

ŠIAULIŲ UNIVERSITETAS
TECHNOLOGIJOS, FIZINIŲ IR BIOMEDICINOS MOKSLŲ FAKULTETAS
ELEKTRONIKOS IR ELEKTROS KATEDRA

Remigijus Vainauskas

NUO TEKSTO NEPRIKLAUSOMAS
KALBĖTOJO ATPAŽINIMAS

Magistro darbas

Vadovas

Prof. Dr. G. Daunys

ŠIAULIAI, 2016

SUMMARY

This work focuses in speaker recognition programs, recognition stages and experiment with the chosen method.

First issue of the programs, which processes speech signals and recognizes speakers. Recognition programs can be controlled by voice and identify the gender. Also in system can be speaker identification and verification.

The following text of the speaker recognition gives theory how extract features of speech signals. Also use pre-emphasis and noise removal of speech signal. Feature extraction of speech signals using for Text Independent Speaker Recognition system model.

In experiment use Matlab 2007b software package. Text Independent Speaker Recognition we use „MSR Identity Toolbox v1.0: A MATLAB Toolbox for Speaker-Recognition Research“ software. Speakers signals use of LibriSpeech ASR corpus, which includes male and female speech. System modeling for speaker recognition using A Gaussian Mixture Model. Which can processing a lot of speakers signals.

During the experiment, we need to get a good percentage of the speaker recognition. The best case than the equal error rate (EER) of speaker recognition, is less than one. Also present the results of speaker recognition by increasing the number of speakers.

SANTRAUKA

Šiame darbe gilinimasi į kalbėtojo atpažinimo programas, atpažinimo etapus ir atliekamas eksperimentas su pasirinktu metodu.

Pirmiausia nagrinėjamos programos, kurios apdoroja kalbos signalus ir atpažįsta kalbėtojus. Atpažinimo programos gali būti valdomos balsu ir identifikuoti lytį. Taip pat sistemoje gali būti kalbėtojų identifikavimas ir verifikavimas.

Toliau nagrinėjama, kaip reikia išskirti kalbos požymius. Taip pat kalbos signalams taikomas pirminis apdorojimas, kaip pradinė filtracija ar tylos pašalinimas. Turint kalbos požymius galime panaudoti kalbėtojo atpažinimo metodą, kuriant nepriklausomo nuo teksto kalbėtojo atpažinimą.

Naudojama programinė Matlab 2007b įranga. Nepriklausomai nuo teksto kalbėtojo atpažinimui naudosime „MSR Identity Toolbox v1.0: A MATLAB Toolbox for Speaker-Recognition Research“ programinį paketą. Kalbos įrašai naudojami iš LibriSpeech ASR corpus, kuriame yra vyrų ir moterų kalbos įrašai. Atpažinimas atliekas Gauso mišinio modelio metodu, kuris gali apdoroti didelis kiekius duomenų.

Atliekant tyrimą, turime gauti gerą kalbėtojo atpažinimo procentą. Geriausiai atpažįstą sistemą, kai lygių klaidų vertė (EER) yra mažiau už vieną. Atiekami tyrimai, kaip atpažįsta kalbėtojus didinant kalbėtojų skaičių, bei pateikiami rezultatai.

TURINYS

IŽANGA	8
1. KALBĖTOJO ATPAŽINIMO SISTEMOS APŽVALGA	9
1.1. Programų apžvalga	9
1.1.1. Phonexia Speaker Identification.....	9
1.1.2. SIFT.....	9
1.1.3. VeriSpeak.....	10
1.2. Kalbančiojo identifikavimas ir verifikavimas.....	11
1.4. Kalbėtojo atpažinimo metodai	13
1.4.1. Vektorių kvantavimo.....	13
1.4.2. Neuroniniai tinklai.....	15
1.4.3. Dirbtiniai neuronų tinklai	16
1.4.4. Tiesinės prognozės	17
1.4.5. Atraminių vektorių mašinos	18
1.4.6. Gauso mišinio modelis	19
2. KALBOS SIGNALŲ POŽYMIAI	23
2.1. Kalba.....	23
2.2. Pirminis signalų apdorojimas.....	23
2.2.1. Pradinė filtracija	24
2.2.2. Kadravimas.....	25
2.3. Melo fazės kepstro koeficientai	26
3. Nepriklausomo nuo teksto kalbėtojo atpažinimas	28
3.1. Triukšmo pašalinimas	28
3.2. Kalbėtojo atpažinimo struktūra.....	29
3.3. Kalbėtojų verifikavimo sistema	30
3.4. Tyrimas su Matlab ToolBox paketu	32
IŠVADOS IR PASIŪLYMAI.....	51

LITERATŪRA	52
PRIEDAI	54

IŽANGA

Kalba tai žmogaus bendravimo įrankis, kuris leidžia dalintis mintimis, jausmais, prisiminimais. Kadangi kiekvienas žmogus turi sau būdingą kalbėjimo manierą, ritmą, intonaciją - tai galime panaudoti kalbėtojo atpažinime. Vartotojas sakydamas tam tikrus žodžius ar sugalvotus žodžius, gali būti atpažintas sistemos, nes sistema bando atpažinti kalbėtoją pagal jo balso požymius. Didžiausia problema, kuriant tokią sistemą tai triukšmas, dėl kurios sistema tampa nepatikima.

Kalbėtojo atpažinimo metodas reikalingas gamybos pramonėje, biuruose bei kompiuterių kambariuose, kur naudojami metodai asmenims atpažinti. Dažniausios problemos naudojant kalbėtojo atpažinimo metodą yra triukšmas atsirandantis signale. Kartais triukšmas atsiranda naudojant netinkamus įrašymo priemones, kaip prastus mikrofonus dėl kurių įrašai būna prastos kokybės. Todėl norint atpažinti asmenis reikia pašalinti ar mažinti atsiradusius triukšmus.

Kalbos technologijos turi gerus pasiekimus moksle, todėl nepriklausomo nuo teksto kalbėtojo atpažinimas, leidžia naudotis netik apsaugos sistemose, bet ir namuose. Vartotojas sakydamas tam tikrus žodžius ar kalbėdamas kokius sugalvotus žodžius, gali būti atpažintas įrenginio.

Darbo objektas – nepriklausomo nuo teksto kalbėtojo atpažinimas naudojant atpažinimo metodus.

Darbo tikslas – realizuoti kalbėtojų atpažinimus panaudojant Gauso mišinio modelį.

Darbo uždaviniai

- Kalbėtojo atpažinimo programų apžvalga.
- Kalbančiojo identifikavimas ir verifikavimas.
- Nepriklausomo nuo teksto atpažinimo metodų analizė.
- Atlikti pirminius kalbos signalų apdorojimus.
- Atlikti tyrimus su pasirinktu metodu.
- Patikrinti atpažinimą naudojanči LibriSpeech garsyną.
- Iširti bei įvertinti gautus rezultatus.

1. KALBĖTOJO ATPAŽINIMO SISTEMOS APŽVALGA

Šiuo metu yra sukurta daug atpažinimo technologijų, kurios naudojamos pasaulyje. Dažniausiai naudojamos tiek pirštų antspaudų, akies rainelės ar veido geometrijos. Tarp anksčiau paminėtų žmogaus bruožų, kuriuos naudoja biometrinės technologijos, yra dar vienas, kuris naudojamas atpažinimui – tai balso atpažinimas. Nepriklausomo nuo teksto kalbėtojo atpažinimą galima naudoti tiek kompiuteriuose ar kituose įrenginiuose, kur reikalingas kalbėtojo atpažinimas. Taigi, šiame skyriuje aptarsime keletą atpažinimo programų, atpažinimo metodų bei kalbėtojo verifikavimą bei indentifikavimą.

1.1. Programų apžvalga

Kalbėtojų atpažinimo programų yra panašių ir skirtingų. Kiekviena programą turi savo metodo modelį, kaip atpažinti kalbėtoją. Dauguma programų naudoja Gauso mišinių modelį, nes geriausia atliekantis kalbėtojų atpažinimą. Taip pat programos naudoja kalbėtojų lyčių indentifikavimą – atpažįsta kalbėtoją pagal lytį.

1.1.1. Phonexia Speaker Identification

Nuo 2006 m Phonexia padeda klientams automatiškai išgauti vertingą informaciją iš šnekamosios kalbos. Jie sukūrė technologijas, kurios gauna informaciją iš kalbos, taip pat kalbos analizę ir balso biometrija. Šios technologijos yra naudojamos skambučių centruose, telekomunikacijose, bankuose, serveriuose ir transliacijos paslaugų teikėjams. Phonexia taip pat suteikia mokslinių tyrimų ir plėtros paslaugas, tokių kaip kalbos technologijų optimizavimas, kuriant naujas kalbines versijas [1].

Tai labai tiksli technologija naudojanti balso požymius, automatiškai atpažins kalbėtojo balsą ir ieškos konkretaus kalbėtojo, kuris yra archyve. Jis leidžia patikrinti kalbėtojo balsą ir palyginti prieš kitus. Technologija taip pat turi integruotą lyčių indentifikaciją.

Programinėje įrangoje naudojama Gauso mišinių modeliai. Šis metodas priklauso stochastiniam modeliavimui ir modeliuoja statistinį požymių kitimą. Tai yra statistinis pavaizdavimas kaip kalbėtojas taria garsus. Taip pat naudojant šį paketą rekomenduojami kompiuterio resursai: 2.6GHz CPU ar didesnis ir 4GB RAM atminties, nes kompiuteriuose su lėtais procesoriais atpažinimas gali būti su didelėmis pauzėmis [1].

1.1.2. SIFT

Speaker Identification Field toolkit (SIFT) yra kalbėtojo indentifikavimo rinkinys, kurį galima įmontuoti į nešiojamąjį kompiuterį. Sukurta remiantis tiksliai ir itin sparčiu ketvirtos kartos

AGNITiO balso atpažinimo technologija. Tai leidžia organizacijoms taikyti balso atpažinimo technologijas ir palaikyti keletą teisėsaugos, kriminalistikos ir žvalgybos operacijas. SIFT funkcijos:

- Valdymas balsu
- Balso įrašymas
- Kalbėtojo atpažinimas

SIFT turi automatinį balso atpažinimo paketą, kuris gali pašalinti artefaktus, triukšmus ir ne balso garsus. Tai leidžia apdoroti ir išsaugoti tam tikros naudingos informacijos apie kalbėtojo balsą. Paketas buvo sukurtas siekiant išsaugoti iki 1000 skirtingų balsų ir išsaugant savo duomenų bazėje. Jis net leidžia apdoroti įrašytus du balsus, siekiant nustatyti atskirus kalbėtojų balsus. Taip pat suteikia galimybę nustatyti tarp nežinomo balso ir žinomo kalbėtojo. SIFT suteikia dideles galimybes kalbėtojo nustatymui, kurios naudojamos net karinėse organizacijose [2].

1.1.3. VeriSpeak

VeriSpeak balso identifikavimo technologija yra skirta biometrinių sistemų kūrėjams. Atpažinimo algoritmas užtikrina sistemų saugumą tikrinant balso ryšiui ir frazės autentiškumą.

Programos paketas yra prieinamas kaip programinės įrangos kūrimo rinkinys, kuris įgalina programų kūrimą. Palaiko Microsoft Windows, Linux ir Mac OS X platformoms. Algoritmas vykdo balso priėmimą ir analizuoja naudojant garso apdorojimo technologijas. Identifikacijos atveju atliekamas vienas su daugeliu arba vienas su N palyginimas

Naudojant VeriSpeak programinę įrangą, kalbėtojo atpažinimas yra priklausomas nuo pasakyto teksto. Tai į atpažinimo sistemą pateikiama frazė yra iš karto žinoma. Siekiant didesnio atpažinimo tikslumo, skirtingos frazės fonetiniu turiniu turi būti panašios, kad būtų galima jas tinkamai palyginti.

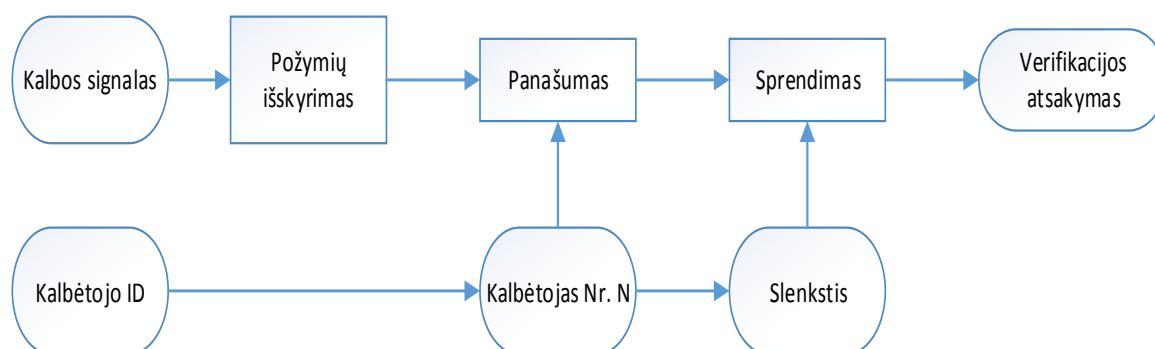
VeriSpeak yra palaikoma SDK [3]:

- VeriSpeak Standartinė SDK palaikoma kompiuterio pagalba ir yra licenzijuota. Turi programavimo pavyzdžių ir naudojimo instrukciją programinės įrangos dokumentacijoje. SDK leidžia naudotis Microsoft Windows, Linux ar Mac OS X operacinėse sistemose.
- VeriSpeak išplėstinė SDK skirtas interneto pagrindams ir tinklo programų kūrimui. Tai apima visas funkcijas ir komponentus su standartiniu SDK, turi programų pavyzdžių ir paruoštus naudojimui programinius paketus.

1.2. Kalbančiojo identifikavimas ir verifikavimas

Kalbėtojo atpažinimas yra skirtomas pagal kalbėtojo indentifikavimą bei jo verifikavimą. Tiek verifikavimas ir indentifikavimas gali būti priklausomas nuo teksto arba nepriklausomas.

Kalbančiojo verifikavimo metu nežinomo asmens balsas yra lyginamas su žinomo asmens balsu. Šio lyginimo metu gautas panašumo laipsnis dar palyginamas su šiam žinomam asmeniui nustatytu slenkščiu. Asmens balso verifikacijos atveju atliekamas palyginimas vienas su vienu. Asmuo pasako kalbos pavyzdį ir tuomet jo balso pavyzdys palyginimas su kitais ankščiau įrašytais to kalbėtojo pavyzdžiais. Kalbėtojo verifikavimas pateiktas 1.1. paveiksle [4].



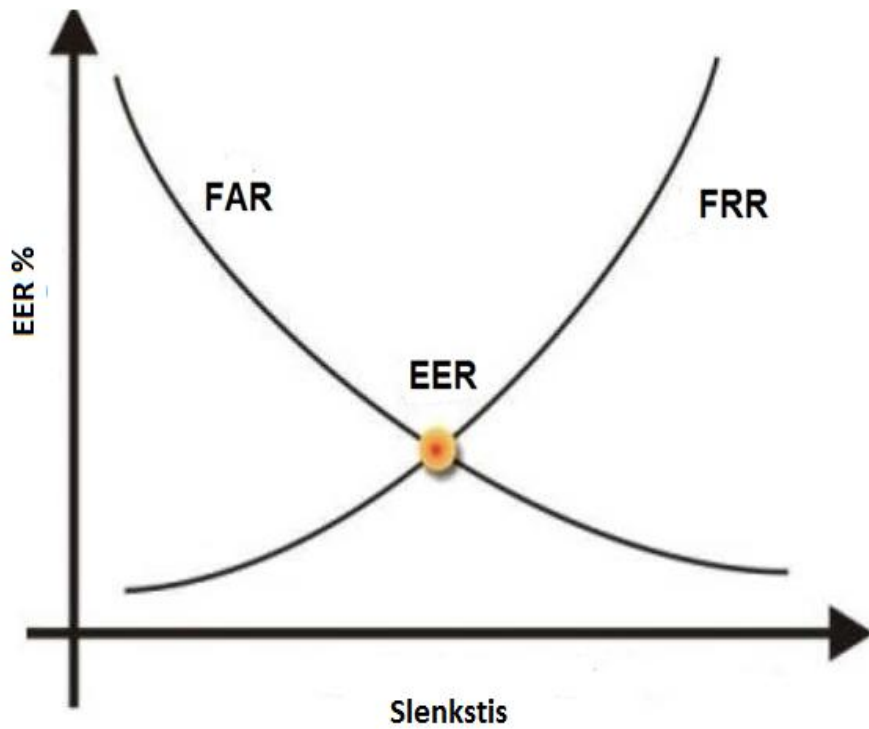
1.1. pav. Kalbėtojo verifikavimas.

Jei pasakyto kalbos pavyzdys viršija pasirinktus slenkčius tai asmens tapatybė patvirtinama, kitu atveju atmetama. Verifikacijos algoritmas daro dviejų rūšių klaidas:

FRR – Klaidingo atmetimo rodyklis

FAR – Klaidingo priėmimo rodyklis

Šių klaidų procentinės reikšmės priklauso nuo pasirinkto sprendimo slenkščio. Atliekant tyrimus svarbiausia, kad lygių klaidų vertė (EER) būtų kuo mažesnė. EER gaunama, kai $FRR=FAR$. Tai reiškia, kad EER gaunama, kai atmetimo ir priėmimo klaidos yra lygios.



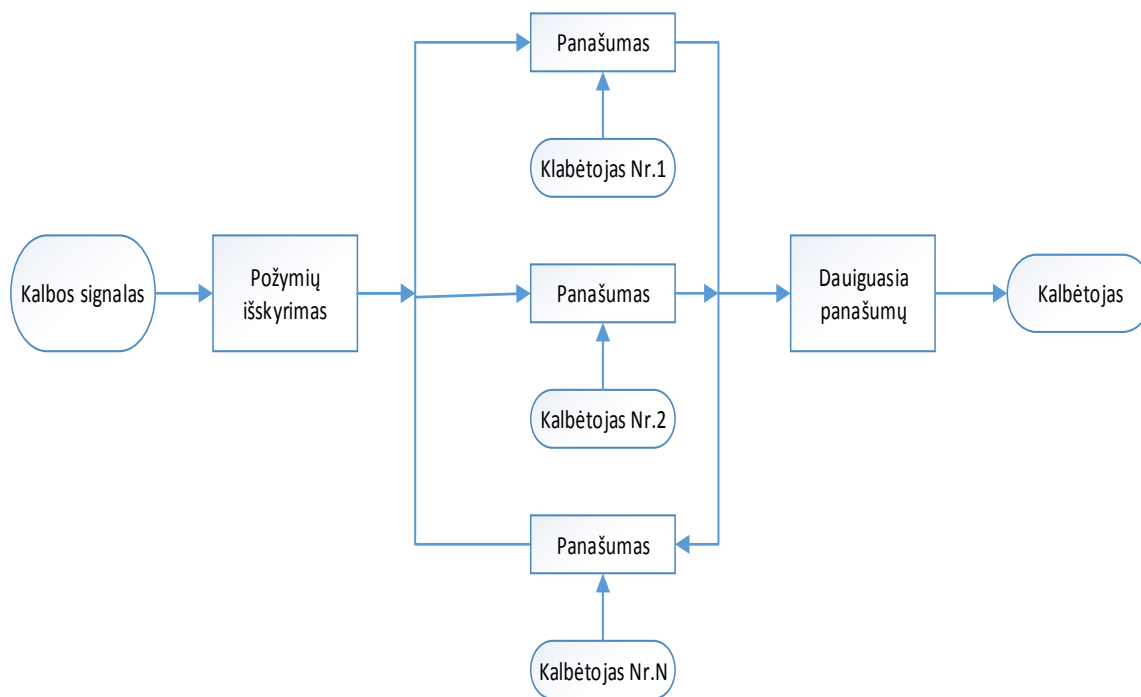
1.2. pav. Equal Error Rate (EER)

Kuo kreivės grafikas mažesnis tuo verifikacijos algortimo kokybė geresnė, pateiktas 1.2 paveiksle.

Kalbančiojo identifikavimas atliekamas vienas su daugeliu N palyginimas. Balso identifikavimo įrenginys turi rasti labiausiai panašaus balso atitikmenį. Atliekamas uždaras ir atviras identifikavimas. Kalbėtojo identifikavimo sistema pateikta 1.3. paveiksle[4].

Atviro identifikavimo atveju reikia nuspręsti ar balso pavyzdys priklauso sistemai. Kai randamas panašus balsas tai jis yra pateikiamas.

Uždaro identifikavimo atveju sistema daro priverstinį sprendimą, paprasčiausiai surasdama labiausiai atitinkantį kalbėtoją, nors jei kalbėtojas neatitinka. Todėl sistema gali būti nekokybiška.



1.3. pav. Kalbėtojo indentifikacija

Taip pat sistemoje gali būti nuo teksto priklausantis ir nepriklausantis kalbančiojo atpažinimas. Galima rinktis tarp priklausančio ar nepriklausančio teksto. Priklausančiamenuo teksto sistemoje, išarta bet kokia frazė yra žinoma, nes buvo apmokyta. Kitu atveju sistemoje, kur nepriklauso nuo teksto, gali būti išarta bet kokia frazė. Nepriklausomos nuo teksto sistemos yra geresnė, nes kiekvienas kalbėtojas turi savo dažniausiai naudojamus žodžius. Toks sistemos patvirtinimas gali trukti net kelias minutes.

1.4. Kalbėtojo atpažinimo metodai

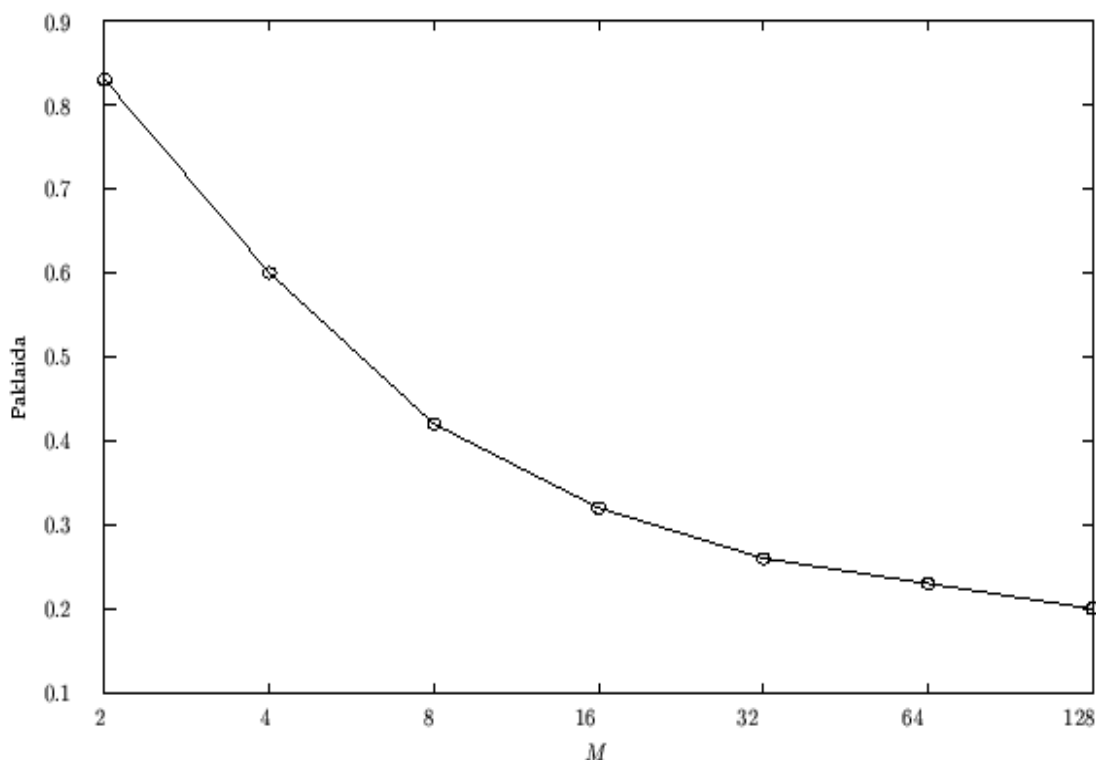
Šiame skyriuje aptarsime keletą kalbėtojo atpažinimo metodų, kuriuos galima naudoti kuriant sistemą. Dažniausiai kalbėtojo atpažinimo sistemose naudojami Gauso mišinių modeliai (GMM) ir vektorinio kvantavimo (VQ) metodai, kurie yra nepriklausomi nuo teksto. Literatūroje teigiama, kad GMM geriau atpažįstą kalbėtojus, bei leidžia pasiekti tikslesnius rezultatus. Vektorinio kvantavimo metodas skirtas mažesniems kiekiams apdoroti. Taigi aptarkime keletą atpažinimo metodų.

