, kur T_n – pjūvių skaičius n –tojoje amokymo imties sekoje.

Tolydinio tankio paslėptieji Markovo modeliai naudojami tais atvejais, kuriais stebėjimai išreiškiami ne diskrečiais simboliais bet požymių vektoriais. Gauso skirstinių mišinio tikimybė aprašoma :

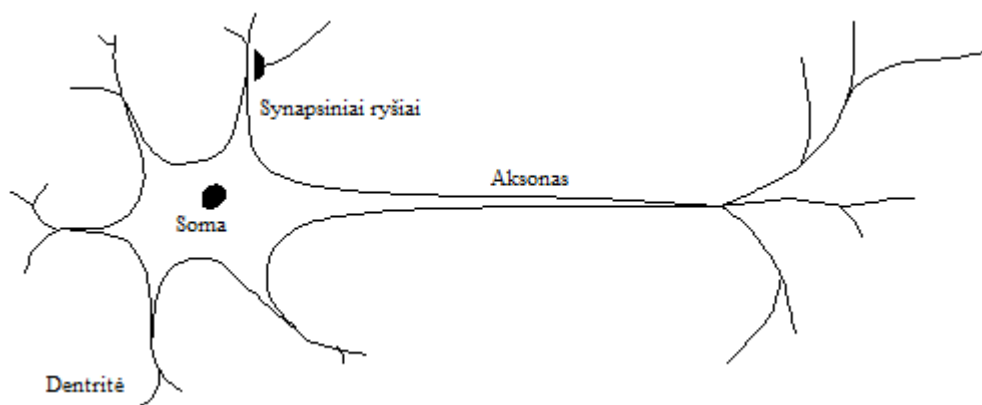
$$b_j(O) = \sum_{k=1}^M c_{jk} N(O, \mu_{jk}, U_{jk}) \quad (1.21)$$

Kur $1 \leq j \leq N$ modeliuojamas stebimas požymių vektorius, c_{jk} - k -tojo Gauso skirstinio mišinio komponento svoris būsenoje j , o N – Gauso skirstinio matematinė išraiška (bendru atveju galima taikyti bet kuri eliptiškai simetrinį skirstinį), μ_{jk} ir U_{jk} yra k -tojo Gauso skirstinio vidurkių vektorius ir kovariacinė matrica būsenoje j atitinkamai [10], [12].

1.1.5.3 Dirbtiniai neuronų tinklai

Dirbtinis neuronų tinklas – tai informacijos apdorojimo modelis, sukurtas pagal biologinę nervų sistemą, tokią kaip smegenys. Modelis yra sudarytas iš didelio kiekio tarpusavyje sujungtų apdorojimo elementų (neuronų) dirbančių tarpusavyje sprendžiant tą pačią problemą.

Vienas biologinis neuronas susideda iš ląstelinio kūno vadinamo soma, kelių dendičių, ir aksono. Dendritėmis surenkamai signalai iš kitų neuronų. Aksonu siunčiami signalai kitiems neuronams. Synapsinės struktūros paverčia dendičių aktyvumą elektriniais efektais, kurie sužadina arba slopina aksono aktyvumą ir signalus perduodamus kitiems neuronams. Mokymosi metu synapsiniai.



1.10 pav. Tipinio biologinio neurono supaprastintas atvaizdas [15]

Tiesioginio sklaidimo dirbtinis neuronų tinklas gali turėti tik vieną sluoksnį svorių, kuriame įvestys yra tiesiogiai sujungtos su išvestimis. Arba tarp įvesties ir išvesties gali būti viena, ar daugiau įterptų paslėptųjų sluoksnių. Paslėptieji neuronai neuroniniuose tinkluose naudojami geresniam įvesties „dėsningumų“ radimui. Buvo parodyta, kad su atitinkamu kiekiu paslėptųjų neuronų įmanoma aproksimuoti bet kokią funkciją [15].

Grįžtamojo ryšio tinkluose signalas gali sklirti abiem kryptimis – į priekį ir atgal, sudarydamas ciklus tinkle. Tai dinaminiai tinklai, kurių būseną kinta tol, kol pasiekiamas pusiausvyra. Pusiausvyroje jie lieka tol kol pasikeičia įvesties duomenys ir ieškomas naujas pusiausvyros taškas [16].

Daugiasluoksnis perceptronas (ang. *Multilayer perceptron*) yra sudarytas iš daug neuronų, kurie yra suskirstyti į įėjimo ir išėjimo sluoksnius bei tarp šių esantį vieną ar daugiau vidinių paslėptuosius sluoksnius. Daugiasluoksnių perceptronų tinkluose greta esantys sluoksniai yra pilnai sujungti, t.y. kiekvienas neuronas yra sujungtas su visais sekančio sluoksnio neuronais. Terminas gilieji neuronų tinklai (ang. *deep neural networks*) originaliai buvo pristatytas kaip

daugiasluoksnis perceptronas su daugiau nei vienu paslėptųjų sluoksnių, bet vėliau buvo išplėstas ir reiškia bet koki neuroninį tinklą su gilia struktūra [17].

Kartu su giliaisiais neuronų tinklais naudojami *i-vektorių* (ang. *i-vector*) metodas ir susiaurėjimo arba BN (ang. *bottleneck*) topologija. *i-vektorių* (ang. *i-vector*) metodas naudojamas įvesties duomenų sumažinimui iki mažesnių dimensijų požymių vektoriaus, išlaikant didžiąją dalį svarbios informacijos. BN neuroniniuose tinkluose reiškia vieno iš paslėptųjų sluoksnių susiaurėjimą – vienas iš paslėptųjų sluoksnių turi mažiau neuronų nei gretimi. BN požymių vektorius gaunamas siunčiant pirminį įvesties požymių vektorius neuroniniu tinklu ir nuskaitant požymių vertes BN sluoksnyje [18].

Diktorius atpažinimo užduotyje gilusis neuronų tinklas naudojamas prognozuoti diktorius klasę vienam kalbos signalo kadrai arba visam signalui. Kadangi visas kalbos signalas priklauso vienai klasei kalbos signalo kadro tikimybės turi būti kombinuojamos viso signalo priskyrimui vienam iš diktorių. Tai gali būti pasiekta vidurkinant gautas tikimybes arba naudojant antrą klasifikatorių – tokį kuris naudoja visą įvestį gautą iš giliųjų neuronų tinklų kaip vieną požymių vektorius. Gilieji neuronų tinklai taip pat naudojami fonetinių vienetų arba senonų prognozavimui kiekvienam kalbos kadre [19],[20].

1.2 Esamų sprendimų analizė

1.1 lentelėje pateikti literatūros šaltiniuose rasti rezultatai. Tikslumu laikoma dalis teisingai atpažintų įrašų. Lygių klaidų lygiu (LKL) taškas, kuriame klaidingo savojo ir klaidingo svetimo yra arčiausiai nulio arba lygios. Naudojamas kaip pagalbiniis pirmos eilės efektyvumo indikatorius 1:1 verifikacijos sistemose [1]. Kuo mažesnė lygių klaidų lygių vertė tuo didesnis biometrinės sistemos tikslumas.

Lentelė 1.1 Diktoriaus atpžinimo sistemų palyginimas

Šaltinis	Metodas	Garso įrašai	Kiti duomenys	Tisklumas, %	LKL,%
[8]	MFCC	Nepateikta	Nepateikta	51.25	Nepateikta
[8]	LPCC	Nepateikta	Nepateikta	37.5	Nepateikta
[8]	PLPC	Nepateikta	Nepateikta	49.5	Nepateikta
[8]	MFCC	Nepateikta	Nepateikta	86.67	Nepateikta
[8]	LPCC	Nepateikta	Nepateikta	80.5	Nepateikta
[8]	PLPC	Nepateikta	Nepateikta	77.4	Nepateikta
[8]	MFCC	Nepateikta	Nepateikta	93.6	Nepateikta
[8]	LPCC	Nepateikta	Nepateikta	76.6	Nepateikta
[8]	PLPC	Nepateikta	Nepateikta	90.4	Nepateikta
[8]	MFCC	Nepateikta	Nepateikta	96.5	Nepateikta
[8]	LPCC	Nepateikta	Nepateikta	65.8	Nepateikta
[8]	PLPC	Nepateikta	Nepateikta	78.5	Nepateikta
[21]	MFCC-Euklidinių atstumų	Sudaryta	40 diktorių. 10 ištarimų "Open the doors"	83.38	Nepateikta
[21]	MFCC-BPANN	Sudaryta	40 diktorių. 10 ištarimų "Open the doors"	96.18	Nepateikta
[22]	VQ-LPCC	RUSBASE	Naudojama viena frazė, 15 ištarimų 41-nam diktoriui (visi vyriški balsai)	Nepateikta	2.76
[22]	GMM-MFCC	RUSBASE	Naudojama viena frazė, 15 ištarimų 41-nam diktoriui (visi vyriški balsai)	Nepateikta	5.86
[22]	GMM-4F3AF0	RUSBASE	Naudojama viena frazė, 15 ištarimų 41-nam diktoriui (visi vyriški balsai)	Nepateikta	5.17
[22]	RAKTAS	RUSBASE	Naudojama viena frazė, 15 ištarimų 41-nam diktoriui (visi vyriški balsai)	Nepateikta	2.32
[22]	VQ-LPCC	PORTICUS P4	39 diktoriai (24 moterys ir 15 vyrų). Naudojamas ištarimas "8-7-2-3-1-5-9-4-6-0"	Nepateikta	6.64
[22]	GMM-MFCC	PORTICUS P4	40 diktoriai (24 moterys ir 15 vyrų). Naudojamas ištarimas "8-7-2-3-1-5-9-4-6-0"	Nepateikta	6.81
[22]	GMM-4F3AF0	PORTICUS P4	41 diktoriai (24 moterys ir 15 vyrų). Naudojamas ištarimas "8-7-2-3-1-5-9-4-6-0"	Nepateikta	7.28
[22]	RAKTAS	PORTICUS P4	42 diktoriai (24 moterys ir 15 vyrų). Naudojamas ištarimas "8-7-2-3-1-5-9-4-6-0"	Nepateikta	10.65
[23]	MFCC-VQ	Sudaryta	Žodis	63	Nepateikta
[23]	MFCC-HMM	Sudaryta	Žodis	90	Nepateikta
[23]	MFCC-VQ	Sudaryta	Paprastas sakinys	48	Nepateikta
[23]	MFCC-HMM	Sudaryta	Paprastas sakinys	88	Nepateikta
[23]	MFCC-VQ	Sudaryta	Pilnas sakinys	60	Nepateikta
[23]	MFCC-HMM	Sudaryta	Pilnas sakinys	65	Nepateikta

[8] šaltinyje pateikiamas atpažinimo tikslumas skirtingiems požymių tipams. Naudoti klasifikatoriai ir duomenys nėra nurodyti.

[21] šaltinyje eksperimentas atliktas nuo teksto priklausančiam vartotojo atpažinimui. Naudota 13 Mel skalės kepstro koeficientų kiekvienam 26ms signalo kadro ilgiui. Kiekvienas koeficientas klasterizuojamas į 5 elementų klasterį kol suformuojamas 65 elementų vektorius. Apmokoma naudojant daugiasluoksnio perceptrono grįžtamojo ryšio tinklą. Mokymo fazėje atliekamas kalbos signalo apdorojimas, išskiriami Mel skalės kepstro koeficientai, atliekama klasterizacija ir apmokomas dirbtinių neuronų tinklas. Testavimo fazėje atliekamas kalbos signalo apdorojimas, išskiriami Mel skalės kepstro koeficientai, atliekama klasterizacija ir gauti požymių vektoriai perduodami apmokytam neuronų tinklui. Tinklo architektūra – 65 neuronai įvesties sluoksnyje, 40 neuronų paslėptajame sluoksnyje ir 40 neuronų išvesties sluoksnyje. Eksperimentas atliktas 40-ties diktorių garso įrašams su 10 ištarimų „open the door“. 5 įrašai naudojami mokymui ir 5 įrašai naudojami testavimui. Neuronų tinklas keičiamas euklidinių atstumų metodu.

[22] Keturios atpažinimo sistemos buvo naudotos kalbančiojo atpažinimo eksperimentams :

1. Kalbančiojo atpažinimo sistema naudojanti 22 eilės LPCC koeficientus kaip požymius ir vektorinį kvantavimą klasifikavimui. Naudoti 32 centroidai.
2. Kalbančiojo atpažinimo sistema naudojanti 13 eilės MFCC požymiams ir GMM kalbėtojo modeliavimui ir šablonų klasifikavimui. Mišinių komponentų skaičius buvo lygus 16-kai.
3. Kalbančiojo atpažinimo sistema naudojanti 4 formantes, 3 antiformantes, aukštumas (ang. *pitch*) požymiams ir GMM klasifikavimui. Mišinių komponentų skaičius buvo lygus 16 - kai.
4. Foneminis metodas (RAKTAS). Požymių sistema susidedanti iš 36 požymių komponentų, atvaizduojančių individualius kalbėtojo požymius : keturių formančių dažnių, dviejų antiformančių dažnių, 4 formančių normalizuotų amplitudžių ir kitų kombinuotų spektrinių porų parametrų. Formančių radimui buvo naudotas linijinių spektrinių porų metodas. Požymių matrica skaičiuojama kiekvienam įrašui apmokymo metu. Požymių matrica susideda iš požymių vektorių skaičiuojamų kiekvienam kadrai. Apmokymo šablonas gaunamas lyginant dvi požymių matricas. Kalbėtojo požymiams nustatyti lyginamos dvi požymių matricos iš skirtingų įrašų. Balso trakto forma skaičiuojama iš pirmų trijų formančių pirmam kadrai ir tas pats daroma antrai požymių matricai. Randamas skirtumas tarp dviejų minėtų požymių matricų narių. Tada skaičiuojamas skirtumas tarp antro kadro antros požymių matricos pirmos požymių matricos pirmo kadro balso trakto formos. Tas pats daroma trečiam ir kitiems kadrams kol randamas artimiausias požymių vektorius antroje požymių matricoje pirmos požymių matricos pirmam vektoriui. Suradus artimiausią vektorių skaičiuojamas absoliutus atstumas visiems vektorių elementams. Randamos skirtumų

variacijos naudojant histogramas. Įvertinama histogramos forma ir kiti statistiniai parametrai, nustatomas statistinis atstumas (ang. *common statistical distance*). Verifikavimo metu požymių matrica lyginama su kalbėtojo modelio požymių matrica ir, tokiu pat būdu kaip ir apmokymo procese. Atitinkantiems kalbėtojams atstumas patenkanta į intervalą, neatitinkantiems nepatenka.

Rezultatai gauti su garso įrašais iš RUSBASE ir "PORTICUS" firmos sukurtos duomenų bazės.

[23] šaltinyje rezultatai gauti 4-iems skirtingiems diktoriaus (3 moterys, vienas vyras). Mokymui naudojami 5 žodžio „*saya*“ ištarimai, testavimui 10 žodžio „*saya*“ ištarimų, 10 žodžio „*saya sedang belajar*“ ištarimų, 10 žodžio „*saya berangkat*“ ištarimų.

1.3 Programiniai paketai

ALIZÉ

ALIZÉ yra atviro kodo platforma skirta kalbančiojo atpažinimui, sukurta C++ kalboje. ALIZÉ projekto tikslas yra suteikti aukšto ir žemo lygio karkasus, kurie leistų bet kam vystyti programinę įrangą, leidžiančią apdoroti įvairias kalbančiojo atpažinimo užduotis : verifikaciją, identifikaciją, segmentaciją ir pan. Pagrindinis sluoksnis yra ALIZE-Core – žemo lygio biblioteka, turinti funkcijas reikalingas Gauso mišiniams ir įvesties-išvesties funkcijas įvairiems failų formatams. Pagrindiniam sluoksniui buvo sukurtas rinkinys LIA_RAL, siūlantis aukštesnio lygio funkcionalumą [24].

LIA_RAL susideda iš kelių komponentų [24] :

- LIA_SpkDet – įrankių rinkinys skirtas užduočių, reikalingų kalbančiojo autentifikacijos sistemoms įvykdymui : modelių apmokymui, požymių normalizavimui, rezultatų normalizavimui;
- LIA_SpkSeg – įrankiai kalbančiojo diarizacijai;
- LIA_Utils – duomenų formatų, naudojamų platformoje ALIZÉ keitimui;
- LIA_SpkTools – aukšto lygio funkcijos ALIZÉ sluoksniui - biblioteka, kurią naudojant sukurtos kitos dalys.

SPEAR

SPEAR paketas yra dalis *bob.bio* paketų, kurie pateikia atvirojo kodo įrankius biometrinio atpažinimo eksperimentams, sukurtų C++ kalba ir turinčių Python sąsają. SPEAR paketas skirtas kalbančiojo atpažinimo eksperimentams. Jame įtraukti : kalbančiojo atpažinimo duomenų bazės ir jų protokolai, balso aktyvumo detektavimas, požymių išskyrimas, atpažinimo bei verifikavimo įrankiai [25].

HTK

HTK yra programinis įrankis paslėptųjų Markovo modelių manipuliavimui. Paprastai naudojamas kalbos atpažinimui. Kuris taip pat buvo naudotas kitoms užduotims : kalbos sintezej, raidžių atpažinimui ir DNR sekoms tirti.

HTK susideda iš kelių programinių įrankių naudojamų : kalbos analizej, paslėptųjų Markovo modelių apmokymui, testavimo ir rezultatų analizej. Palaiko tolydžius Gauso mišinių tankių ir diskrečius pasiskirstymus [26].

Dirbtinių neuronų tinklų paketas

Dirbtinių neuronų tinklų paketas (ang. *Neural network toolbox*) pateikia įrankius reikalingus kurti, mokyti ir vizualizuoti dirbtinių neuronų tinklus [27].

Statistikos ir mašininio mokymo paketas

Statistikos ir mašininio mokymo paketas (ang. *Statistics and Machine Learning Toolbox*) pateikia funkcijas duomenų analizei, aprašymui ir modeliavimui. Regresijos ir klasifikavimo algoritmus prognozavimo modelių kūrimui [28].

2 PROJEKTAVIMAS

Šiame skyriuje aprašomi tiriami metodai. Pateikiamos duomenų srautų diagramos skirtingiems metodams.

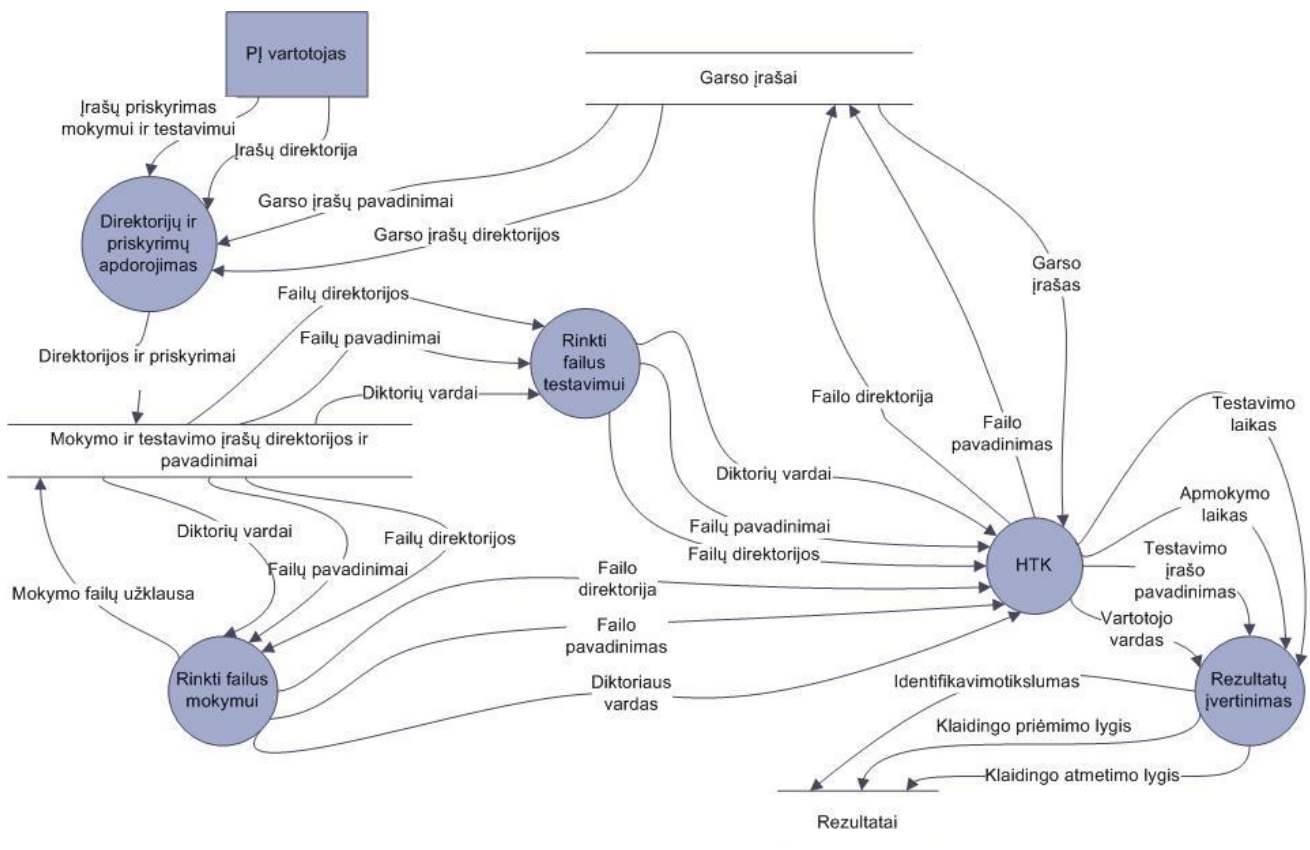
2.1 pav. ir 2.2 pav. procesas „*direktorijų ir priskyrimų apdorojimas*“ turi sukurti ir išsaugoti mokymui priskirtų įrašų direktorijas, įrašų pavadinimus atskiruose failuose (pagal diktoriaus vardą arba failo pavadinimą). Tas pats procesas turi sukurti ir išsaugoti testavimo failų įrašų pavadinimus ir direktorijas naujame faile. Failų priskyrimas mokymo ir testavimo failams priklauso nuo vartotojo pasirinktų parametrų.

Procesas „*rinkti failus mokymui*“ turi nuskaityti garso įrašų direktorijas, garso įrašų pavadinimus ir diktorių vardus iš atskirų failų.

Procesas „*rinkti failus testavimui*“ turi nuskaityti testavimo failų direktorijas, failų pavadinimus ir priskyrimą diktoriui iš vieno failo.

Procesas paslėptųjų Markovo modelių paketas (*HTK*) susideda iš terminalo komandų.

Procesas „*Rezultatų įvertinimas*“ pagal „*HTK*“ proceso sukurtą rezultatų failą turi įvertinti ir išsaugoti identifikavimo tikslumą, klaidingo atmetimo lygį ir klaidingo priėmimo lygį.



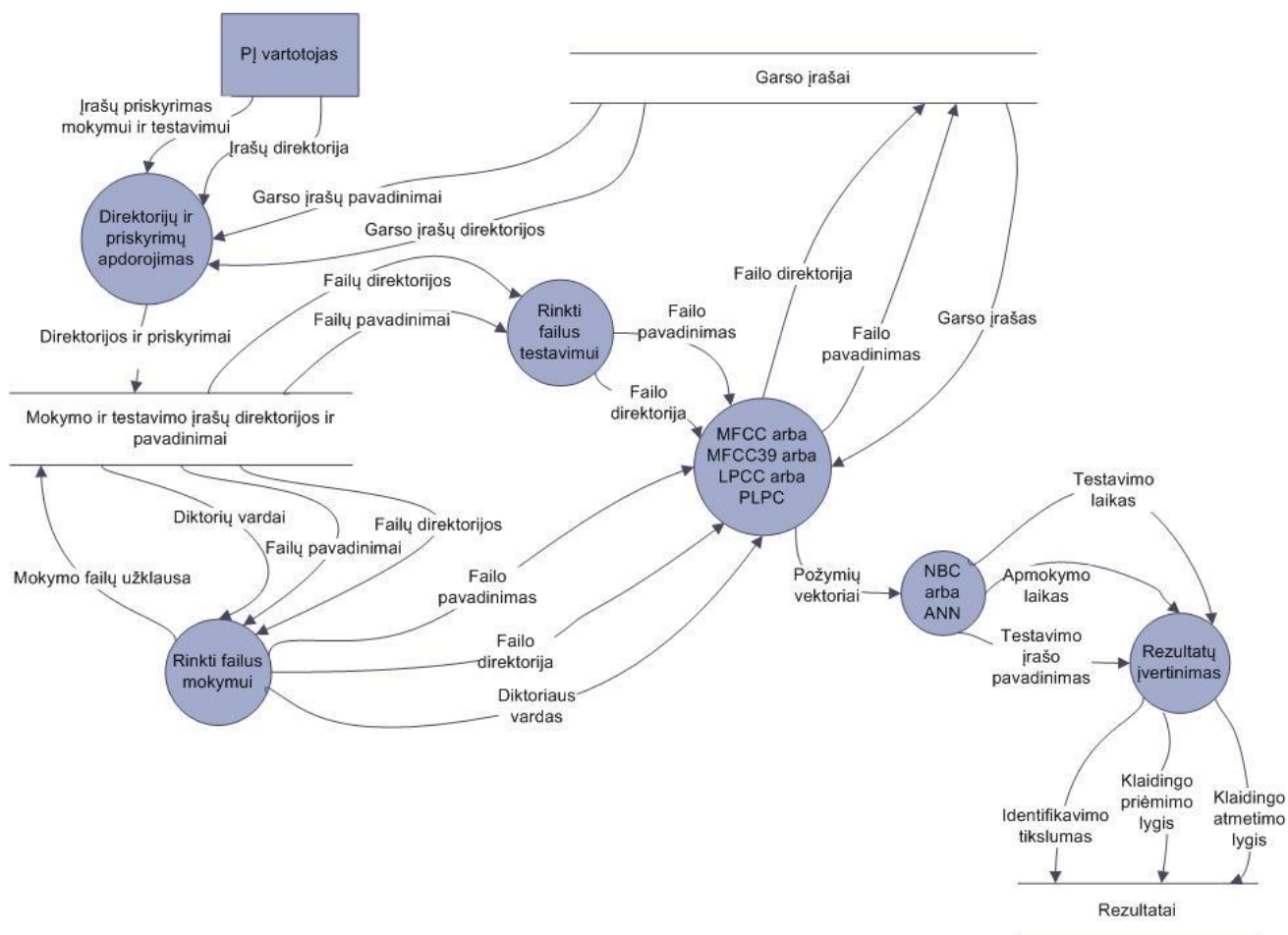
2.1 pav. Paslėptųjų Markovo modelių paketo (HTK) duomenų srauto diagrama.