1.4.1. Vektorių kvantavimo

Vektorių kvantavimo metode naudojami kodų knygos, kuriose yra surašyti kalbėtojų požymiai. Kai kodavimo lentelė yra gauta, tai atvaizdavimas tolydžių vektorių kodavimo lentelės indeksais yra ne kas kita, o paprasčiausias artimiausio kaimyno radimas, t.y. tolydžiam vektoriui

priskiriamas indeksas artimiausio (spektrinio skirtumo prasme) kodavimo lentelės vektoriaus. Taigi, svarbiausia yra sudaryti tinkamą kodavimo lentelę [5].

Vektoriaus kvantavimas veda prie paklaidų atsiradimo, nes visas vektorių erdvės regionas yra atvaizduojamas vienu vektoriumi. Norint, kad paklaidos būtų kuo mažesnės, reikia didinti kodavimo lentelės dydį. 1.4 paveiksle pavaizduota vektoriaus kvantavimo vidutinės paklaidos priklausomybė nuo kodavimo lentelės dydžio M [5].



1.4. pav. Vidutinės paklaidos priklausomybė

Iš pateikto grafiko matome, kad didėjant M , mažėja paklaidų skaičius. Kodinių knygų kūrimas būna dviejų tipų:

- Prižiūrėtos – nustatoma tarpusavio koreliacijos.
- Neprižiūrėtos – kūrimos atskirai.

Yra žinomas Lindo, Buzo ir Gray (LBG) vektorių kvantavimo algoritmas. LBG algoritmo žingsniai:

1. Sukuriama vieno vektoriaus kodų knyga (nevykdoma iteracija).
2. Keičiamas kodų knygos dydis du kartus, dalinant y_n .

$$y_n^+ = y_n(1 + \alpha)$$

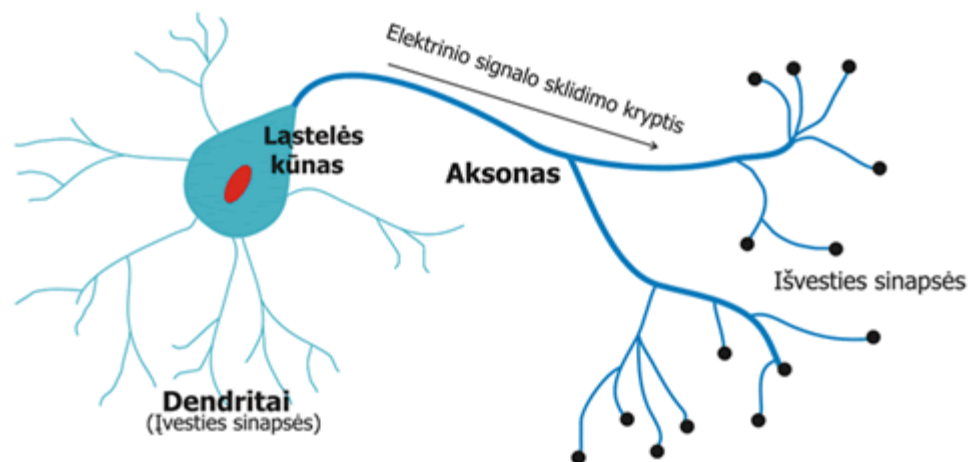
$$y_n^- = y_n(1 - \alpha)$$

Kur n kinta nuo 1 kodų knygos dydžio, α - dalinimo koeficientas.

3. Artimiausio kaimyno paieška: surasti tokia kodų knygą, kuri yra artimiausia (panašus parametrai) ir priskirti tą vektorių (panašiausias kodinis žodis).
4. Atnaujinti centroidus.
5. Iteracija 1: kartoti 3 ir 4 žingsnius, kol vidutinis atstumas bus mažesnis už nustatytą slenkstį.
6. Iteracija 2: kartoti 2,3 ir 4 žingsnius, kol sukuriama kodų knyga M dydžio. Pirmiausia sukuriama vieno vektoriiaus knyga, paremta dalijimosi technika. Vėliau sukuriama dviejų vektorių knyga, kol pasiekama M dydžio kodų knyga[6].

1.4.2. Neuroniniai tinklai

Žmogaus smegenys susideda iš daugelio (apie 10 000 000 000 000) neuronų, sujungtų vienu su kitais. Kiekvienas neuronas turi vidutiniškai keletą tūkstančių jungčių. Tai ląstelė, galinti generuoti elektrocheminį signalą. Neuronas turi išsišakojusią įėjimo struktūrą, vadinamuosius dendritus, ląstelės kūną, vadinamąją somą, ir besišakojančią išėjimo struktūrą – aksoną [8]. Pavyzdys pateiktas 1.5 paveiksle.



1.5. pav. Biologinio neurono atvaizdas

Šios ataugos dalyvauja perduodant nervinį impulsą iš vieno neurono į kitą:

- dendritai perduoda impulsą į neuroną,
- aksonas perduoda impulsą iš neurono.

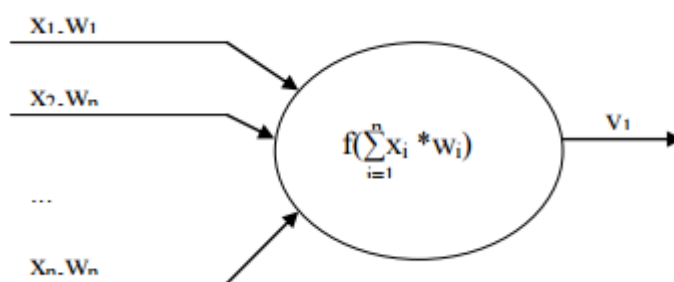
Kai sužadinama pakankamai neuronų, yra perduodamas elektrocheminis impulsas .

- Signalai perduodami kitiems neuronams, kurie gali būti sužadinti.
- Naudojamas slenkstinis lygis, kuris leidžia išsiūsti signalą.

1.4.3. Dirbtiniai neuronų tinklai

Dirbtinis neuronas yra labai supaprastintas biologinio neurono modelis. Neuronas yra užrašomas iš pirmo žvilgsnio labai paprasta matematine išraiška, tačiau dėka joje dažniausiai naudojamos netiesinės funkcijos jo galimybės yra pakankamai didelės. Kita vertus, dirbtinio neurono sąvoka yra pakankamai plati ir nevienareiškmė [8].

Neuronas yra sudėtuvas, nes į jį ateina daug įėjimo ryšių, bet turi vieną išėjimo ryšį. Neurono struktūra pateikta 1.6 paveiksle.



1.6. pav. Neurono modelis

Tai galingas modeliavimo aparatas, kuris gali pats apsimokyti iš jam pateiktų įėjimų. Neurono tinklo sudėtingumas yra suprantamas, kai kai pradeda modeliuoti dirbtinius neuronų tinklus. Tinklas susideda iš įėjimų, išėjimų, ir svorių reikšmių.

Dirbtinis neuronas turi:

- įėjimo reikšmių x_1, x_2, \dots, x_n .
- Kiekviena iš įėjimo reikšmių turi savo svorį w_1, w_2, \dots, w_n .
- Kiekvienas neuronas turi savo slenksčio reikšmę w_0 .
- Perdavimo funkcijas.

Daugiasluoksnis perceptronas

DSP sluoksniai skirstomi į tris grupes:

- Įvesties sluoksnis – gautą informaciją siunčia tinklui.
- Paslėptas sluoksnis – gauna informaciją iš įvesties sluoksnio arba iš kito sluoksnio.
- Išėjimo sluoksnis – perduoda apdorotą informaciją [9].

Neuroninių tinklų apmokymo metodai gali būti suskirstyti į tris kategorijas:

- Mokymą su mokytoju
- Hibridinis mokymas
- Mokymą be mokytojo

Mokymui su mokytoju reikalingas išorinis mokytojas, valdantis mokymosi procesą ir teikiantis informaciją. Metodo tikslas – priversti neuroninį tinklą pakeisti neuroninių jungčių svorius pagal pavyzdinius įėjimus ir išėjimus.

Mokymas be mokytojo neturi išorinio mokytojo, nes sistema turi pati save suderinti. Sistemai apteikiama tik įėjimų pavyzdžiai pagal kuriuos turi gauti įėjimus.

Hibridinis mokymas tai dalis anksčiau minėtų mokymų svorių dydis [10].

Trūkumai:

1. Persimokymas.
2. Kuo daugiau duomenų tuo ilgesnis skaičiavimas.
3. Gali pasiekti nestabilią būseną.
4. Norint apmokyti tinklą turime žinoti išėjimus.

Privalumai:

- Sugebėjimas mokytis.
- Netiesiškumas.
- Didelis apskaičiavimo pajėgumas.

1.4.4. Tiesinės prognozės

Tiesinė prognozė (angl. LPC – Linear Prediction Coefficients arba LinearPredictionCoding) yra vienas galingiausių signalų analizės metodų. Ypač jis gerai užsirekomendavo kalbos analizėje, tiesine prognoze paremtas kalbos kodavimas duoda geriausius kalbos ir kompresijos kokybės parametrus. Šiuo metodu gaunami kalbą aprašantys parametrai, jų skaičius yra nedidelis ir jie apskaičiuojami greitai [11]. Todėl LPC galima pritaikyti kalbėtojo atpažinime. Pagrindinė idėja – reikia surasti ir įvertinti gautas formantes. Formantės aprašomo yra tam tikromis lygtimis. Tiesinės prognozės koeficientai naudojami su kitais modeliais kalbėtojų atpažinime. Dažniausiai naudojama neuronų tinklai, kurie pritaikyti apdoroti daug informacijos.

Diferencialinės išraiškos koeficientai vadinami prognozės (tiesinės) koeficientais. Šie koeficientai aprašo formantes, taigi metodo esmė yra apskaičiuoti tiesinės prognozės koeficientus. Formulės apskaičiuoti LPC koeficientus yra gaunamos minimizuojant liekamojo signalo vidutinę kvadratinę paklaidą. Rezultate yra gaunama atžvilgiu prognozės koeficientų tiesinė algebrinių lygčių sistema [11].

Trūkumai:

- Lengva gauti nerealius parametrus.
- Ilgas skaičiavimas.

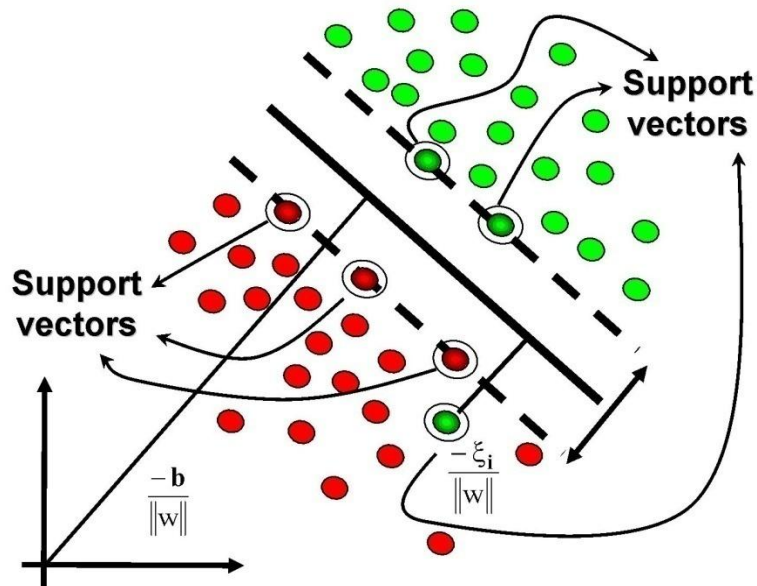
Norint gauti LPC koeficientus taip pat reikia atlikti kalbos signalo pirminius apdorojimus (pradinė filtracija, kadravimas, lango taikymas). Toliau galima taikyti autokoreliacijos, kovariacijos ir kitus metodus. Dažniausiai yra taikoma autokoreliacijos metodai, kurie padeda išgauti LPC koeficientus. Autokoreliacija galima išreikšti:

$$r_i(m) = \sum_{n=0}^{N-1-m} \tilde{x}_i(n) \tilde{x}_i(n+m) \quad m=0,1,\dots,p$$

Kur p yra didžiausia autokoreliacijos reikšmė, \tilde{x}_i - reikšmės po lango funkcijos pritaikymo, N – kadro dydis [12].

1.4.5. Atraminių vektorių mašinos

Atraminių vektorių klasifikatoriai (AVK) (angl Support Vector Machines SVM), panaudodami mokymo duomenis sukuria klasifikavimo taisykles. Pagrindinė vektorių mašinų idėja yra dviejų skirtingų klasių mokymo duomenų projektavimas i hiperplokštumas, atskiriančias dvi klases. Hiperplokštuma konstruojama maksimizuojant atstumą tarp dviejų artimiausių mokymo duomenų. SVM modeio pavyzdys pateiktas 1.7 paveiksle. Matome kaip pasiskirsto vektoriai, kai yra skirtingos klasės. Kairėje pusėje klasė yra lygi 1 ($f(x)<0$), o dešinėje klasė lygi 0 ($f(x)>0$).



1.7. pav. SVM pavyzdys

Naudojama Kernelio funkcija, kurią galima išreikšti:

$$f(x) = \sum_{i=1}^N \alpha_i t_i K(x_i x_i) + d$$

Kur t_i yra idealus išėjimai, $\sum_{i=1}^N \alpha_i = 0$, ir $\alpha_i > 0$. Vektoriai x_i yra atraminiai vektoriai gauti iš mokymo procesų. Idealus išėjimai yra -1 arba 1, priklausomai koks atramos klasės vektorius 0 ar 1. Kernelio išraiška turi tam tikras savybes (Mercerio būklę), tai galima išreikšti:

$$K(x,y) = b(x) \cdot b(y)$$

Kur $b(x)$ yra įvestis iš žemėlapiu (koordinatė)[13].

1.4.6. Gauso mišinio modelis

Populiarus stochastinis metodas, kuris gerai modeliuoja statistinį požymių kitimą. Tai statistinis pavaizdavimas, kaip kalbėtojas taria garsus. Gauso mišinių metodas, dažnai naudojamas, nes jis yra nepriklausomas nuo teksto asmens indentifikavimui. Tai suteikia tam tikro technikos, kuri leidžia sukurti geresnę sistemą.

Gauso mišinių tankis yra pateiktas kaip jo komponentių tankių vidurkių vektoriais ir mišinių svoriais. Taigi, kiekvienas kalbėtojas aprašomas tam tikrų parametrų rinkiniais.

Panaudoti Gauso mišinių tankius, siekiant sukurti kalbančiojo modelį:

- Individualus komponentių tankiai gali modeliuoti tam tikrų akustinių klasių rinkinį.
- Gauso funkcijų tiesinė kombinacija gali atvaizduoti didelę pavyzdžių pasiskirstymų klasę.

Gauso mišinių modelio tankio funkcija aprašoma :

$$p(\vec{x}) = \sum_{i=1}^M p_i b_i(\vec{x}),$$

Kur \vec{x} - D-dimensinis vektorius, kuris yra atsitiktinis, $b_i(\vec{x})$ – komponentių tankiai ir p_i , $i=1...M$ – mišinių svoriai. Galima išreikšti Gauso funkcija kiekvienai komponentei $b_i(\vec{x})$:

$$b_i(\vec{x}) = \frac{1}{\|\Sigma_i^{1/2}\| (2\pi)^{\frac{D}{2}}} \exp^{-\frac{1}{2}(\vec{x} - \vec{\mu}_i)' \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i)},$$

Kur $\vec{\mu}_i$ yra vidurkių vektorius, Σ_i^{-1} - kovariacinė matrica [14].

Gauso mišinių tankis yra pateiktas kaip jo komponentių tankių vidurkių vektoriais, mišinių svoriais bei kovariacinėmis matricomis:

$$\lambda = \{p_i \vec{\mu}_i, \Sigma_i\} \quad i = 1, 2, \dots, M.$$

Naudojant GMM, priklausomai nuo kovariacinės matricos, gali turėti kelias formas:

- Kiekvieną mišinio komponentę gali turėti savo kovariacinę matricą,
- Modelyje gali būti viena kovariacinė matrica visoms Gauso komponentėms,
- Viena kovariacinė matrica visiems kalbėtojams [14].

Maksimalaus tikėtinumo parametrų įvertinimas

Turint kalbėtojų požymių vektorius, modelio tikslas yra įvertinti GMM parametrus λ , kurie geriausiai atitinka mokymo metu gautus vektorių pasiskirstymus. Labiausiai paplitęs metodas yra maksimalaus tikėtinumo (ML) algoritmas. Reikia rasti modelio parametrus ($\lambda = \{p_i, \vec{\mu}_i, \Sigma_i\} i = 1, 2, \dots, M$), kurie maksimizuoja GMM modelio tikėtinumą. Turint tokią seką $x = \vec{x}_1, \dots, \vec{x}_T$, GMM tikėtinumas užrašomas taip:

$$P(X|\lambda) = \prod_{t=1}^T p(\vec{x}_t|\lambda)$$

- Turint pradinį modelį λ , reikia įvertinti naują modelį $\bar{\lambda}$, kad tenkintų lygybę – $p(X|\bar{\lambda}) \geq p(X|\lambda)$. Tada naujasis modelis tampa pradinio modeliu ir vykdoma sekanti iteracija, kol bus pasiektas tam tikras konvergencijos slenkstis. Dažniausiai iteracija apsisuka 10 kartų. Iteracija vykdoma panaudojant matematinės vilties maksimizavimą [14].

Atliekant iteraciją yra perskaičiuojami koeficientai.

Svorių koeficientai:

$$\bar{p}_i = \frac{1}{T} \sum_{t=1}^T p(i|\vec{x}_t, \lambda)$$

Vidurkiai:

$$\bar{\mu}_i = \frac{1}{T} \frac{\sum_{t=1}^T p(i|\vec{x}_t, \lambda) \vec{x}_t}{\sum_{t=1}^T p(i|\vec{x}_t, \lambda)}$$

Dispersija:

$$\bar{\sigma}_i^2 = \frac{\sum_{t=1}^T p(i|\vec{x}_t, \lambda) \vec{x}_t^2}{\sum_{t=1}^T p(i|\vec{x}_t, \lambda)} - \bar{\mu}_i^2$$

Paskutinis etapas, reikia naudoti aposteriorinę tikimybę. Tikslas rasti kalbačiojo modelį, turintį maksimalią aposteriorinę tikimybę [14].

$$p(i|\vec{x}_t, \lambda) = \frac{p_i b_i(\vec{x}_t)}{\sum_{k=1}^M p_k b_k(\vec{x}_t)}$$

Kalbėtojo identifikavimui turime S kalbėtojų $S = \{1, 2, \dots, S\}$, kurie atitinkamai yra GMM $\lambda_1, \lambda_2, \dots, \lambda_S$. Pagrindinis tikslas yra surasti kalbėtojo modelį, turint maksimalią aposteriorinę tikimybę:

$$S = \arg \max_{1 \leq k \leq S} Pr(\lambda_k | X) = \arg \max_{1 \leq k \leq S} \frac{p(X | \lambda_k) \Pr(\lambda_k)}{p(X)},$$

Pritaikant Bajeso taisyklę ($\Pr(\lambda_k) = 1/S$), tai visi kalbėtojai yra vienodai tikėtini ir $p(X)$ yra visiems vienodas tai galima supaprastinti:

$$S = \arg \max_{1 \leq k \leq S} p(X | \lambda_k)$$

Skaičiavimams galima naudoti logaritmus:

$$S = \arg \max_{1 \leq k \leq S} \sum_{t=1}^T \log p(\vec{x}_t | \lambda_k).$$

Kur $p(\vec{x}_t | \lambda_k)$ yra tikėtinumas [12].

Vienas iš pagrindinių GMM metodų trūkumas, kad didėjant kalbėtojų skaičiui taip pat mažėja atpažinimo procentas. Lentelėje, galime pamatyti, kaip keičiasi tikslumas [15].

1. Lentelė. GMM atpažinimo palyginimas

Kalbėtojų skaičius	Atpažinimas procentais (%)
10	98
20	95
30	91
40	89
50	83

Kadangi apdorotas duomenų skaičius kinta nuo kalbos įrašų - tai įrašų trukmės turi priklausomybę atpažinimo procentui. Užsienio lietratūroje teigiama, kad didėjant kalbos įrašų trukmei, turi didėti atpažinimo procentas.

2. Lentelė. EER palyginimas pagal įrašų trukmę (daugiklis 0,1)

Požymiai	Mišinių svoris	EER (%)			
		Įrašo trukmė			
		3 sec	7 sec	10 sec	20 sec
MFKK	4	50	39	28,4	18
	8	40	27,3	17,2	5,8
	16	69	44,5	26	12,6
	32	75	56	42	27,5

Galime pamatyti, kaip keičiasi lygių klaidų vertė didėjant įrašų trukmei. Didėjant trukmei, taip didėja atpažinimo koeficientas, nes mažėja lygių klaidų vertė (EER). Kai EER yra mažas (idealus atvejis, kai $EER < 1$), tai tuoj tikslensė sistema. Taip pat matome, kad

keičiant mišinių svorį keičiasi EER. Tai rodo, kad svoris skaičius įtakoja atpažinimo kokybę. Pateiktoje lentelėje, matome pateiktą Indų studentų tyrimą, naudojant Gold Wave programinę įrangą. Atlikta su 120 įrašų anglų kalba [16].

2. KALBOS SIGNALŲ POŽYMIAI

Šiame skiriuje pateiksime, kaip yra gaunami kalbos požymiai. Kalbėtojo atpažinimo sistemoje yra reikalingi kalbos požymiai, nes pagal juos yra atpažstamas kalbėtojas.

2.1. Kalba

Gerklose esančios balso klostės, sukurią žmogaus balsą – garsą. Taip pat prisidedant lūpos, dantims bei liežiui galima sukurti kitų garsų. Tai gali būti tiek smulkūs pakitimai individualiuose garsuose, tiek ir bendras tonas, kuriuo kalbama, kuris gali parodyti, jog ištarta frazė yra klausimas, nors to nesimato gramatiškai, taip pat išreikšti kalbėtojo požiūrį apie kalbėjimo objektą[17].

Atsižvelgus, kaip žmogus išgauna balsą, galima teigti kad:

- plaučiai - energijos šaltinis;
- balso stygos generuoja tam tikro dažnio bangas;
- Kalbos padargai – sukurią garsą bei jį modifikuoja.

Vienas iš svarbiausių padargų yra burnos ertmė, nes keisdami jos formą galima keisti garsą. Taip pat svarbi ir nosies ertmė, nes išgaunami nosiniai garsai.

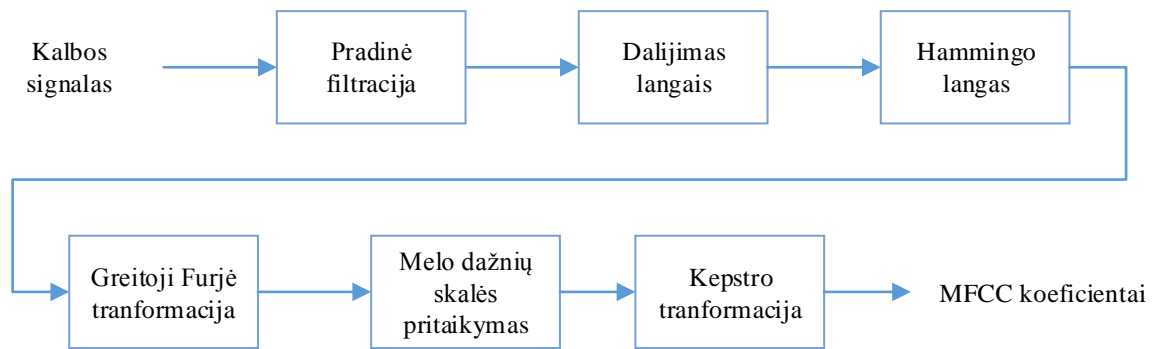
Kiekviename signale galime rasti triukšmą, dėl kurio suprastėja garso kokybė. Nepageidaujamas garsas atsiranda dėl aplink sąlygų, todėl turint signalą su triukšmu, sunku suprasti perteiktą informaciją.

Triukšmas, kurio stiprumas kinta, taip pat suvokiamas kaip stipresnis nei pastovaus intensyvumo garsas. Triukšmas atrodo silpnesnis, kai girdintysis gali lokalizuoti triukšmo šaltinį. Taip pat didesniu triukšmu laikomi garsai, kurie tą patį stiprumą pasiekia per trumpesnę laiką (staigus garsumo padidėjimas) [18].

Aukšto dažnio garsas suvokiamas, kaip labiau triukšmingas, bet žemo dažnio triukšmas yra pavojingesnis sveikatai.

2.2. Pirminis signalų apdorojimas

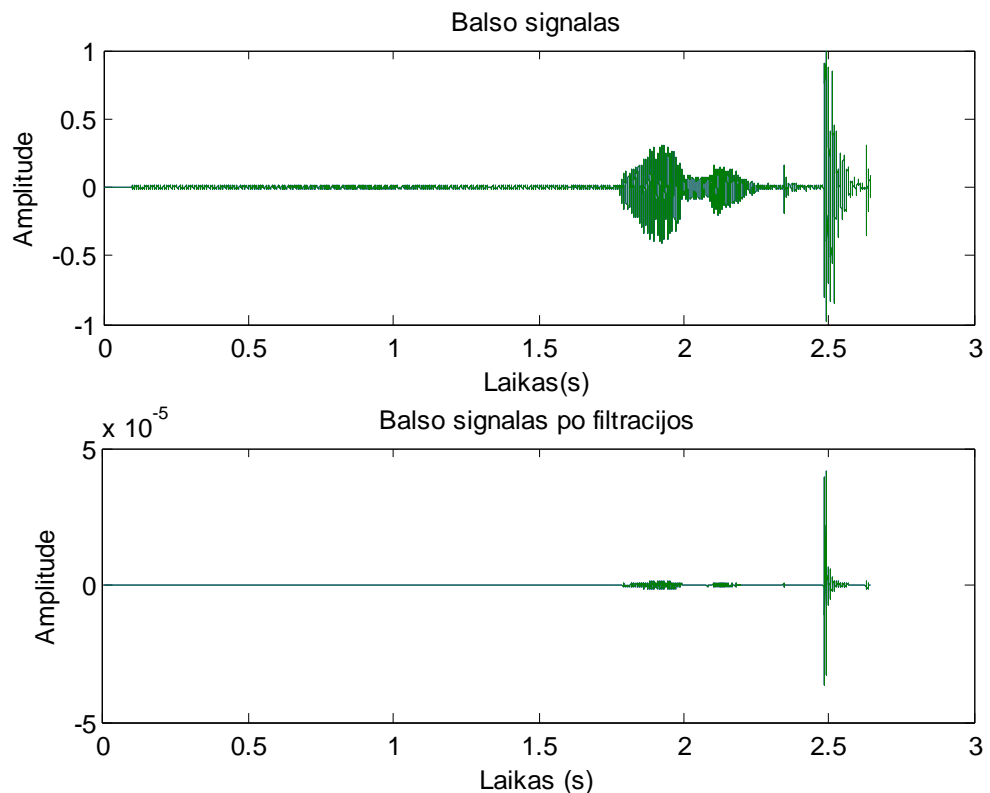
Šiame skyriuje aptarsime keletas medotų, kuriais apdoroti kalbos signalus. Pagal pateikta 4.1 paveikslą, visi kalbos signalai bus apdoroti šiais etapais, norint išgauti kalbos požymius. Išgauti kalbos požymiai leidžia panaudoti atliekant kalbėtojo sistemos modeliavimui. Gautus požymius galima apdoroti su programiniu paketu bei atlikti tyrimą kaip atpažįstą kalbėtoją. Taigi, kiekvieną etapą trumpai aptarsime.



2.1. pav. MFCC koeficientų gavimas

2.2.1. Pradinė filtracija

Aukštesniuose dažniuose kalbos signalo spektras nėra tiek daug išsidėstęs nei žemų dažnių srityje. Todėl dažniausiai išryškinašamos žemų dažnių formantės, nors aukštos formantės taip pat turi daug informacijos apie kalbą. Pradinės filtracijos tikslas iškelti aukštesnes formantes, kurios turi daug informacijos apie kalbantįjį asmenį, bei pagerinti naudojamų garso signalo kokybę [19]. Kaip keičiasi garso signalas matome 4.2. paveiksle.



2.2. pav. Pradinė filtracija

Norint išvengti triukšmo įtakos naudojama pradinė filtracija. Dažniausiai aukšto dažnio filtras naudojamas su $\alpha=0,95$ koeficientu:

$$\text{kai } y(t)=x(t)-0.95x(t-1)$$

kur α - koeficientas, apsprendžiantis šnekos signalo spektro išlyginimo laipsnį, jis yra parenkamas intervale $0,9 \leq \alpha \leq 1,0$; $x(t)$ – įėjimo duomenys; $y(t)$ išėjimo.

2.2.2. Kadrovimas

Kalbos signalas dalinamas į kadrus labai svarbus kuriant kalbos ir kalbančiojo atpažinimo sistemas, nes daroma prielaida, kad per trumpą laiko tarpą žmogaus balso trakto parametrai nespėja pasikeisti. Trumpame laiko tarpe galima aprašyti žmogaus balso parametrus. Spektriniai kalbos signalų požymiai yra išryškunami iš intervalų. Kadru trukmės yra apie 20-30 ms (dažniausiai 25 ms). Atskiras kadras gali būti išreikštas:

$$S(j,n) = s(j*(N-O)+n),$$

čia $s(n)$ – originalus signalas, $s(j,n)$ – j -tasis kadras, N – kadro ilgis atskaitomis,

O – gretimų kadru persiklojimo ilgis atskaitomis.

2.5.3. Lango funkcijos taikymas

Pritaikant lango funkciją siekiama pagerinti signalo kokybę paskirstant triukšmą tolygiai. Kalbos signalo kadrai yra dauginami is tam tikro lango funkcijos $w(n)$. Taigi signalą po lango funkcijos galime išreikšti:

$$s_w(n) = s(n)*w(n).$$

Lango funkcija dažniausiai pasirenkama Hammingo, Hanningo arba Gauso funkcija. Abu langai sumažina kraštines kalbos kadro reikšmes, o tai padidina signalo/triukšmo santykį/dažnių srityje [19].

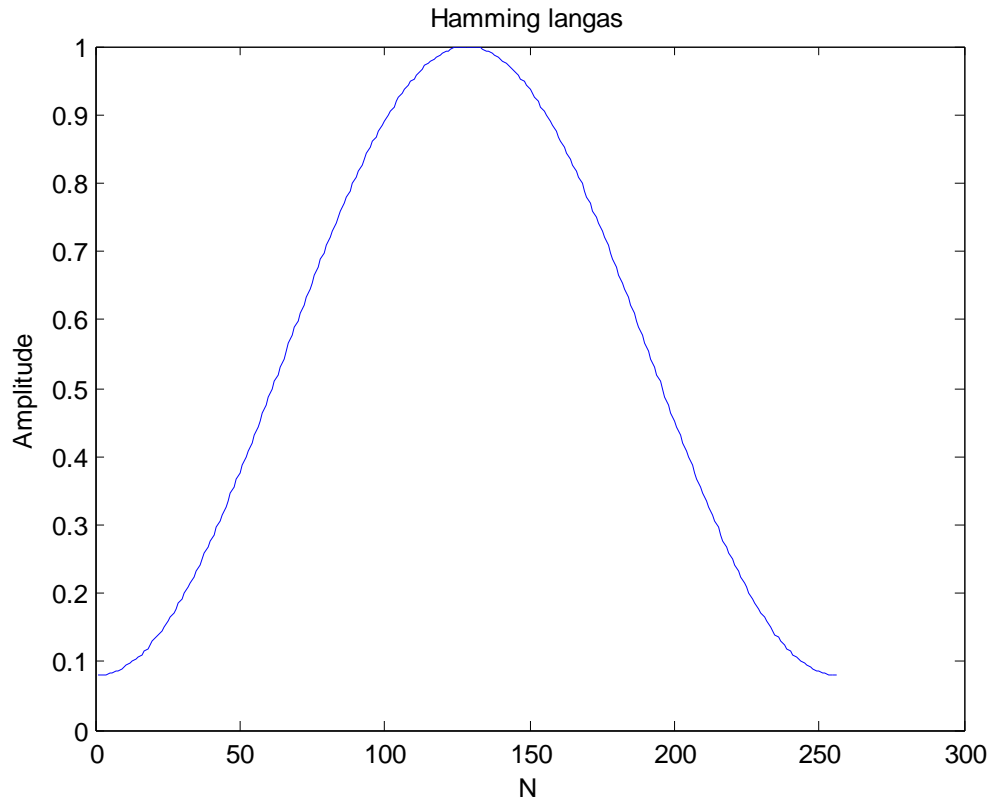
Hemingo lango funkcija:

$$W(n)=\begin{cases} 0.54 - 0.46 * \cos\left(2\pi \frac{n}{N-1}\right), & 1 \leq n \leq N - 1; \\ 0 & \end{cases}$$

Haningo lango funkcija:

$$W(n)=\begin{cases} 0.5 - 0.5 * \cos\left(2\pi \frac{n}{N}\right), & 1 \leq n \leq N; \\ 0 & \end{cases}$$

Kalbėtojo atpažinime yra naudojamas Hamming langas, kuris pateiktas 4.3. paveiksle.



2.3. pav. Hamming langas

2.3. Melo fazės kepsro koeficientai

Norint gauti kalbėtojo kalbos požymius reikalinga gauti melo fazės kepsro koeficientus. Juos galima gauti atliekant keletą punktų.

Pirmiausia tai greitosios Furjė transformacijos panaudojimas, kuris pateikia dažnių spektrą. Tai gana populiarus įrankis, norint užfiksuoti konkrečias kalbėtojo charakteristikas, kuris mums padės atliekant kalbėtojo atpažinime.

Tai daug greitesnė transformacija negu paprasta Furjė transformacija, kuri kiekviena kadrą išskleidžia pagal dažnį.

$$X_k = \sum_{n=0}^{N-1} X_n e^{-j2\pi kn/N}, k = 0, 1, 2, \dots, N-1$$

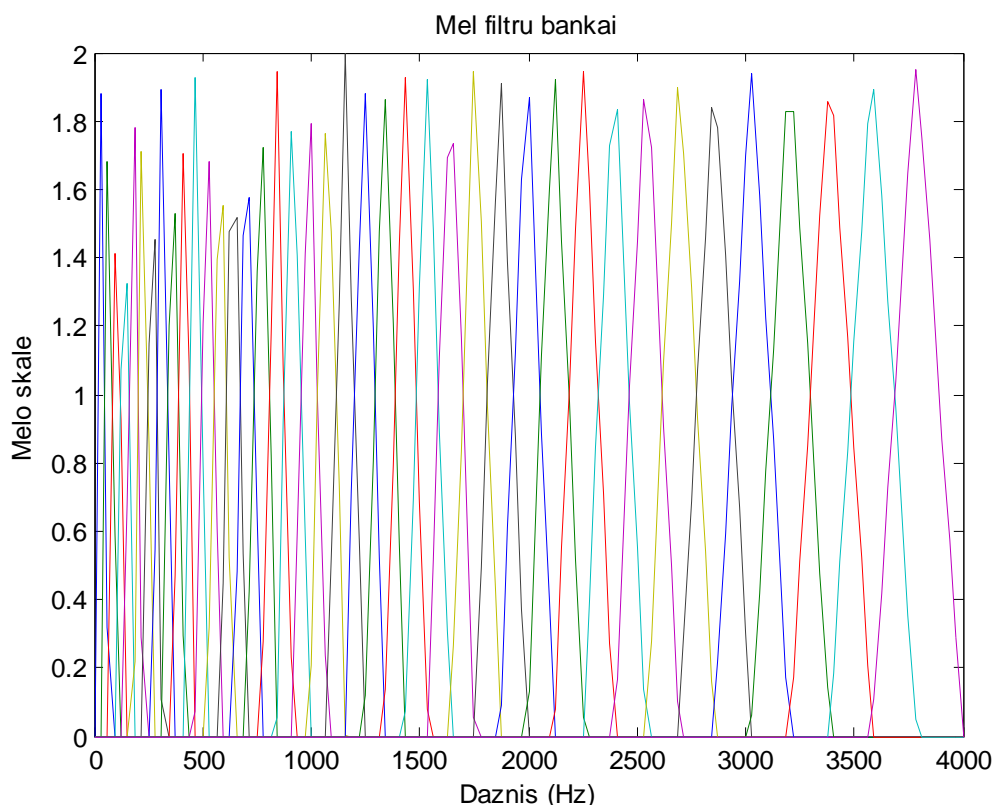
Kur, $k=0, 1, 2, \dots, N-1$ [19].

Atlikus greitąją Furjė transformaciją reikia naudoti filtrų bankus, nes nėra pašalinamas informacijos perteklius. Taip pat filtrų bankai padeda sumažinti koduojama bitų skaičių bei surandama kiekvienai juostai tenkanti energija.

Prieš projektuojant filtrų bankus, reikėtų pagalvoti, kad žmogaus klausa yra jautresnė žemesnių dažnių energijai. Todėl dažniausiai yra naudojama Melo (Mel) skalė. Dažnio f atvaizdavimas į Melo skalę atliekamas pagal pateiktą lygybę:

$$\text{Mel}(f) = 1125 * \ln(1 + f/700)$$

Melo skalėje filtrai turi vienodus intervalus. Todėl tiesinėje skalėje žemesniuose dažniuose filtro plotis žymiai mažesnis, negu aukštesniuose dažniuose. Melo skalė iki 1000 Hz yra lėjinis išdėstymas, o virš 1000 Hz yra logaritminis. Dažniausiai filtro forma parenkama trikampė, pateikta 1.3 paveiksle.



2.4. pav. Melo skalė

Paskutiniame žingsnyje reikia naudoti diskrečiąją kosinusų transformaciją. Gautus melų dažnių skalės kepsro koeficientus galime naudoti atliekant kalbėtojo atpažinimą [19].

$$c(j) = \sum_{i=0}^{I-1} C(i) * \cos\left(\pi n \frac{i-1}{2I}\right), j = 0, 1, \dots, I.$$

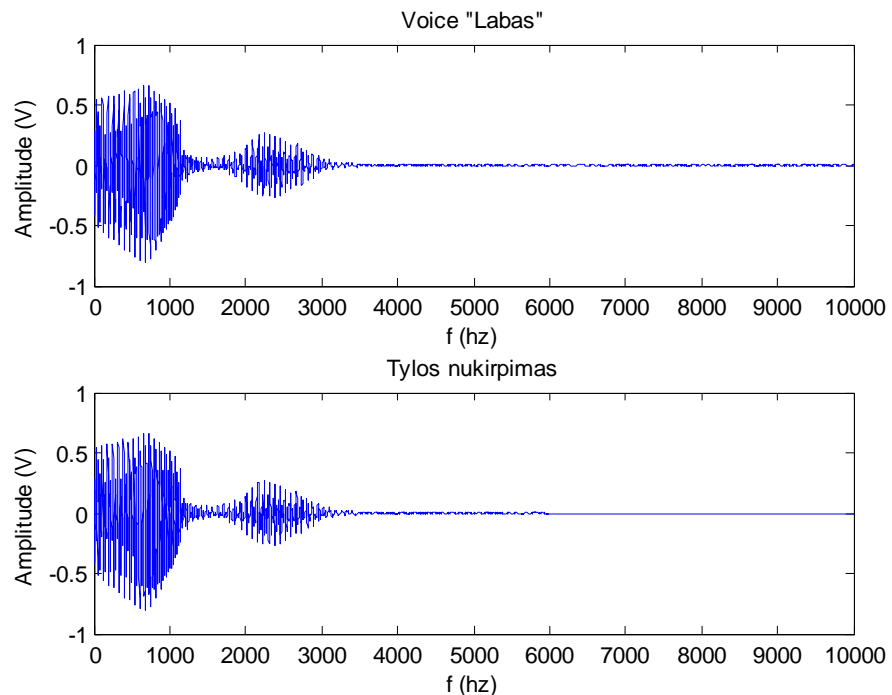
Kur I yra filtrų skaičius, kuris paprastai kinta nuo 24 iki 40 [12]. Kalbėtojo atpažinime nauduosime 13.

3. NEPRIKLAUSOMO NUO TEKSTO KALBĖTOJO ATPAŽINIMAS

Šiame skyriuje atliksime tyrimus su GMM, bei naudosime nemokamus garso įrašus. Taip pat atliksime tyrimus su skirtingais įrašais. Atliksime „labas“ žodžio triukšmo bei tylos pašalinimą.

3.1. Triukšmo pašalinimas

Programiniu paketu „Matlab“ įrašome žmogaus balsą, kur bus ištariamas žodis „labas“. Įrašę balsą galime atlikti pagrindinius veiksmus balso požymiams išreikšti ar iškirpti tylos atkarpas. Pirmiausia atazduodami balso įrašą vektoriais ar matricos pavidalu turime iškirpti atkarpas, kuriose žmogus tylėjo (pradžioje ar pabaigoje). Tada galėsime naudoti kitas operacijas. Įrašytas balsas su programa „SoundRecorder“, garso įrašas nėra „švarus“, nes įrašytas kompiuterio pagalba. Tylos signalo nukirpimas pateiktas 3.1 paveiksle.

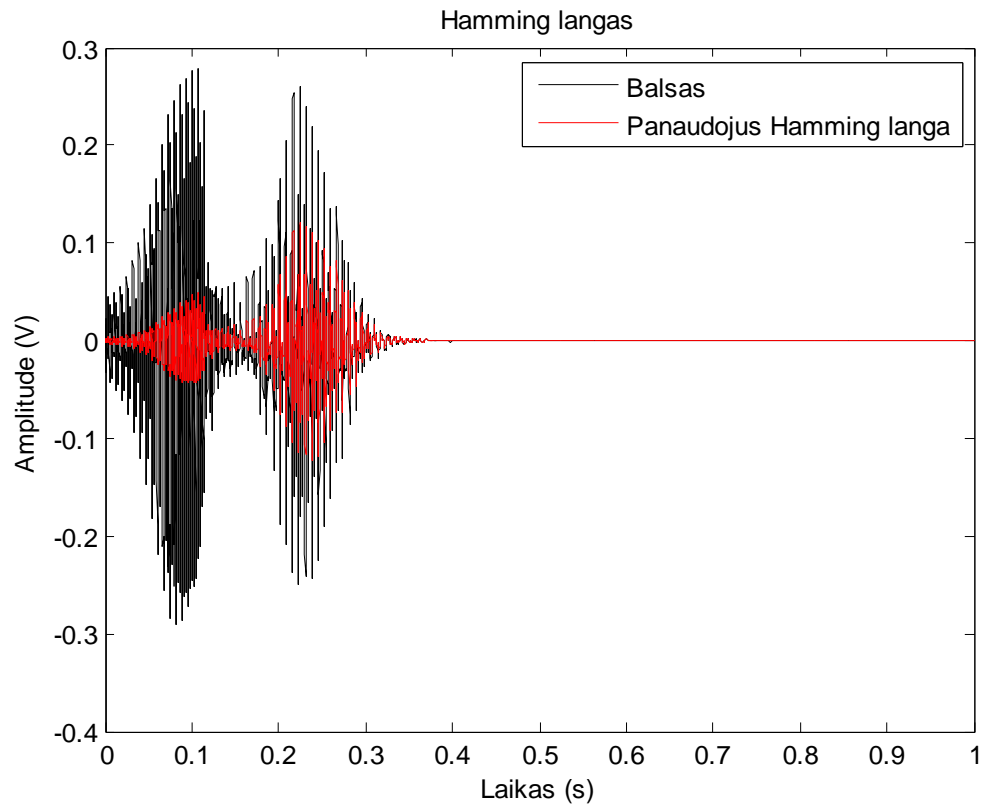


3.1. pav. Tylaus signalo nukirpimas.

Paveiksle matome, jog signalas yra su tam tikrais triukšmais. Pirmame signalo yra tam tikras fono triukšmas, bet panaudojus tylos panaikinimą jos nebelieka. Toliau galime atlikti kitus veiksmus.

Sekantis žingsnis procese yra langų pritaikymas kiekvienam freimui. Tikslas yra minimizuoti signalo nutrūkimus freimo pradžioje ir gale. Tam yra naudojamas langas skirtas sumažinti spektrinį iškraipymą, kuris mažina signalą iki nulio freimo pradžioje ir pabaigoje. Po

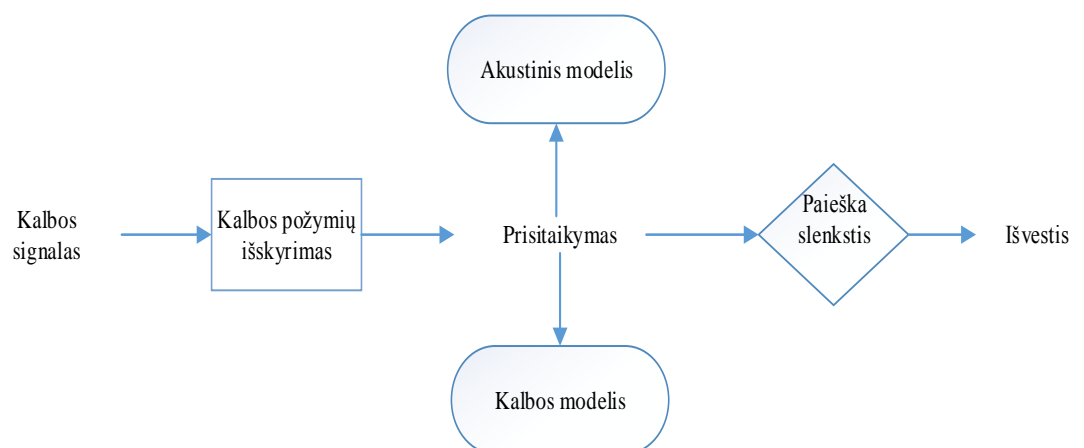
hamming funkcijos signalas pateiktas 3.2. paveiksle. Paveiksle matome, jog hamming funkcija signalą sumažino, nebėra tokios didelio spektrinio iškraipymo.



3.2. pav. Hamming funkcijos pritaikymas

3.2. Kalbėtojo atpažinimo struktūra

Pirmiausia reikia turėti keletą žmonių, jog galėtume atlikti tyrimą. Kalbėtojo atpažinimui galima naudoti ir kitokias žodžių junginių su tais pačiais žmonėmis ir stebėti kaip atpažinama. GMM atpažinimo sistemos struktūra pateikta 3.3. paveiksle. Kalbos požymių išskyrimas reikalingas sistemai, kad galėtume atlikti eksperimentą.