2.2 pav. procesas „*MFCC arba 39MFCC arba LPC arba PLP*“ atitinka vieną iš keturių procesų :

1. *MFCC* – turi gražinti pasirinktą Mel skalės kepstro koeficientų požymių skaičių.
2. *MFCC39* – turi gražinti 12 Mel skalės kepstro koeficientų verčių, 12 delta Mel skalės kepstro koeficientų verčių, 12 delta-delta Mel skalės kepstro koeficientų verčių, vieną kadro energijos logaritmo vertę, vieną delta kadro energijos logaritmo vertę, vieną delta-delta kadro energijos logaritmo vertę. Toks požymių rinkinys laikomas standartu [23].
3. *LPC* – turi gražinti pasirinktą skaičių tiesinės prognozės koeficientų.
4. *PLP* – turi gražinti pasirinktą skaičių suvokimo (percepinės) prognozės koeficientų.

Procesas “NBC arba ANN” atitinka viena iš dviejų procesų :

1. *NBC* – mokymo metu turi sudaryti parametrus, kurie testavimo metu naudojami identifikavimui.
2. *ANN* – mokymo metu sudaryti neuron tinklą ir nustatyti jo parametrus, kurie testavimo metu naudojami identifikavimui.



2.2 pav. MFCC-NBC, MFCC39-NBC, LPCC-NBC, PLPC-NBC, MFCC-ANN, MFCC39-ANN, LPCC-ANN ir PLPC-ANN duomenų srauto diagrama.

3 REALIZACIJA

Skyriuje pateikiama programinės įrangos naudotos eksperimentui atlikti realizacija. Mokymo ir testavimo failai skirstomi pagal vartotojo parinktus parametrus ir surašomi į *.mat* failus, naudojant *MATLAB* skriptus ir funkcijas. Failai reikalingi paslėptųjų Markovo modelių paketui kuriami *MATLAB* skriptais ir funkcijomis iš tų pačių *.mat* failų.

3.1 Paslėptųjų Markovo modelių paketas

Diktoriaus atpažinimui reikalingi failai :

1. „*gram.dict*“ faile diktorių vardai atskiriami | simboliu.

```
1 $speaker = FAGNGRA | FAGNVIN | FAISIZI | FAISZYM | FAUSNEM | FDALLOI | FGINGED | FIEVJUR | FIEVSAB | FIEV
2 VIS | FJUSKIN | FKAMMOS | FLAUZET | FRAISAV | FRUTNAN | FSIMMEI | FUGNBUC | FUGNNOV | FVAIVAI | FVANPEC | K;
3 ($speaker)
```

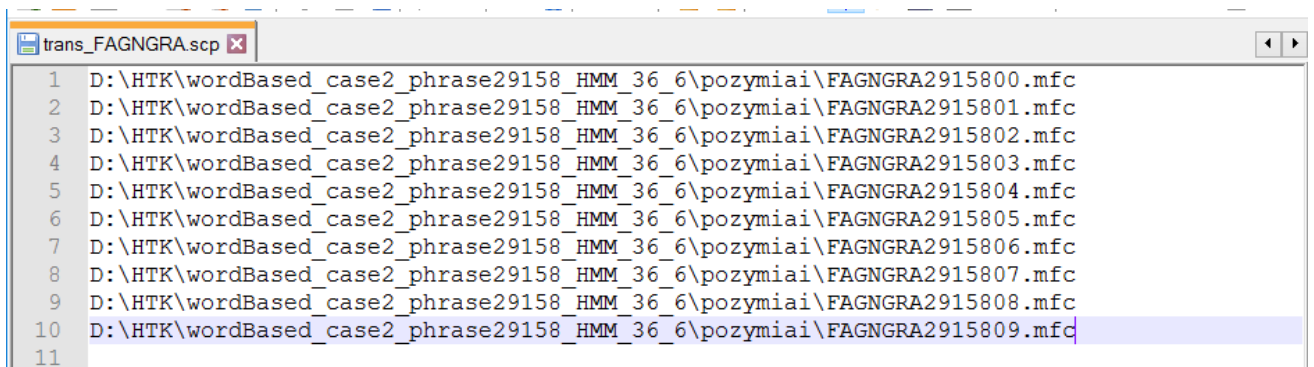
3.1 pav. „*gram.dict*“ failo pavyzdys

2. „*failai.scp*“ nurodomi garso įrašai ir failai, kuriuose bus talpinami požymiai.

```
1 C:\Users\rouge\Desktop\SpeakerZINskDU\irasai\FAGNGRA200.wav C:\Users\rouge\Desktop\SpeakerZINskDU\pozymiai\FAGNGRA200.mfc
2 C:\Users\rouge\Desktop\SpeakerZINskDU\irasai\FAGNGRA201.wav C:\Users\rouge\Desktop\SpeakerZINskDU\pozymiai\FAGNGRA201.mfc
3 C:\Users\rouge\Desktop\SpeakerZINskDU\irasai\FAGNGRA202.wav C:\Users\rouge\Desktop\SpeakerZINskDU\pozymiai\FAGNGRA202.mfc
4 C:\Users\rouge\Desktop\SpeakerZINskDU\irasai\FAGNGRA203.wav C:\Users\rouge\Desktop\SpeakerZINskDU\pozymiai\FAGNGRA203.mfc
5 C:\Users\rouge\Desktop\SpeakerZINskDU\irasai\FAGNGRA204.wav C:\Users\rouge\Desktop\SpeakerZINskDU\pozymiai\FAGNGRA204.mfc
6 C:\Users\rouge\Desktop\SpeakerZINskDU\irasai\FAGNVIN200.wav C:\Users\rouge\Desktop\SpeakerZINskDU\pozymiai\FAGNVIN200.mfc
7 C:\Users\rouge\Desktop\SpeakerZINskDU\irasai\FAGNVIN201.wav C:\Users\rouge\Desktop\SpeakerZINskDU\pozymiai\FAGNVIN201.mfc
8 C:\Users\rouge\Desktop\SpeakerZINskDU\irasai\FAGNVIN202.wav C:\Users\rouge\Desktop\SpeakerZINskDU\pozymiai\FAGNVIN202.mfc
9 C:\Users\rouge\Desktop\SpeakerZINskDU\irasai\FAGNVIN203.wav C:\Users\rouge\Desktop\SpeakerZINskDU\pozymiai\FAGNVIN203.mfc
10 C:\Users\rouge\Desktop\SpeakerZINskDU\irasai\FAGNVIN204.wav C:\Users\rouge\Desktop\SpeakerZINskDU\pozymiai\FAGNVIN204.mfc
```

3.2 pav. „*failai.scp*“ failo pavyzdys

3. „*trans_*.scp*“ – failuose, kurių pavadinimai prasideda „*trans_*“ ir baigiasi plėtiniu „*.scp*“ nurodomi mokymui skirti požymių failai. Failų pavadinimai atitinka diktorių vardus. Tokių failų skaičius atitinka diktorių skaičių sistemoje.



```
trans_FAGNGRA.scp
1 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai\FAGNGRA2915800.mfc
2 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai\FAGNGRA2915801.mfc
3 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai\FAGNGRA2915802.mfc
4 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai\FAGNGRA2915803.mfc
5 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai\FAGNGRA2915804.mfc
6 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai\FAGNGRA2915805.mfc
7 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai\FAGNGRA2915806.mfc
8 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai\FAGNGRA2915807.mfc
9 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai\FAGNGRA2915808.mfc
10 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai\FAGNGRA2915809.mfc
11
```

3.3 pav. „*trans_FAGNGRA.scp*“ failo pavyzdys

4. „*hmm_**“ – failai, kurie prasideda „*hmm_*“ naudojami kaip modelių failai. Taip pat pilnas failo pavadinimas atitinka diktoriaus vardui – pvz. „*hmm_FAGNGRA*“ ir failų skaičius atitinka diktorių skaičių sistemoje. Modelio failo pavyzdys pateiktas 3.4 pav., modelio failo parametrų aprašymas 3.1 lentelėje.

```

1 ~o
2 <STREAMINFO> 1 39
3 <VECSIZE> 39<NULLD><MFCC_E_D_A>
4 ~h "hmm_FAGNGRA"
5 <BEGINHMM>
6 <NUMSTATES> 3
7 <STATE> 2
8 <MEAN> 39
9 -1.804065e+001 1.822792e-001 -5.504993e+000 -3.084137e+000 -8.302935e-001 -1.075914e+000 -3.220609e+00
10 0 -2.169371e+000 -2.888519e+000 -2.875683e+000 -3.929880e+000 -3.982161e+000 5.133442e-001 -1.597592e-
11 001 -1.023607e-001 -6.501031e-002 -7.424914e-003 1.162796e-001 1.661122e-001 2.086510e-001 1.879907e-0
12 01 7.306663e-002 5.799584e-002 1.143509e-001 9.368838e-002 -4.867257e-003 -3.693496e-003 8.087719e-004
13 9.183593e-004 4.125852e-003 -5.605795e-003 -1.784145e-002 -1.409231e-002 -3.455685e-003 -2.951380e-003
14 3.270557e-003 -8.228377e-003 -1.172888e-002 -1.151007e-004
15 <VARIANCE> 39
16 3.854116e+001 1.921158e+001 2.638009e+001 5.835318e+001 1.165135e+001 1.929613e+001 4.677795e+001 2.46
17 7102e+001 2.012724e+001 2.270684e+001 1.216588e+001 1.579007e+001 6.502423e-002 1.030196e+000 1.064985
18 e+000 1.065678e+000 2.368123e+000 9.835215e-001 1.225612e+000 1.591962e+000 1.073853e+000 1.391295e+00
19 0 1.813913e+000 9.145826e-001 9.396029e-001 1.308009e-003 1.466247e-001 1.380796e-001 1.392069e-001 3.
20 106321e-001 1.716268e-001 2.372055e-001 2.667109e-001 1.938590e-001 2.306694e-001 2.612837e-001 1.7579
21 45e-001 1.794921e-001 1.457760e-004
22 <GCONST> 7.431341e+001
23 <TRANSP> 3
24 0.000000e+000 1.000000e+000 0.000000e+000
25 0.000000e+000 9.842767e-001 1.572327e-002
26 0.000000e+000 0.000000e+000 0.000000e+000
27 <ENDHMM>

```

3.4 pav. Modelio failo pavyzdys

Lentelė 3.1 Modelio failo prametrų reikšmės [29]

Parametras	Reikšmė1	Reikšmė2
~o	Pradžios simbolis	-
<STREAMINFO>	Eilučių skaičius, kuriose bus išdėstytos požymių koeficientų vertės	Požymių koeficientų verčių skaičius
<VECSIZE>	Požymių vektoriaus dydis	Požymių vektoriaus tipas
~h "hmm_*"	Failo pavadinimas	
<BEGINHMM>	Modelio pradžios žyma	
<NUMSTATES>	Būsenų skaičius	
<NUMMIXES>	Gauso mišinių skaičius	
<MIXTURE>	Mišinio numeris	Svorio vertė
<MEAN>	Vidutinės požymių vertės diktoriaus modeliui	-
<VARIANCE>	Požymių dispersijų vertės diktoriaus modeliui	-
<TRANSP>	Matricos dydis (lygus būsenų skaičiui)	Perėjimo matrica, kitaip vadinama stochastine, tikimybių arba Markovo matrica.
<ENDHMM>	Modelio pabaigos žyma	-

5. „CONFIG“ failas naudojamas požymių išskyrimui. „CONFIG“ failo turinys pateiktas 3.5 pav.


```

1 SOURCEKIND=WAVEFORM
2 SOURCEFORMAT=WAVE
3 TARGETKIND=MFCC_D_A_E
4 SAVEWITHCRC=F
5 SOURCERATE=625
6 TARGETRATE=100000.0
7 WINDOWSIZE=250000.0
8 USEHAMMING=T
9 PREEMCOEF=0.96
10 NUMCEPS=12
11 NUMCHANS=20

```

3.5 „CONFIG“ failo turinys

3.2 „CONFIG“ parametrų aprašymas [29]

Parametras	Vertė	Reikšmė
SOURCEKIND	WAVEFORM	Visi įvesties failai atitinka kalbos signalą (ang. <i>speech waveform</i>)
SOURCEFORMAT	WAVE	Signalų failo formatas
TARGETKIND	MFCC_D_A_E	Išvesties (požymių failai) susideda iš 12 Mel skalės kepstro koeficientų, 12 delta Mel skalės kepstro koeficientų, 12 delta-delta skalės kepstro koeficientų, energijos logaritmo, delta energijos logaritmo ir delta-delta energijos logaritmo
SAVEWITHCRC	F	Nepridedama kontrolinė suma (ang. <i>checksum</i>) išvesties parametro failui
SOURCERATE	625	Dažninė rezoliucija 100ns vienetais
TARGETRATE	100000	Periodas 100ns vienetais
WINDOWSIZE	250000	Lango dydis 100ns vienetais
USEHAMMING	T	Naudojamas Hamming langas
PREEMCOEF	0.96	Filtro koeficiento vertė
NUMCEPS	12	Kepstro koeficientų skaičius
NUMCHANS	20	Filtrų rinkinių skaičius

6. „*test.scf*“ papildomas failas požymių failų sudarymui. Analogiškas „*failai.scf*“ failui.

7. „*test_mfc.scf*“ nuorodų į testinius požymių failus failas :

```

1 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNGRA2915810.mfc
2 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNGRA2915811.mfc
3 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNGRA2915812.mfc
4 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNGRA2915813.mfc
5 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNGRA2915814.mfc
6 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNGRA2915815.mfc
7 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNGRA2915816.mfc
8 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNGRA2915817.mfc
9 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNGRA2915818.mfc
10 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNGRA2915819.mfc
11 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNVIN2915810.mfc
12 D:\HTK\wordBased_case2_phrase29158_HMM_36_6\pozymiai2\FAGNVIN2915811.mfc

```

3.6 pav. „test_mfc.scf“ failo pavyzdys

8. „*hmmlist*“ faile – modelių pavadinimų sąrašas.

```

1 hmm_FAGNGRA
2 hmm_FAGNVIN
3 hmm_FAISIZI
4 hmm_FAISZYM
5 hmm_FAUSNEM
6 hmm_FDAILOI
7 hmm_FGINGED
8 hmm_FIEVJUR
9 hmm_FIEVSAB
10 hmm_FIEVVIS
11 hmm_FJUSKIN
12 hmm_FKAMMOS
13 hmm_FLAUZET
14 hmm_FRAISAV
15 hmm_FRUTNAN
16 hmm_FSIMMEI
17 hmm_FUGNBUC
18 hmm_FUGNNOV
19 hmm_FVAIVAI
20 hmm_FVANPEC
21 hmm_K

```

3.7 pav. „hmmlist“ failo turinio pavyzdys

9. „dict“ faile nurodoma kokios klasės turėtų būti atvaizduotos rezultatų faile

```

1 FAGNGRA hmm_FAGNGRA
2 FAGNVIN hmm_FAGNVIN
3 FAISIZI hmm_FAISIZI
4 FAISZYM hmm_FAISZYM
5 FAUSNEM hmm_FAUSNEM
6 FDAILOI hmm_FDAILOI
7 FGINGED hmm_FGINGED
8 FIEVJUR hmm_FIEVJUR
9 FIEVSAB hmm_FIEVSAB
10 FIEVVIS hmm_FIEVVIS
11 FJUSKIN hmm_FJUSKIN
12 FKAMMOS hmm_FKAMMOS
13 FLAUZET hmm_FLAUZET
14 FRAISAV hmm_FRAISAV
15 FRUTNAN hmm_FRUTNAN
16 FSIMMEI hmm_FSIMMEI
17 FUGNBUC hmm_FUGNBUC
18 FUGNNOV hmm_FUGNNOV
19 FVAIVAI hmm_FVAIVAI
20 FVANPEC hmm_FVANPEC
21 K hmm_K

```

3.8 pav. „dict“ failo turinio pavyzdys

10. „testref.mlf“ faile nurodoma kokiam kalbėtojų turėtų būti priskirtas įrašas :

```

1  #!MLF!#
2
3  "*FAGNGRA2915810.lab"
4  FAGNGRA
5  .
6  "*FAGNGRA2915811.lab"
7  FAGNGRA
8  .
9  "*FAGNGRA2915812.lab"
10 FAGNGRA
11 .
12 "*FAGNGRA2915813.lab"
13 FAGNGRA
14 .
15 "*FAGNGRA2915814.lab"
16 FAGNGRA

```

3.9 pav. „testref.mlf“ failo turinio pavyzdys

```

1  HParse.exe gram.dict wordnet.txt
2  HCopy -C CONFIG -S failai.scp
3
4  HRest.exe -T 1 -S trans_FAGNGRA.scp hmm_FAGNGRA>>kk1
5  HRest.exe -T 1 -S trans_FAGNVIN.scp hmm_FAGNVIN>>kk2
6  HRest.exe -T 1 -S trans_FAISIZI.scp hmm_FAISIZI>>kk3
7  HRest.exe -T 1 -S trans_FAISZYM.scp hmm_FAISZYM>>kk4
8  HRest.exe -T 1 -S trans_FAUSNEM.scp hmm_FAUSNEM>>kk5
9  HRest.exe -T 1 -S trans_FDAILOI.scp hmm_FDAILOI>>kk6
10 HRest.exe -T 1 -S trans_FGINGED.scp hmm_FGINGED>>kk7
11 HRest.exe -T 1 -S trans_FIEVJUR.scp hmm_FIEVJUR>>kk8
12 HRest.exe -T 1 -S trans_FIEVSAB.scp hmm_FIEVSAB>>kk9
13 HRest.exe -T 1 -S trans_FIEVVIS.scp hmm_FIEVVIS>>kk10
14 HRest.exe -T 1 -S trans_FJUSKIN.scp hmm_FJUSKIN>>kk11
15 HRest.exe -T 1 -S trans_FKAMMOS.scp hmm_FKAMMOS>>kk12
16 HRest.exe -T 1 -S trans_FLAUZET.scp hmm_FLAUZET>>kk13
17 HRest.exe -T 1 -S trans_FRAISAV.scp hmm_FRAISAV>>kk14
18 HRest.exe -T 1 -S trans_FRUTNAN.scp hmm_FRUTNAN>>kk15
19 HRest.exe -T 1 -S trans_FSIMMEI.scp hmm_FSIMMEI>>kk16
20 HRest.exe -T 1 -S trans_FUGNBUC.scp hmm_FUGNBUC>>kk17
21 HRest.exe -T 1 -S trans_FUGNNOV.scp hmm_FUGNNOV>>kk18
22 HRest.exe -T 1 -S trans_FVAIVAI.scp hmm_FVAIVAI>>kk19
23 HRest.exe -T 1 -S trans_FVANPEC.scp hmm_FVANPEC>>kk20
24 HRest.exe -T 1 -S trans_K.scp hmm_K>>kk21
25
26 HCopy -C CONFIG -S test.scp
27 HVite -n 3 3 -T 1 -S test_mfc.scp -i results -w wordnet.txt dict hmmlist
28 HResults -p -I testref.mlf dict results>>rez

```

3.10 pav. HTK terminalo komandų pavyzdžiai

HTK apmokymas ir atpažinimas susideda iš kelių komandų (3.10 pav.)

1. *HParse.exe gram.dict wordnet.txt* komanda naudoja „gram.dict“ failą naudojamas *wordnet.txt* failo sukūrimui
2. *HCopy -C CONFIG -S failai.scp*. Sukuria .mfc plėtinio požymių failus. -C – config failo nustatymui (config failo pavyzdys pateiktas 3.4 pav., jo parametrų aprašymai pateikti 3.1 lentelėje), -S - .scp failo nurodymui.
3. *HRest.exe -T 1 -S trans_FAGNGRA.scp hmm_FAGNGRA>>kk1*. HRest yra naudojamas mokymui - atlieka Baum-Welch įvertinimą naudojant priskirtus požymių failus kalbėtojiui (pvz. „trans_FAGNGRA.scp“, kuriame nurodyti požymių failai atitinkamam diktoriui) ir modelių failus (pvz. „hmm_FAGNGRA“)
4. *HVite -n 3 3 -T 1 -S test_mfc.scp -i results -w wordnet.txt dict hmmlist*. Sukuria rezultatų „results“ failą. Kuriame matomi pirmi trys panašiausi diktoriai testavimo failui.

5. *HResults -p -I testref.mlf dict results>>rez.* Sukuria rezultatų failą, pateiktą 3.11 pav.

```

===== HTK Results Analysis =====
Date: Fri Mar 24 15:37:47 2017
Ref : testref.mlf
Rec : results
----- Overall Results -----
SENT: %Correct=80.61 [H=399, S=96, N=495]
WORD: %Corr=80.61, Acc=80.61 [H=399, D=0, S=96, I=0, N=495]
----- Confusion Matrix -----

```

	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	K
	A	A	A	A	A	D	G	I	I	I	J	K	L	R	R	S	U	U	V	V
	G	G	I	I	U	A	I	E	E	E	U	A	A	A	U	I	G	G	A	A
	N	N	S	S	S	I	N	V	V	V	S	M	U	I	T	M	N	N	I	N
	G	V	I	Z	N	L	G	J	S	V	K	M	Z	S	N	M	B	N	V	P
	Del [%c / %e]																			
FAGN	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FAGN	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FAIS	0	0	14	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
FAIS	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FAUS	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FDAI	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FGIN	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0
FIEV	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0
FIEV	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0
FIEV	0	1	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0
FJUS	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	1
FKAM	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0
FLAU	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0
FRAI	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0
FRUT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	1
FSIM	5	0	0	0	0	0	0	3	0	0	0	0	0	0	0	4	0	0	0	3
FUGN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0
FUGN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0
FVAI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0
FVAN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15
K	4	0	1	0	0	0	0	11	10	6	2	3	1	1	14	0	1	3	1	23
Ins	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

3.11 pav. Rezultatų failo pavyzdys

3.2 MATLAB paketai

3.2.1 Požymių išskyrimas

Mel skalės kepstro koeficientų išskyrimui, delta Mel skalės kepstro koeficientų, delta-delta Mel skalės kepstro koeficientų, energijos logaritmo vertė, delta energijos logaritmo vertė, delta-delta energijos logaritmo vertė ir tiesinės prognozės koeficientų išskyrimui naudojamo *VOICEBOX* [30] paketo funkcijos.

Mel skalės kepstro koeficientai išskiriami naudojant funkciją *melcepst VOICEBOX* paketo funkciją. Išvesties parametras atitinka koeficientų vertes. Įvesties parametru reikšmės pateiktos 3.3 lentelėje.

```
c = melcepst(s, fs, ['M', 'E', 'd', 'D'], nc, 22, 256, 100);
```

3.12 pav. *melcepst* funkcijos kvietimo pavyzdys

Lentelė 3.3 *melcepst* funkcijos parametų reikšmės

Įvesties parametrai	Reikšmė
s	Signalų amplitudė
fs	Signalų dažninė rezoliucija
'M'	Hamming langų naudojimui
'E'	Pridedamos energijos logaritminės reikšmės
'd'	Pridedami delta koeficientai
'D'	Pridedami delta-delta koeficientai
nc	Požymių koeficientų skaičius
22	Filtrų rinkinių skaičius
256	Kadro ilgis
100	Kadro persiklojimas

LPC koeficientai išskiriami naudojant funkciją *VOICEBOX* paketo funkciją. Iš įvesties parametų naudojami autokoreliacijos koeficientai. 3.13 pav. s – signalo amplitudė, nc – autokoreliacijos koeficientų skaičius, 256 – kadro dydis.

$$c = \text{lpcauto}(s, nc, 256);$$

3.13 pav. *lpcauto* funkcijos kvietimo pavyzdys

RASTA-PLP koeficientai išskiriami naudojant *rastapl* funkciją [31].

$$c = \text{rastapl}(s, fs, 0, nc);$$

3.14 pav. *rastapl* funkcijos kvietimo pavyzdys

3.14 pav. c – percepcinės tiesinės prognozės koeficientų skaičius, s – signalo amplitudė, fs – dažninė rezoliucija, 0 – suvokimo tiesinės prognozės koeficientų skaičiavimas be *RASTA* filtro, nc – koeficientų skaičius.

3.2.2 Dirbtiniai neuronų tinklai

Dirbtiniams neuronų tinklams naudojamas integruotas MATLAB *Neural Network* paketas. Funkcijos *patternnet()* naudojamas parametras yra neuronų skaičius paslėptajame sluoksnyje ir paslėptųjų sluoksnių skaičius. *patternnet()* sudaro objektą, kurio savybės atitinka neuronų tinklo parametrus. Mokymui naudojama *train()* funkciją, kurios pirmas įvesties parametras yra neuronų tinklo objektas, antras – požymių matrica, trečias – „taikinio matrica“, kurioje nurodomi požymių vektorius priskyrimai atitinkamai klase. Testavimas atliekamas *net()* funkcija, kurios įvesties parametras yra testavimo įrašų požymių matrica. Išvesties parametras – tikimybių matrica.

3.2.3 Bajeso klasifikatorius

Bajeso klasifikatoriui naudojamas integruotas MATLAB *Statistics and machine Learning* paketas. Funkcijos *NaiveBayes.fit()* pirmas įvesties parametras yra požymių matrica, antras – klasių

vardų vektorius. *predict()* įvesties parametras yra testuojamų garso įrašų požymių matrica išvesties – priskirtų klasių vardų vektorius.