3.3. Pav. Kalbėtojo atpažinimo struktūra

Kitas etapas, kalbos požymius reikia įvesti į Matlab ToolBox sistemą, kur bus atliekamas tyrimas. Automatinė atpažinimo sistema tai pagrindinis programos scriptas, pagal kurį yra nurodoma kaip įvertinamos balso požymiai ir atliekamas atpažinimas. Gauso mišinių tankių panaudojimas kalbančiojo atpažinimui yra tai, kad funkcijų tiesinė kombinacija gali atvaizduoti didelę pavyzdžių pasiskirstymų klasę. Kalbančiojo požymių pasiskirstymą apibūdina vidurkio vektoriumi ir kovariacine matrica. Taigi, sistema naudoja verifikavimo metodą.

3.3. Kalbėtojų verifikavimo sistema

Šiame skyriuje aptarsime verifikavimo sistemą. Verifikavimo sistema turi du uždavinius, kuriuos reikia išspręsti:

H_0 : X yra kalbėtojo hipotezė s ;

H_1 : X nėra kalbėtojo hipotezė s ;

Nulinė hipotezė H_0 atstovauja kalbėtojo modeliui, išreikštam λ_s . Kita vertus, alternatyvi hipotezė H_1 atstovauja kalbėtojo modeliui λ_{alt} . Optimalus sprendinys tarp dviejų hipotezių yra tikimybės testas:

$$\frac{p(X|\lambda_s)}{p(X|\lambda_{alt})} \begin{cases} \geq \theta \\ < \theta \end{cases}$$

Taikant logaritmus statistikai, galime išreikšti:

$$\log p(X|\lambda_s) - \log p(X|\lambda_{alt})$$

Alternatyvi hipotezė λ_{alt} gauta iš didelio kiekio kalbėtojų modelių ir treniruota vienam GMM. Todėl ją galima vadinti universaliu fono modeliu (Universal Background Model - UBM) [20].

Toliau taikant šį modelį, visi treniruojami vektoriai yra sudėti į viena klasterį ir suskaičiuojama vidutinė vertė. Kiekvieną kartą dalijantis, klasterių vidurkis papildomas nedidele reikšme ε ir atsiranda du nauji klasteriai:

$$\mu_i^+ = \mu_i(1 + \varepsilon)$$

$$\mu_i^- = \mu_i(1 - \varepsilon)$$

Po dalijimosi, modelis perskaičiuoja kelis kartus naudodamas maksimalaus tiketinum algoritmą.

Verifikavimo sistemoje, kalbėtojo modelis yra sukuriamas pritaikant parametrus iš universalus modelio, naudojant treniravimo parametrus. Taip pat žinoma, kad Bajeso pritaikymas arba maksimalus aposteriorinis pritaikymas, leidžia greičiau vykdyti skaičiavimus. Toliau yra skaičiuojami mišinių svoriai, vidurkis bei kovariacijos, kurios yra paminėtos GMM metode.

$$n_m = \sum_{t=1}^T \gamma_{t,m} X_t, \quad 1 \leq m \leq M$$

$$E_m(X) = \frac{\sum_{t=1}^T \gamma_{t,m} X_t}{\sum_{t=1}^T \gamma_{t,m}} \quad 1 \leq m \leq M$$

$$E_m(X)^2 = \frac{\sum_{t=1}^T \gamma_{t,m} X_t^2}{\sum_{t=1}^T \gamma_{t,m}} \quad 1 \leq m \leq M$$

Toliau yra reikalinga atnaujinti senas statistikas iš UBM, sukuriant pritaikytas GMM mišinių svorius, vidurkius ir kovariacijas:

$$\hat{p}_m = \left[\frac{\alpha_m^p n_m}{T} + (1 - \alpha_m^p) p_m \right] \gamma, \quad 1 \leq m \leq M$$

$$\hat{\mu}_m = \alpha_m^\mu E_m(x) + (1 - \alpha_m^\mu) \mu_m, \quad 1 \leq m \leq M$$

$$\hat{\sigma}_m^2 = \alpha_m^\sigma E_m(x)^2 + (1 - \alpha_m^\sigma) (\sigma_m^2 + \mu_m^2) - \hat{\mu}_m^2, \quad 1 \leq m \leq M$$

Skalės faktorius γ skaičiuojamas per visiems pritaikytiems mišinių svoriams, kad susidėtų į vieną t.y. $\sum_{m=1}^M \hat{p}_m = 1$. Kur $\{\alpha_m^p, \alpha_m^\mu, \alpha_m^\sigma\}$ yra pritaikyti koeficientai svoriams, vidurkiams ir kovariacijoms [20].

Kiekvieno derinimo koeficientais išreiškiamas taip:

$$\alpha_m^p = \frac{n_m}{n_m + r^p}$$

Kur p priklauso $\{p, \mu, \sigma\}$ ir r^p yra fiksuotas parametrai $p, r=16$.

Pabaigoje, mums reikia surasti S pritaikytus GMM parametrus $\lambda_1 \lambda_2 \dots \lambda_S$ ir λ_{ubm} . Tarkime, kad T vektoriaus dimensija D yra nežinomo kalbėtojo testas, duotas $X = \{X_t \in \mathbb{R}^D; 1 \leq t \leq T\}$. Tai reikia padaryti sprendimą. Sudaroma hipotezė:

- $H_0: X$ hipotezė s
- $H_1: X$ nėra hipotezė s

Ir galiausiai pritaikant logaritminę tikėtinumo vertę (log likelihood ratio):

$$L(x) = \Phi_s - \Phi_{ubm}$$

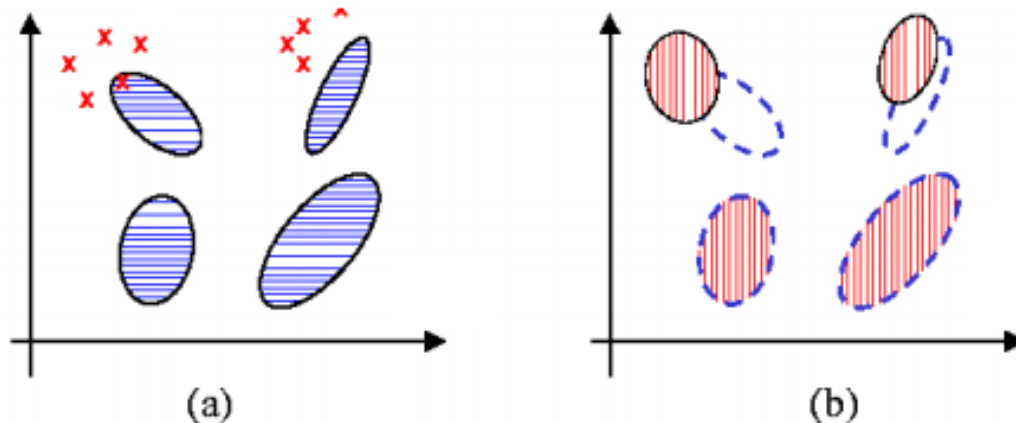
kur Φ_s ir Φ_{ubm} yra panašumo balai iš požymių vektorių, jie apskaičiuojami:

$$\Phi_s = \sum_{t=1}^T \log p(x_t | \lambda_s)$$

$$\Phi_{ubm} = \sum_{t=1}^T \log p(x_t | \lambda_{ubm})$$

Ir jei tikimybės santykis viršija pasirinktą sprendimų ribą, tai pasirenkama hipotezė ir atvirkščiai [20].

Pateiktas 3.4 paveikslas, kuriame matome parametrų prisitaikymą.



3.4. pav. Klasterių susigrupavimas

Tikslas yra surasti kiekvienam kalbėtojui slenkstį, pagal kurį būtų atliekamas atpažinimas. Mokymo metu, požymių vektoriai yra lyginami su kiekvieno kalbėtojo modeliu, skaičiuojamas tikėtinumai. Problema, kad tie tikėtinumai yra labai skirtingi, todėl surasti vieną slenkstį, kuris atitiktų visiems kalbėtojams yra netikslinga. Taigi, kiekvienam kalbėtojui yra apskaičiuojamas individualus slenkstis, pagal kurį yra lyginamas.

3.4. Tyrimas su Matlab Toolbox paketu

Tyrimas atliekamas programine įranga „Matlab“, kurios pagalba galima koreguoti programinį kodą ir atlikti kalbėtojo atpažinimą. Taip pat naudojamas MATLAB Toolbox for Speaker-Recognition Research, kuris padeda nagrinėti sistemos veikimą. Matlab toolbox pagalba galime sukurti tam tikrų kalbėtojų balso požymius ir juos atpažinti. Kadangi MATLAB Toolbox for Speaker-Recognition Research sistema treniravimo požymius sukuria „random“ funkcija, bei testas atliekamas su tais pačiais požymiais – kalbėtojo atpažinimas būna idealus. Todėl mums reikia programinę įrangą patobulinti, kad galėtume naudoti savo garsyno požymius.

Atpažinimui naudojami nemokami įrašai iš Open Speech and Language Resources. Tyrimo eksperimentui naudojama „švari“ kalbėtojų įrašai, kurie leidžia gauti geresnius rezultatus. Įrašuose yra vyrų ir moterų įrašyti balsai. Įrašai naudojami atsitiktine tvarka, jie nėra grupuojami atskirai į moterų ar vyrų grupes. Įrašams apdoroti naudojama Voice box for Speaker recognition Analysis. Melspectrum pagalba galime gauti balso įrašo požymius, kurie vėliau padės kuriant kalbėtojo atpažinimo sistemą. Įrašui įkelti į „Matlab“ scriptą naudojame funkcija „wavread“.

...

```
[s, fs]=wavread(failoKelias);
```

...

- wavread funkcija leidžia įkelti kalbėtojų kalbos įrašus tik wav formatu. Duomenis perduoda kintamajam „s“ ir kalbos dažnis „Fs“.

Taip pat naudojama funkcija „melcepst“, dėl kalbos požymių išskyrimo iš VoiceBox. Šioje funkcijoje reikia pasirinkti pradinis parametrus, nes kitaip bus naudojami pradiniai tiekėjo nustatymai.

...

```
m=melcepst(s,fs);
```

...

- „melcepst“ funkcija leidžia išskleisti kalbėtojų požymius, kurie vadinasi melų fazės kepstro kefcientai. Pradiniai nustatymai būna 12 MFKK ir 256 kadrai. Kalbėtojo atpžinime naudojama 13 MFKK – tai dažniausias parenkamas dydis. Išėjimas pateikiamas, kaip MFKK matrica, kurios dydis išreiktas taip: duomenys matricoje (LxT) pavidalu, kai L= skaičius MFKK (tai 13) ir T= skaičius kalbos vektorių.

. Tyrimo tikslas gauti kuo mažesnę reikšmę tarp FPR ir tarp FNR. Gauta reikšmė tarp teisingų ir neteisingų priimtų reikšmė yra to modelio požymis, kaip jis gerai apmoko.

$$FNR = FN ./ (TP + FN);$$

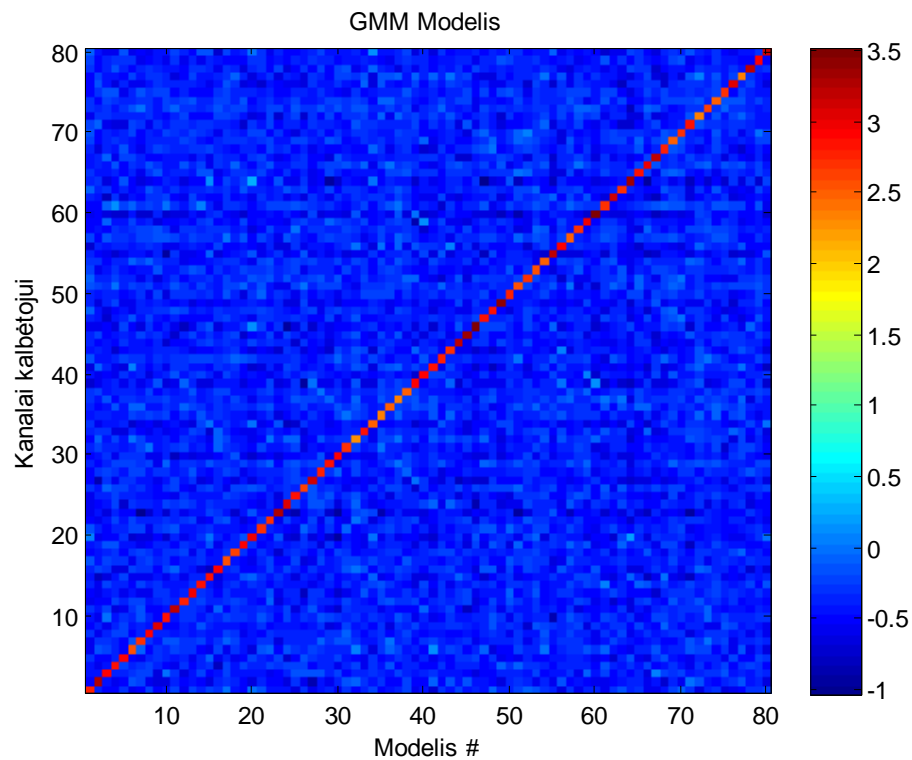
$$FPR = FP ./ (TN + FP);$$

Kur FP - skaičius klaidingai priimtų skaičius, TN skaičius teisingai atmetų.

Kur FN klaidinai atmetų skaičius, TP teisingai priimtų skaičius.

Atliekant tyrimus reikia, kad ERR – lygių kaidų lygis, būtų kuo mažesnis. Jei EER yra mažas dydis tai atpažinimo sistemos kokybė yra gera.

Pirmiausia reikia apdoroti garso failus. Tai reikia paleisti funkcija MFKK.m, kurį yra iš VoiceBox. Po funkcijos gauname MFKK koeficientus, kuriuos galime naudoti toliau siekiant atpažinti kalbėtojus. Toliau reikia paleisti iš Toolbox katalogo funkcija TRENIRAVIMAS.m. Ten reikia isikelti gautus MFKK, kad alétume pradėti apmokymus. Koeficientus įkeliame į TrainSpeakerData. Kadangi atliksime skirtingų kalbėtojų atpažinimą, tai reikia kitus kalbėtojų MFKK įkelti į TestSpeakerData. Taip pat galima leisti sistemai pačiai sukurti testavimo požymius.

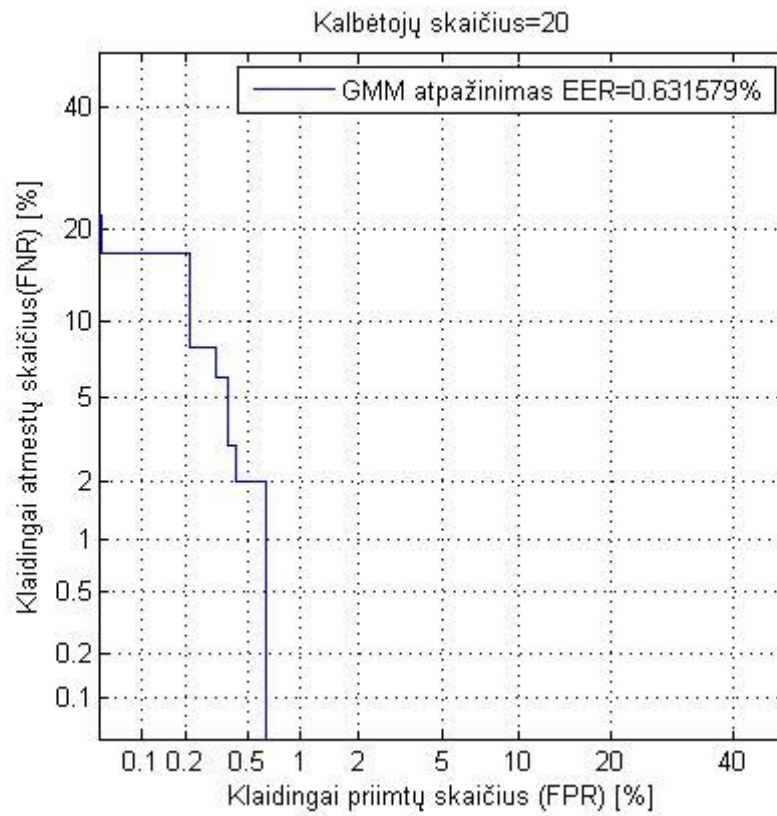


3.5. pav. Spalvų (colobar) palyginimas

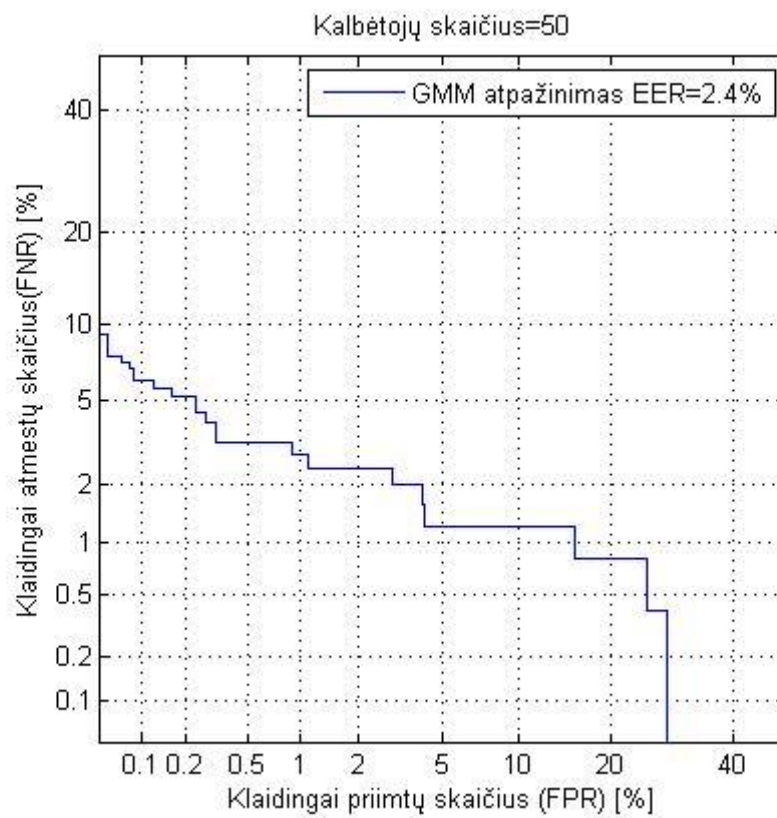
Kaip matome iš 3.5. paveikslo raudonos spalvos tiesė rodo idealų kalbėtojų atpažinimą. Toks idealus atvejis realiai negali būti gaunamas, nes atpažinimo sistemos nėra tokios tikslios kad taip gerai atpažintų visus kalbėtojus. GMM modelis turi trukumą, kad atpažinant kuo daugiau kalbėtojų (įvedus per daug informacijos), pradeda pristaikyti prie duomenų ir tai lemia mažas tikslumas.

Toliau tyrimas atliekamas su LibriSpeech įrašais, kai parenkami keli ar keliolika kalbėtojų ir naudojamas GMM modelis jų atpažinimui. Pavyzdžiui yra naudojama 10 ir 20 nepriklausomų nuo žodžių kalbėtojų ir jie yra treniruojami ir atpažįstami.

Galime pateikti keletą pavyzdžių, kaip sistema veikia ir atpažįstą kalbėtojus. Atliekant tyrimus galėsime keisti netik kalbėtojus, bet ir įrašų trukmes. Didėjant požymių skaičiui, taip pat didėja duomenų skaičius, kurį reikia apdoroti. Pirmiausia pradedant atlikti tyrimus, reikia surasti mišinių svorių skaičių su kuriuo geriausiai veikia sistema.



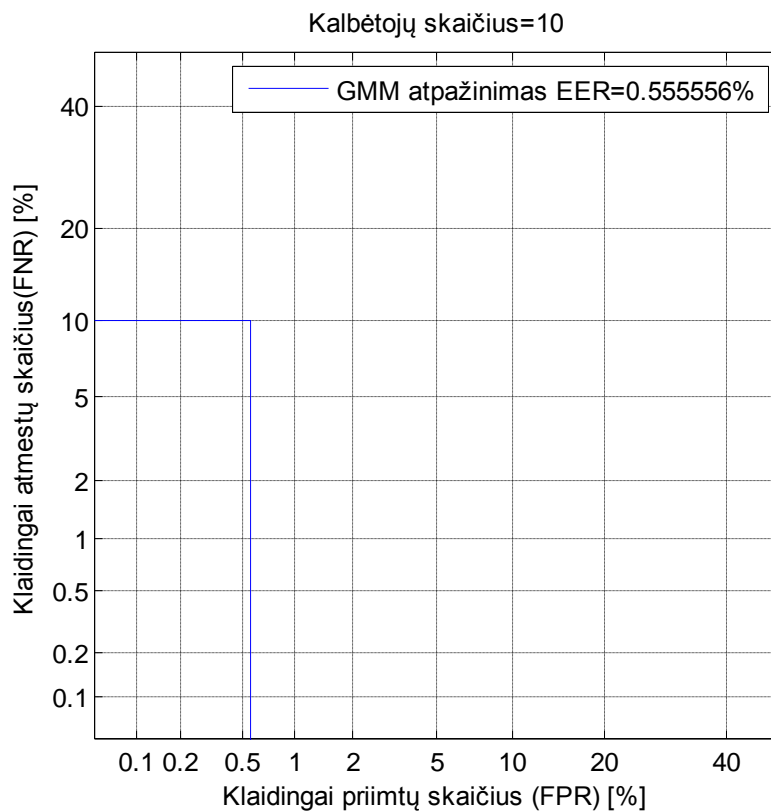
3.6.pav. ERR 20 kalbėtojų.



3.7.pav EER 50 kalbėtojų.

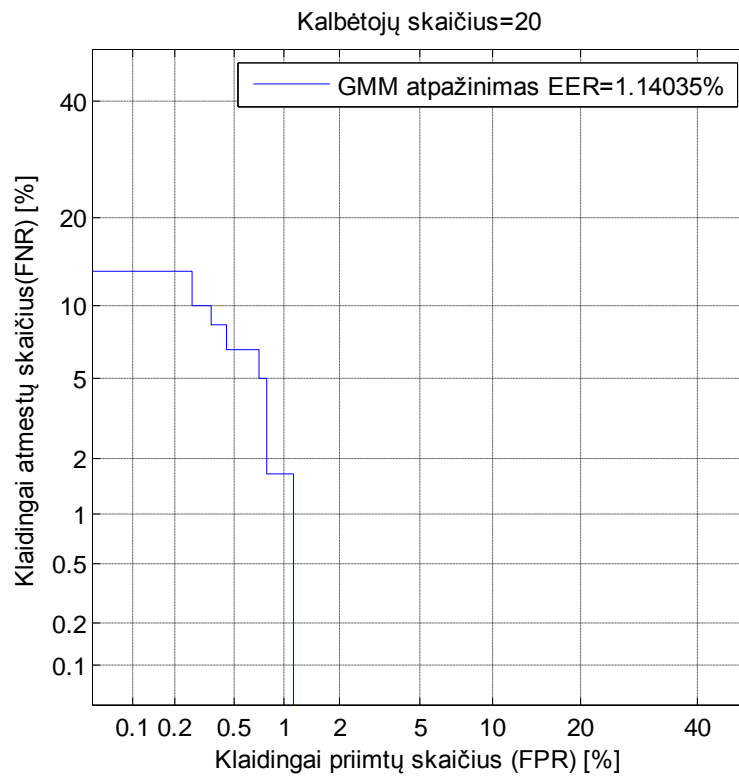
Pateiktuose paveiksluose matome, kaip sistema veikia. Esant daugiau kalbėtojų sistema atpažįsta prasčiau.

Taip pat pateiksime ir daugiau grafikų, kad pamatytume kaip keičiasi kalbėtojo atpažinimas, keičiantis kalbėtojų skaičiui. Įrašų trukmės yra įvairios, todėl jas suskirstome kelias grupes, kad tyriamą įrašai būtų vienodos trukmės. Visų garso failų dažnis 16000Hz.

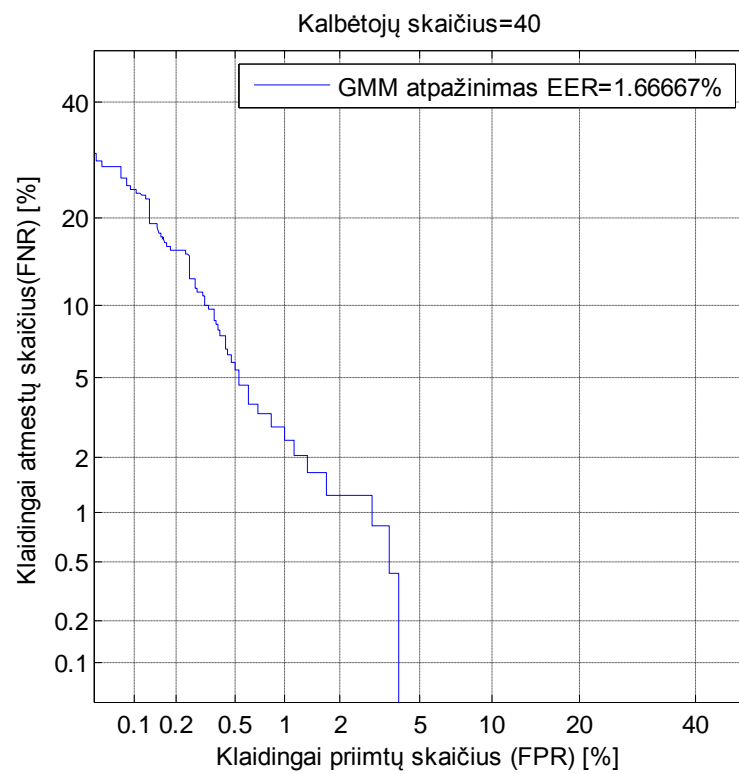


3.8. pav. Atpažinimas su 10 kalbėtojų

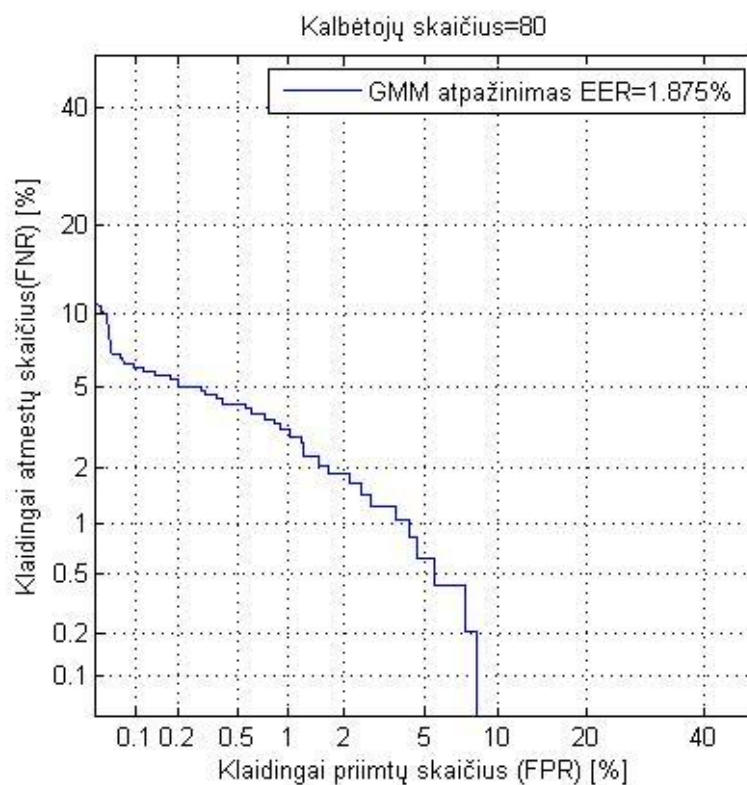
Matome, kad EER yra lygus 0,56 proc. Tai yra geras EER skaičius, nes esant mažiau kalbėtojų geriau veikia sistema.



3.9. pav. Atpažinimas su 20 kalbėtojų



3.10. pav. Atpažinimas su 40 kalbėtojų

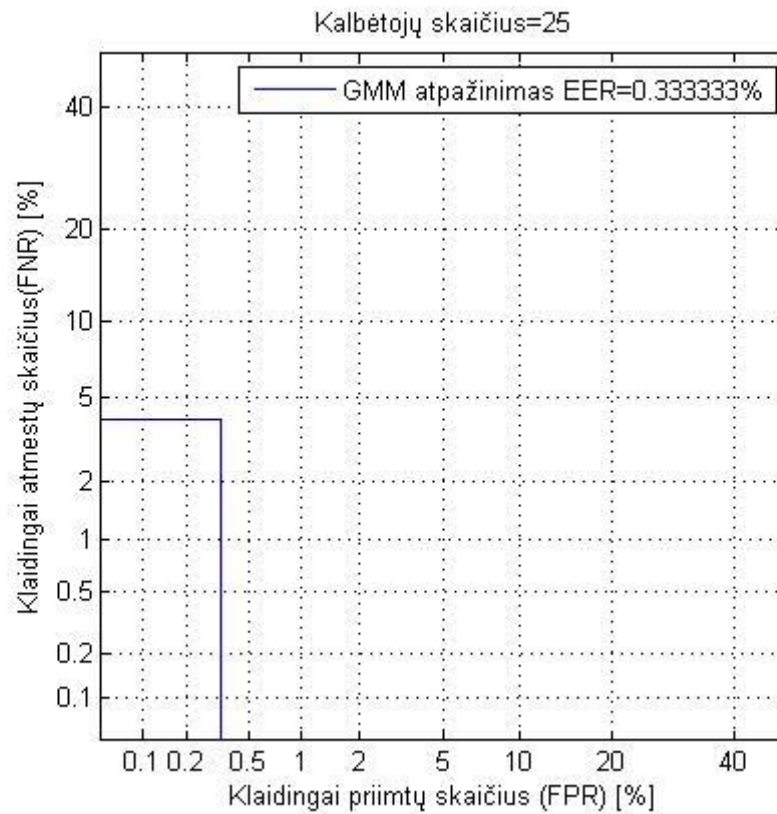


3.11. pav. Atpažinimas su 80 kalbėtojų

3.Lentelė. ERR palyginimas

Kalbėtojų skaičius	EER
10	1,1%
20	1,14%
40	1,67%
80	1,88%

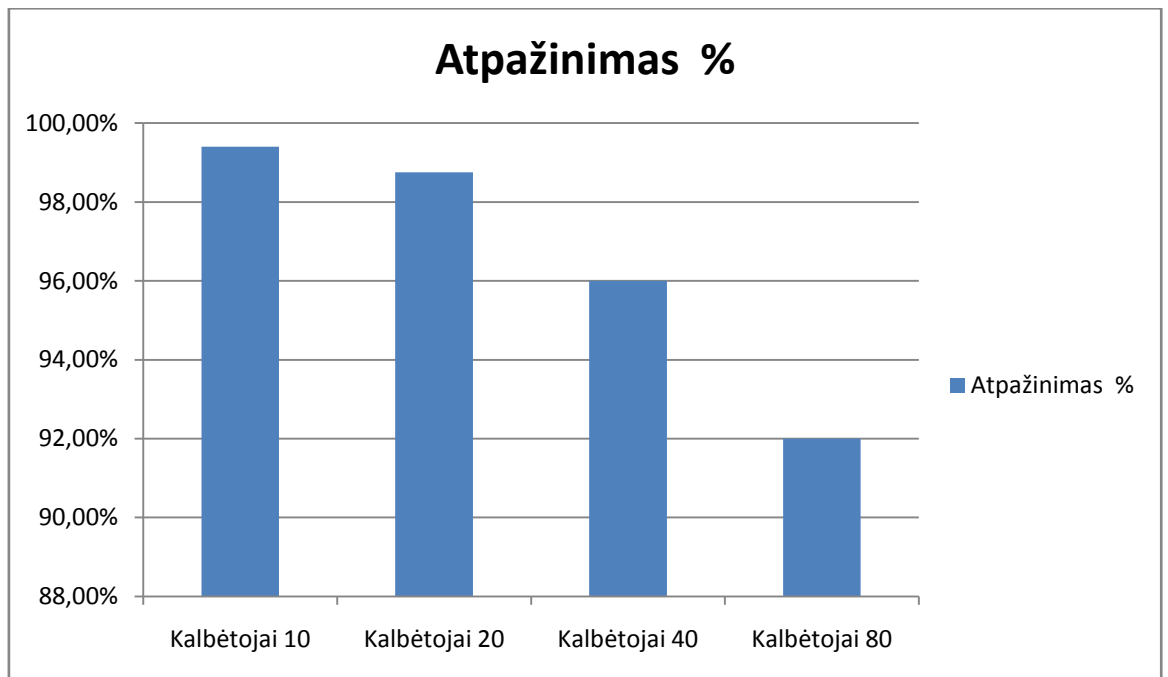
Pamėginkime atlikti su 9 sekundėmis trukmės įrašais, kai yra geros kokybės.



3.12.pav. EER su 25 kalbėtojais.

Matome, kad su geros kokybės įrašais atpažinimas yra geras.

Toliau galime gautus rezultatus išnagrinėti ir pateikti pirmas išvadas apie kalbėtojo atpažinimo sistemą. Lentelėje surašyti kalbėtojų skaičius bei jų vidutinė kvadratinė vertė (EER), kuri parodo kaip atpažįsta sistema kalbėtojus. Taigi, matome kad geriausiai sistema atpažįsta kai yra ne daug kalbėtojų įrašų, kuo daugiau sistemai tenka apdoroti duomenų tuo sistema tampa prastesnė. Histogramoje matoje, kaip mažėja atpažinimo patikimumas didėjant kalbėtojų skaičiui.



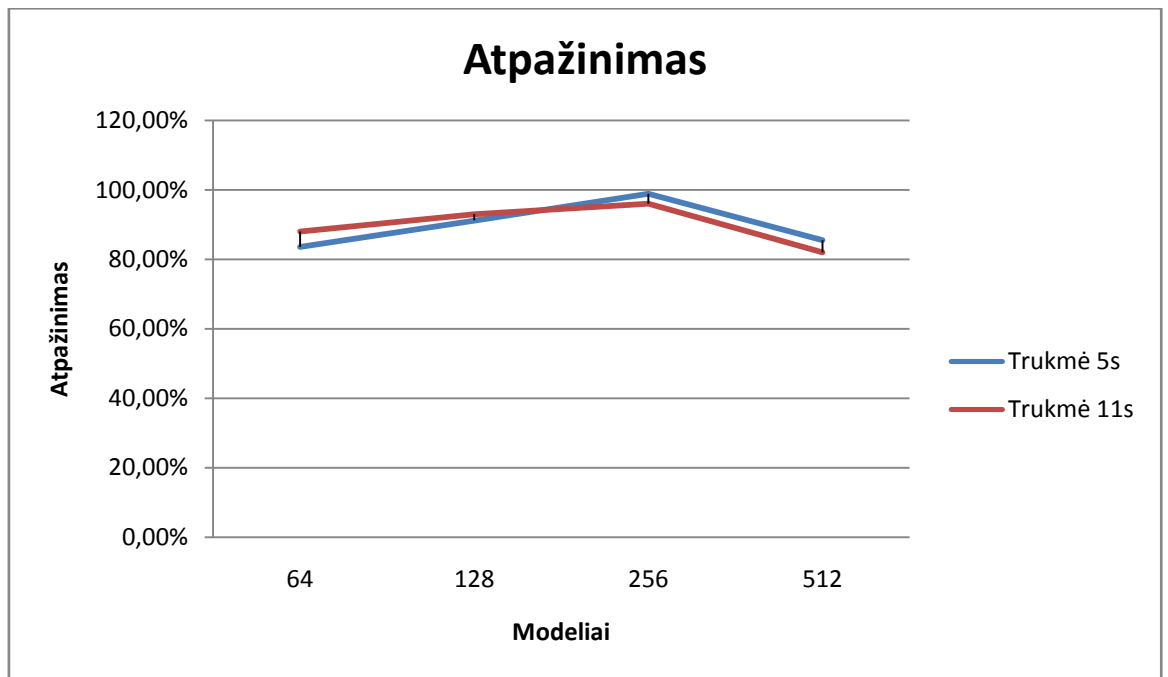
3.13.pav. Kalbėtojų atpažinimas

Matlab ToolBox programiniame pakete galima keisti Gauso mišinių skaičių, nuo kurio keičiasi atpažinimo sistemos kokybė. Literatūroje rašoma, kad pasiekus reikiamą dydį, toliau didinant mišinių skaičių, sistema pradeda prasčiau atpažinti kalbėtojus.

4. Lentelė. Atpažinimo palyginimas

Mišinių skaičius	Atpažinimas 5s proc.	Atpažinimas 11s proc.
64	83,6%	88%
128	91,2%	93%
256	98,9%	96%
512	85,5%	82%

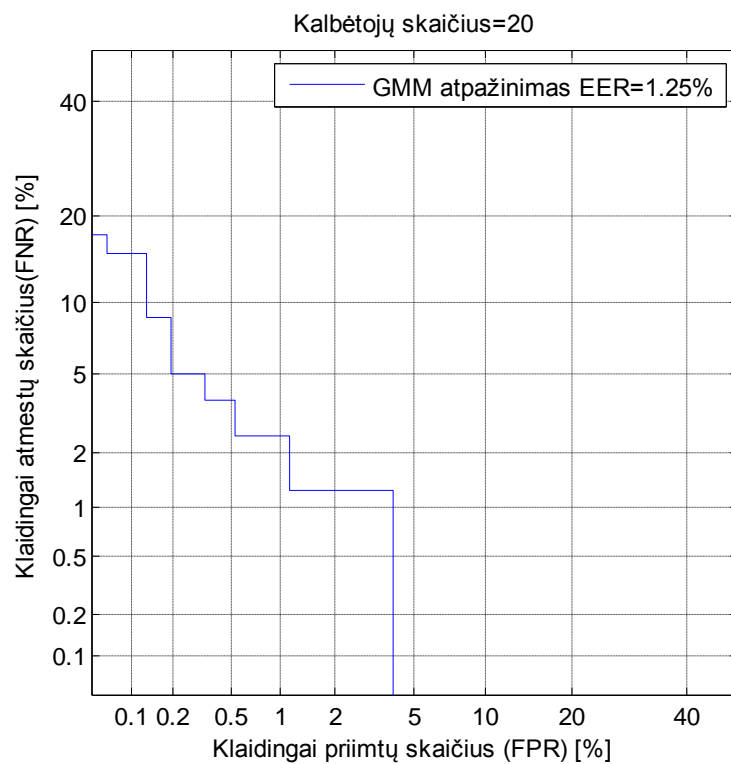
Lentelėje pateikiame, kaip atpažįsta kalbėtojus keičiant Gauso mišinių skaičių. Eksperimentui pasirinkta 10 skirtingų 5s trukmės įrašai bei 11s įrašai.



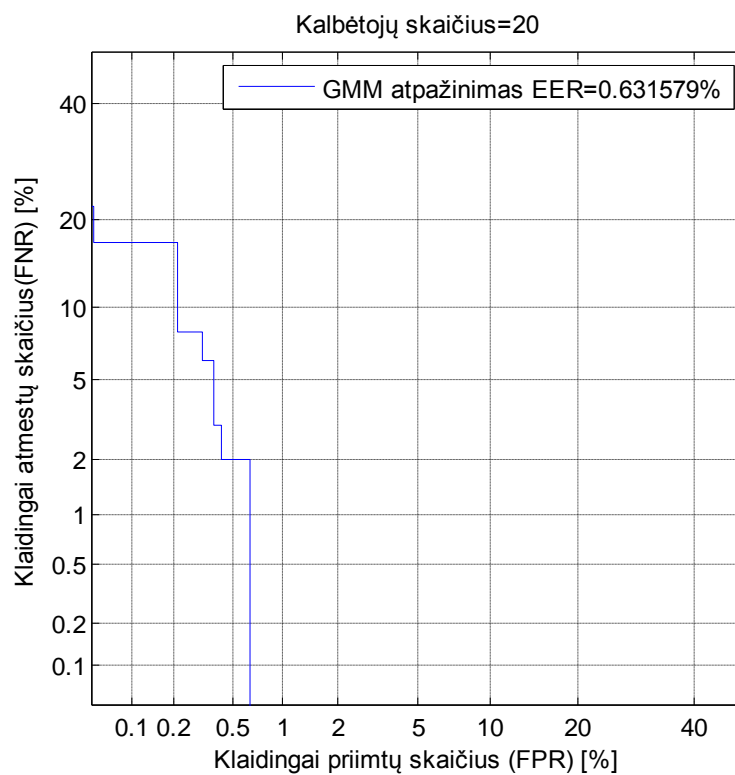
3.14.pav. Palyginimas svorių skaičiaus įtaka

Atlikus bandymus, galima teigti, kad didėjant įrašų trukmei galima didinti mišinių skaičių, nes atpažinimas ties 256 bei 512 mišinių skaičiaus procentais panašus. Kartais atliekant bandymą su 11s trukmės įrašų, buvo vienodas atpažinimas.

Kadangi kalbėtojų skaičius yra didelis tai galima susiskirstyti pagal tam tikrą skaičių kalbėtojų. Kai atliekamas eksperimentas su daug kalbėtojų sistema labai ilgai apdoroja visą informaciją, todėl eksperimente yra suskirstyti kalbėtojai. Pirmiausiai atlikime iki 7s trukmės įrašų atpažinimus, kurių yra 500 kalbėtojų įrašų. Visi įrašų kalba yra angliška ir visi yra skirtingi – kalbos tekstai bei intonacijos skirtingos. Prieš tai atlikus bandymus nustatėme, kad geriausiai sistema atpažįsta kalbėtojus kai mišinių skaičius lygus 256 ir 128. Tai toliau atliksime bandymus su šiais dydžiais ir stebėsime kaip keičiasi atpažinimo dydis.



3.15. pav. Atpažinimas su 512 skaičiumi



3.16. pav. Atpažinimas su 256 skaičiumi

Iš pateiktų grafikų matome, kad geriausiai atpažįsta kalbėtojus su 256 skaičiumi. Toliau galime atlikti daugiau bandymų didindami kalbėtojų skaičių ir keisdami kitus įrašus. Visus duomenis surašysime į lentelę.

5. Lentelė. Atpažinimo palyginimas

Įrašai	Mišinių svoriai	Atpažinimas proc.
60	512	89%
60	256	92%
80	256	89%
80	512	82%
100	256	78%
100	512	73%

Lentelėje surašyti rezultatai vėl gi parodo, kad geriausiai sistema atpažįsta kalbėtojus su 256 mišinių skaičiumi. Lentelėje pataikti 500 įrašų atpažinimas išskaidant po 50 įrašų. Atpažinimo laikas su 100 kalbėtojų yra ilgas ir atpažinimo kokybė prastesnė, nes sistema persimoko dėl didelių duomenų kiekių.

6. Lentelė. EER palyginimas

Kalbėtojų skaičius	50	50	50	50	50	50	50	50	50	50
EER proc.	1,2%	1,5%	2%	2,5%	2,4%	3,3%	3,64%	3,67%	2,8%	2,6%

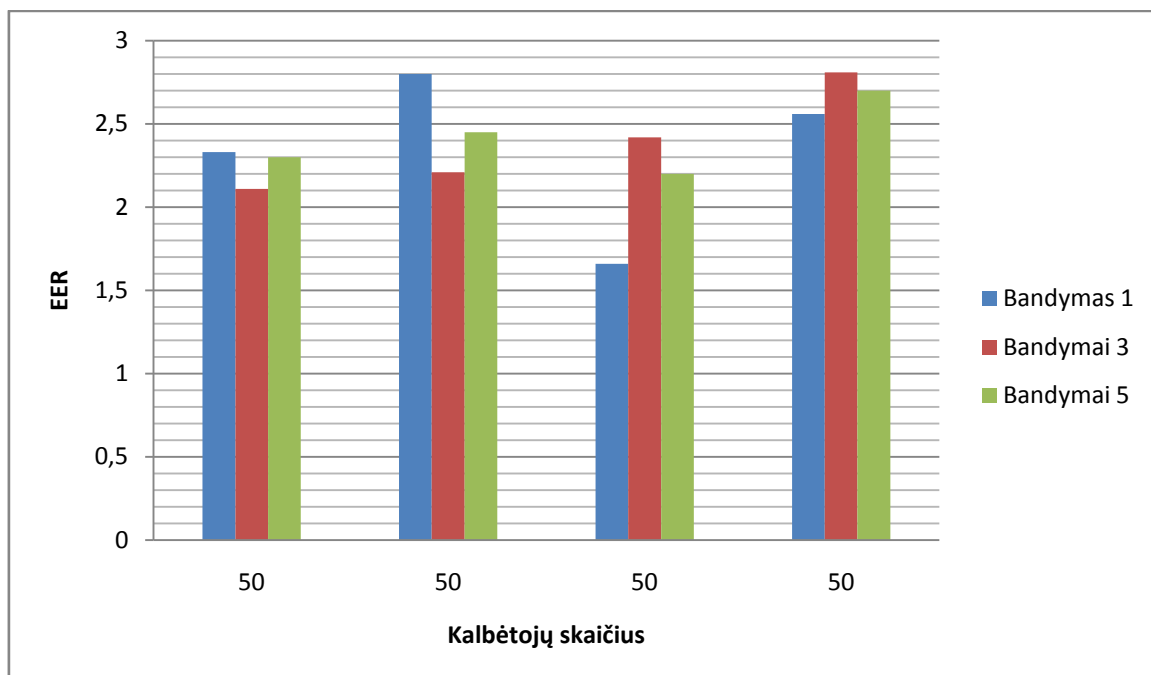
Iš pateiktų rezultatų matome, kaip atpažįsta sistema kalbėtojus. Kuo ilgesnė kalbėtojo įrašo trukmė tuo daugiau kalbos požymių. Požymiai lemia, kaip atpažins sistema kalbėtojus. Kai yra per daug požymių sistema gali veikti nepatikimai.

Toliau nagrinėsime kitus 200 garso įrašus, bei pateiksime atpažinimo rezultatus. Taip pat atliksime kelis kartus atpažinimus su tais pačiais įrašais ir apskaičiuosime vidurkius.

7. Lentelė. EER palyginimas

Kalbėtojų skaičius	Kartai	50	50	50	50
EER proc.	1	2,33%	2,8%	1,66%	2,56%
EER proc.	3	2,11%	2,21%	2,42%	2,81%
EER proc.	5	2,3%	2,45%	2,2%	2,70%

Pagal gautus rezultatus, galime juos palyginti pasinaudoje histograma. Matome, kad pirmais dviem atvejais, pirmas atpažinimo procesas atliekant 1 bandymą buvo šiek tiek blogesnis negu už 3 ir 5 bandymus. Kituose dviejuose bandymuose matome, kad pirmasis bandymas buvo geriausias..



3.17. pav. Bandymų histograma

Ilgų trukmių įrašai užima daug laiko tyrimui, nes didelis duomenų kiekis apsunkina sistemos veikimą. Vieno atpažinimo metu reikia laukti net 5 minutes, kol sistema parodo rezultatus.

Toliau atliekant tyrimus, kalbos įrašus sukirstėme pagal trukmes ir žiūrėsime, kaip keisis EER. Literatūroje teigiama, kad didėjant įrašų trukmei, didėja atpažinimo kokybė. Gauti rezultatai turėtų parodyti, jog EER mažėja. Įrašus programai užkrausime į treniravimo ir testavimo masyvus.