4 EKSPERIMENTAS

4.1 Eksperimento duomenys

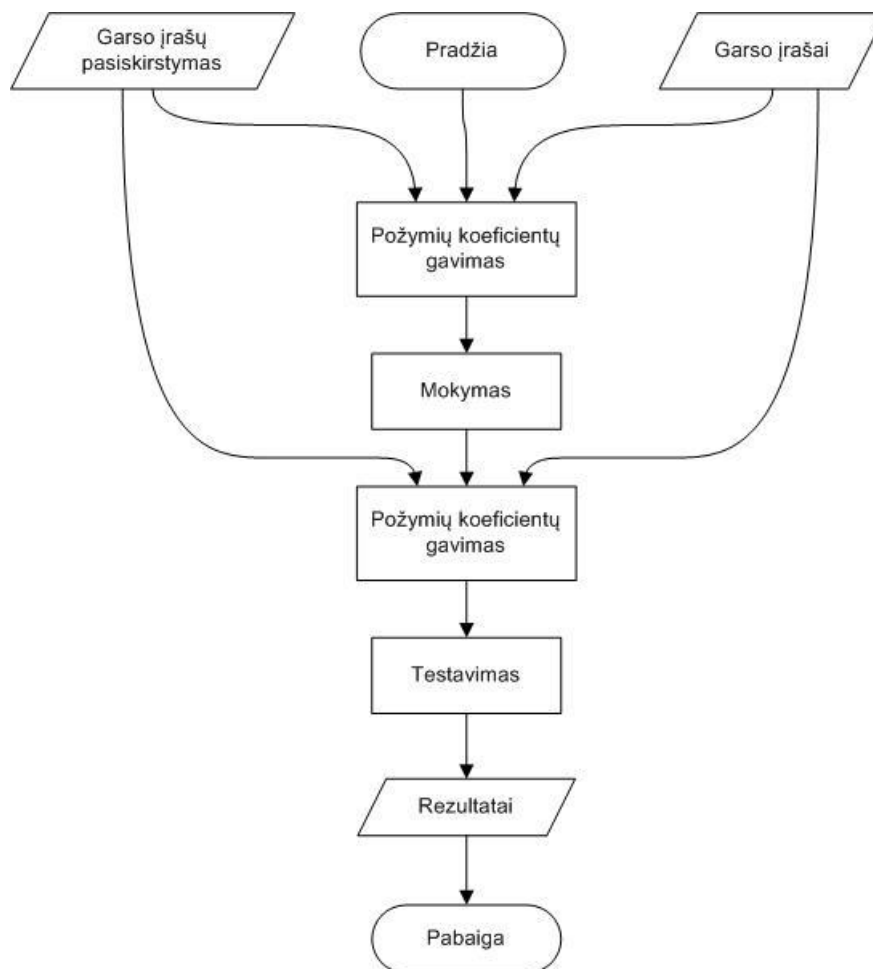
Eksperimento duomenys susideda iš 6000 .wav formato įrašų. Įrašai sudaryti 30-ties diktorių 7 iš jų vyrai, 23 moterys. Kiekvienas diktorius yra ištaręs tą pačią frazę 20 kartų. Tariamų frazių skaičius lygus 10. Tiriamos frazės : „nulis“, „vienas“, „du“, „trys“, „keturi“, „penki“, „šeši“, „septyni“, „aštuoni“, „devyni“. Minėti ištarimai toliau žymimi skaitmenimis.

4.2 Eksperimento eiga

1. Gauti rezultatus su pradinėmis reikšmėmis mokymo ir testavimo įrašus priskiriant pagal pirmą atvejį pateiktą priede 7.1 lentelėje (17.5% įrašų prisikiriant mokymo įrašams, 82.5 % testavimo įrašams)
2. Gauti rezultatus su pradinėmis reikšmėmis mokymo ir testavimo įrašus priskiriant pagal antrą atvejį pateiktą priede 7.1 lentelėje (41.6% įrašų priskiriant mokymo įrašams, 58.4% testavimo įrašams)
3. Gauti rezultatus su pradinėmis reikšmėmis mokymo ir testavimo įrašus priskiriant pagal trečią atvejį pateiktą priede 7.2 lentelėje (62.5% įrašų priskiriant mokymo įrašams, 37.5% testavimo įrašams)
4. Gauti rezultatus su pradinėmis reikšmėmis mokymo ir testavimo įrašus priskiriant pagal ketvirtą atvejį pateiktą priede 7.2 lentelėje (75% įrašų priskiriant mokymo įrašams, 25% testavimo įrašams)
5. Gauti rezultatus keičiant Mel skalės kepstro, tiesinės prognozės kepstro, suvokimo tiesinės prognozės kepstro koeficientų skaičių
6. Gauti rezultatus keičiant Gauso mišinių ir būsenų skaičių paslėptųjų Markovo modelių programiniame pakete
7. Gauti rezultatus keičiant paslėptojo sluoksnio neuronų skaičių dirbtiniams neuronų tinklams

4.3 Eksperimento rezultatai

Eksperimento metu rezultatai gaunami iš skirtingų požymių ir klasifikatorių konfigūracijų : tiesinės prognozės kepstro koeficientų su dirbtiniais neuronų tinklais (LPCC-ANN), Mel skalės kepstro koeficientų su dirbtiniais neuronų tinklais (MFCC-ANN), suvokimo tiesinės prognozės kepstro koeficientų su dirbtiniais neuronų tinklais (PLPC-ANN), Mel skalės kepstro koeficientų su Bajeso klasifikatoriumi (MFCC-NBC), tiesinės prognozės kepstro koeficientų su Bajeso klasifikatoriumi (LPCC-NBC), suvokimo tiesinės prognozės su Bajeso klasifikatoriumi (PLPC-NBC), 12- ka Mel skalės kepstro koeficientų, 12 delta – Mel skalės kepstro koeficientų, 12 delta-delta Mel skalės kepstro koeficientų, 1- na logaritmuota energijos vertė, 1-na logaritmuota delta-energijos vertė, 1-na logaritmuota delta-delta energijos vertė su dirbtiniais neuronų tiklais, Bajeso klasifikatoriumi ir paslėptaisiais Markovo modeliais (MFCC39-ANN, MFCC39-NBC, MFCC39-HTK). Skaičiai rašomi prie MFCC, LPCC, MFCC ir PLPC reiškia koeficientų skaičių. Naudojant programą, kurios apibendrinta blokinė schema vienam metodui pateikta 4.1 pav. gali skirtis priklausomai nuo reikalingų parametrų gavimo ar metodų parametrų keitimo, reikalingų skirtingiems eksperimento atvejams visi metodai leidžiami nuosekliai.

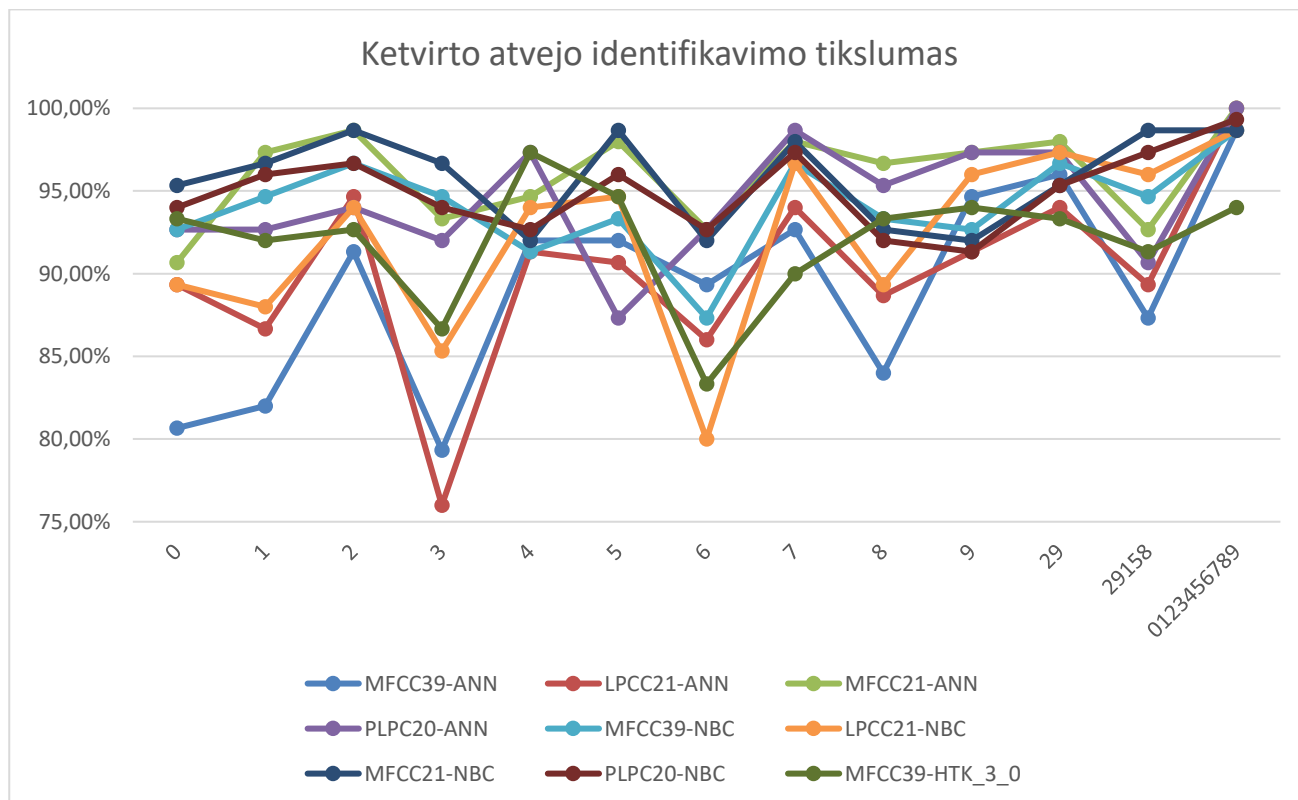


4.1 pav. Vieno metodo eksperimento schema

4.3.1 Rezultatai su pradinėmis reikšmėmis

Pradinėmis reikšmėmis vadinama 3 būsenos ir 0 mišinių – paslėptųjų Markovo modelių paketui, 10 neuronų ir vienas paslėptasis sluoksnis – dirbtiniams neuronų tinklams, 21 Mel skalės kepstro koeficientas, 21 tiesinės prognozės koeficientas, 20 suvokimo tiesinės prognozės koeficientų.

Rezultatai su pradinėmis reikšmėmis skirti identifikavimo tikslumo, klaidingų priėmimo lygių, klaidingo atmetimų lygių, apmokymo laikų, vieno įrašo testavimo laikų įvertinimui ir eksperimento garso įrašų paskirstymo atvejų pateiktų priede 7.1 ir 7.2 lentelėse įvertinimui.



4.2 pav. Metodų identifikavimo tikslumai ketvirto atvejo garso įrašo paskirstymui

4.2 pav. vertikaloje ašyje pateikti identifikavimo tikslumai, horizontalioje – skirtingi ištarimai, skirtingiems metodams. Horizontalioje ašyje simbolis „0“ atitinka frazę „nulis“, simbolis „1“ atitinka frazę „vienas“, simbolis „2“ atitinka frazę „du“, simbolis „3“ atitinka frazę „trys“, simbolis „4“ atitinka frazę „keturi“, simbolis „5“ atitinka frazę „penki“, simbolis „6“ atitinka frazę „šeši“, simbolis „7“ atitinka frazę „septyni“, simbolis „8“ atitinka frazę „aštuoni“, simbolis „9“ atitinka frazę „devyni“, simbolis „29“ atitinka frazę „du, devyni“, simbolis „29158“ atitinka frazę „du, devyni, vienas, penki, aštuoni“ ir simbolis „0123456789“ atitinka frazę „nulis, vienas, du, trys, keturi, penki, šeši, septyni, aštuoni, devyni“. Metodų trumpinių lentelė pateikta priede 8.1 lentelėje. Sistema yra apmokoma ir testuojama su vienodomis frazėmis. Garso įrašų paskirstymui pagal ketvirtą atvejį labiausiai tinkama frazė yra „0123456789“ – gaunamas didžiausias vidutinis identifikavimo tikslumas visiems metodams – 98,67%, mažiausiai tinkama frazė yra „6“ – gaunamas minimalus vidutinis tikslumas – 88,44 %. Tiksliausiai veikiantis metodas MFCC21-ANN, kuriam gaunamas maksimalus vidutinis tikslumas

siekia 96 %. Netiksliausiai veikiantis metodas – MFCC39-ANN, gautas vidutinis tikslumas – 89.23 %.

Minimalios ir maksimalios identifikavimo tikslumo vertės skirtingoms frazėms įrašų paskirstymo pagal pirmą ir antrą atvejį pateiktos 4.1 lentelėje. Visos identifikavimo tikslumo vertės pirmo ir antro atvejo garso įrašų paskirstymui pateiktos 8.1 pav. ir 8.2 pav. priede. Pirmo atvejo garso įrašų paskirstymui tinkamiausia frazė „0123456789“, gautas vidutinis tikslumas visiems metodams – 82.31 %, prasčiausias vidutinis identifikavimo tikslumas gautas frazei „3“ – 65.54%. Vidutiniškai geriausiai suveikęs metodas MFCC21-NBC (vidutinis identifikavimo tikslumas visiems žodžiams 84.44%), prasčiausiai suveikęs metodas MFCC39-ANN (vidutinis identifikavimo tikslumas visiems žodžiams 48.11%). Įrašų paskirstymui pagal antrą atvejį labiausiai tinkama frazė „0123456789“ (vidutinė tikslumo vertė visiems metodams 95.17%), minimali vidutinė tikslumo vertė visiems metodams – 77.52 % frazei „3“. Daugiausiai teisingų įrašų priskyrimų diktoriams gražinantis metodas – PLP20-NBC (vidutinis tikslumas visiems ištarimams 94.31 %). Daugiausiai neteisingų įrašų priskyrimų diktoriams gražinantis metodas yra MFCC39-ANN (vidutinis tikslumas visiems ištarimams 72 %).

Lentelė 4.1 Maksimalios ir minimalios identifikavimo tikslumo vertės įrašų paskirstymams pagal pirmą ir antrą atvejį

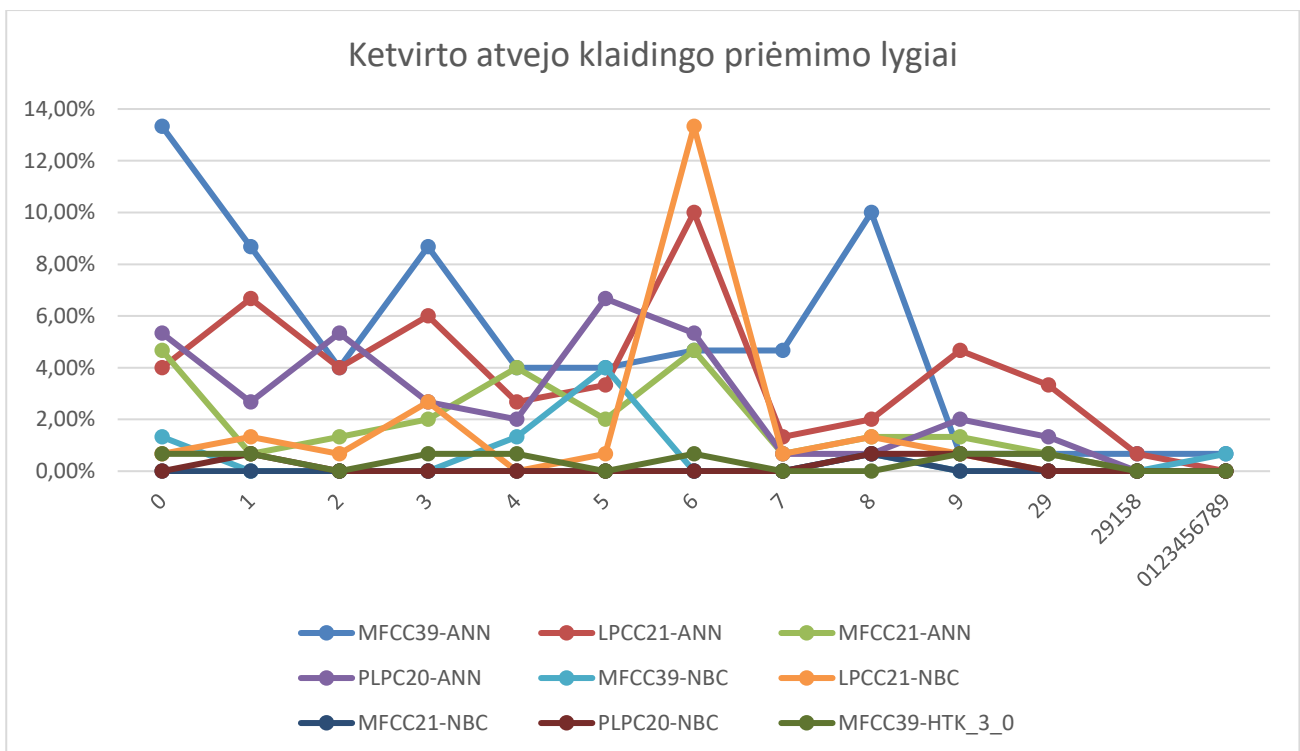
	Pirmas atvejis		Antras atvejis	
	Maksimali vertė	Minimali vertė	Maksimali vertė	Minimali vertė
MFCC39-ANN	60.00%	32.73%	92.57%	54.00%
LPCC21-ANN	84.24%	47.47%	97.14%	62.29%
MFCC21-ANN	84.85%	57.17%	99.14%	75.71%
PLPC20-ANN	90.71%	59.80%	96.29%	68.57%
MFCC39-NBC	88.69%	52.73%	97.43%	84.00%
LPCC21-NBC	89.49%	71.11%	97.14%	72.57%
MFCC21-NBC	90.10%	76.36%	97.14%	88.00%
PLPC20-NBC	88.89%	75.76%	99.14%	88.57%
MFCC39-HTK_3_0	88.89%	70.10%	96.00%	77.14%

Minimalios ir maksimalios identifikavimo tikslumo vertės skirtingoms frazėms įrašų paskirstymo pagal trečią ir ketvirtą atvejį pateiktos 4.2 lentelėje. Ketvirto atvejo vidutinės maksimalios ir minimalios vertės aprašytos 4.3.1 skyriaus pradžioje. Trečio atvejo garso įrašų paskirstymui geriausias vidutinis tikslumo rezultatas gautas frazei „0123456789“ (vidutinis visų metodų identifikavimo tikslumas frazei „0123456789“ - 92.30%), prasčiausias – frazei „6“ (vidutinis visų metodų identifikavimo tikslumas frazei „6“ – 78.62%). Didžiausias vidutinis tikslumas visoms frazėms 97.30%, metodui MFCC21-NBC. Mažiausias vidutinis tikslumas visoms frazėms 72.24 %, metodui LPCC21-ANN.

Lentelė 4.2 Maksimalios ir minimalios identifikavimo tikslumo vertės įrašų paskirstymams pagal trečią ir ketvirtą atvejį

	Trečias atvejis		Ketvirtas atvejis	
	Maksimali vertė	Minimali vertė	Maksimali vertė	Minimali vertė
MFCC39-ANN	87.56%	55.11%	98.67%	79.33%
LPCC21-ANN	86.67%	53.78%	100.00%	76.00%
MFCC21-ANN	95.56%	67.11%	100.00%	90.67%
PLPC20-ANN	95.56%	62.67%	100.00%	87.33%
MFCC39-NBC	98.22%	85.33%	98.67%	87.33%
LPCC21-NBC	99.11%	74.22%	98.67%	80.00%
MFCC21-NBC	99.56%	94.67%	98.67%	92.00%
PLPC20-NBC	99.56%	95.11%	99.33%	91.33%
MFCC39-HTK_3_0	95.56%	66.22%	97.33%	83.33%

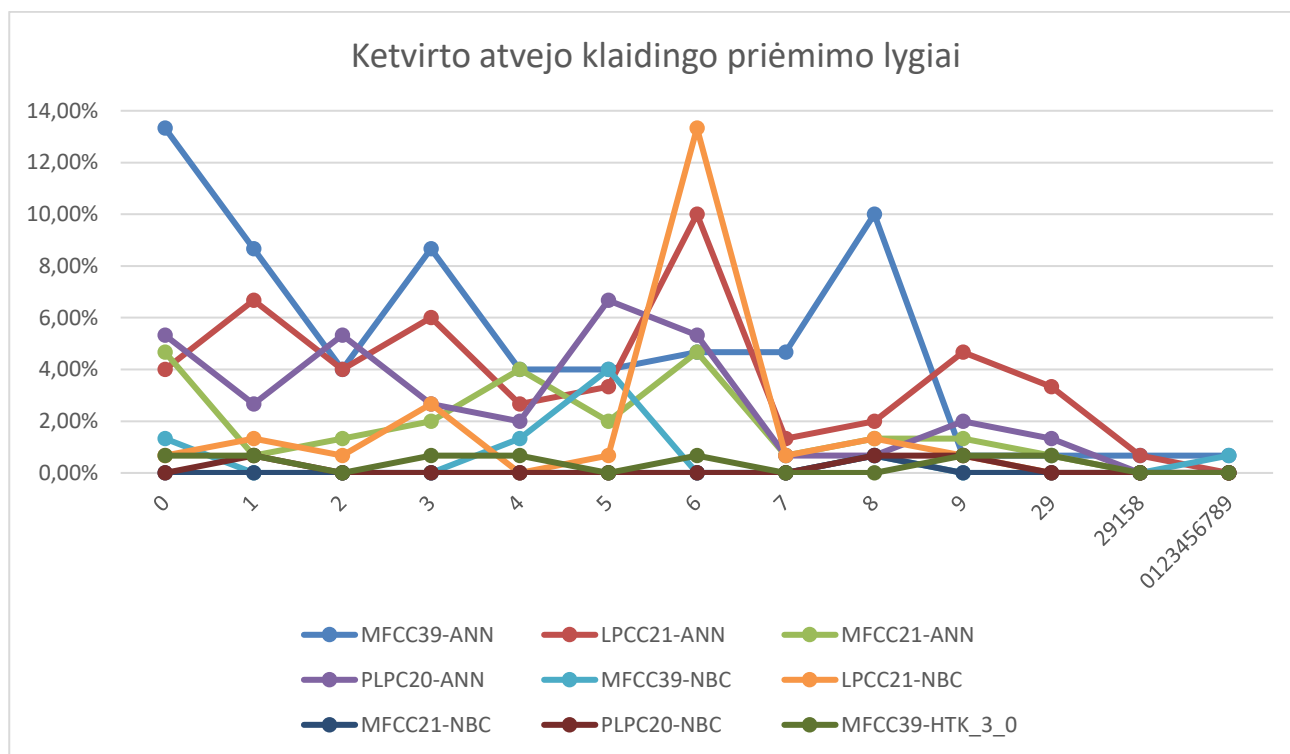
Apibendrinant identifikavimo tikslumo priklausomybę nuo skirtingų tirtų frazių, galima teigti, kad tinkamiausia frazė yra „0123456789“ nepriklausomai nuo garso įrašų skirtų mokymui ir testavimui skaičiaus. Frazė „3“ duoda prasčiausius identifikavimo tikslumo rezultatus, kai mokymo duomenų ir testavimo duomenų santykis yra mažesnis nei 50%. Frazė „6“ duoda prasčiausius rezultatus kai mokymo ir testavimo duomenų santykis yra didesnis nei 50%.



4.3 pav. Metodų klaidingo priėmimo lygiai ketvirto atvejo garso įrašo paskirstymui

4.3 pav. vertikaloje ašyje pateikti klaidingo priėmimo lygiai įrašų paskirstymui pagal ketvirtą atvejį, horizontalioje – skirtingi ištarimai, skirtingiems metodams. Įrašų paskirstymai pagal pirmą, antrą, trečią ir ketvirtą atvejį pateikti priede 7.1 ir 7.2 lentelėse. Klaidingo priėmimo lygis apibūdina kiek įrašų, kurie turėjo būti atmesti kaip sistemoje neegzistuojantys diktoriai, testavimo metu buvo priskirti sistemoje egzistuojantiems diktoriais. Didžiausias vidutinis klaidingo priėmimo lygis visiems metodams gautas frazei „6“ (4.30%), mažiausias vidutinis klaidingo priėmimo lygis visiems metodams

gautas frazėms „29158“ ir „0123456789“ (0.15%). Didžiausias vidutinis klaidingo priėmimo lygis visoms frazėms gautas metodui MFCC39-ANN (4.98%). Mažiausias klaidingo priėmimo lygis visoms frazėms gautas metodui MFCC21-NBC (0.05%).



4.4 pav. Metodų klaidingo priėmimo lygiai ketvirto atvejo garso įrašo paskirstymui

4.3 lentelėje pateiktos pirmo ir antro garso įrašų paskirstymo atvejų minimalios ir maksimalios klaidingų priėmimų lygių vertės skirtingiems metodams. Visos pirmo ir antro garso įrašų paskirstymo atvejo klaidingų priėmimo lygių vertės skirtingoms frazėms ir metodams pateiktos priede 8.4 pav. ir 8.5 pav. Didžiausia klaidingo priėmimo lygio vidutinė vertė pirmam garso įrašų paskirstymo atvejui, visiems metodams gauta frazei „3“ (7.86%), įrašų paskirstymui pagal antrą atvejį – frazei “3” (5%) ir “5” (5%). Mažiausia klaidingo priėmimo lygio vidutinė vertė pirmam garso įrašų paskirstymo atvejui, visiems metodams gauta (1.21%) frazei „0123456789“, įrašų paskirstymui pagal antrą atvejį mažiausia klaidingo priėmimo lygio vidutinė vertė gauta (0%) frazėms „29158“ ir „0123456789“. Didžiausia vidutinė klaidingo priėmimo lygio vertė visoms frazėms garso įrašo paskirstymui pagal pirmą atvejį gauta MFCC39-ANN metodui (19.27%), mažiausia – MFCC39-HTK_3_0 (0.26%). Mažiausia vidutinė klaidingo priėmimo lygio vertė visoms frazėms garso įrašo paskirstymui pagal antrą atvejį gauta MFCC39-ANN metodui (8%), mažiausia 0% – metodams MFCC21-NBC, PLP20-NBC ir MFCC39-HTK_3_0.

Lentelė 4.3 Maksimalios ir minimalios klaidingo priėmimo lygio vertės įrašų paskirstymams pagal pirmą ir antrą atvejį

	Pirmas atvejis		Antras atvejis	
	Maksimali vertė	Minimali vertė	Maksimali vertė	Minimali vertė
MFCC39-ANN	30.91%	8.28%	13.43%	1.71%
LPCC21-ANN	16.97%	1.21%	17.43%	0.29%
MFCC21-ANN	7.68%	0.00%	4.00%	0.00%
PLPC20-ANN	9.70%	0.00%	5.14%	0.00%
MFCC39-ANN	10.10%	0.00%	3.14%	0.00%
LPCC21-ANN	13.33%	0.00%	14.00%	0.00%
MFCC21-ANN	3.43%	0.00%	0.57%	0.00%
PLPC20-ANN	1.01%	0.00%	2.00%	0.00%
MFCC39-HTK_3_0	0.81%	0.00%	0.86%	0.00%

4.4 lentelėje pateiktos trečio ir ketvirto garso įrašų paskirstymo atvejų minimalios ir maksimalios klaidingų priėmių lygių vertės skirtingiems metodams. Visos trečio garso įrašų paskirstymo atvejo klaidingų priėmimo lygių vertės skirtingoms frazėms ir metodams pateiktos priede 8.6 pav. Didžiausia vidutinė klaidingo priėmimo vertė visiems metodams garso įrašo paskirstymui pagal trečią atvejį gauta frazei „0“ 4.30%, mažiausia – frazei „0123456789“ (0.15%). Didžiausia vidutinė klaidingo priėmimo vertė visoms frazėms garso įrašų paskirstymui pagal trečią atvejį gauta 4.07% – metodui MFCC39-ANN. Mažiausia vidutinė klaidingo priėmimo lygio vertė visoms frazėms garso įrašų paskirstymui pagal trečią atvejį gauta 0.03% – metodui MFCC21-NBC.

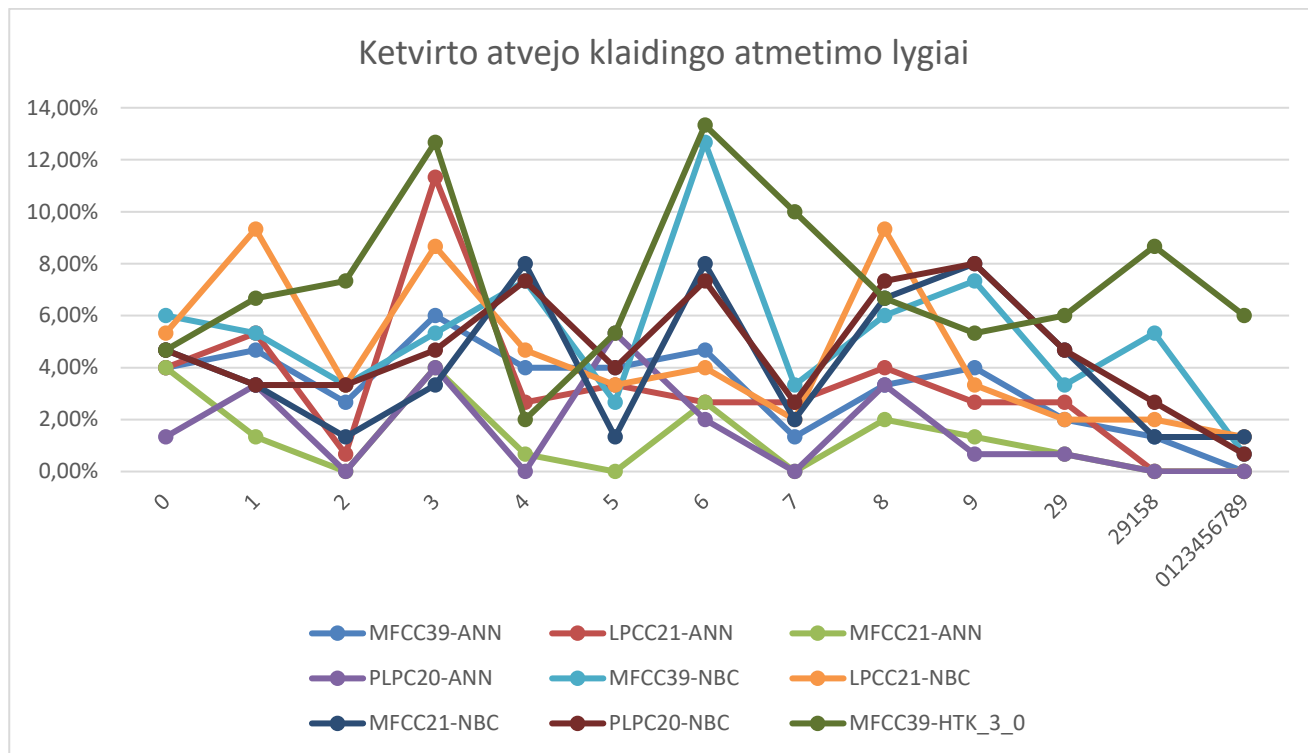
Lentelė 4.4 Maksimalios ir minimalios klaidingo priėmimo lygio vertės įrašų paskirstymams pagal trečią ir ketvirtą atvejį

	Trečias atvejis		Ketvirtas atvejis	
	Maksimali vertė	Minimali vertė	Maksimali vertė	Minimali vertė
MFCC39-ANN	8.44%	0.00%	13.33%	0.67%
LPCC21-ANN	8.00%	0.00%	10.00%	0.00%
MFCC21-ANN	3.11%	0.00%	4.67%	0.00%
PLPC20-ANN	3.11%	0.00%	6.67%	0.00%
MFCC39-ANN	3.56%	0.00%	4.00%	0.00%
LPCC21-ANN	8.89%	0.00%	13.33%	0.00%
MFCC21-ANN	0.44%	0.00%	0.67%	0.00%
PLPC20-ANN	0.44%	0.00%	0.67%	0.00%
MFCC39-HTK_3_0	0.89%	0.00%	0.67%	0.00%

Apibendrinant klaidingo priėmimo lygio vertes galima teigti kad visais atvejais, nepriklausomai nuo apmokymo duomenų kiekio metodas MFCC39-ANN daugiausiai diktorių nepriklausančių sistemai sumaišytų su diktoriais, priklausančiais sistemai.

4.4 pav. vertikalioje ašyje pateikti klaidingo atmetimo lygiai įrašų paskirstymui pagal ketvirtą atvejį, horizontalioje – skirtingi ištarimai, skirtingiems metodams. Įrašų paskirstymai pagal pirmą, antrą, trečią ir ketvirtą atvejį pateikti priede 7.1 ir 7.2 lentelėse. Klaidingo atmetimo lygis apibūdina kiek įrašų, kurie turėjo būti priimti kaip sistemoje egzistuojantys diktoriai, testavimo metu buvo atmesti, kaip diktoriai nepriklausantys sistemai. Didžiausias vidutinis klaidingo atmetimo lygis visiems metodams gautas frazei „3“ (6.67%), mažiausias vidutinis klaidingo atmetimo lygis visiems metodams

gautas frazei „0123456789“ (1.11%). Didžiausias vidutinis klaidingo atmetimo lygis visoms frazėms gautas metodui MFCC39-HTK_3_0 (7.28%). Mažiausias klaidingo priėmimo lygis visoms frazėms gautas metodui MFCC21-ANN (1.28%).



4.5 pav. Klaidingo atmetimo lygiai ketvirto atvejo garso įrašų paskirstymui

4.5 lentelėje pateiktos pirmo ir antro garso įrašų paskirstymo atvejų minimalios ir maksimalios klaidingų atmetimų lygių vertės skirtingiems metodams. Visos pirmo ir antro garso įrašų paskirstymo atvejų klaidingų atmetimų lygių vertės skirtingoms frazėms ir metodams pateiktos priede 8.7 pav. ir 8.8 pav. Didžiausia klaidingo atmetimo lygio vidutinė vertė pirmam garso įrašų paskirstymo atvejui, visiems metodams gauta frazėms „8“ ir „9“ (12%), įrašų paskirstymui pagal antrą atvejį – frazei „1“, „3“ ir „6“ (7%). Mažiausia klaidingo atmetimo lygio vidutinė vertė pirmam garso įrašų paskirstymo atvejui, visiems metodams gauta (7%) frazei „0123456789“, įrašų paskirstymui pagal antrą atvejį mažiausia klaidingo atmetimo lygio vidutinė vertė gauta (2%) frazei „0123456789“. Didžiausia vidutinė klaidingo atmetimo lygio vertė visoms frazėms garso įrašų paskirstymui pagal pirmą atvejį gauta MFCC39-HTK_3_0 metodui (19%), mažiausia – MFCC21-ANN metodui (0%). Didžiausia vidutinė klaidingo atmetimo lygio vertė visoms frazėms garso įrašų paskirstymui pagal antrą atvejį gauta MFCC39-HTK_3_0 metodui (12%), mažiausia 1% – metodams MFCC21-ANN ir PLPC20-ANN.

Lentelė 4.5 Klaidingo atmetimo lygių maksimalios ir minimalios vertės įrašų paskirstymams pagal pirmą ir antrą atvejį

	Pirmas atvejis		Antras atvejis	
	Maksimali vertė	Minimali vertė	Maksimali vertė	Minimali vertė
MFCC39-ANN	3.84%	0.40%	6.00%	0.57%
LPCC21-ANN	3.03%	0.00%	5.71%	0.29%
MFCC21-ANN	2.42%	0.00%	2.29%	0.00%
PLPC20-ANN	0.81%	0.00%	3.71%	0.00%
MFCC39-ANN	27.88%	10.91%	13.43%	2.29%
LPCC21-ANN	23.43%	7.07%	11.14%	2.57%
MFCC21-ANN	20.40%	9.90%	12.00%	2.86%
PLPC20-ANN	24.04%	11.11%	9.71%	0.86%
MFCC39-HTK_3_0	26.46%	10.10%	22.29%	3.71%

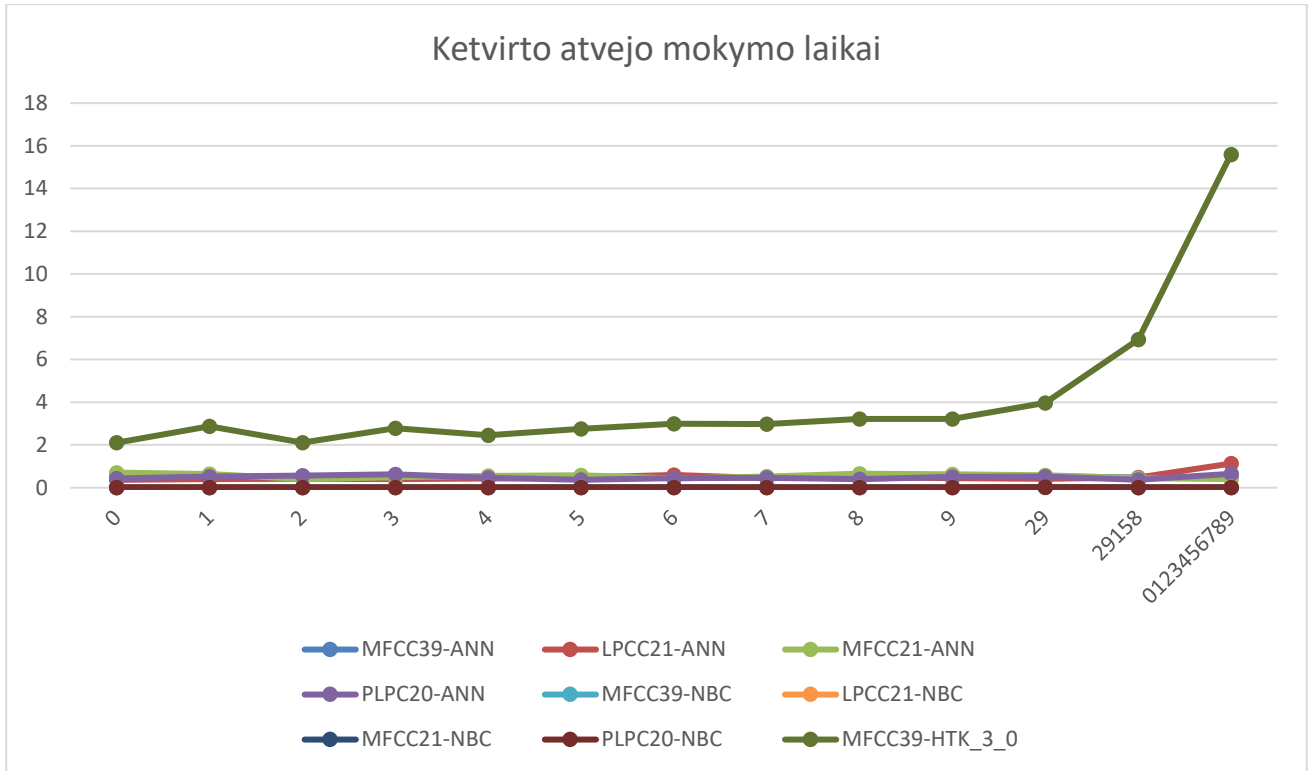
4.6 lentelėje pateiktos trečio ir ketvirto garso įrašų paskirstymo atvejų minimalios ir maksimalios klaidingų atmetimų lygių vertės skirtingiems metodams. Visos trečio garso įrašų paskirstymo atvejo klaidingų atmetimų lygių vertės skirtingoms frazėms ir metodams pateiktos priede 8.9 pav. Didžiausia vidutinė klaidingo atmetimo lygio vertė visiems metodams garso įrašo paskirstymui pagal trečią atvejį gauta frazei „3“ 6.12%, mažiausia – frazei „0123456789“ (1.04%). Didžiausia vidutinė klaidingo atmetimo lygio vertė visoms frazėms garso įrašų paskirstymui pagal trečią atvejį gauta 17.98% – metodui MFCC39-HTK_3_0. Mažiausia vidutinė klaidingo atmetimo lygio vertė visoms frazėms garso įrašų paskirstymui pagal trečią atvejį gauta 0.44% – metodui PLPC20-ANN.

Lentelė 4.6 Klaidingo atmetimo lygių maksimalios ir minimalios vertės įrašų paskirstymams pagal trečią ir ketvirtą atvejį

	Trečias atvejis		Ketvirtas atvejis	
	Maksimali vertė	Minimali vertė	Maksimali vertė	Minimali vertė
MFCC39-ANN	3.11%	0.00%	6.00%	0.00%
LPCC21-ANN	4.00%	0.00%	11.33%	0.00%
MFCC21-ANN	1.33%	0.00%	4.00%	0.00%
PLPC20-ANN	1.33%	0.00%	5.33%	0.00%
MFCC39-ANN	9.33%	1.78%	12.67%	0.67%
LPCC21-ANN	5.33%	0.44%	9.33%	1.33%
MFCC21-ANN	5.33%	0.44%	8.00%	1.33%
PLPC20-ANN	4.89%	0.44%	8.00%	0.67%
MFCC39-HTK_3_0	33.78%	4.44%	13.33%	2.00%

Pagal klaidingo atmetimo lygio rezultatus galima teigti kad daugiausiai diktorių, kurie turėtų būti identifikuoti yra atmesti naudojant MFCC39-HTK_3_0 metodą. 3 būsenos buvo imtos kaip minimalus galimų būsenų skaičius, neatsižvelgiant į žodį. 4.3.3 skyriuje pateikti rezultatai keičiant būsenų skaičių pagal fonemų (raidžių) skaičių žodžiuose.

4.5 pav. vertikalioje ašyje pateikti mokymo laikai sekundėmis, horizontalioje frazės. Iš grafiko matyti, kad daugiausiai laiko sueikvojantis metodas yra paslėptųjų Markovo modelių metodas. Maksimali vertė pasiekama ištarimui “0123456789”. Vidutinis mokymo laikas visiems ištarimams paslėptųjų Markovo modelių paketui – 4.1566s, Bajeso klasifikatoriui (visiems požymių tipams ir ištarimams) – 0.0173s, dirbtiniams neuronų tinklams – 0.5074s.



4.6 pav. Mokymo laikai ketvirto atvejo garso įrašų paskirstymui

4.7 ir 4.8 lentelėse pateiktos maksimalios ir minimalios mokymo laikų vertės visiems metodams. Priede 8.10 pav., 8.11 pav. ir 8.12 pav. pateikti mokymo laikai pirmam, antram ir trečiam garso įrašų paskirstymo atvejui. Galima teigti, kad mokymo laikas paslėptųjų Markovo modelių paketui didėja didėjant mokymo duomenų skaičiui. Vidutinis mokymo laikas garso įrašų paskirstymui pagal pirmą atvejį 2.1271s, pagal antrą – 2.4505s, trečią - 3.2835s ir ketvirtą – 4.1566s.

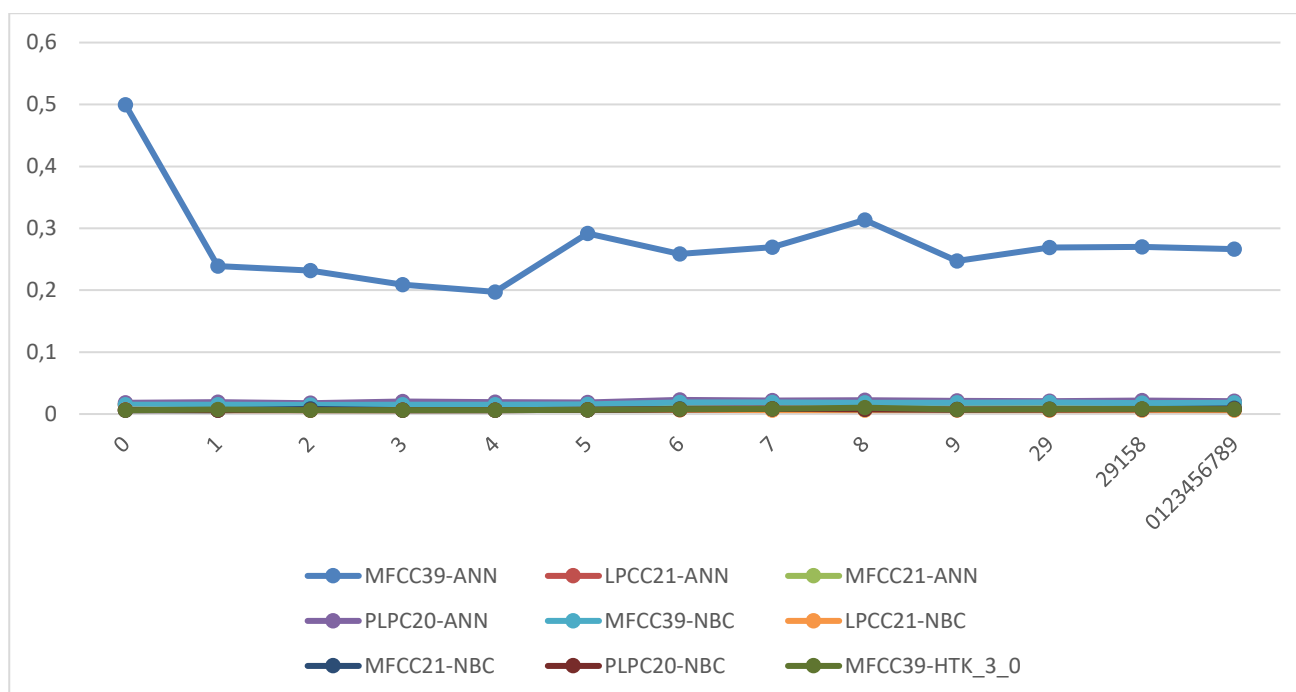
Lentelė 4.7 Maksimalios ir minimalios mokymo laikų vertės įrašų paskirstymams pagal pirmą ir antrą atvejį

	Pirmas atvejis		Antras atvejis	
	Maksimali vertė	Minimali vertė	Maksimali vertė	Minimali vertė
MFCC39-ANN	1.0721	0.2675	0.6558	0.373
LPCC21-ANN	0.7509	0.2377	0.6193	0.3197
MFCC21-ANN	2.3879	0.2832	2.722	0.2785
PLPC20-ANN	1.3761	0.2497	1.3056	0.3409
MFCC39-ANN	0.3573	0.0157	0.0241	0.0157
LPCC21-ANN	0.0621	0.0123	0.0203	0.0134
MFCC21-ANN	0.0181	0.0129	0.0207	0.0134
PLPC20-ANN	0.0204	0.013	0.0231	0.0144
MFCC39-HTK_3_0	4.734	1.6528	7.1464	1.5538

Lentelė 4.8 Maksimalios ir miniamlaivos mokymo laikų vertės įrašų paskirstymams pagal trečią ir ketvirtą atvejį

	Trečias atvejis		Ketvirtas atvejis	
	Maksimali vertė	Minimali vertė	Maksimali vertė	Minimali vertė
MFCC39-ANN	0.5532	0.3866	0.544	0.4179
LPCC21-ANN	0.681	0.2914	1.1289	0.3689
MFCC21-ANN	0.6975	0.3559	0.7045	0.4037
PLPC20-ANN	0.5584	0.3269	0.6499	0.3692
MFCC39-ANN	0.0198	0.0154	0.021	0.0156
LPCC21-ANN	0.0202	0.0129	0.0356	0.0132
MFCC21-ANN	0.0173	0.0131	0.0213	0.0131
PLPC20-ANN	0.0177	0.0137	0.0206	0.0135
MFCC39-HTK_3_0	11.6156	1.9841	15.5999	2.1056

4.7 pav. pateiktas vieno įrašo testavimo laikas sekundėmis skirtingiems metodams su pradinėmis reikšmėmis. Maksimali vidutinė įrašo testavimo laiko trukmė – 0.274 s MFCC39-ANN metodui, minimali vidutinė įrašo testavimo laiko trukmė – 0.0065s LPCC21-NBC metodui.

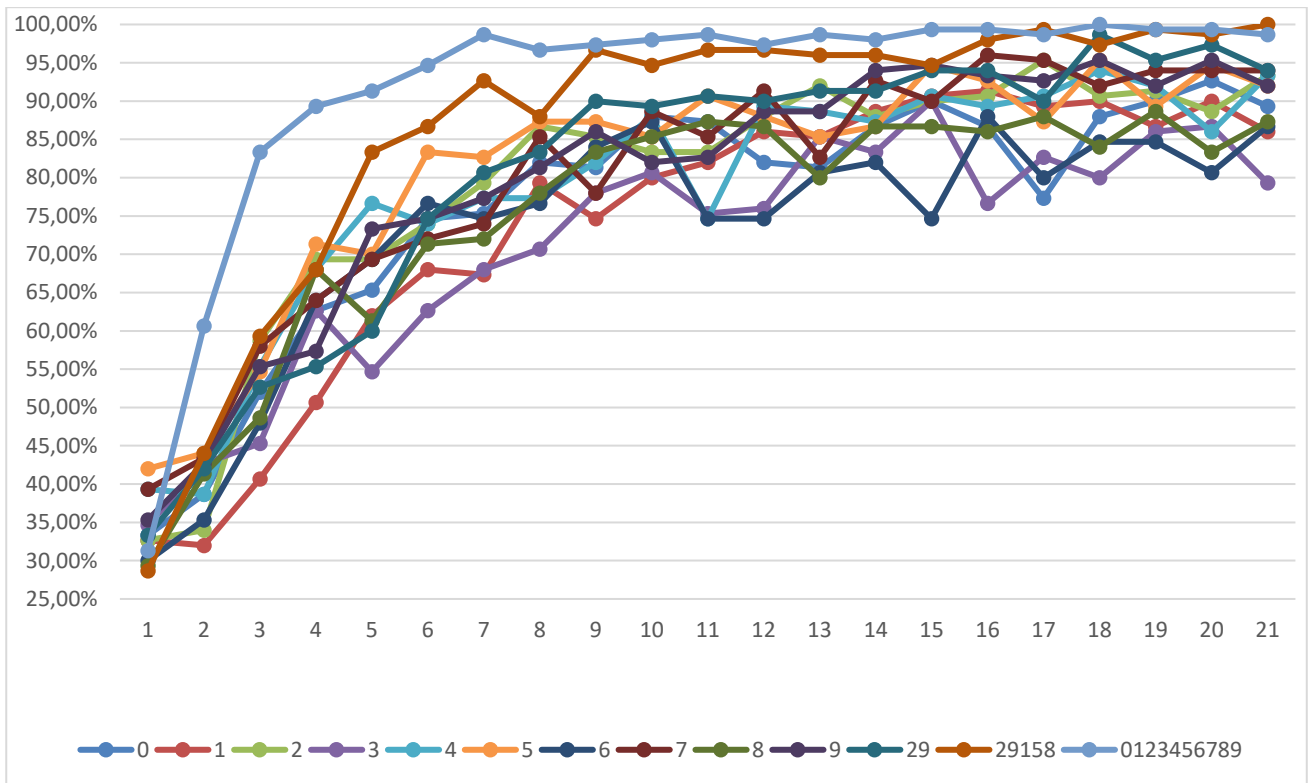


4.7 pav. Vieno įrašo testavimo laikas (sekundėmis) garso įrašų paskirstymui pagal ketvirtą atvejį

Vidutinis identifikavimo tikslumas su pradinėmis reikšmėmis visiems metodams ir ištarimams garso įrašų paskirstymui pagal pirmą atvejį – 72.21%, pagal antrą atvejį – 86.24 %, pagal trečią atvejį 85.62%, pagal ketvirtą atvejį – 93.22%.