8. Lentelė. EER su 50 kalbėtojų

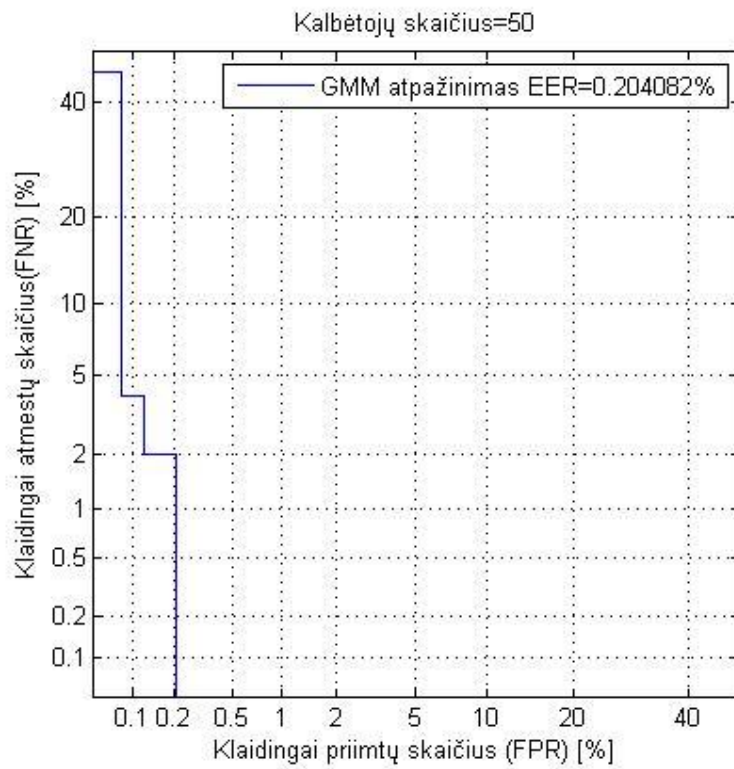
Požymiai	Mišinių svoris	EER (%)			
		Įrašo trukmė (sekundėmis)			
		5 s	7 s	10 s	20 s
MFKK	32	2,2	2,08	2,2	0,83
	64	2	1,59	2	1.5
	128	1,38	1,15	1,67	0.67
	256	0,82	1,02	0,89	0.83

Pateiktoje lentelėje galime įžvelgti, kad didėjant mišinių svorių skaičiui mažėja EER, bet kai kuriuose bandymuose EER yra didesnis. Geriausias atliktas bandymas, kai EER mažiausias tai kai yra 20 sekundžių trukmės įrašas. Kartais atliekant tyrimą buvo keičiama mišinių svorių skaičius net iki 1024 – siekiant surasti mažiausią EER. Jei sistema surado mažiausią EER, tai bandant keisti mišinių skaičių į 1024, rezultatas išlieka toks pat. Sistema negali išgauti geresnės kokybės.

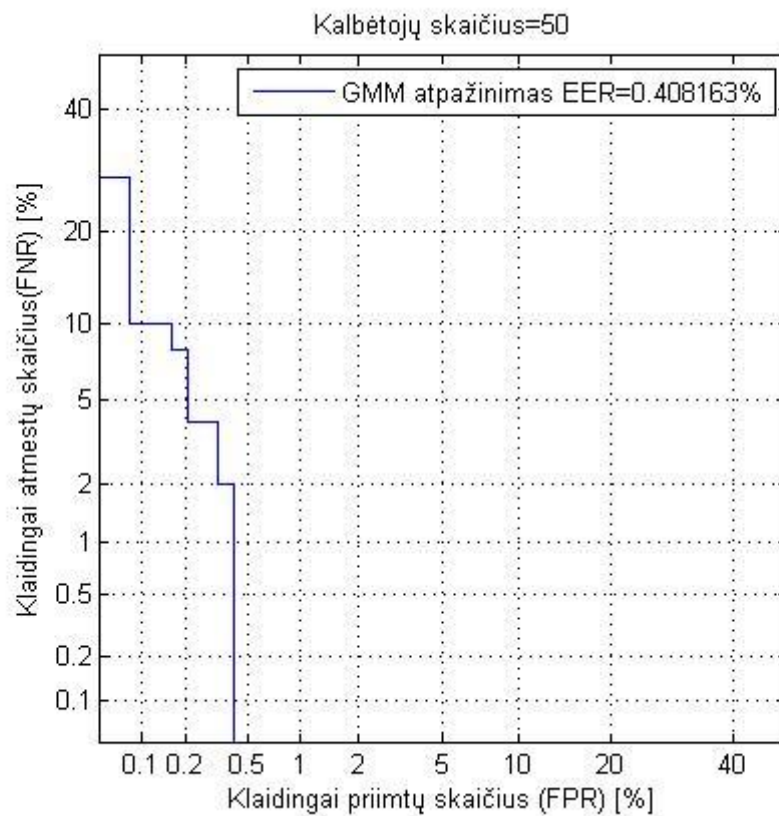
Kadangi naudojame geros kokybės įrašus, tai EER turėtų būti gana mažas. Toliau atliekamas tyrimas su 50 kalbėtojų, kai skiriasi jų požymių skaičius. Įrašų trukmės yra 4 sekundžių, bet požymių skaičius ženkliai skiriasi. Kadangi ieškant geresnių rezultatų, pastebėta, kad keičiantis požymių skaičiui, kartais reikia didinti mišinių svorių skaičių, norint gauti geriausią rezultatą. Toks bandymas surasti geriausią rezultatą užima daug laiko, todėl dažniausiai yra parenkama vienas ar keli svorių skaičiai. Kai mišinių svoris nustatomas 1024 ir turint ilgesnių trukmių įrašų, tenka laukti ilgai, kol gaunamas rezultatas. Todėl atliekant tyrimus reikia gana galingo kompiuterio, kad tilptų visi duomenys, kol apsimoko.

9. Lentelė. 50 kalbėtojų ERR

Mišinių svoris	EER (%)	
	Įrašo trukmė (sekundėmis)	
	Daugiau požymių	Mažiau požymių
32	1,55	1,3
64	0,65	0,98
128	0,41	0,61
256	0,204	0,41

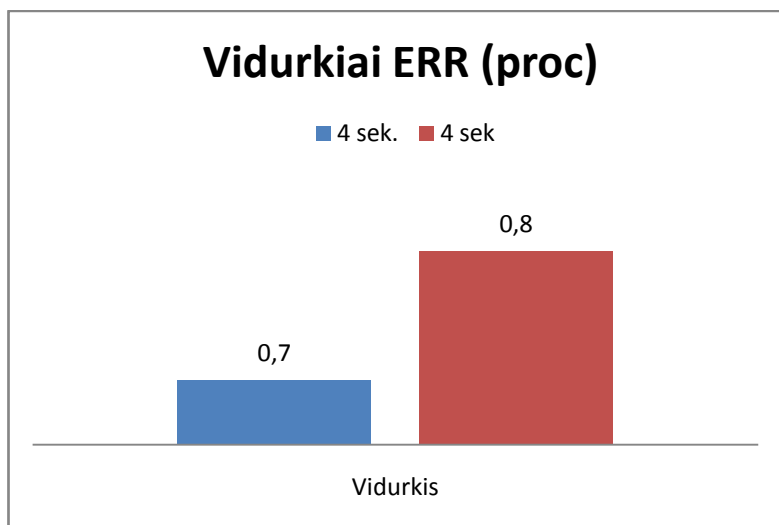


3.18.pav. EER su 4s trukmė.



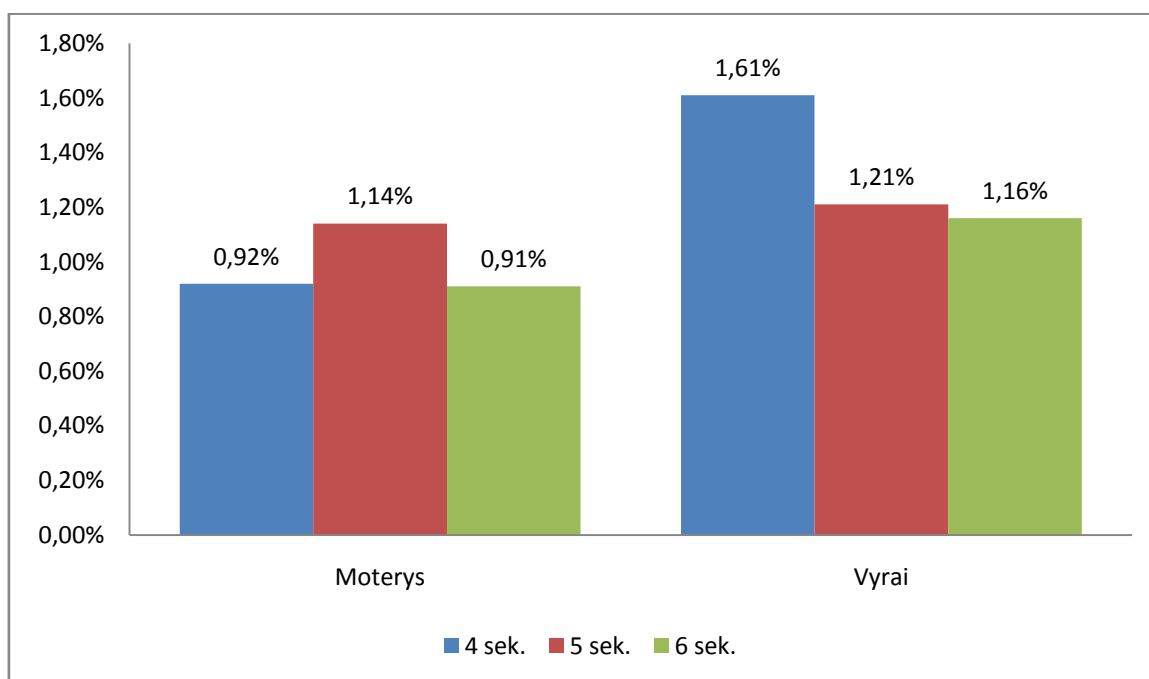
3.19. pav. ERR su 4s trukme.

Lentelėje matome, kad didėjant požymių skaičiui (didėjant įrašų trukmei), didėja atpažinimo kokybė. Galime pamatyti, kaip EER vidurkiai skiriasi. Matome, kad EER yra mažesnis, kai požymių daugiau.



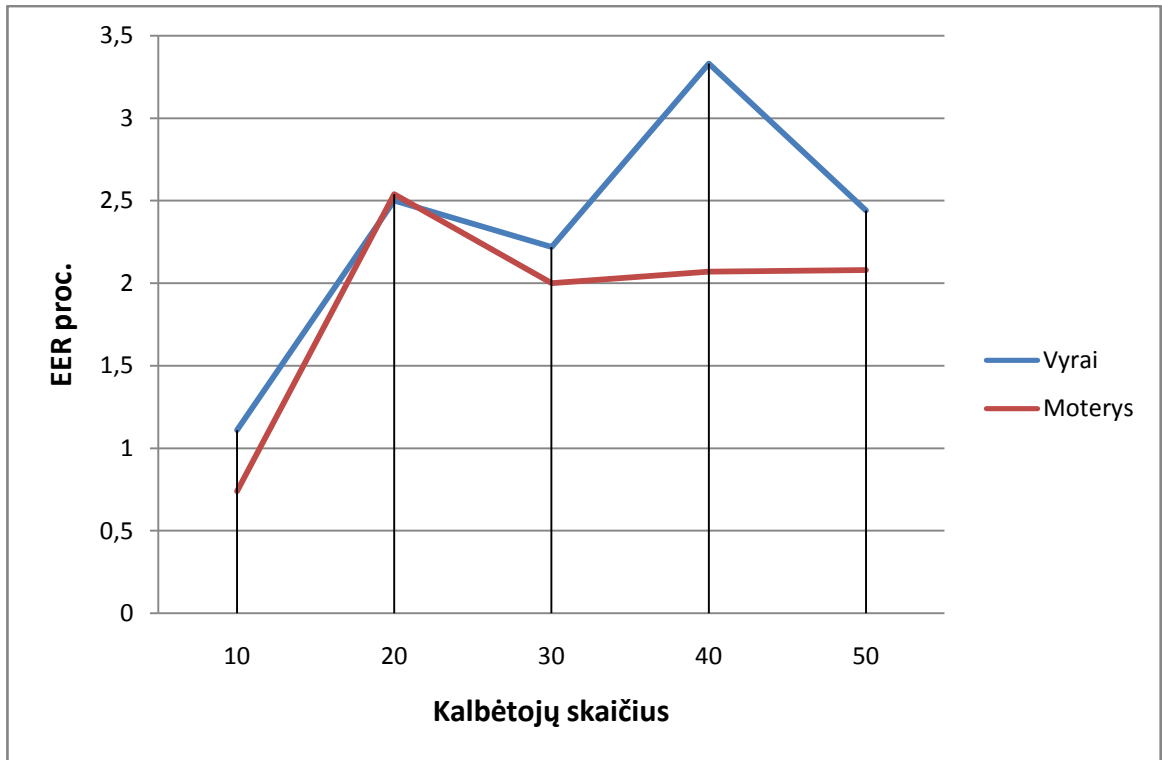
3.20. pav. Vidurkių palyginimas

Taip pat atliekama vyrų ir moterų balsų atpažinimas pagal įrašų trukmes. Siekiama pastebėti, kaip sistema atpažįsta kalbėtojus pagal lytis.



3.21. pav. Vidurkių palyginimas

Histogramoje matome, kad moterų balsus atpažįsta geriau. Galbut dėl to, kad naudojami įrašai yra kiti nei prieš tai naudoti. Naudojami „test“ įrašai, kurie yra blogesnės kokybės, siekiant įvertinti sistemą. EER mažesnis moterims galėjo įtakoti, kad parenkat įrašus buvo šiek tiek geresnės kokybės ar geresnis tarimas. Dar vienas palyginimas pateiktas didinant kalbėtojų skaičių.



3.22.pav. Vyrų ir moterų atpažinimas

Iš grafiko matome, kad tiek vyrų tiek moterų balsai yra skirtingi, jų atpažinimas taip pat skiriasi. Vyrų basų atpažinimas tiek 40 įrašų yra blogiau atpažinamas, todėl jų EER yra didesnis. Taip pat matome, kad ties 20 kalbėtojų EER yra panašus.

Toliau atliekant tyrimus su įrašais, galime naudoti funkcija, kur prideda baltą Gauso triukšmą. Galime atlikti tyrimą, kaip sistema veikia su dirbtinai įvestu triukšmu.

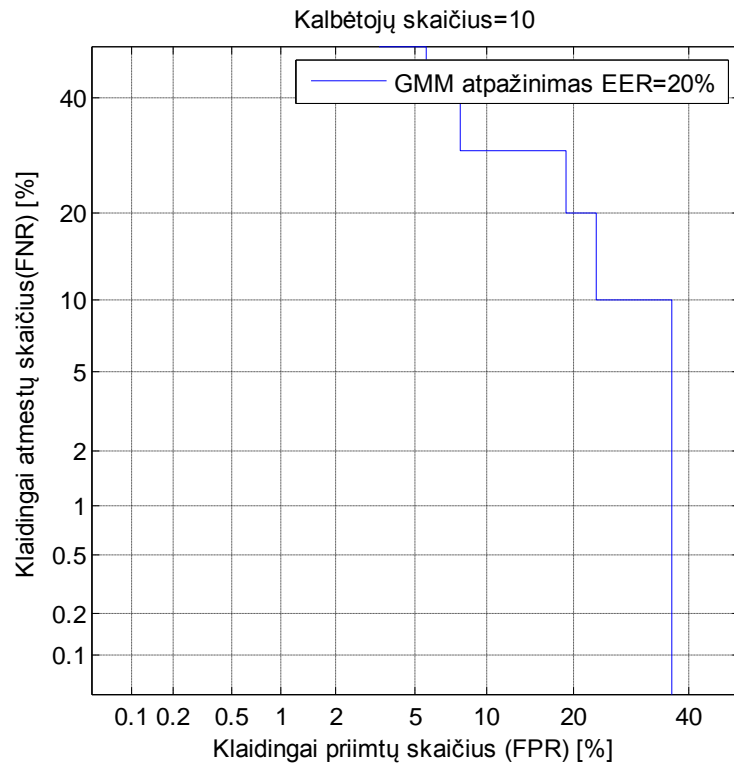
...

$$y = \text{awgn}(x, \text{snr})$$

....

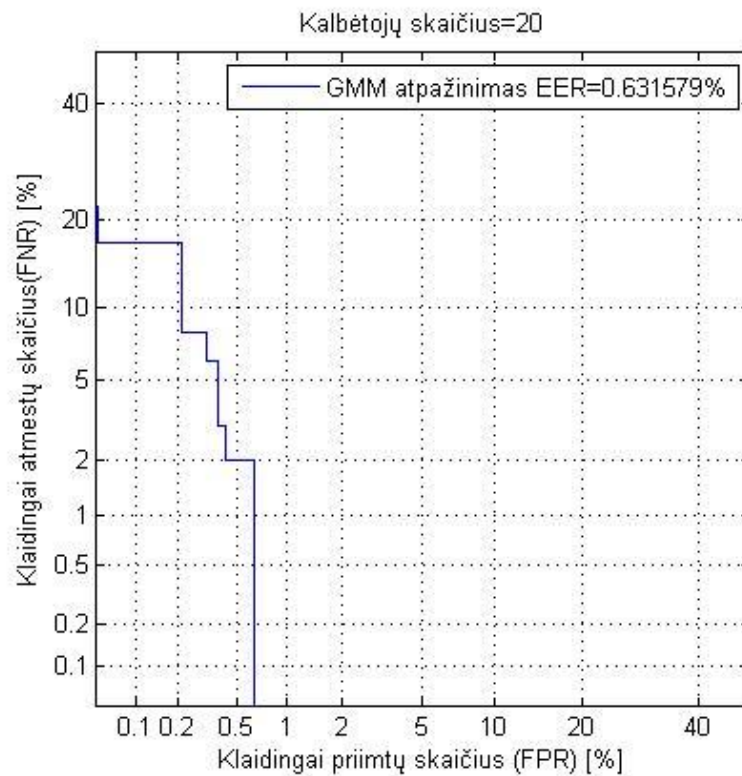
Funkcija „awgn“ sukuria baltą triukšmą signalui, x –vektorius, snr – signalui parenkama triukšmo vertė, decibelais dB.. Tyrime parenkama 10.

Galime pateikti keletą pavyzdžių kaip veikia sistema su dirbtiniu triukšmu. Dirbtinis triukšmas kalbėtojo modeliui pablogina kokybę ir sistema turi prasčiau atpažinti kalbėtojus.



3.23.pav. EER su 7 sek. trukme.

Matome, kaip prastai yra atpažistami kalbėtojai. EER yra labai didelis, todėl kalbėtojų atpažinimas yra neatikimas. Tokia sistema su triukšmais negali būti apsaugos priemonėse, nes veikia nestabiliai.



3.24.pav. EER su 7 sekundėmis

Palyginimui pateikiame 7 sekundžių trukmės įrašus su 20 kalbėtojais. Matome, kad EER yra labai mažas, tai sistema veikia puikiai, kai nėra triukšmų. Pagrindinis tikslas atliekant tyrimus su GMM metodų naudojant geros kokybės įrašus yra gauti labai mažą EER. Idealiausiu atveju, kai yra mažiau už vieną, bet literatūroje aprašoma, kad kai naudojami geros kokybės įrašai, tai EER būna net iki 0,1. Toks sistemos kokybiškumas yra labai geras ir galima naudoti visose apsaugos priemonėse.

IŠVADOS IR PASIŪLYMAI

Darbe išnagrinėtos kalbos signalų apdorojimo metodai. Taip pat peržvelgta sistemos pagrindiniai modeliavimo įrankiai, kurie vėliau leis sukurti atpažinimo sistemą. Taigi, kalbėtojo atpažinimo sistema leidžia apsaugoti namus ar kitus daiktus. Sistemos saugomus nėra gana aukštas, nes šaltiniuose rasta, jog tokias sistemas galima apeiti naudojant garso įrašymo įrenginius. Jei sukurta sistema nėra pakankamai gera, tai ji nėra saugi.

Aptartuose modeliuose vienas iš svarbesnių dalykų vektoriniame kvantavime yra kodinės knygos dydžio parinkimas, nes didinant kodinę knygą mažėja paklaidų skaičius. Gauso mišinio modelis turi gana gerą atpažinimo kokybę. Didžiausia problema, kad esant duomenų prisitaikymui mažėja atpažinimo kokybė. Taigi, sistema gali būti naudojama kur žmonių grupės yra didelės, bet su sąlyga, kad nebus duomenų perkrovimo.

Atlikus eksperimentą galime teigti, kad duomenys turi būti paruošti modelio treniravimui, kad būtų galima atlikti teisingai eksperimentą. Kalbėtojų įrašai turi būti parinkti tokiu skaičiumi, kad atitiktų tam tikrą skaičių požymių. Kuo daugiau požymių tuo gali geriau atpažinti, bet jei duomenų jau yra per daug tai modelis prisitaiko prie duomenų ir mažėja tikslumas. Taigi reikia atsižvelgti į norimos sistemos dydį, jei reikalingi nedideli kiekiai kalbėtojų tai galima požymių skaičių didinti, nes sistema su daugiau požymių (bet ne per daug) lengviau atpažins kalbėtoją. Tokia sistema leidžia efektyviai išnaudoti saugius kalbėtojų atpažinimus ir pakeisti kitas saugumo sistemas.

Tyrimui panaudoti kalbos įrašai parodo, kad GMM modelis gerai atpažįsta kalbėtojus, kai yra daugiau požymių. Kiekvieną kalbos įrašą reikia apdoroti, kad gautume melų fazės kepstro koeficientus. Taip pat svarbus pasirinkimas, kiek bus pasirenkama kalbėtojui atpažinti MFKK skaičius. Įrašai, kurie buvo geros kokybės buvo atpažinti geriau, negu prastos kokybės. Tai rodo, kad atpažinimo sistema veikia geriau, kai balso įrašai yra įrašomi gera kokybe. Taip pat pastebėta, kad geriausiai kalbėtojus atpažįsta, kai mišinių svorių skaičius yra 256.

Norint gauti geriausią rezultatą, reiktų prieš atliekant tyrimą atlikti keletą eksperimentų keičiant pradinius parametrus.