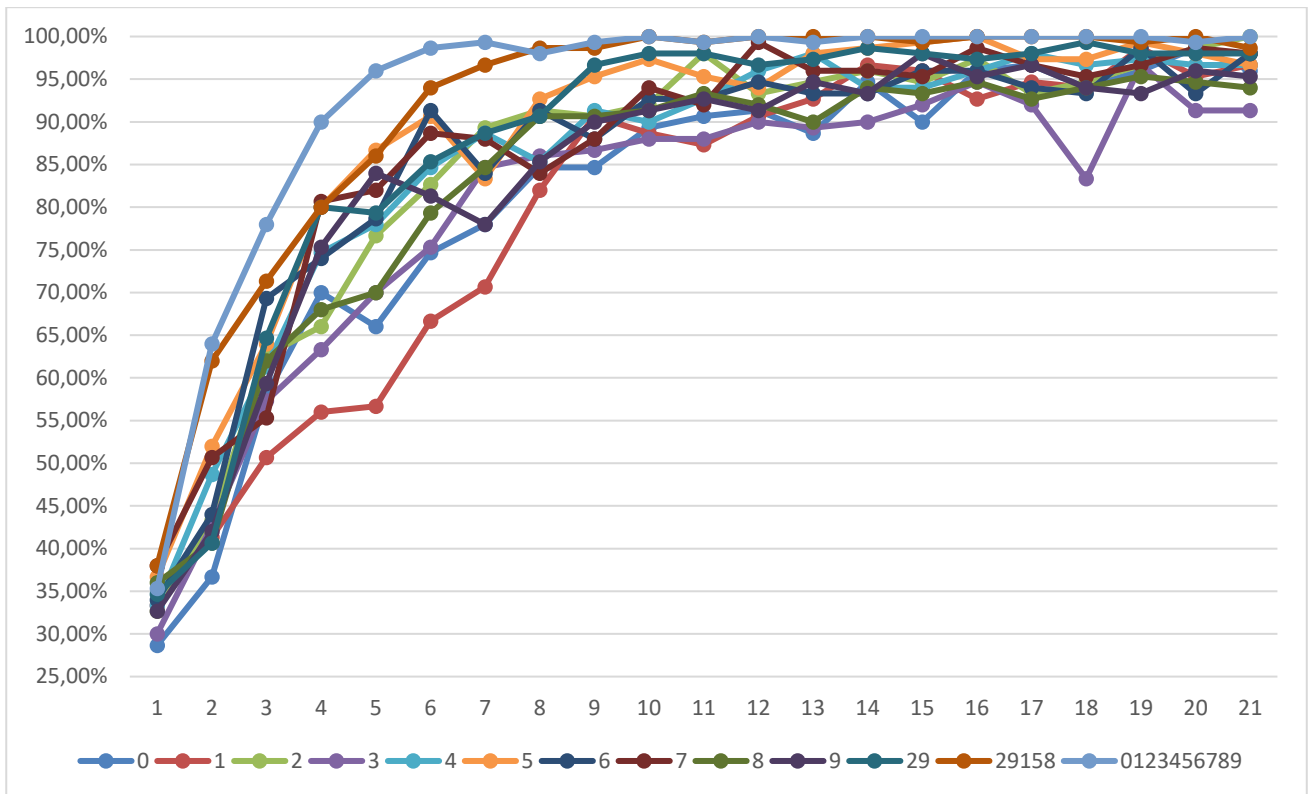
4.3.2 Rezultatai skirtingiems požymių koeficientų skaičiams

Identifikavimo tikslumų priklausomybė nuo požymių skaičiaus tiesinės prognozės kepstro koeficientų skaičiaus dirbtiniams neuronų tinklams, skirtingiems ištarimams, įrašų paskirstymui pagal ketvirtą atvejį pateikta 4.8 pav. Matoma kad didesniai nei 50 % tikslumui pasiekti pakanka 5 tiesinės prognozės koeficientų. Su visais koeficientų skaičiais, išskyrus 17 ir 21 didžiausias identifikavimo tikslumas gaunamas frazei “0123456789”.



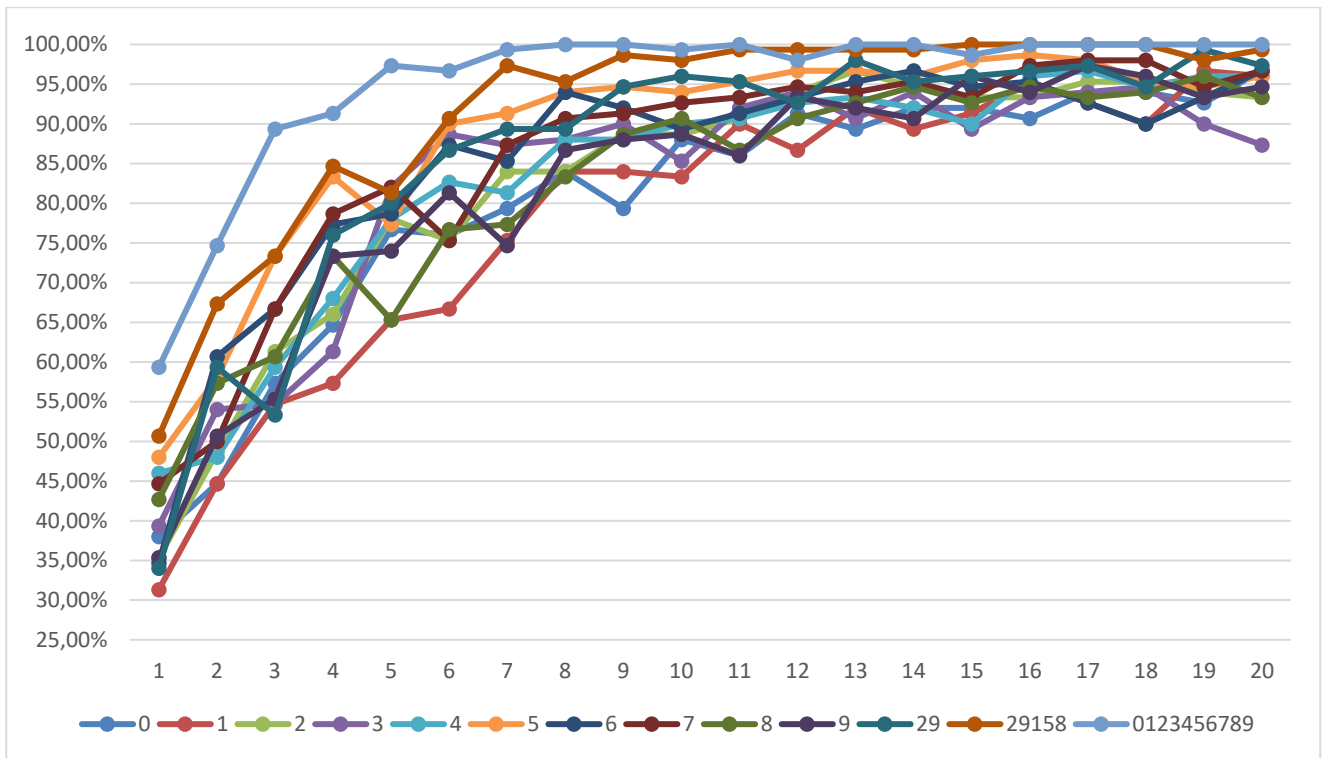
4.8 pav. Tiesinės prognozės kepstro koeficientų ir dirbtinių neuronų tinklų (LPCC-ANN) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį

Identifikavimo tikslumų priklausomybė nuo Mel skalės kepstro koeficientų skaičiaus dirbtiniams neuronų tinklams, skirtingiems ištarimams, įrašų pasiskirstymui pagal ketvirtą atvejį pateikta 4.9 pav. Matoma kad didesniai nei 50 % tikslumui pasiekti pakanka 4 Mel skalės kepstro koeficientų. Daugumoje atvejų geriausias rezultatas gaunamas frazei “0123456789”.



4.9 pav. Mel skalės kepsro koeficientų ir dirbtinių neuronų tinklų (MFCC-ANN) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį

Identifikavimo tikslumų priklausomybė nuo suvokimo tiesinės prognozės kepsro koeficientų skaičiaus dirbtiniams neuronų tinklams, skirtingiems ištarimams, įrašų paskirstymui pagal ketvirtą atvejį pateikta 4.10 pav. Matoma kad didesniai nei 50 % tikslumui pasiekti pakanka 3 suvokimo tiesinės prognozės kepsro koeficientų. Daugumoje atvejų geriausias rezultatas gaunamas frazei “0123456789”.



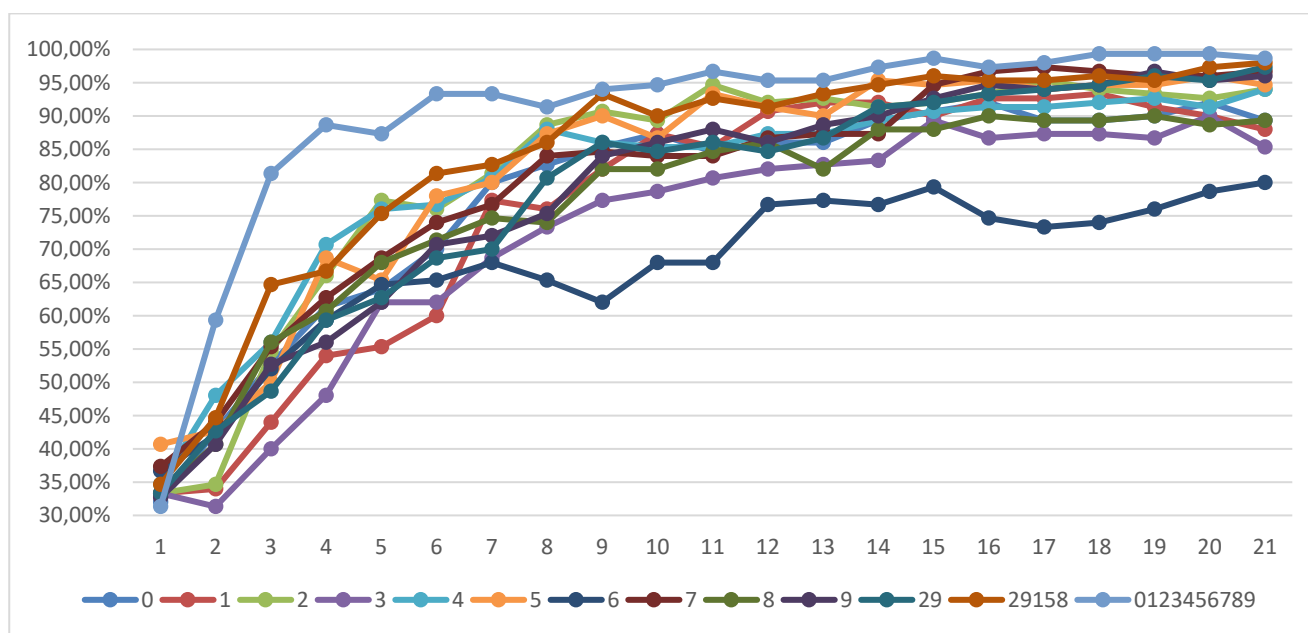
4.10 pav. Suvokimo tiesinės prognozės ir dirbtinių neuronų tinklų (PLPC-ANN) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį

Maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams su dirbtiniais neuronų tinklais (ANN) ketvirto atvejo garso įrašų paskirstymui pateiktos 4.9 lentelėje. Iš minėtos lentelės galima matyti, kad frazei „29158“ reikia 21 tiesinės prognozės kepstro koeficiento pasiekti, 10 Mel skalės kepstro koeficientų ir 15 suvokimo tiesinės prognozės koeficientų pasiekti 100% tikslumui. Visoms kitoms frazėms, išskyrus „0123456789“ pasirinktų koeficientų skaičių nepakanka pasiekti 100% tikslumui.

Lentelė 4.9 Maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams su dirbtiniais neuronų tinklais (ANN)

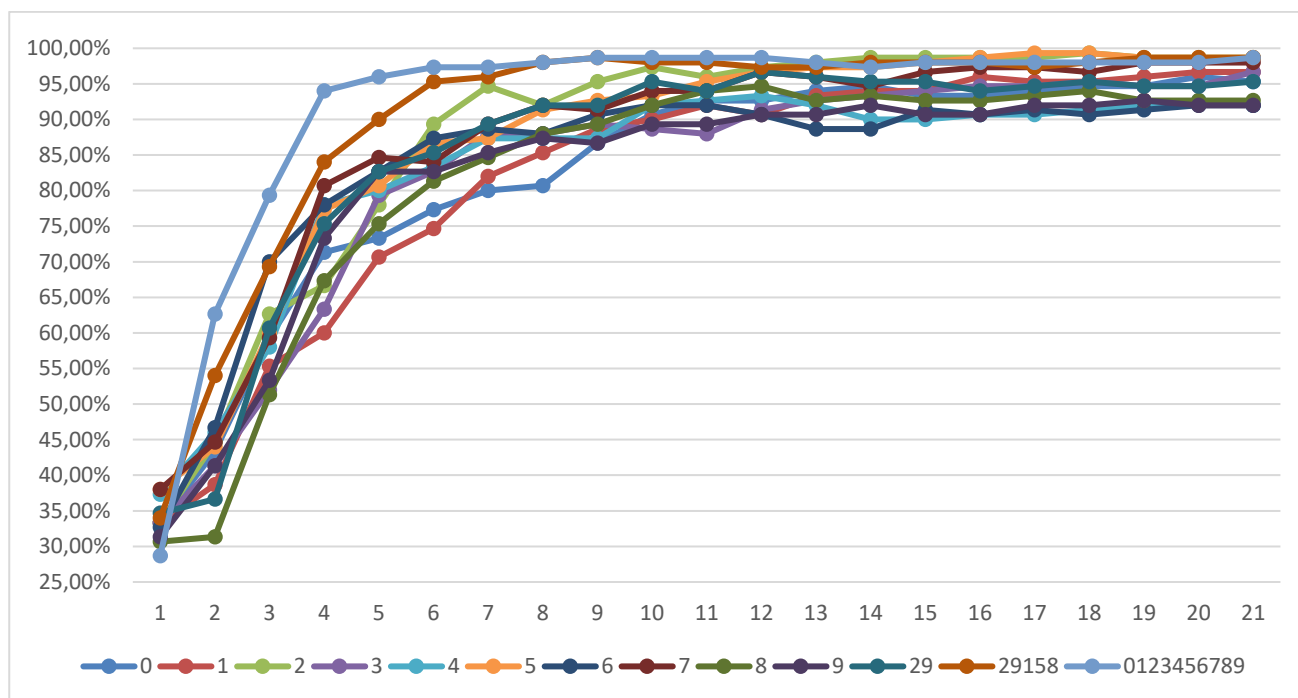
Ištarimas	LPCC-ANN		MFCC-ANN		PLPC-ANN	
	Koeficientų skaičius	Maksimali vertė	Koeficientų skaičius	Maksimali vertė	Koeficientų skaičius	Maksimali vertė
0	20	92.67%	20	96.67%	20	96.67%
1	16	91.33%	19	96.67%	19	96.67%
2	17	95.33%	21	96.67%	13	96.67%
3	15	90.00%	19	94.67%	18	94.67%
4	18	94.00%	13;17	96.67%	17	96.67%
5	18	95.33%	16	98.67%	16	98.67%
6	16	88.00%	19	96.67%	14;20	96.67%
7	16	96.00%	12	98.00%	17;18	98.00%
8	19	88.67%	19	96.00%	19	96.00%
9	18; 20	95.33%	15	97.33%	17	97.33%
29	18	98.67%	18	98.67%	19	99.33%
29158	21	100.00%	10;12;13;14; 16;17;18;20	100.00%	15;16;17;18	100.00%
0123456789	18	100.00%	10;12;14;15; 16;17;18;19; 21;21	100.00%	8;9;11;13;14; 16;17;18;19; 20	100.00%

Tiesinės prognozės kepstro koeficientų ir Bajeso klasifikatoriaus (LPCC-NBC) identifikavimo tikslumai skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį pateikti 4.11 pav. Matoma kad didesniai nei 50 % tikslumui pasiekti pakanka 4 tiesinės prognozės kepstro koeficientų. Daugumoje atvejų geriausias rezultatas gaunamas frazei “0123456789”.



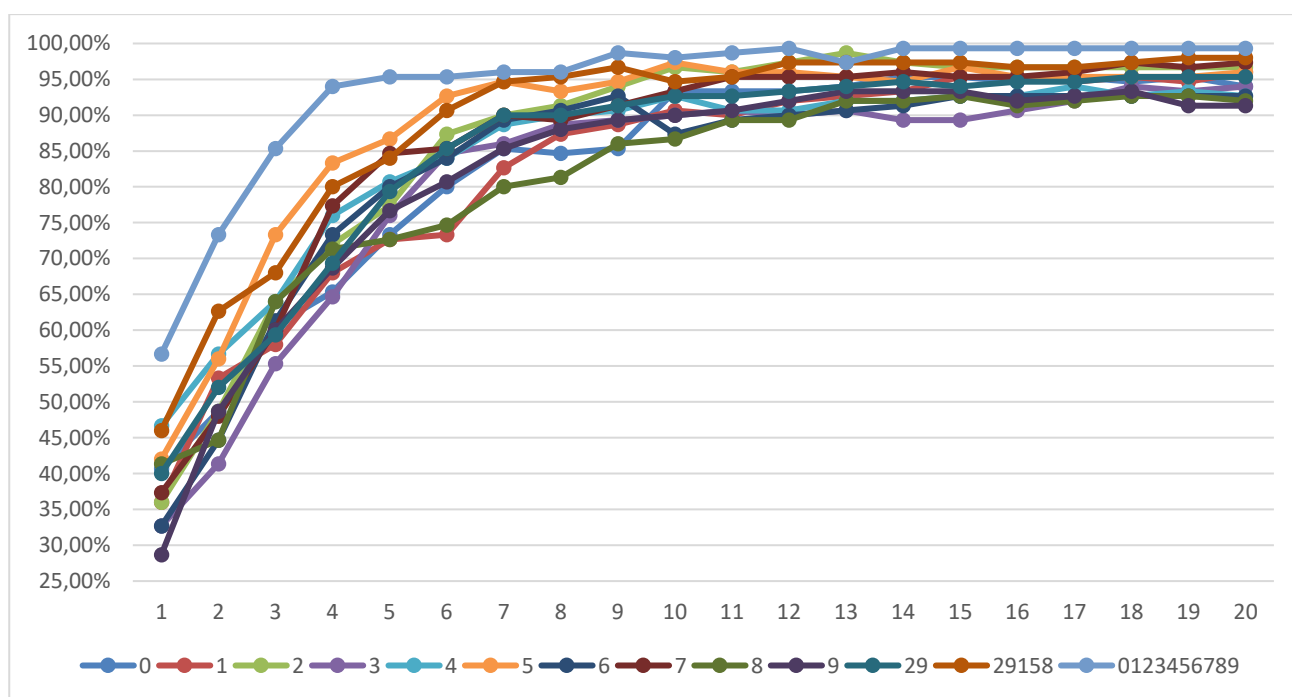
4.11 pav. Tiesinės prognozės kepstro koeficientų ir Bajeso klasifikatoriaus (LPCC-NBC) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį

Mel skalės kepstro koeficientų ir Bajeso klasifikatoriaus (MFCC-NBC) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį 4.12 pav. Matoma kad didesniai nei 50 % tikslumui pasiekti pakanka 3 Mel skalės kepstro koeficientų.



4.12 pav. Mel skalės kepstro koeficientų ir Bajeso klasifikatoriaus (MFCC-NBC) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį

Suvokimo tiesinės prognozės ir Bajeso klasifikatoriaus (PLPC-NBC) identifikavimo tikslumai skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį pateikti 4.13 pav. Matoma kad didesniai nei 50 % tikslumui pasiekti pakanka 3 suvokimo tiesinės prognozės koeficientų.



4.13 pav. Suvokimo tiesinės prognozės ir Bajeso klasifikatoriaus (PLPC-NBC) identifikavimo tikslumas skirtingam koeficientų skaičiui, įrašų pasiskirstymui pagal ketvirtą atvejį

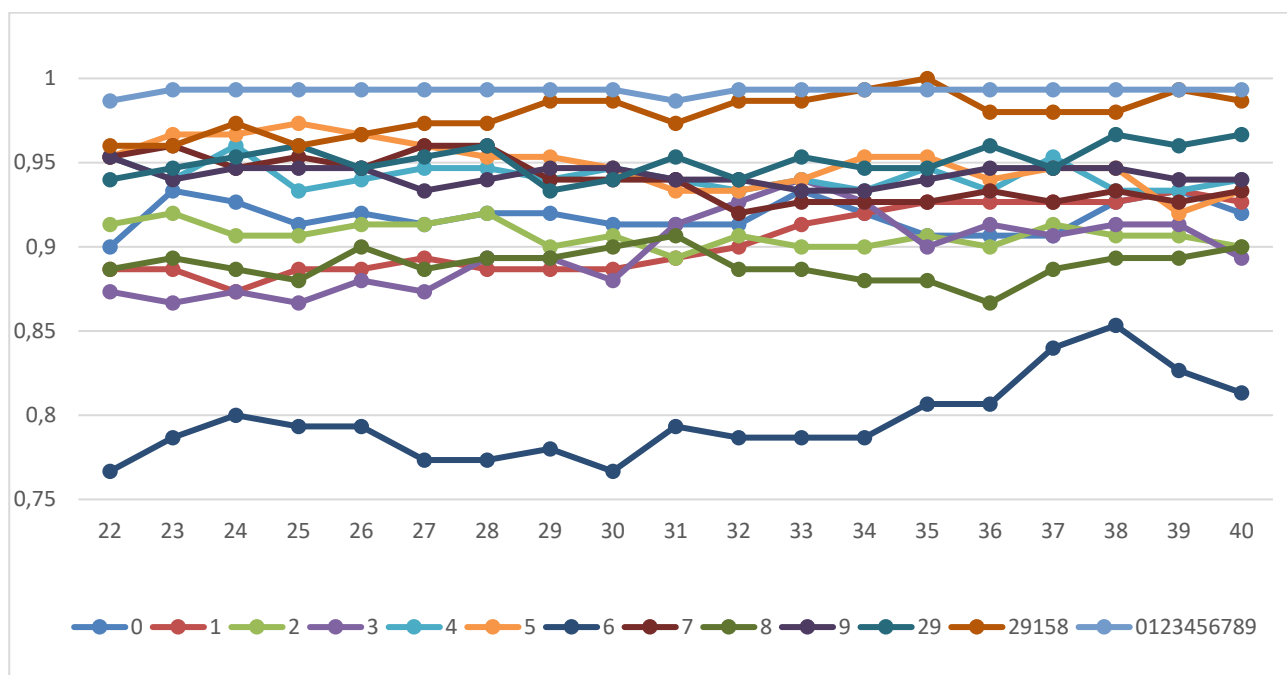
Maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams su Bajeso kladifikatoriumi (NBC) pateiktos 4.10 lentelėje. Lentelėje galima matyti, kad didžiausias tikslumas pasiekiamas su fraze „0123456789“, 18 tiesinės prognozės koeficientų ir 12 tiesinės suvokimo prognozės koeficientų. Su 9 Mel skalės kepstro koeficientų maksimalus rezultatas 98.67% frazėms „29158“ ir „0123456789“. Būtų galima spėti, kad koeficientų skaičiaus parinkimas priklauso nuo ištartos frazės tačiau taip nėra. 9.1 ir 9.2 lentelėse priede pateiktos maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams su dirbtiniais neuronų tinklais (ANN) ir Maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams (nuo 1 iki 20 arba nuo 1 iki 21) su Bajeso kladifikatoriumi (NBC) garso įrašų paskirstymui pagal antrą atvejį. Pavyzdžiui 9.1 lentelėje minimalus reikalingas tiesinės prognozės koeficientų skaičius reikalingas gauti maksimaliam identifikavimo tikslumui „0123456789“ yra 15, 4.9 lentelėje – 18.

Lentelė 4.10 Maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams (nuo 1 iki 20 arba nuo 1 iki 21) su Bajeso kladifikatoriumi (NBC)

Ištarimas	LPCC-NBC		MFCC-NBC		PLPC-NBC	
	Koeficientų skaičius	Maksimali vertė	Koeficientų skaičius	Maksimali vertė	Koeficientų skaičius	Maksimali vertė
0	16;20	92.00%	20	96.00%	14;15;17;19	95.33%
1	18	93.33%	20;21	96.67%	20	96.00%
2	17	95.33%	18	99.33%	13	98.67%
3	20	90.00%	21	96.67%	18; 20	94.00%
4	21	94.00%	12	93.33%	17	94.00%
5	20	96.00%	17; 18	99.33%	10	97.33%
6	21	80.00%	10;11;20;21	92.00%	9;15;16;17;18;19;20	92.67%
7	17	97.33%	19;20;21	98.00%	18; 20	97.33%
8	16;19	90.00%	12	94.67%	15;18;19	92.67%
9	19	96.67%	19	92.67%	13;14;15;18	93.33%
29	21	97.33%	12	96.67%	18;19;20	95.33%
29158	21	98.00%	9;19;20;21	98.67%	19;20	98.00%
0123456789	18; 19; 20	99.33%	9;10;11;12;21	98.67%	12;14;15;16;17;18;19;20	99.33%

Maksimali identifikavimo tikslumo vertė 100% Bajeso klasifikatoriui pasiekama ištarimu „29158“ ir 35 tiesinės prognozės kepstro koeficientais. 4.14 pav. pateiktos tiesinės prognozės kepstro koeficientų ir Bajeso klasifikatoriaus (LPCC-NBC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams, kai koeficientų skaičiai imami nuo 22 iki 40, garso įrašų paskirstymui pagal

ketvirtą atvejį. Kitų – suvokimo tiesinės prognozės koeficientų skaičiaus didinimas virš 20 ir Mel skalės kepstro koeficientų skaičiaus didinimas virš 21 negalimas dėl pasirinktų funkcijų realizacijos.



4.14 pav. Tiesinės prognozės kepstro koeficientų ir Bajeso klasifikatoriaus (LPCC-NBC) identifikavimo tikslumas koeficientų skaičiui nuo 22 iki 40, garso įrašų paskirstymui pagal ketvirtą atvejį

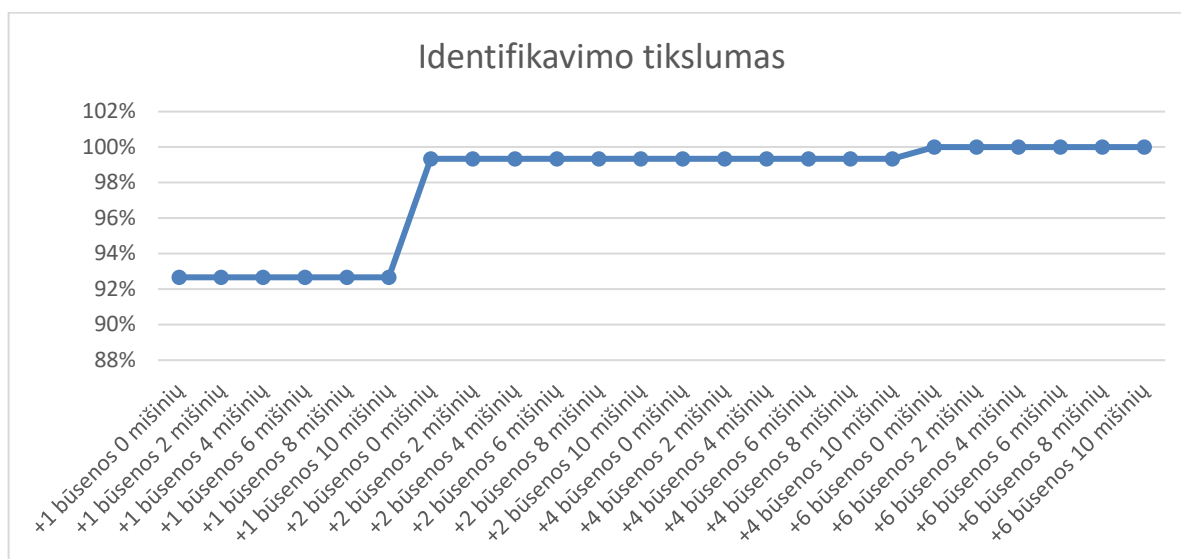
4.3.3 Paslėptųjų Markovo modelių paketo rezultatai

Paslėptųjų Markovo modelių paketų tyrimui naudojamas žodžiais grįstas modeliais metodais. Naudojamų būsenų skaičius parenkamas priklausomai nuo žodžio (raidžių žodyje skaičio). 4.11 lentelėje pateiktos „+2 būsenos“, „+4 būsenos“ ir „+6 būsenos“ paaiškinimai.

Lentelė 4.11 Būsenų skaičius skirtingoms frazėms

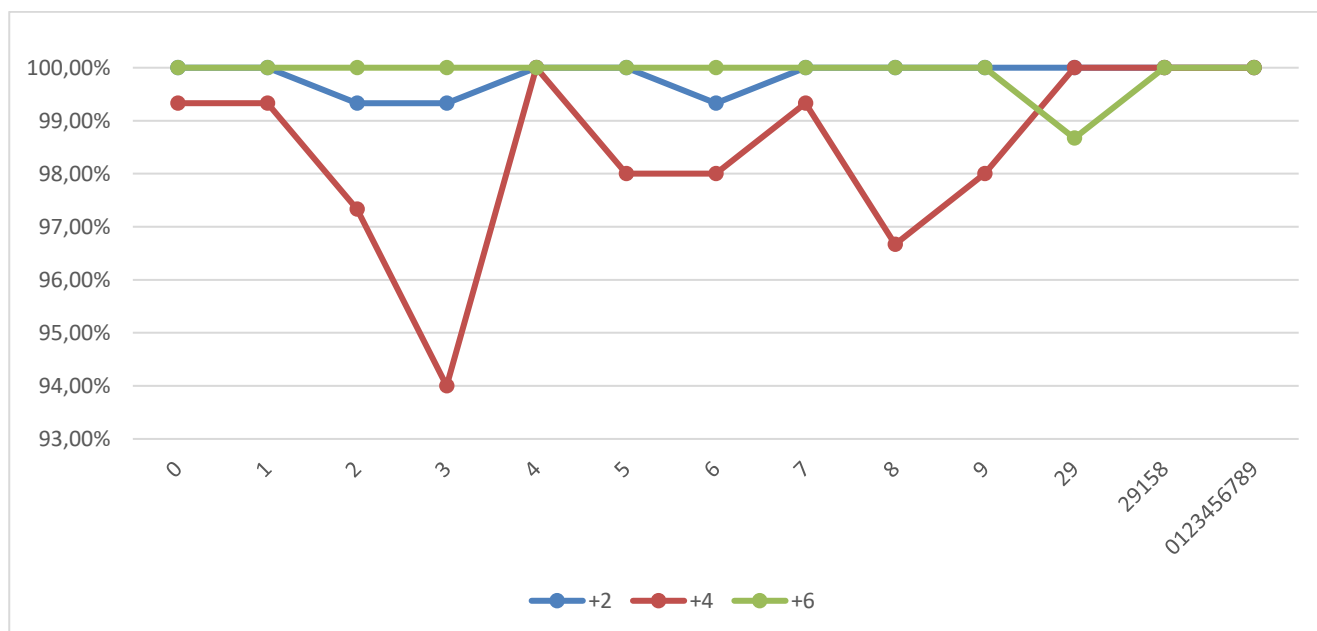
Ištarimas	Fonemų (raidžių) skaičius	+2 būsenos	+4 būsenos	+6 būsenos
Nulis	5	7	9	11
Vienas	6	8	10	12
Du	2	4	6	8
Trys	4	6	8	10
Keturi	6	8	10	12
Penki	5	7	9	11
Šeši	4	6	8	10
Septyni	7	9	11	13
Aštuoni	7	9	11	13
Devyni	6	8	10	12
Du devyni	8	10	12	14
Du devyni vienas penki aštuoni	26	28	30	32
Nulis vienas du trys keturi penki šeši septyni aštuoni devyni	52	54	56	58

Identifikavimo tikslumo priklausomybė nuo skirtingų būsenų ir mišinių kiekio frazei „du“ pateikta 4.15 pav. Iš grafiko matyti, kad gauso mišinių skaičius identifikavimo tikslumui įtakos neturi.



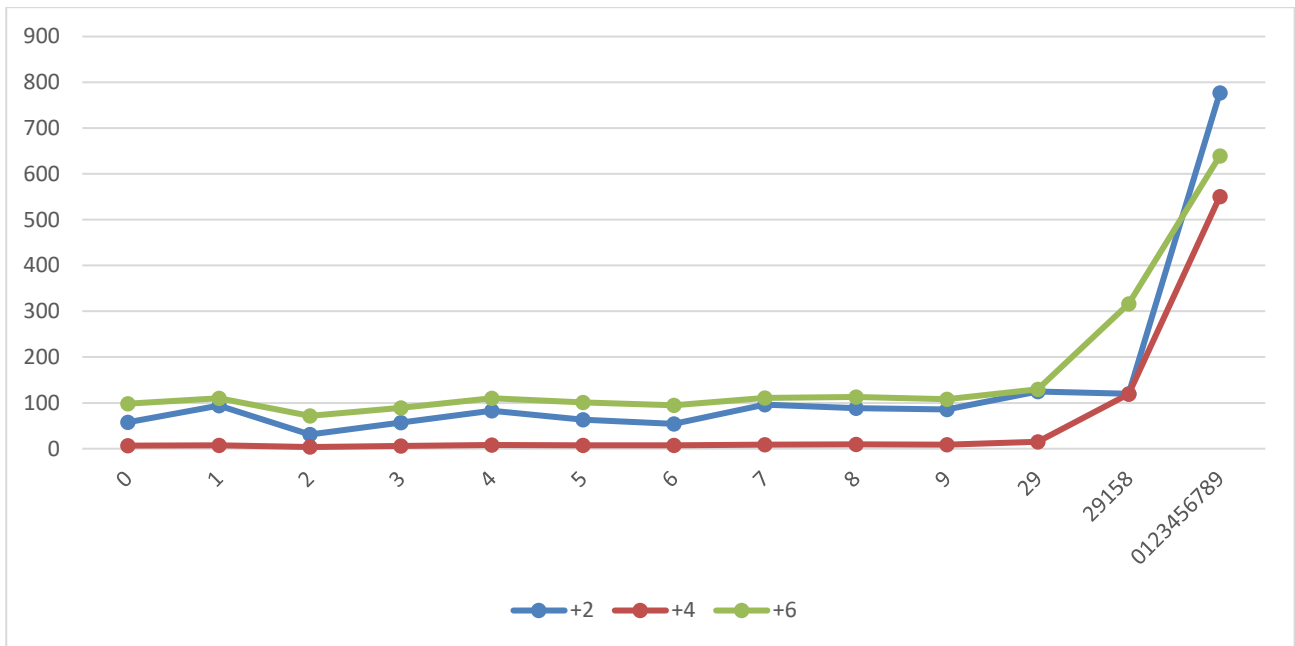
4.15 pav. Ketvirto atvejo rezultatai skirtingoms būsenoms ir mišiniams

4.16 pav. pateiktas ketvirto atvejo identifikavimo tikslumas skirtingam būsenų skaičiui. Geriausias identifikavimo tikslumas visiems ištarimams gaunamas su „+6 būsenom“. Išimtis frazei – „du, devyni“ aukštesni identifikavimo tikslumai gauti „+2“ ir „+4“ būsenom. Priede 10.1 pav. pateikti identifikavimo tikslumai skirtingoms frazėms ir būsenų skaičiui. Iš 10.1 pav matyti, kad 100% tikslumą (ne visiems ištarimams) galima pasiekti ir 17.5% įrašų prisikiriant mokymo įrašams, 82.5 % testavimo įrašams.



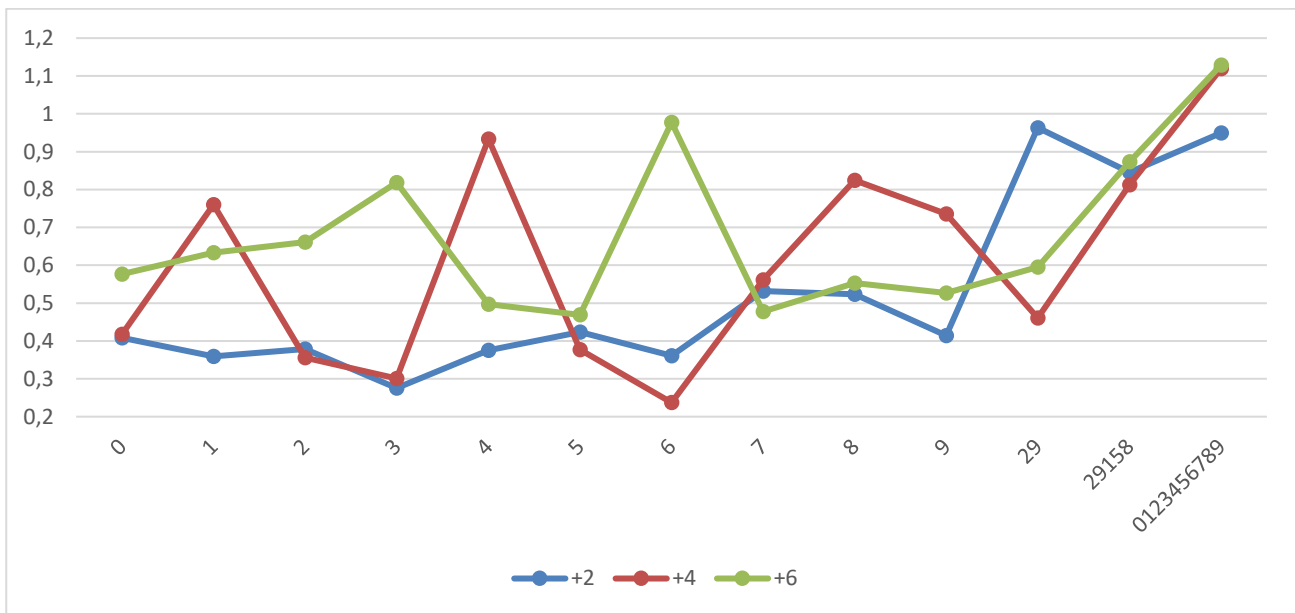
4.16 pav. Ketvirto atvejo identifikavimo tikslumas skirtingoms būsenoms

4.17 pav. pateiktas mokymo laikas (sekundėmis) „+2 būsenoms“, „+4 būsenoms“ ir „+6 būsenoms“. Maksimali vidutinė mokymo laiko vertė gauta „+6 būsenoms“ 160.963s, minimali vidutinė mokymo laiko vertė gauta „+4 būsenoms“ – 58.315s.



4.17 pav. Ketvirto atvejo mokymo laikas (sekundėmis) skirtingam būsenų skaičiui

4.18 pav. pateikti vieno įrašo testavimo laikai sekundėmis, įrašų paskirstymui pagal ketvirtą atvejį. Maksimali vidutinė testavimo laiko vertė vienam įrašui gauta “+6 būsenoms” – 0.67s, minimali vidutinė testavimo laiko vertė gauta “+2 būsenoms” – 0.52s.



4.18 pav. Įrašo testavimo laikai (sekundėmis) skirtingiems būsenų skaičiams įrašų paskirstymui pagal ketvirtą atvejį

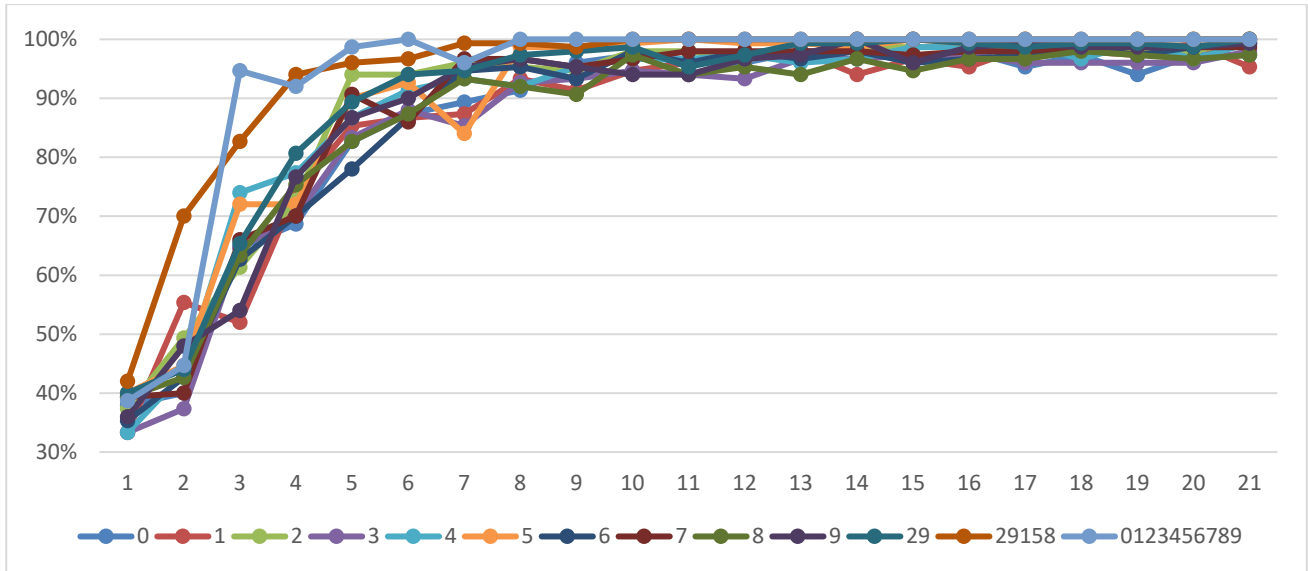
4.3.4 Rezultatai skirtingoms dirbtinio neuronų tinklo architektūroms

Neuronų skaičiui paslėptajame sluoksnyje nustatyti yra daugelis metodų tokių kaip nykščio (ang. *rule-of-thumb*) [32] :

- Neuronų skaičius paslėptajame sluoksnyje turėtų būti ne didesnis už įvesties ir išvesties sluoksnius

- Neuronų skaičius paslėptajame sluoksnyje turėtų būti $2/3$ įvesties sluoksnio dydžio pridėjus išvesties sluoksnio dydį
- Paslėptų neuronų skaičius turėtų būti mažesnis nei dvigubas skaičius neuronų įvesties sluoksnyje.

4.19 pav. pateiktas identifikavimo tikslumas 19-kai Mel skalės kepstro koeficientų skirtingam neuronų skaičiui paslėptajame sluoksnyje. Maksimali vidutinio tikslumo vertė gauta su 21 neuronu.



4.19 pav. Identifikavimo tikslumas su 19 Mel skalės kepstro koeficientų, skirtingam neuronų skaičiui paslėptajame sluoksnyje, garso įrašų paskirstymui pagal ketvirtą atvejį

5 IŠVADOS

1. Ištyrus diktoriaus identifikavimo tikslumo priklausomybę nuo požymių tipo ir požymių skaičiaus nustatyta, kad didžiausią identifikavimo tikslumą galima pasiekti su 19 Mel skalės kepstro koeficientų (MFCC), naudojant dirbtinius neuronų tinklus (gautas vidutinis identifikavimo tikslumas, kai 75% duomenų skirta mokymui, 25% testavimui – 97.38%).
2. Analizuojant diktoriaus identifikavimo tikslumo priklausomybę nuo klasifikatoriaus tipo nustatyta, kad 100 % identifikavimo tikslumą frazei „0123456789“ užtikrina dirbtiniai neuronų tinklais su 8 neuronais paslėptajame sluoksnyje. Bajeso klasifikatoriui 100% tikslumas gautas tik vienu atveju - su 35 tiesinės prognozės kepstro koeficientais naudojant frazę „29158“. Bajeso klasifikatoriaus mokymo ir įrašo testavimo laikai yra nežymiai trumpesni nei dirbtinių neuronų tinklų.
3. Išnagrinėjus diktoriaus identifikavimo tikslumo priklausomybę nuo paslėptųjų Markovo modelių (PMM) būsenų ir Gauso mišinių skaičiaus pastebėta, kad su HTK paketu naudojant žodžiais grįstus PMM pasiekiamas didžiausias diktoriaus identifikavimo tikslumas lyginant su Bajeso ir neuroninių tinklų klasifikatoriais: 100 % identifikavimo tikslumas gaunamas net trumpoms frazėms, kai būsenų skaičius viršija fonemų skaičių nuo 2 iki 6.

6 LITERATŪRA

1. Biometrics Metrics Report v3.0. 2012. [interaktyvus][žiūrėta 2015 balandžio 30] . Prieiga per internetą: <http://www.usma.edu/ietd/docs/BiometricsMetricsReport.pdf>
2. Kamarauskas J. Asmens atpažinimas pagal balsą. 2009. [interaktyvus]. [žiūrėta 2015 birželio 12]. Prieiga per internetą: http://donelaitis.vdu.lt/disertacijos/Kamarauskas_2009.pdf
3. Bolle, R.M.; Connell, J.H. ; Pankanti, S. ; Ratha, N.K. ; Senior, A.W. The relation between the ROC curve and the CMC. *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*, 17-18 Oct. 2005, p. 15 - 20. [interaktyvus][žiūrėta 2015 gegužės 06] . Prieiga per internetą: <http://ieeexplore.ieee.org/xpl/abstractAuthors.jsp?tp=&arnumber=1544394&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F10364%2F32967%2F01544394.pdf%3Farnumber%3D1544394>
4. Challenges of Speaker Recognition Applications. [interaktyvus][žiūrėta 2015 gegužės 06]. Prieiga per internetą: <http://www.ncmmsc.org/upload%5C121107-Challenges%20of%20VPR%20Applications-Thomas.pdf>
5. Marinov, S. Text Dependent and Text Independent Speaker Verification Systems. *Technology and Applications*. 2003. [interaktyvus][žiūrėta 2015 rugsėjo 9] . Prieiga per internetą: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=DDEC1D3B682E5C4F15925F037E0015F6?doi=10.1.1.134.1529&rep=rep1&type=pdf>
6. Ngo, K. Digital signal processing algorithms for noise reduction, dynamic range compression, and feedback cancellation in hearing aids. 2011. [interaktyvus][žiūrėta 2015 balandžio 30]. Prieiga per internetą: ftp://ftp.esat.kuleuven.be/sista/kngo/reports/phd_KimNgo.pdf
7. Ramírez, J.; Górriz J.M; Segura, J.C. Voice Activity Detection. *Fundamentals and Speech Recognition System Robustness. INTECH, ISBN 987-3-90213-08-0, pp.460*. 2007. [interaktyvus][žiūrėta 2015 balandžio 30] . Prieiga per internetą: <http://cdn.intechopen.com/pdfs-wm/104.pdf>
8. Dawande, Y.; Dhopeswarkar, M. Analysis of different feature extraction techniques for speaker recognition system: A review. *ISSN No: 2250-3536, Issue 1*. 2015. [interaktyvus]. [žiūrėta 2015 birželio 12]. Prieiga per internetą: http://ijater.com/Files/e1417deb-a27c-4b60-ba5e-bb99c5fd3412_IJATER_20_02.pdf
9. Kinnunen, T.; Li, H. An Overview of Text-Independent Speaker Recognition: from Features to Supervectors. 2009. [interaktyvus][žiūrėta 2016 sausio 18] . Prieiga per internetą : http://www.cs.joensuu.fi/pages/tkinnu/webpage/pdf/speaker_recognition_overview.pdf
10. Muller, C. *Speaker Classification I*. 2007.
11. The fundamental of feature extraction in speaker recognition : a review. [interaktyvus]. [žiūrėta 2015 birželio 12]. Prieiga per internetą: <http://comp.utm.my/pars/files/2013/04/THE-FUNDAMENTAL-OF-FEATURE-EXTRACTION-IN-SPEAKER-RECOGNITION-A-REVIEW.pdf>
12. Pranevičius H.; Raudys Š.; Rudžionis A.; Rudžionis V.; Ratkevičius K.; Sakaluskaitė J.; Makackas D. *Agentinių sistemų modeliai*, 2008, ISBN: 9789955591559
13. Equal loudness filter. [interaktyvus]. [žiūrėta 2017 balandžio 02]. Prieiga per internetą: http://replaygain.hydrogenaud.io/proposal/equal_loudness.html

14. Features and techniques for speaker recognition. Singh, S.K. [interaktyvus]. [žiūrėta 2015 gegužės 13]. Prieiga per internetą: https://www.ee.iitb.ac.in/~esgroup/es_mtech03_sem/sem03_paper_03307409.pdf
15. Dennis, S. Introduction to Neural Networks. 1997. [interaktyvus]. [žiūrėta 2017 balandžio 02]. Prieiga per internetą: <http://staff.itee.uq.edu.au/janetw/cmc/chapters/Introduction/>
16. Stergiou C.; Siganos D. NEURAL NETWORKS. [interaktyvus][žiūrėta 2017 balandžio 21] http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Feed-forward%20networks
17. Dong, Y.; Li, D. Automatic speech recognition. A deep learning approach. 2015. ISBN 978-1-4471-5779-3
18. Matějka, P.; Glembek, O.; Novotný, O.; Plhot, O.; Grézl, F.; Burget, L.; Černocký, J. Analysis of DNN approaches to speaker identification. *ICASSP*. 2016. [interaktyvus]. [žiūrėta 2017 balandžio 02]. Prieiga per internetą: http://www.fit.vutbr.cz/research/groups/speech/publi/2016/matejka_icassp2016_0005100.pdf
19. Richardson, F.; Reynolds, D.; Dehak, N. Deep Neural Network Approaches to Speaker and Language Recognition. *IEEE SIGNAL PROCESSING LETTERS, VOL. 22, NO. 10*. 2015 [interaktyvus]. [žiūrėta 2017 balandžio 02]. Prieiga per internetą: https://groups.csail.mit.edu/sls/publications/2015/Dehak_IEEE-2015.pdf
20. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.; Kingsbury, B. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Process. Mag.*, pp. 82–97. 2012. [interaktyvus]. [žiūrėta 2017 balandžio 02]. Prieiga per internetą: <https://static.googleusercontent.com/media/research.google.com/lt/pubs/archive/38131.pdf>
21. Praveen, N. Text Dependent Speaker Recognition using MFCC features and BPANN. *International Journal of Computer Applications (0975 –8887)*, 2014. [interaktyvus][žiūrėta 2016 sausio 18] . Prieiga per internetą: <http://research.ijcaonline.org/volume74/number5/pxc3889788.pdf>
22. Šalna, B.; Kamarauskas, J. Evaluation of Effectiveness of Different Methods in Speaker Recognition. *ELECTRONICS AND ELECTRICAL ENGINEERING*. 2010. ISSN 1392 – 1215
23. Handaya, D.; Fakhuroja, H.; Muhammad, E.; Hidayat I.; Machbub, C. Comparison of Indonesian Speaker Recognition Using Vector Quantization and Hidden Markov Model for Unclear Pronunciation Problem. *IEEE 6th International Conference on System Engineering and Technology (ICSET)*. 2016
24. ALIZÉ open source speaker recognition. [interaktyvus]. [žiūrėta 2017 kovo 11]. Prieiga per internetą: <http://mistral.univ-avignon.fr/>
25. Khoury, E. and El Shafey, L. and Marcel, S. SPEAR: AN OPEN SOURCE TOOLBOX FOR SPEAKER RECOGNITION BASED ON BOB. *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. 2014.
26. HTK Speech recognition toolkit. [interaktyvus]. [žiūrėta 2017 kovo 02]. Prieiga per internetą: <http://htk.eng.cam.ac.uk/>
27. Neural Network Toolbox. [interaktyvus]. [žiūrėta 2017 kovo 02]. Prieiga per internetą : <https://se.mathworks.com/products/neural-network.html>
28. Statistics and Machine Learning Toolbox. [interaktyvus]. [žiūrėta 2017 kovo 02]. Prieiga per internetą : <https://se.mathworks.com/products/statistics.html>

29. The HTK Book. [interaktyvus]. [žiūrėta 2017 kovo 08]. Prieiga per internetą: http://speech.ee.ntu.edu.tw/homework/DSP_HW2-1/htkbook.pdf
30. VOICEBOX: Speech Processing Toolbox for MATLAB. [žiūrėta 2017 kovo 02]. Prieiga per internetą : <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>
31. PLP and RASTA (and MFCC, and inversion) in Matlab using melfcc.m and invmelfcc.m. [interaktyvus]. [žiūrėta 2017 kovo 08]. Prieiga per internetą: <http://www.ee.columbia.edu/ln/rosa/matlab/rastamat/>
32. Panchal, G. Ganatra, A., Kosta, Y., Pancha, D. Behaviour Analysis of Multilayer Perceptrons with Multiple Hidden Neurons and Hidden Layers. *ISSN: 1793-8201*. 2017. [interaktyvus]. [žiūrėta 2017 balandžio 02]. Prieiga per internetą: <http://ijcte.org/papers/328-L318.pdf>

7 PRIEDAS. Garso įrašų paskirstymo atvejų lentelė

Lentelė 7.1 Pirmo ir antro eksperimento atvejų garso įrašų pasiskirstymas

	Kalbėtojo vardas	Iš viso įrašų skirtingoms frazėms	1-mo atvejo mokymui skirtų įrašų skaičius	1-mo atvejo testavimui skirtų įrašų skaičius	2-o atvejo mokymui skirtų įrašų skaičius	2-o atvejo testavimui skirtų įrašų skaičius		
Moteriški balsai	FAGNGRA	20	5	15	10	10		
	FAGNVIN	20	5	15	10	10		
	FAISIZI	20	5	15	10	10		
	FAISZYM	20	5	15	10	10		
	FAUSNEM	20	5	15	10	10		
	FDAILOI	20	5	15	10	10		
	FGINGED	20	5	15	10	10		
	FIEVJUR	20	5	15	10	10		
	FIEVSAB	20	5	15	10	10		
	FIEVVIS	20	5	15	10	10		
	FJUSKIN	20	5	15	10	10		
	FKAMMOS	20	5	15	10	10		
	FLAUZET	20	5	15	10	10		
	FRAISAV	20	5	15	10	10		
	FRUTNAN	20	5	15	10	10		
	FSIMMEI	20	5	15	10	10		
	FUGNBUC	20	5	15	10	10		
	FUGNNOV	20	5	15	10	10		
	FVAIVAI	20	5	15	10	10		
	FVANPEC	20	5	15	10	10		
Vyriški balsai	FVILVAI	20	1	Atmestiniai klasei	19	10	Atmestiniai klasei	
	FVIONAB	20	1		19	10		
	FZIVSTA	20	1		19	10		
	MDAIGUS	20	1		19	10		
	MEDGVOL	20	1		19	10		
	MKAZANU	20	0		20	-		20
	MLINJUR	20	0		20	-		20
	MMODSLE	20	0		20	-		20
MRIMAPA	20	0	20	-	20			
MVYGVAI	20	0	20	-	20			