LITERATŪRA

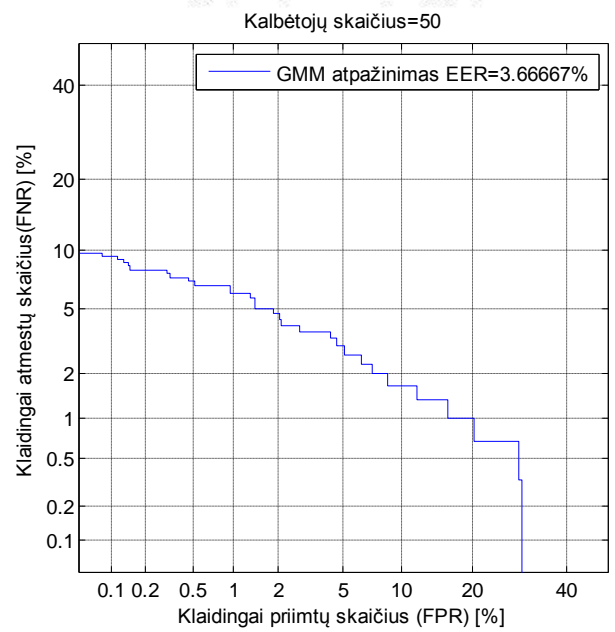
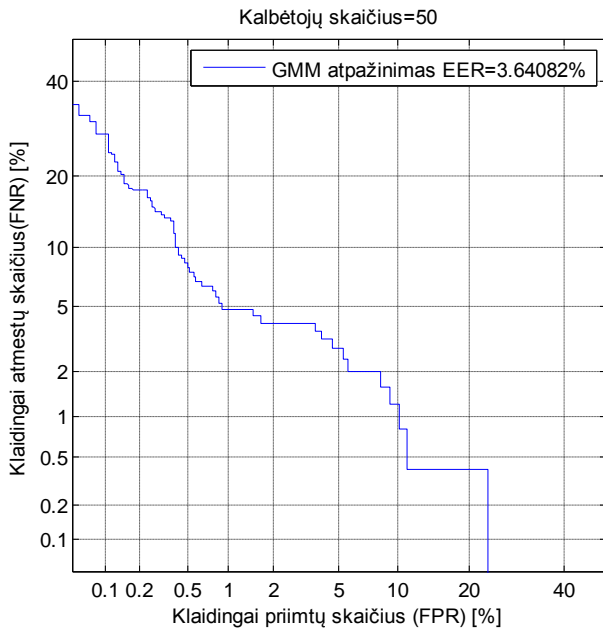
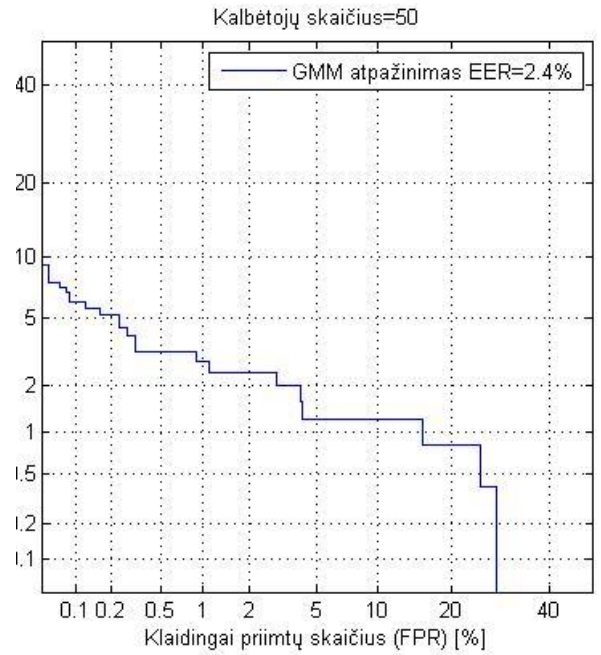
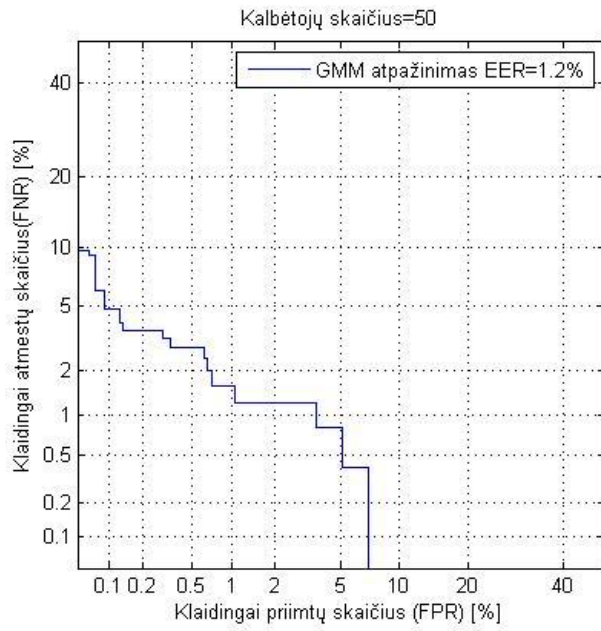
1. Phonexia Speaker Identification. Introduction [Interaktyvus]. [žiūrėta 2016.04.01]. Prieiga per internetą: <<https://www.phonexia.com/technologies/sid>>.
2. SIFT [Interaktyvus]. [žiūrėta 2016.02.15]. Prieiga per internetą: <<http://www.agnitio-corp.com/products/government/speaker-recognition>>.
3. VeriSpeak [Interaktyvus]. [žiūrėta 2016.02.15]. Prieiga per internetą: <<http://www.neurotechnology.com/verispeak.htm>>.
4. An Automatic Speaker Recognition System [Interaktyvus]. [žiūrėta 2016.04.01]. Prieiga per internetą: <http://minhdo.ece.illinois.edu/teaching/speaker_recognition/speaker_recognition.html>.
5. Vektorias kvantavimas [Interaktyvus]. [žiūrėta 2015.11.16]. Prieiga per internetą: <<http://www2.el.vgtu.lt/ssa/sB5node4.html>>.
6. Hemlata Eknath Kamale, Dr. R. S. Kawitkar. Vector Quantization Approach for Speaker Recognition [Interaktyvus]. [žiūrėta 2015.11.26]. Prieiga per internetą <http://www.ijctee.org/NSPIRE2013/IJCTEE_0313_Special_Issue_27.pdf>.
7. Vector Quantization In Text Dependent Automatic Speaker Recognition Using Mel-frequency Cepstrum Coefficient [Interaktyvus]. [žiūrėta 2015.03.16]. Prieiga per internetą <<http://wseas.us/e-library/conferences/2007egypt/papers/568-216.pdf>>.
8. Neuronas ir jo modelis [Interaktyvus]. [žiūrėta 2016.03.21]. Prieiga per <<http://www2.el.vgtu.lt/ssa/sA3node1.html>>.
9. Stergiou Christos. Neural Networks. [Interaktyvus]. [žiūrėta 2016.02.21]. Prieiga per <https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Network%20layers>.
10. Sheen Charlie. Neural Networks. [Interaktyvus]. [žiūrėta 2016.02.21]. Prieiga per <<http://natureofcode.com/book/chapter-10-neural-networks>>.
11. Tiesinės prognozės metodas ir jo taikymai [Interaktyvus]. [žiūrėta 2016.03.14]. Prieiga per internetą <http://www.mif.vu.lt/~bastys/academic/ATE/LPC/LPC_Kodavimas.htm>.
12. Wijoyo S. Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot [Interaktyvus]. [žiūrėta 2016.03.21]. Prieiga per <<http://www.ipcsit.com/vol6/36-E091.pdf>>.
13. Campbell, W. M. Support Vector Machines for Speaker and Language Recognition [Interaktyvus]. [žiūrėta 2016.02.14]. Prieiga per internetą : <https://www.ll.mit.edu/mission/cybersec/publications/publication-files/full_papers/0604_CampbellW-JA.pdf>.
14. Speaker Recognition using universal background model on YOHO database [Interaktyvus]. [žiūrėta 2016.04.21]. Prieiga per internetą:

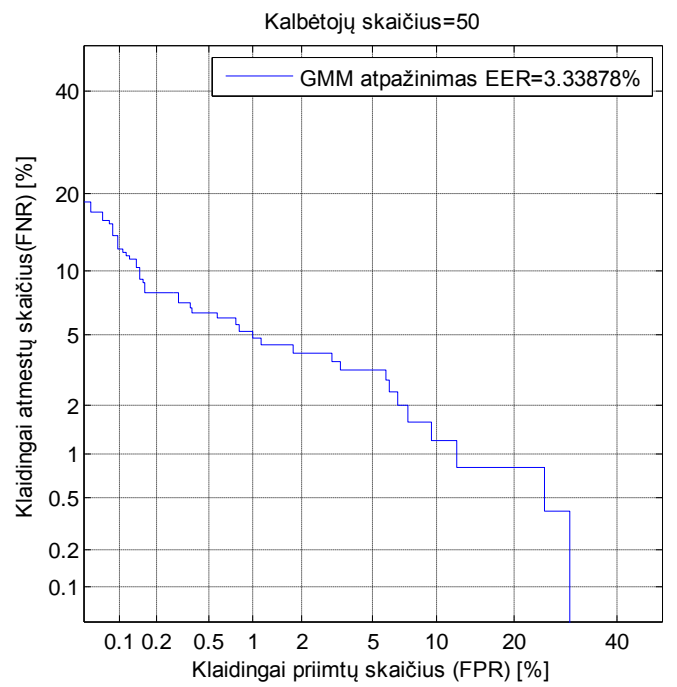
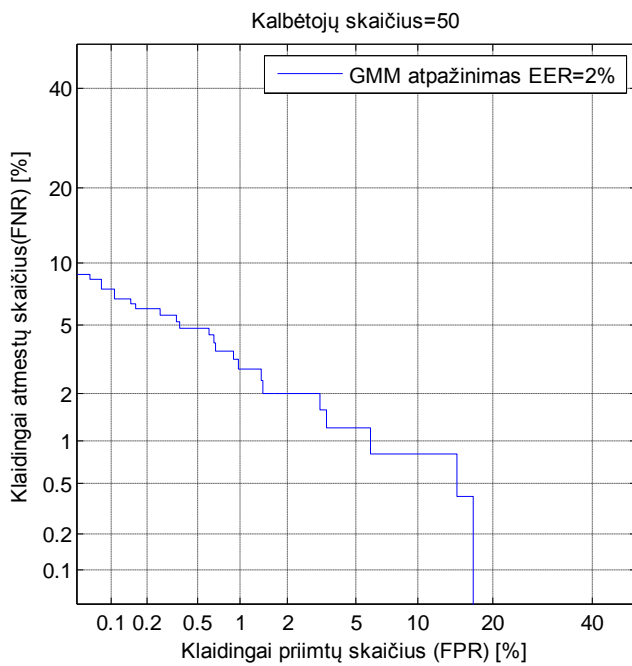
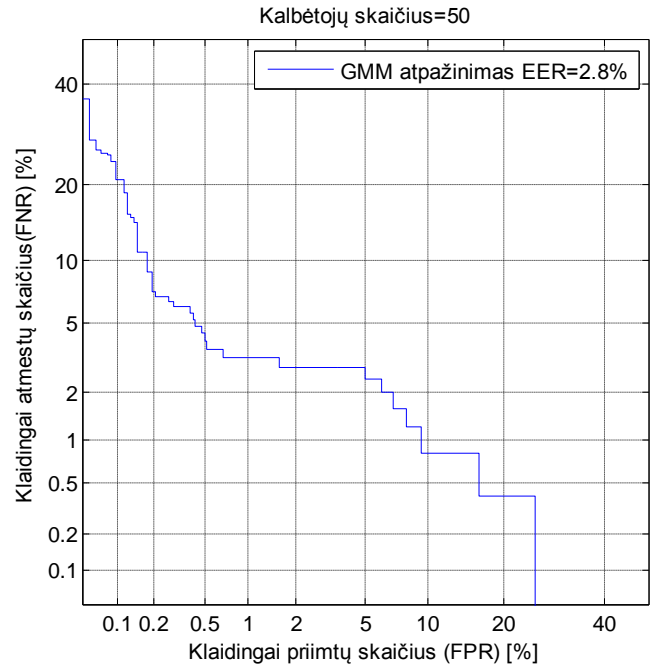
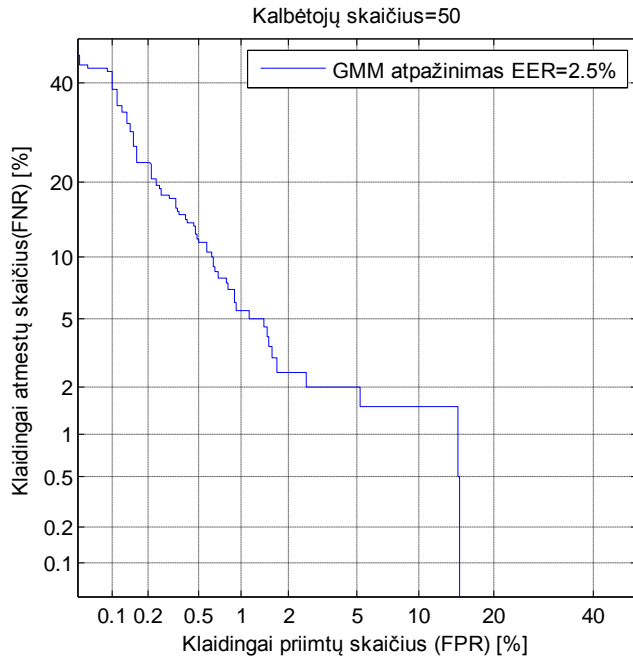
- <http://projekter.aau.dk/projekter/files/52688806/speaker_recognition_using_ubm_on_yoho_database.pdf>.
15. Shende Archana. Comparison of different parameters used in GMM based automatic speaker recognition [Interaktyvus]. [žiūrėta 2016.04.01]. Prieiga per internetą: <http://www.ijscce.org/attachments/File/Vol-1_Issue-3/C043051311.pdf>.
 16. Language and Text-Independent Speaker Identification System Using GMM [Interaktyvus]. [žiūrėta 2016.04.07]. Prieiga per internetą: <<http://www.wseas.org/multimedia/journals/signal/2013/105714-135.pdf>>.
 17. Žmogaus balsas [Interaktyvus]. [žiūrėta 2016.03.11]. Prieiga per internetą: <https://lt.wikipedia.org/wiki/Žmogaus_balsas>.
 18. Triukšmas [Interaktyvus]. [žiūrėta 2016.02.01]. Prieiga per internetą: <<https://lt.wikipedia.org/wiki/Triukšmas>>.
 19. Muda Lindasalwa. Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques [Interaktyvus]. [žiūrėta 2016.04.01]. Prieiga per internetą: <<https://arxiv.org/ftp/arxiv/papers/1003/1003.4083.pdf>>.
 20. Raman Pujita. Speaker Identification and Verification Using Line Spectral Frequencies [Interaktyvus]. [žiūrėta 2016.04.21]. Prieiga per internetą: <https://vtechworks.lib.vt.edu/bitstream/handle/10919/52964/Raman_P_T_2015.pdf?sequence=1>.

PRIEDAI

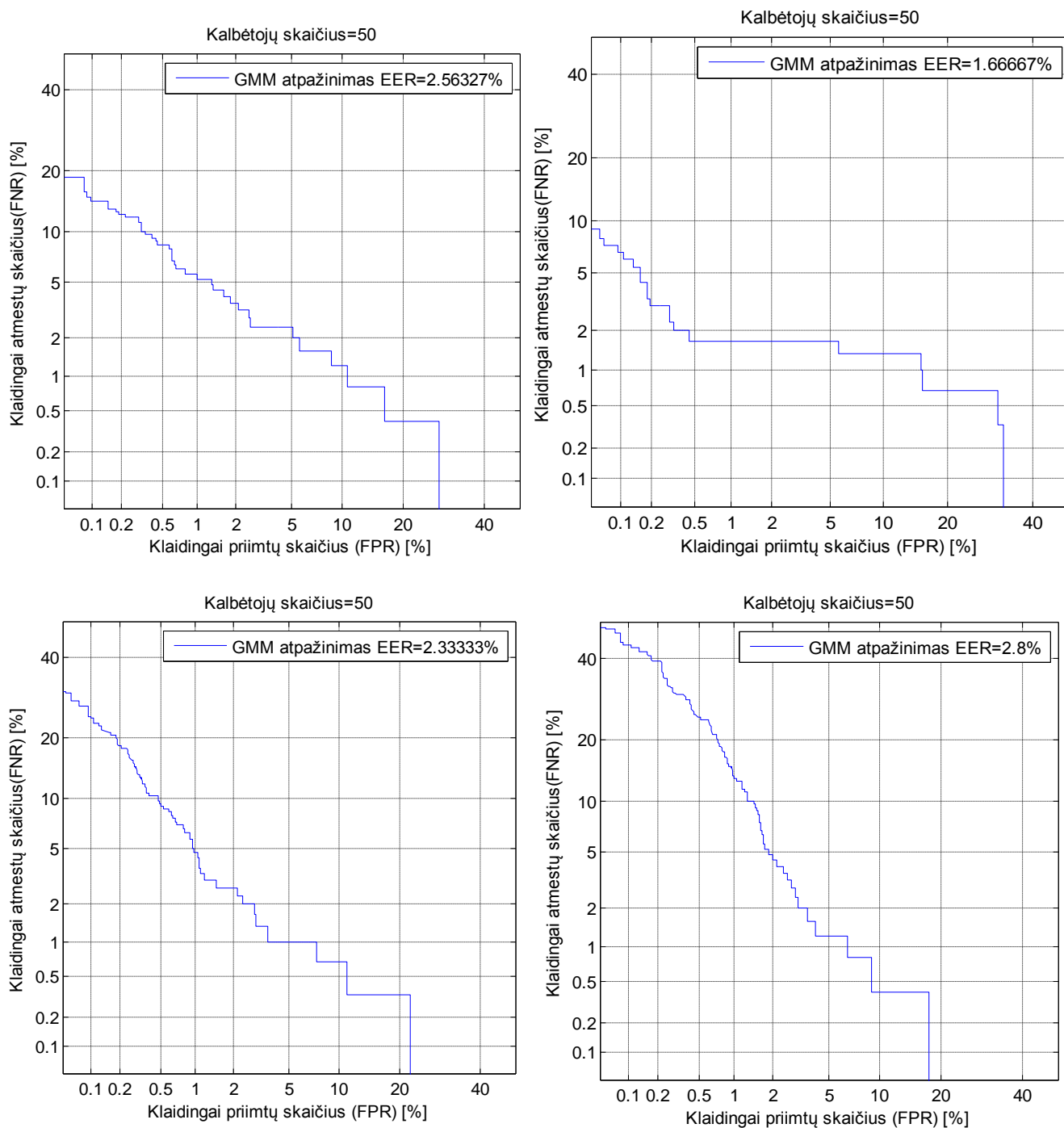
- 1 PRIEDAS Atpažinimo grafikai naudojant iki 7s trukmės 500 įrašus.
- 2 PRIEDAS Atpažinimo grafikai naudojant iki 11 trukmės 200 įrašus.
- 3 PRIEDAS Atpažinimo grafikai naudojant moterų ir vyrų balsus.
- 4 PRIEDAS Kiti atpažinimo grafikai.
- 5 PRIEDAS Atpažinimas su 20 kalbėtojų.
- 6 PRIEDAS Atpažinimas su 64 ir 128 mišinių svoriais.
- 7 PRIEDAS Atpažinimas pagal kalbėtojus.
- 8 PRIEDAS „melcepst“ funkcija.
- 9 PRIEDAS Failo įkėlimas

Atpažinimo grafikai naudojant iki 7s trukmės 500 įrašus

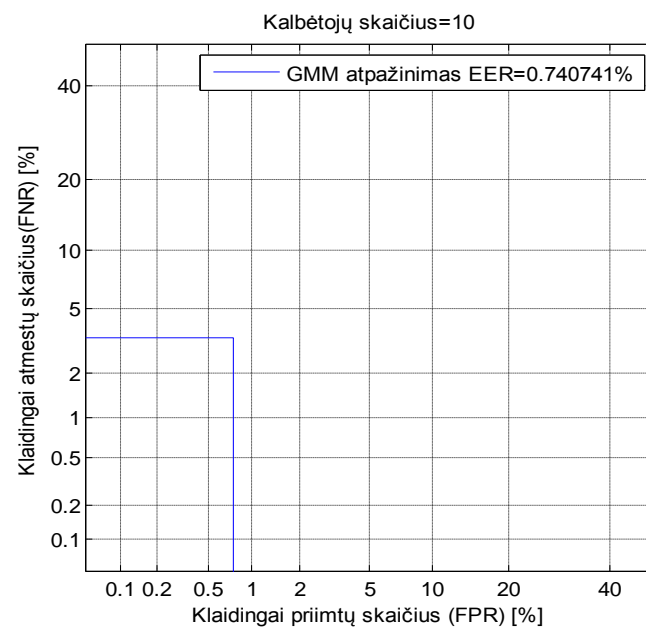
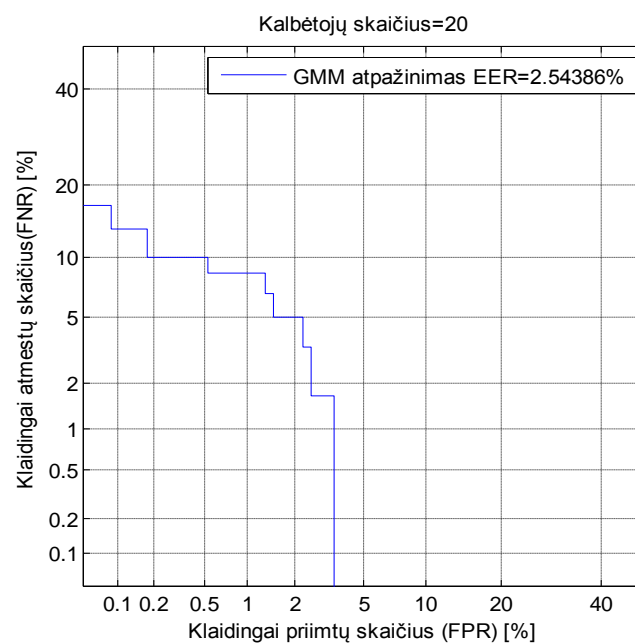
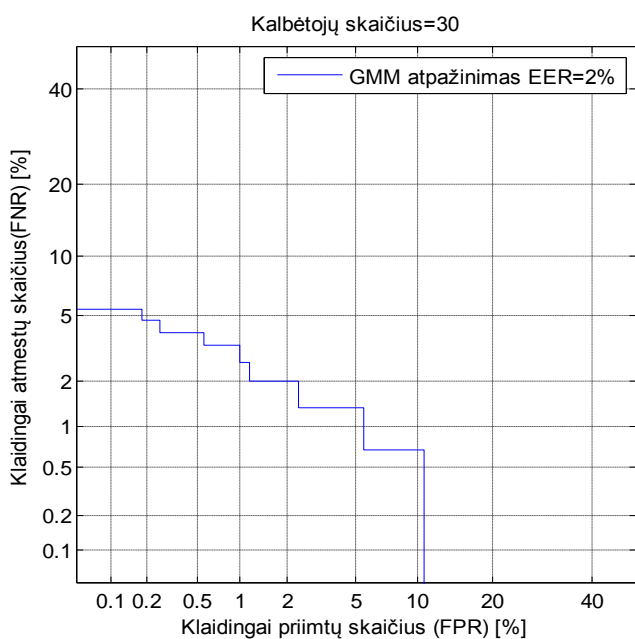
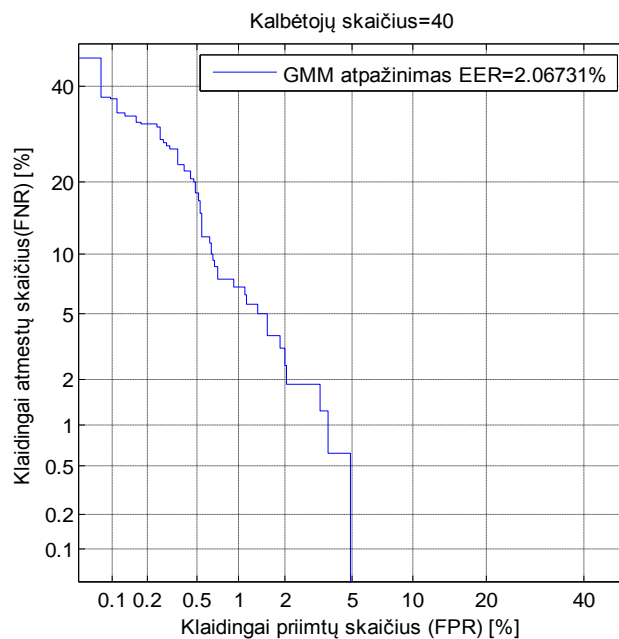
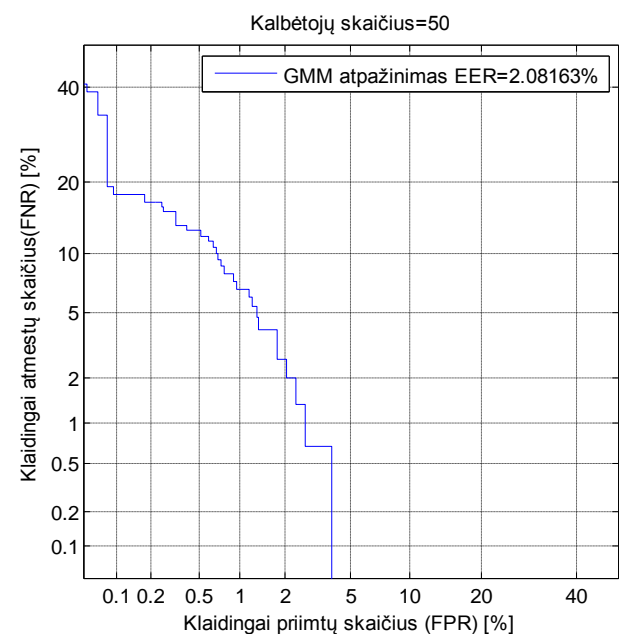