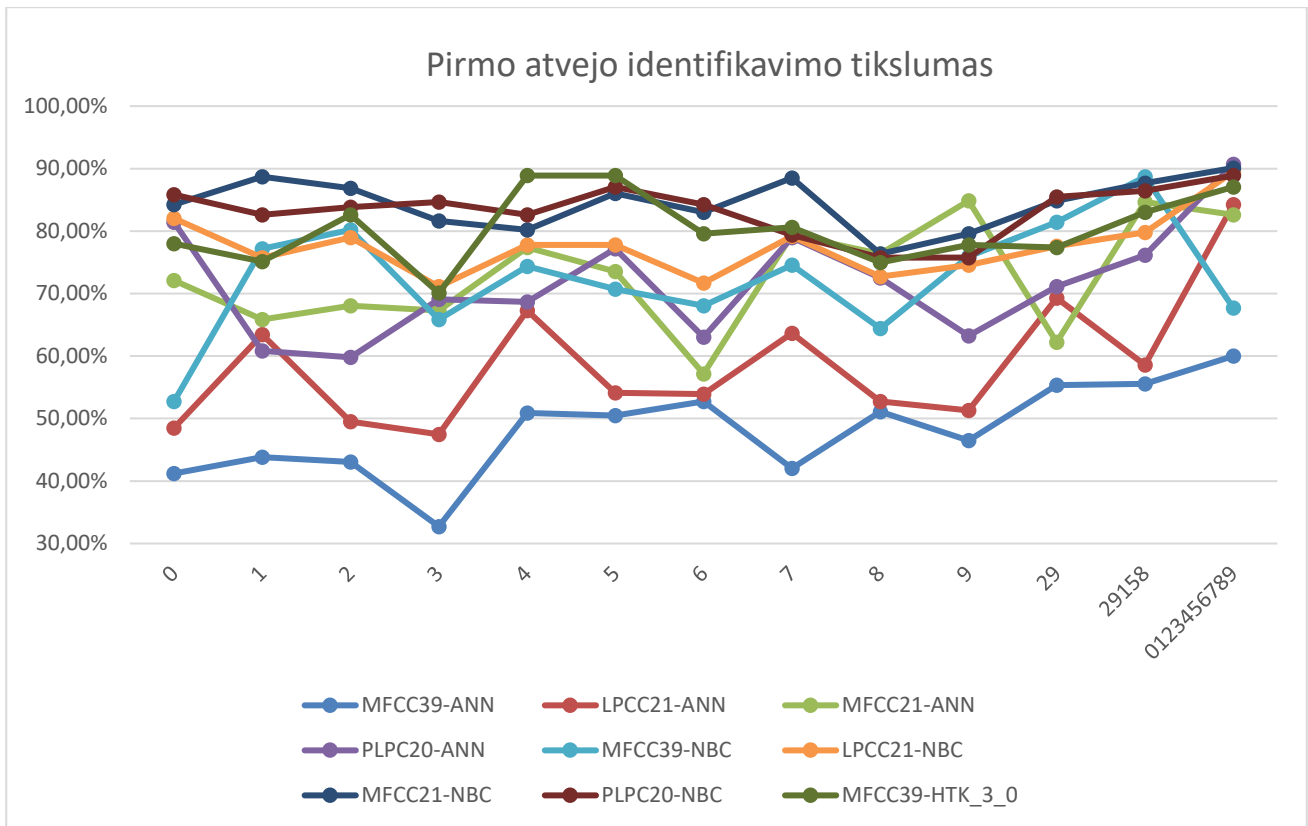
Lentelė 7.2 Trečio ir ketvirto eksperimento atvejų garso įrašų paskirstymas

	Kalbėtojo vardas	Iš viso įrašų skirtingoms frazėms	3-o atvejo mokymui skirtų įrašų skaičius	3-o atvejo testavimui skirtų įrašų skaičius	4-o atvejo mokymui skirtų įrašų skaičius	4-o atvejo testavimui skirtų įrašų skaičius
Moteriški balsai	FAGNGRA	20	15	5	15	5
	FAGNVIN	20	15	5	15	5
	FAISIZI	20	15	5	15	5
	FAISZYM	20	15	5	15	5
	FAUSNEM	20	15	5	15	5
	FDAILOI	20	15	5	15	5
	FGINGED	20	15	5	15	5
	FIEVJUR	20	15	5	15	5
	FIEVSAB	20	15	5	15	5
	FIEVVIS	20	15	5	15	5
	FJUSKIN	20	15	5	15	5
	FKAMMOS	20	15	5	15	5
	FLAUZET	20	15	5	15	5
	FRAISAV	20	15	5	15	5
	FRUTNAN	20	15	5	15	5
	FSIMMEI	20	15	5	15	5
	FUGNBUC	20	15	5	15	5
	FUGNNOV	20	15	5	15	5
	FVAIVAI	20	15	5	15	5
	FVANPEC	20	15	5	15	5
	Vyriški balsai	FVILVAI	20	15	5	15
FVIONAB		20	15	5	15	5
FZIVSTA		20	15	5	15	5
MDAIGUS		20	15	5	15	5
MEDGVOL		20	15	5	15	5
MKAZANU		20	-	20	15	5
MLINJUR		20	-	20	15	5
MMODSLE		20	-	20	15	5
MRIMAPA		20	-	20	15	5
MVYGVAI		20	-	20	15	5

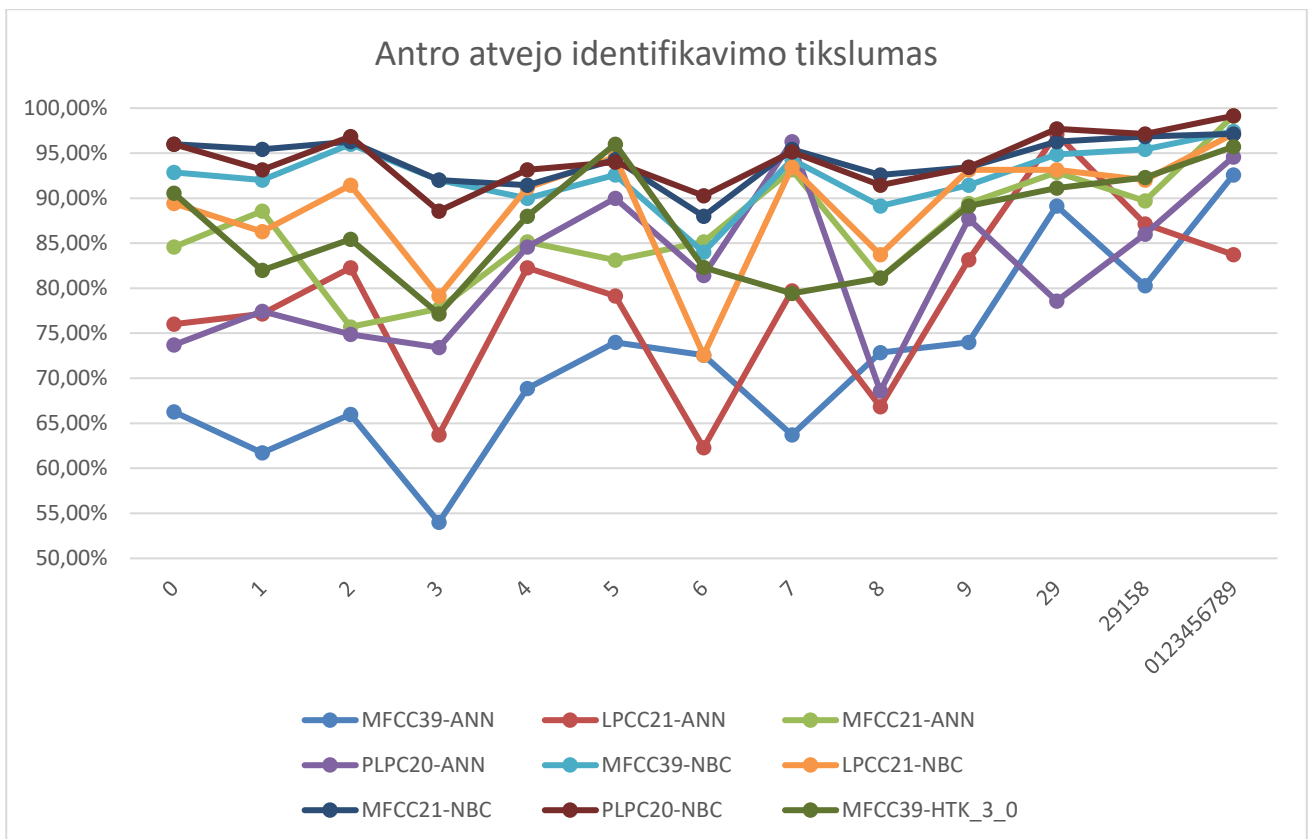
8 PRIEDAS. Rezultatai su pradinėmis reikšmėmis

Lentelė 8.1 Metodų trumpinių lentelė

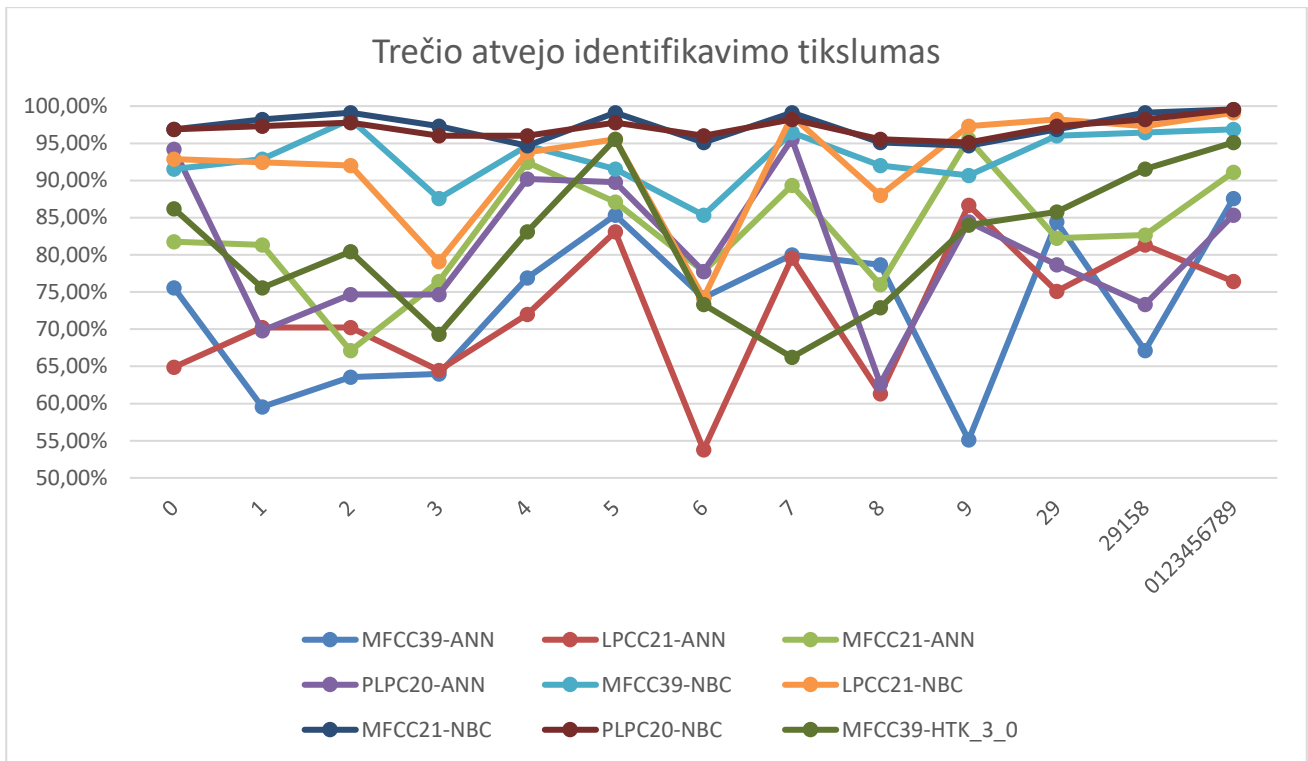
Trumpinys	Metodas
MFCC39-ANN	12 Mel skalės kepstro koeficientų, 12 delta Mel skalės kepstro koeficientų, 12 delta-delta Mel skalės kepstro koeficientų, vieną kadro energijos logaritmo vertę, vieną delta kadro energijos logaritmo vertę, vieną delta-delta kadro energijos logaritmo vertę su dirbtiniais neuronų tinklais
MFCC39-NBC	12 Mel skalės kepstro koeficientų, 12 delta Mel skalės kepstro koeficientų, 12 delta-delta Mel skalės kepstro koeficientų, vieną kadro energijos logaritmo vertę, vieną delta kadro energijos logaritmo vertę, vieną delta-delta kadro energijos logaritmo vertę su Bajeso klasifikatoriumi
MFCC21-ANN	21 Mel skalės kepstro koeficientas su dirbtiniais neuronų tinklais
MFCC21-NBC	21 Mel skalės kepstro koeficientas su Bajeso klasifikatoriumi
LPC21-ANN	21 tiesinės prognozės koeficientas su dirbtiniais neuronų tinklais
LPC21-NBC	21 tiesinės prognozės koeficientas su Bajeso klasifikatoriumi
PLP20-ANN	20 suvokimo tiesinės prognozės koeficientų su dirbtiniais neuronų tinklais
PLP20-NBC	20 suvokimo tiesinės prognozės koeficientų su Bajeso klasifikatoriumi
MFCC39-HTK_3_0	12 Mel skalės kepstro koeficientų, 12 delta Mel skalės kepstro koeficientų, 12 delta-delta Mel skalės kepstro koeficientų, vieną kadro energijos logaritmo vertę, vieną delta kadro energijos logaritmo vertę, vieną delta-delta kadro energijos logaritmo vertę su paslėptaisiais Markovo modeliais



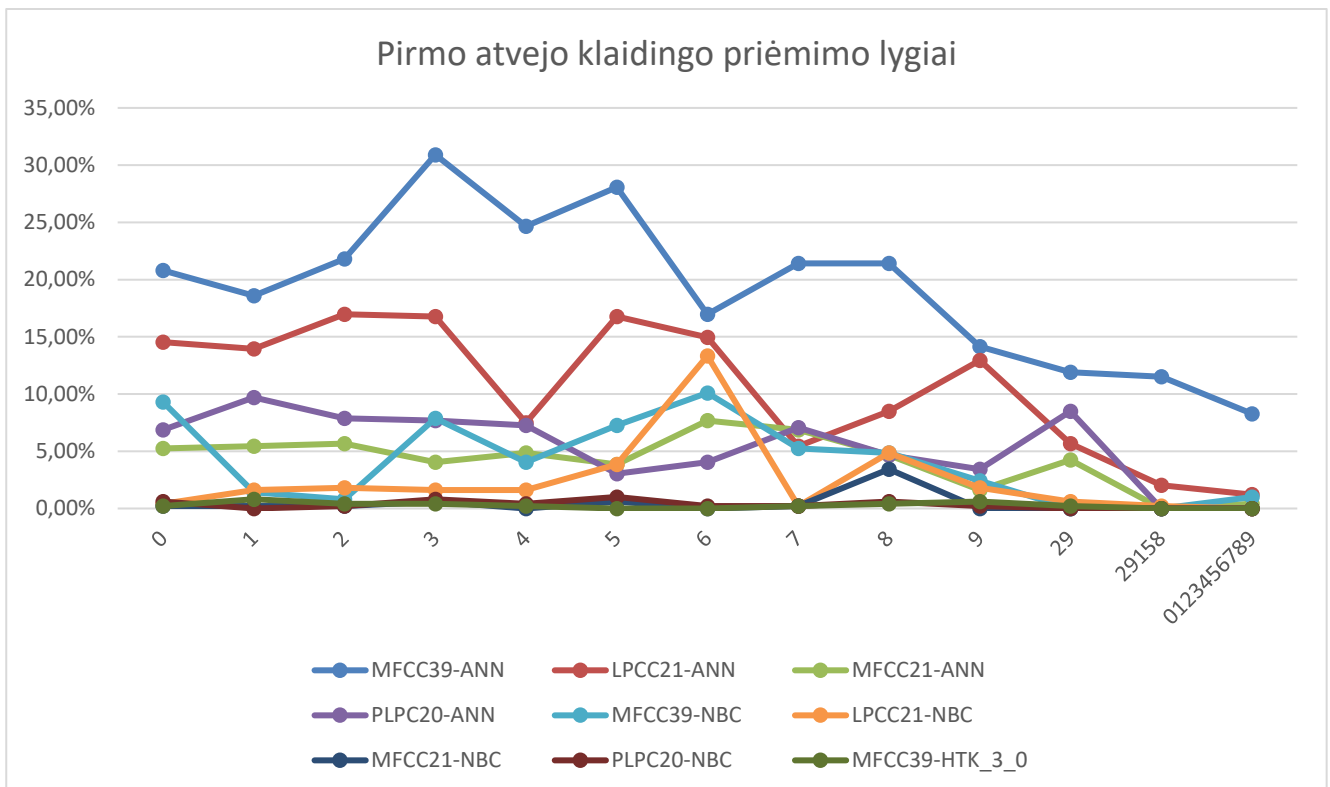
8.1 pav. Metodų identifikavimo tikslumai pirmo atvejo garso įrašo paskirstymui



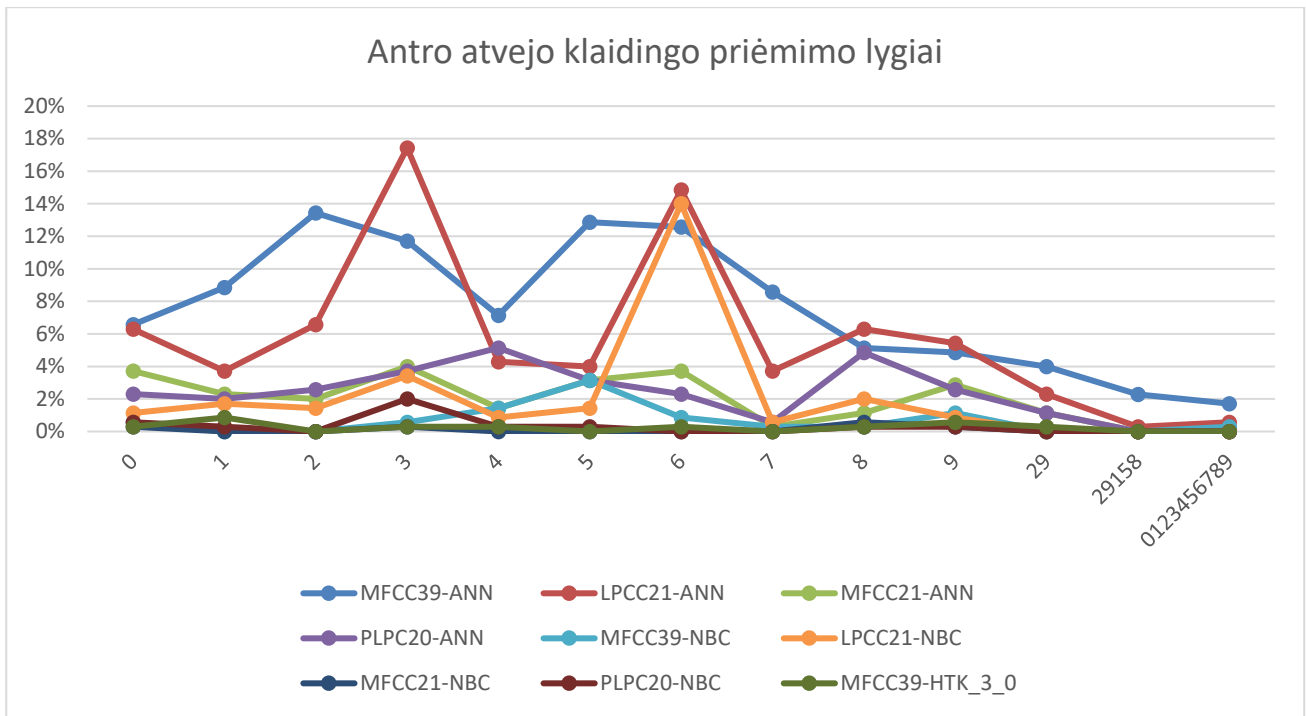
8.2 pav. Metodų identifikavimo tikslumai antro atvejo garso įrašo paskirstymui



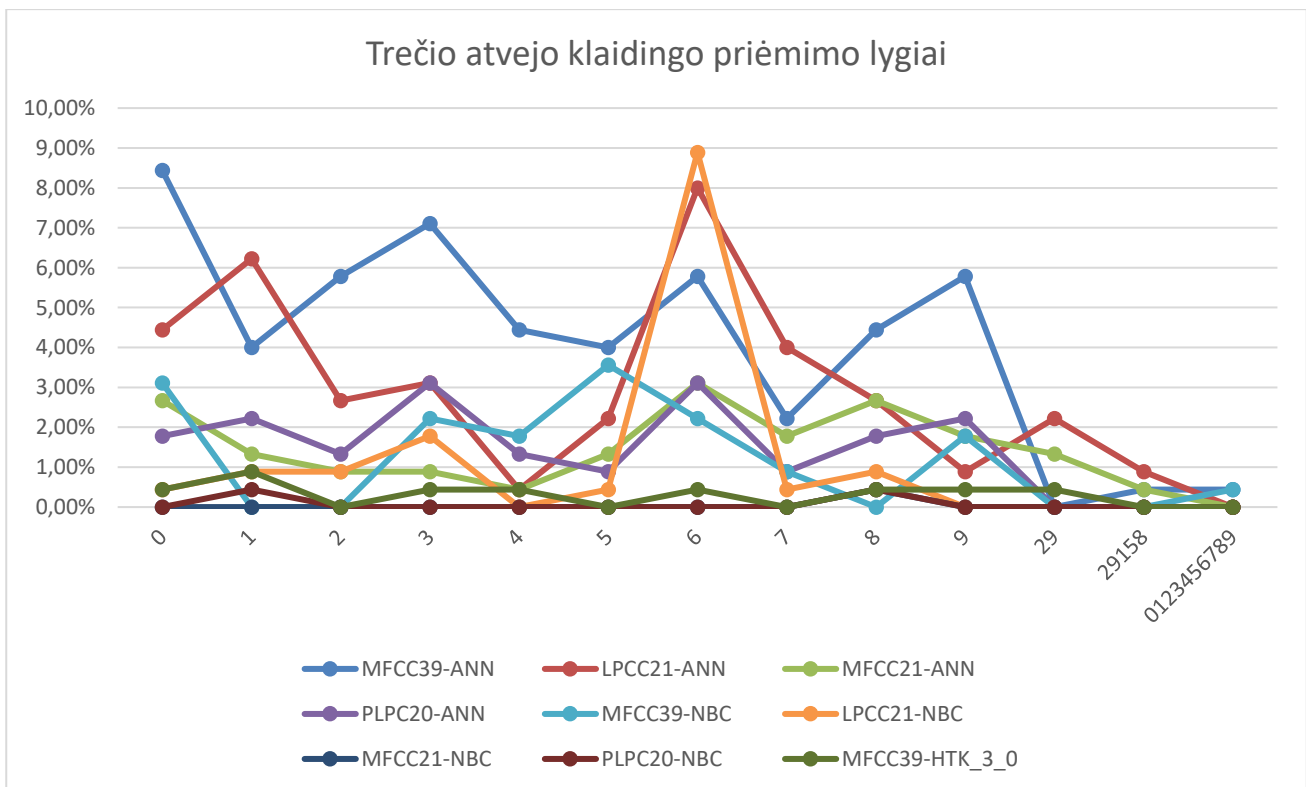
8.3 pav. Metodų identifikavimo tikslumai trečio atvejo garso įrašo paskirstymui



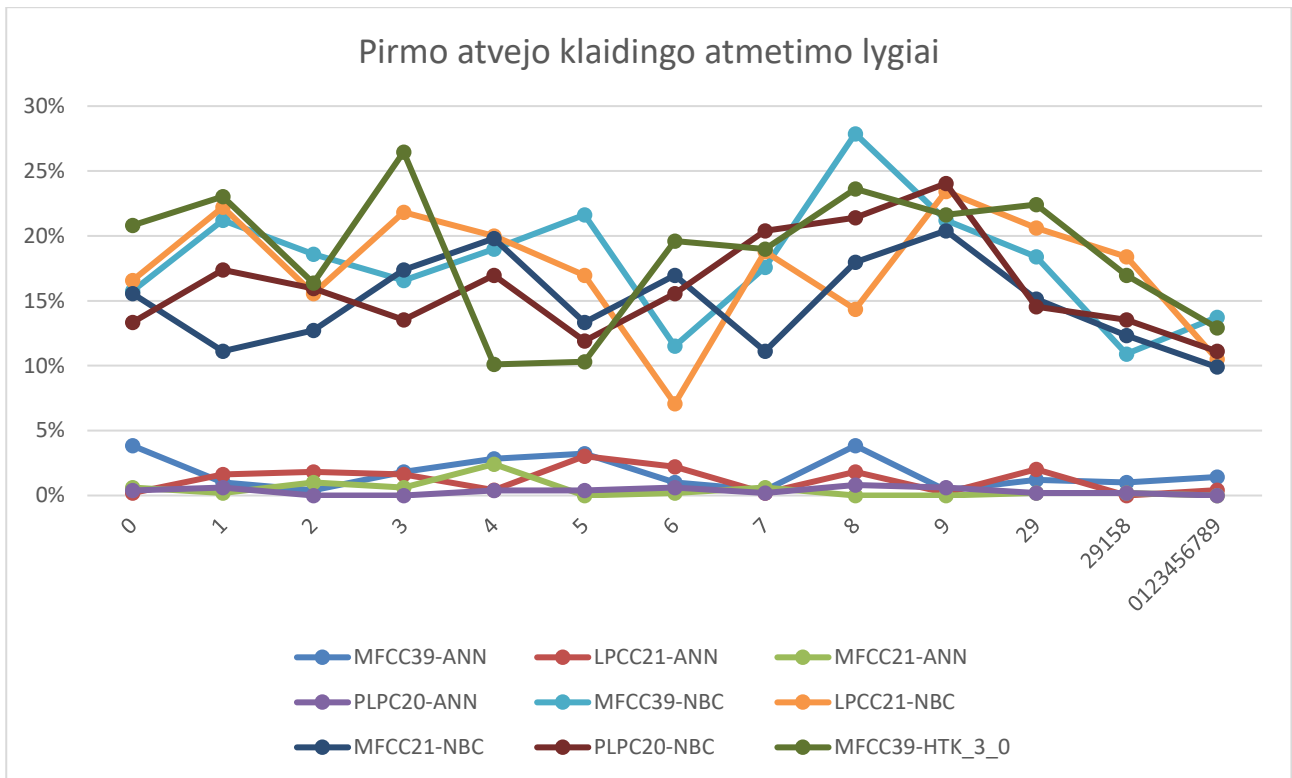
8.4 pav. Metodų klaidingo priėmimo lygiai pirmo atvejo garso įrašų paskirstymui