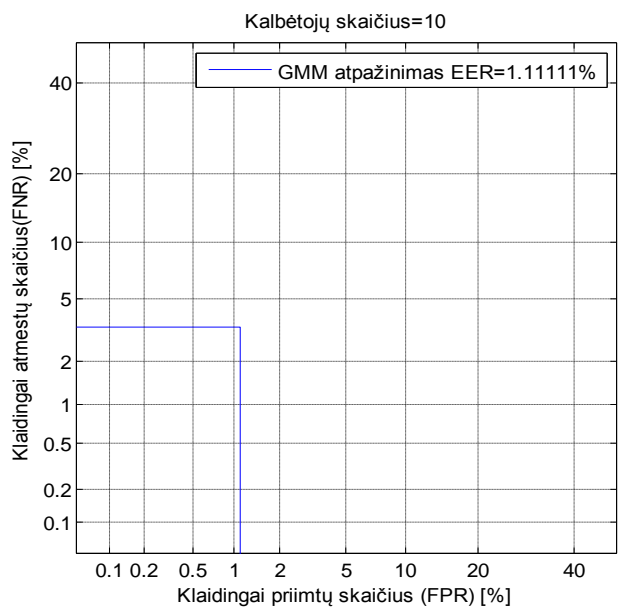
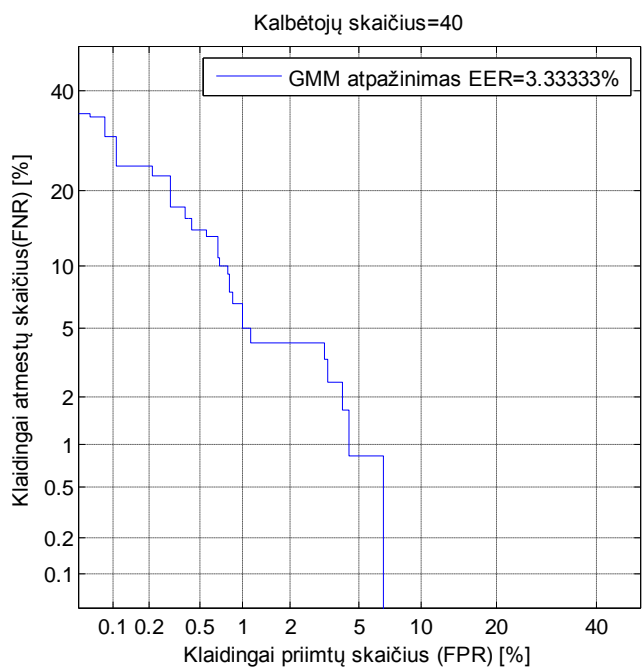
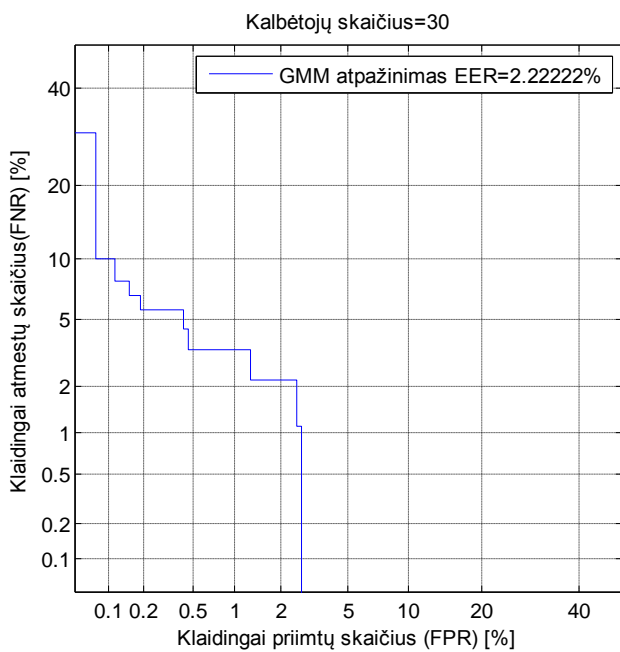
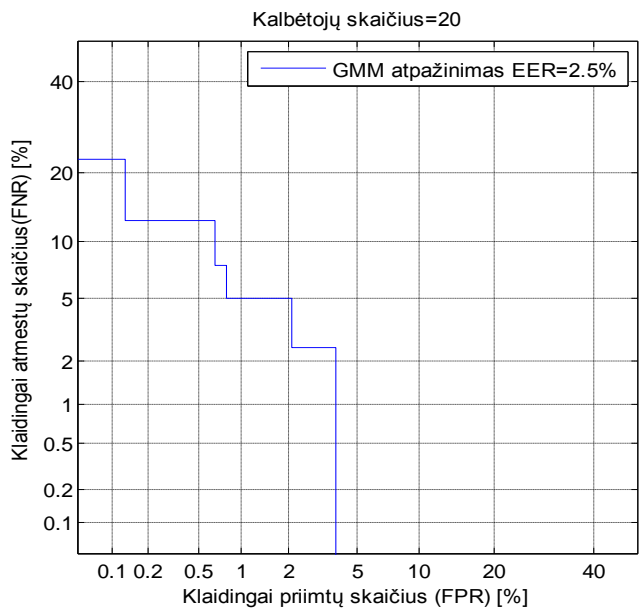
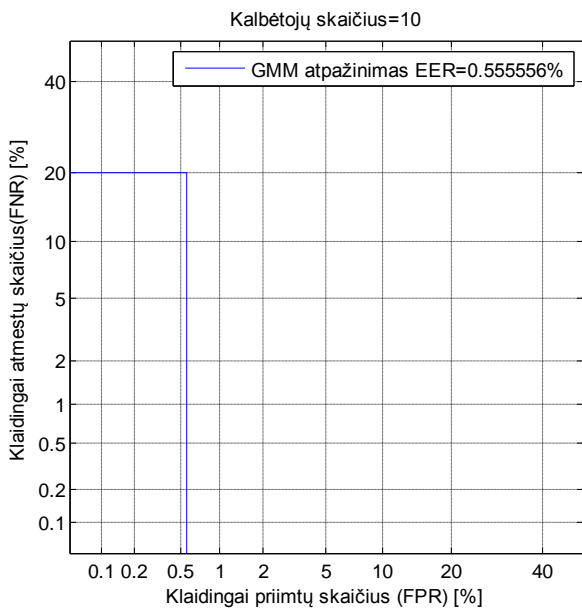


Atpažinimo grafikai naudojant iki 11 trukmės 200 įrašus

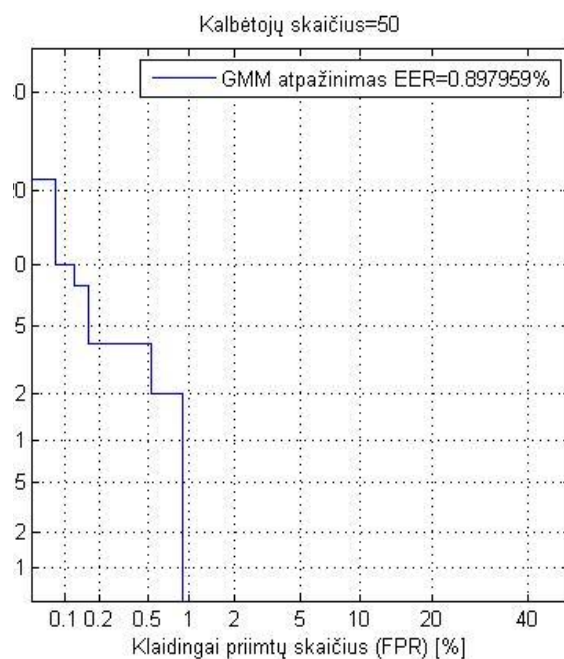
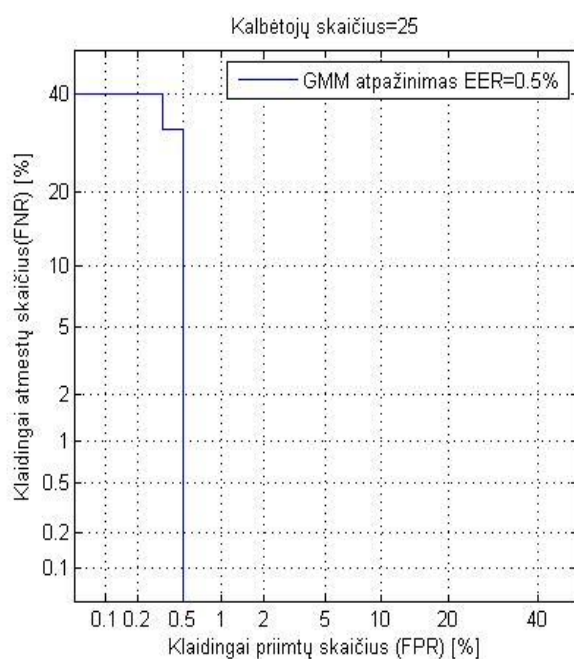
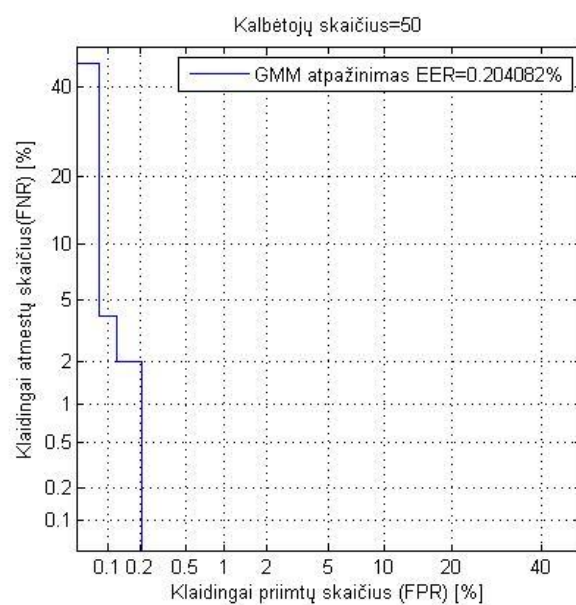
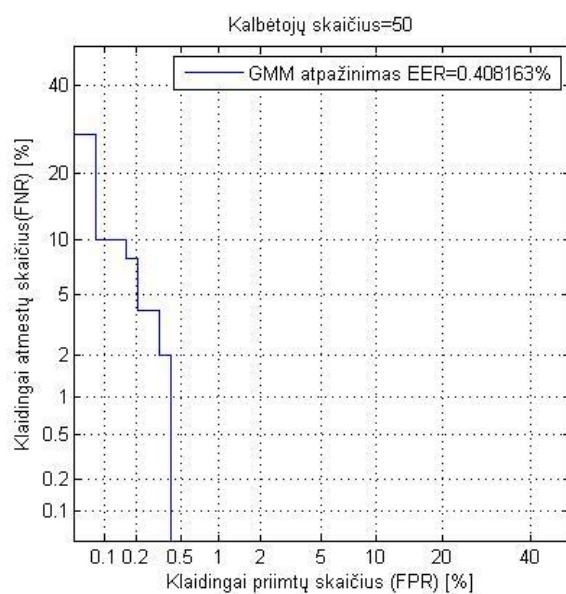


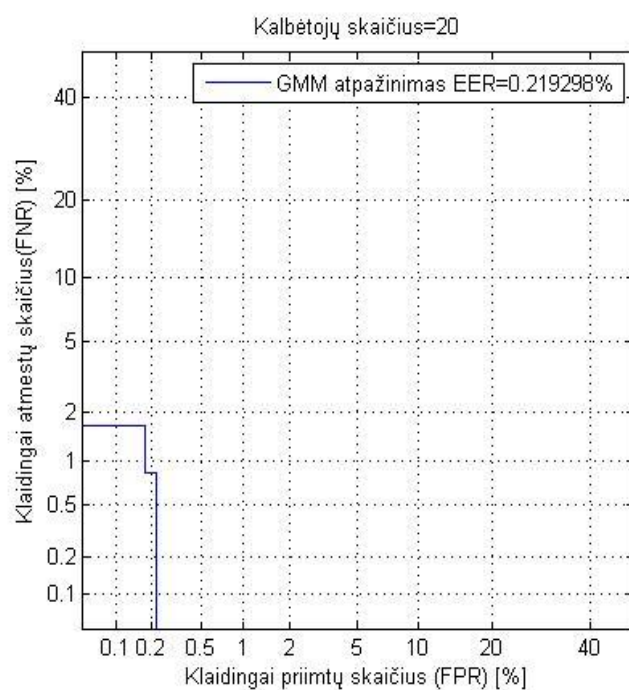
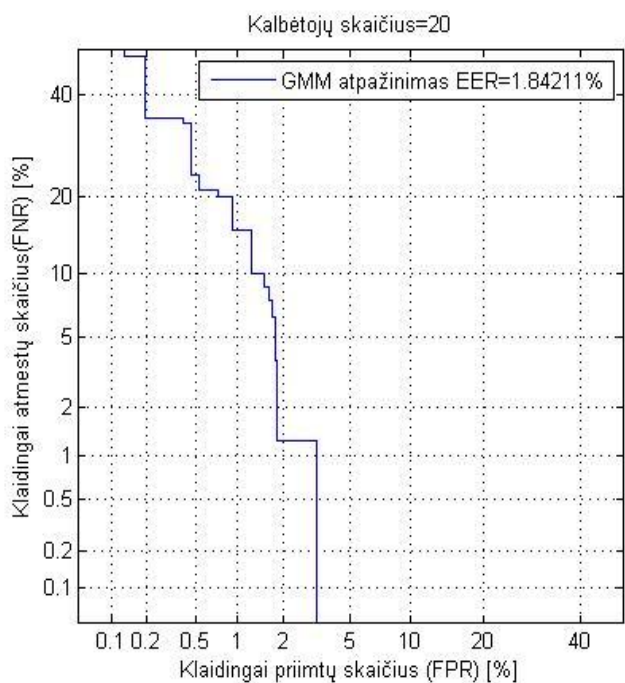
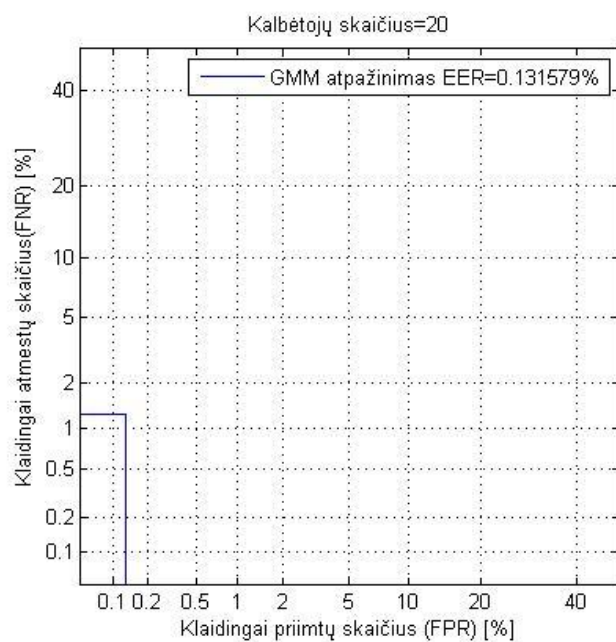
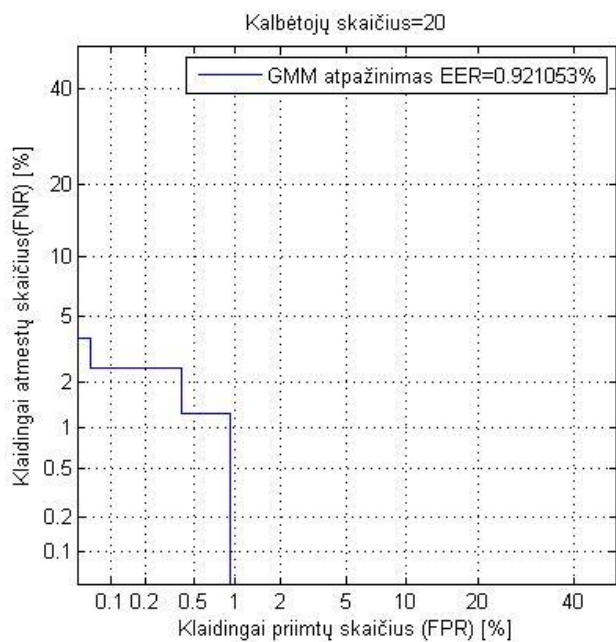
Atpažinimo grafikai naudojant moterų ir vyrų balsus



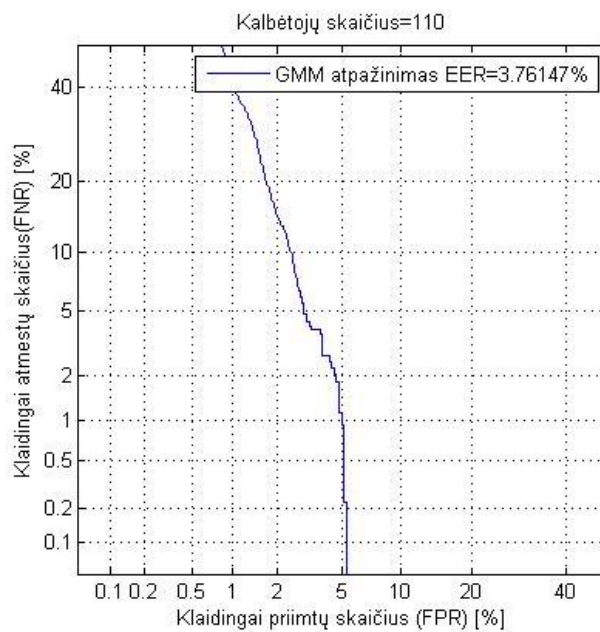
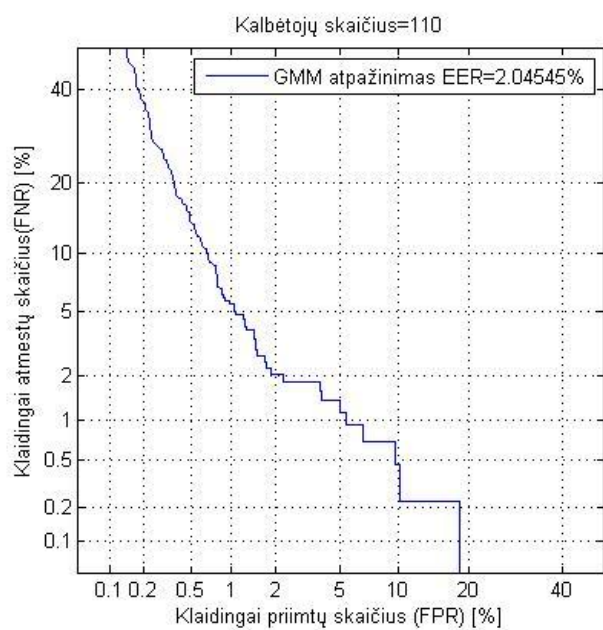


Kiti atpažinimo grafikai

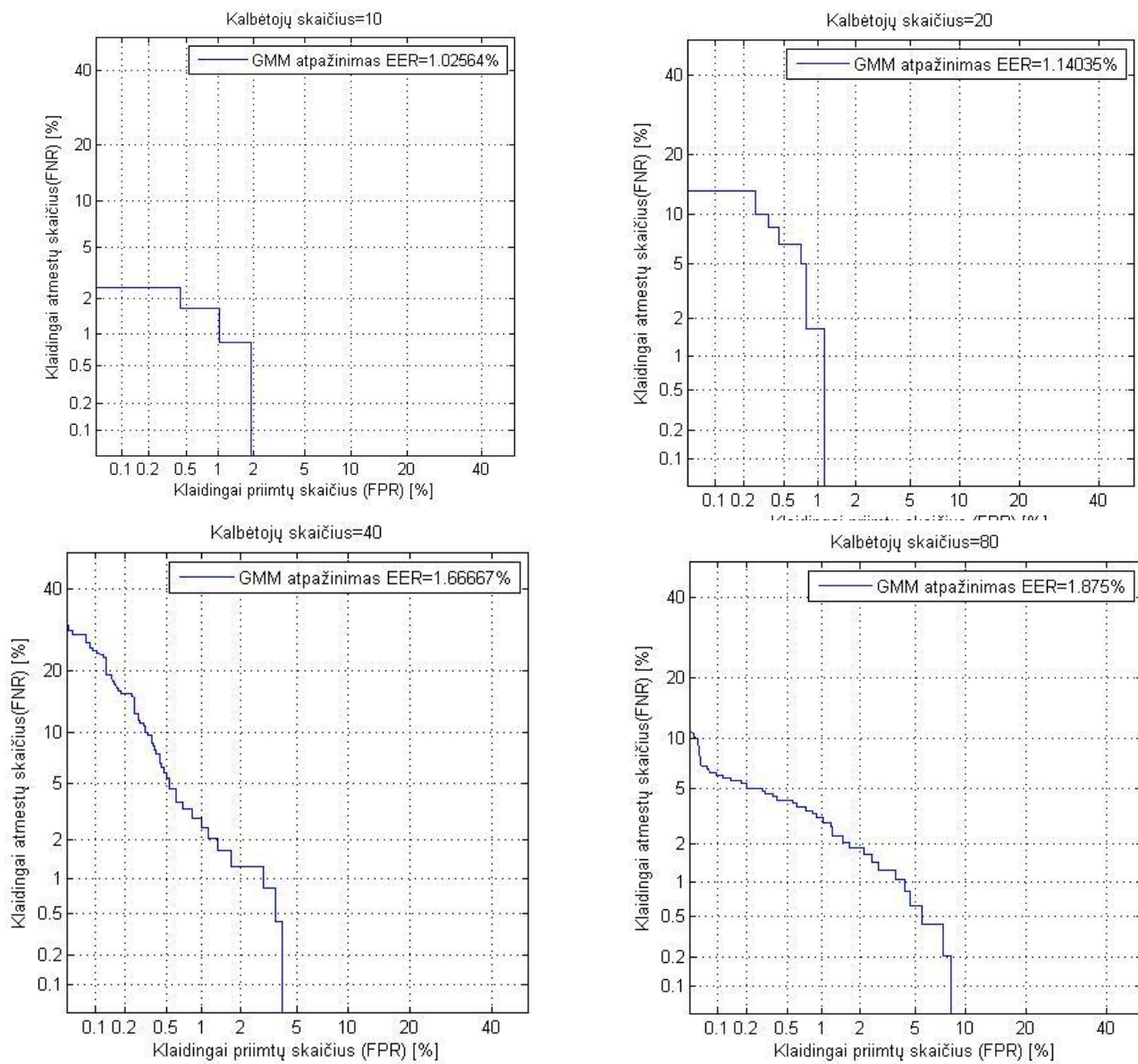




Atpažinimas su 64 ir 128 mišinių svoriais



Atpažinimas pagal kalbėtojus



„melcepst“ funkcija

MELCEPST.m

```
function c=melcepst(s,fs,w,nc,p,n);
```

s – garso failo vektorius;

fs – dažnis;

w – lango funkcija;

nc – MFKK skaičius;

p – kadro trukmė;

n –kadų ilgis atskaitomis.

Garso failo įkėlimas

TRENIRAVIMAS.m

...

```
[s,Fs]=wavread(failo_kelias/pavadinimas.wav');
```

...