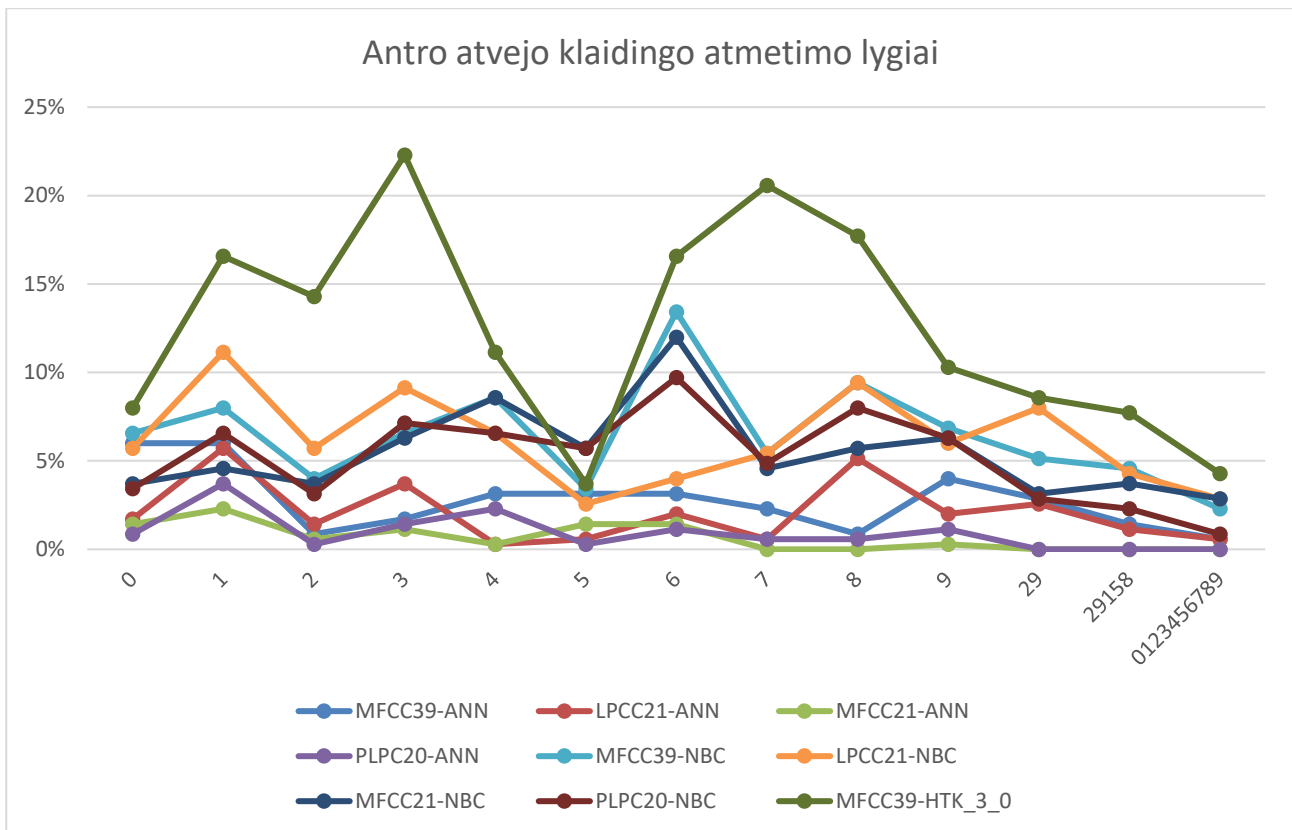
8.5 pav. Metodų klaidingo priėmimo lygiai antro atvejo garso įrašo paskirstymui



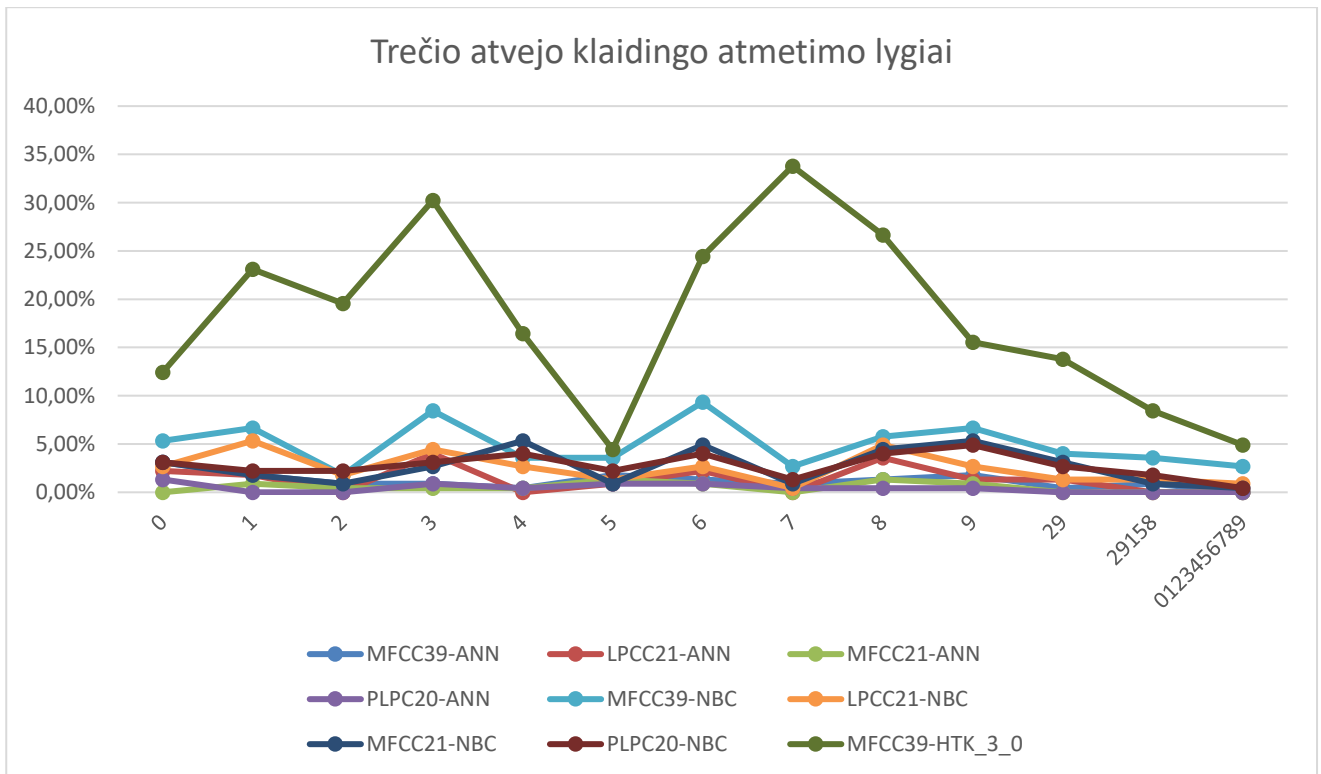
8.6 pav. Metodų klaidingo priėmimo lygiai trečio atvejo garso įrašų paskirstymui



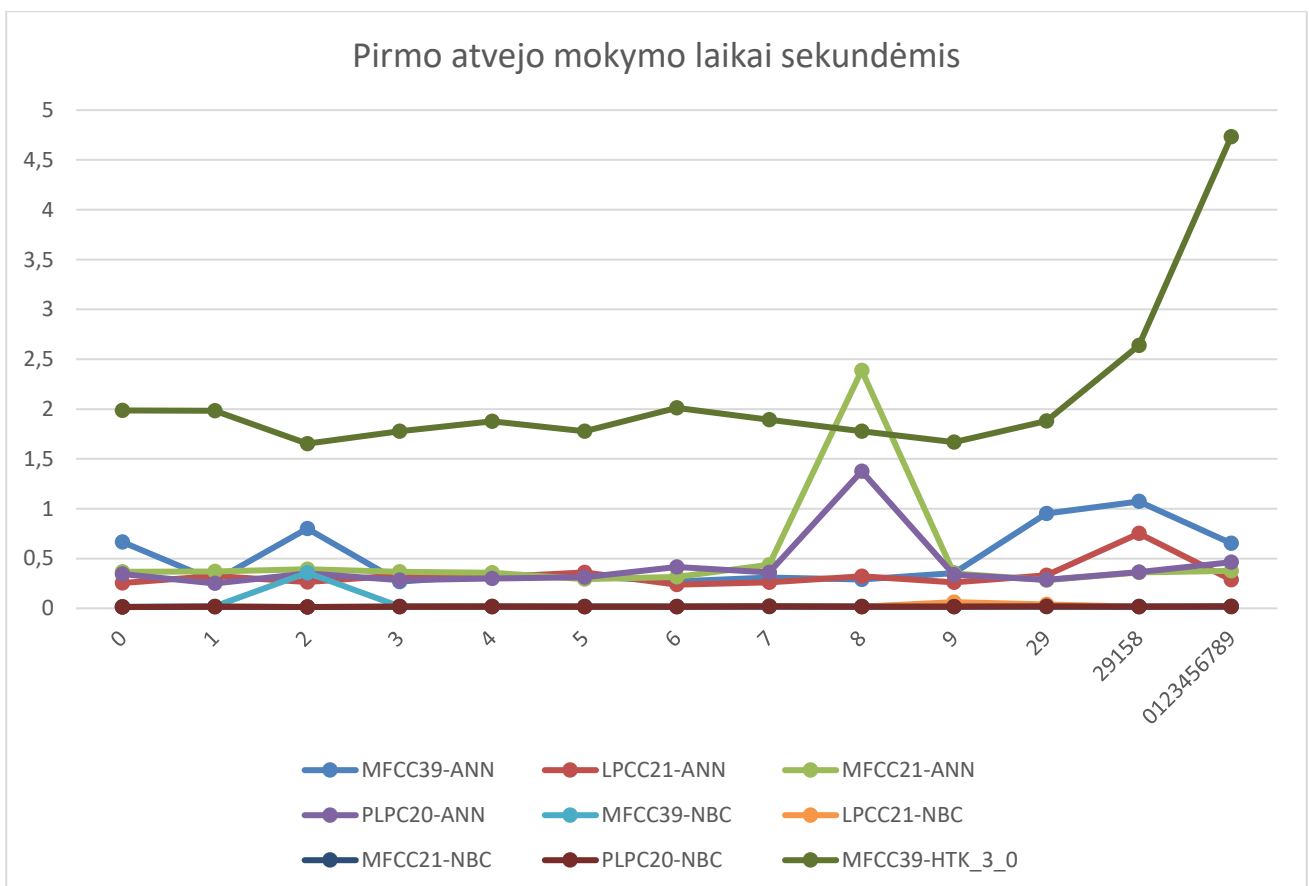
8.7 pav. Klaidingo atmetimo lygiai pirmo atvejo garso įrašų paskirstymui



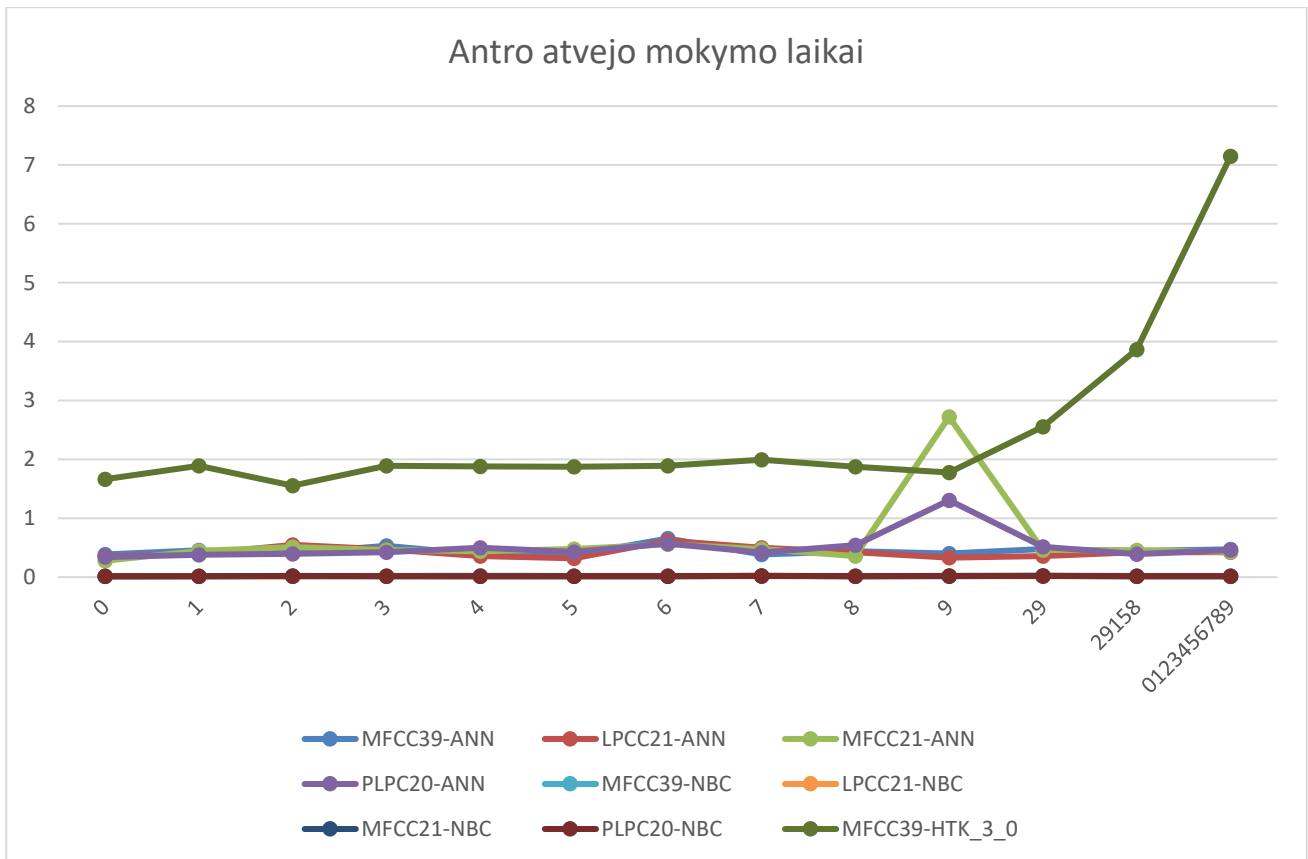
8.8 pav. Klaidingo atmetimo lygiai antro atvejo garso įrašų paskirstymui



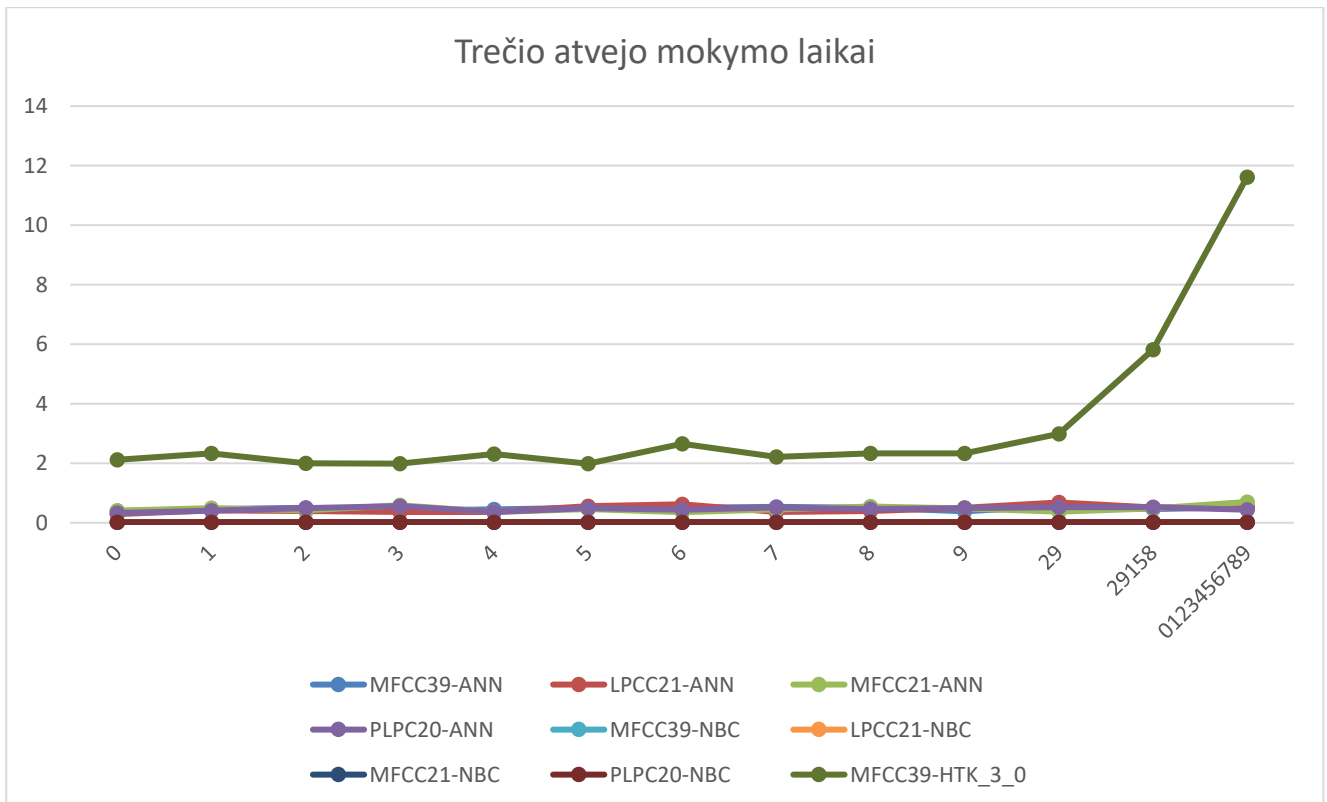
8.9 pav. Klaidingo atmetimo lygiai trečio atvejo garso įrašų paskirstymui



8.10 pav. Mokymo laikai pirmo atvejo garso įrašų paskirstymui



8.11 pav. Mokymo laikai antro atvejo garso įrašų paskirstymui



8.12 pav. Mokymo laikai trečio atvejo garso įrašų paskirstymui

9 PRIEDAS. Rezultatai skirtingiems požymių koeficientams

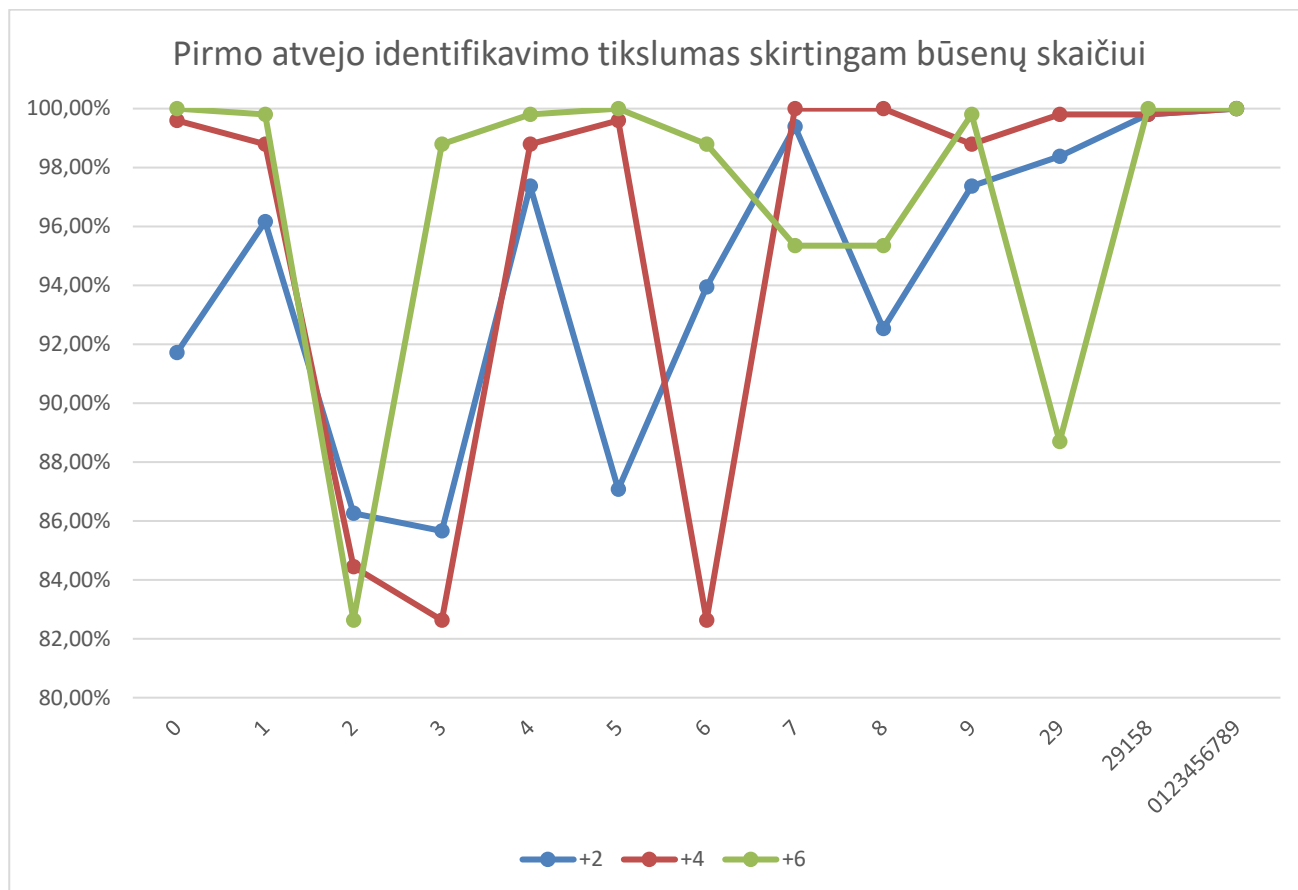
Lentelė 9.1 Maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams su dirbtiniais neuronų tinklais (ANN) antro atvejo garso įrašų paskirstymui

Ištarimas	LPCC-ANN		MFCC-ANN		PLPC-ANN	
	Koeficientų skaičius	Maksimali vertė	Koeficientų skaičius	Maksimali vertė	Koeficientų skaičius	Maksimali vertė
0	19	92.86%	21	97.14%	20	97.43%
1	15	91.71%	14	97.71%	20	96.29%
2	17	94.57%	14	97.14%	17	96.86%
3	15;17	86.57%	21	96.86%	16	94.00%
4	16	92.86%	16	100.00%	16	98.86%
5	14	95.14%	17	99.14%	13	98.57%
6	15;16;20;21	86.57%	20	98.57%	13	97.14%
7	21	97.14%	20	98.57%	16	97.43%
8	15	86.57%	20	98.29%	18	97.14%
9	20	95.71%	15;16;21	98.86%	16;17	98.29%
29	18	96.57%	13;21	99.71%	18;20	98.86%
29158	16	98.86%	11;12;13;14; 15;16;17;18; 19;20;21	100.00%	10;12;13;14; 17;19	100.00%
0123456789	15;18;19;20	100.00%	6;8;9;10;12;14; 15;16;17;18; 19;20;21	100.00%	7;11;12;13;15; 16;18;19;20	100.00%

Lentelė 9.2 Maksimalios tiesinės prognozės kepstro koeficientų (LPCC), Mel skalės kepstro koeficientų (MFCC) ir suvokimo tiesinės prognozės koeficientų (PLPC) identifikavimo tikslumo vertės skirtingiems koeficientų skaičiams (nuo 1 iki 20 arba nuo 1 iki 21) su Bajeso klaidifikatoriumi (NBC) antro atvejo garso įrašų paskirstymui

Ištarimas	LPCC-NBC		MFCC-NBC		PLPC-NBC	
	Koeficientų skaičius	Maksimali vertė	Koeficientų skaičius	Maksimali vertė	Koeficientų skaičius	Maksimali vertė
0	14;16;21	90.00%	21	95.71%	20	96.57%
1	16	90.29%	21	95.14%	19;20	93.43%
2	17	90.86%	21	96.86%	20	96.86%
3	15	88.00%	21	92.29%	10	90.00%
4	21	91.43%	13	93.14%	16;19	93.43%
5	20	95.14%	17	95.43%	20	94.57%
6	20	78.29%	18;19	89.14%	17	91.43%
7	20	95.43%	21	95.43%	17	96.00%
8	16	86.00%	18	94.57%	19	92.57%
9	21	93.14%	16	95.14%	16	94.86%
29	19;20	92.86%	14;15	97.43%	18	96.29%
29158	18;20	95.43%	10	97.14%	20	97.71%
0123456789	19	96.86%	11	98.00%	15;18;20	98.86%

10 PRIEDAS. Paslėptųjų Markovo modelių paketo rezultatai



10.1 pav. Pirmo atvejo identifikavimo tikslumas skirtingam būsenų skaičiui