

**VILNIAUS UNIVERSITETAS**  
**KAUNO HUMANITARINIS FAKULTETAS**  
**INFORMATIKOS KATEDRA**

Verslo informatikos studijų programa  
Kodas 62109P101

**AURELIJUS MORKEVIČIUS**  
**MAGISTRO BAIGIAMASIS DARBAS**

**ALGORITMAS SKIRTAS BALSŲ IR PRIEBALSŲ DIFERENCIACIJAI**  
**ŠNEKAMOSIOS KALBOS SIGNALE**

Kaunas 2009

**VILNIAUS UNIVERSITETAS**  
**KAUNO HUMANITARINIS FAKULTETAS**  
**INFORMATIKOS KATEDRA**

**AURELIJUS MORKEVIČIUS**  
**MAGISTRO BAIGIAMASIS DARBAS**

**ALGORITMAS SKIRTAS BALSŲ IR PRIEBALSŲ DIFERENCIACIJAI**  
**ŠNEKAMOSIOS KALBOS SIGNALE**

Leidžiama ginti \_\_\_\_\_

Magistrantas \_\_\_\_\_  
(parašas)

Darbo vadovas \_\_\_\_\_  
(parašas)

Doc. dr. Pranas Žvinys  
(darbo vadovo mokslo laipsnis, mokslo  
pedagoginis vardas, vardas ir pavardė)

Darbo įteikimo data \_\_\_\_\_

Registracijos Nr. \_\_\_\_\_

# TURINYS

<b>TURINYS</b> .....	<b>3</b>
<b>PAVEIKSLŲ SĄRAŠAS</b> .....	<b>5</b>
<b>LENTELIŲ SĄRAŠAS</b> .....	<b>7</b>
<b>SANTRUMPŲ SĄRAŠAS</b> .....	<b>8</b>
<b>SUMMARY</b> .....	<b>9</b>
<b>ĮVADAS</b> .....	<b>10</b>
<b>1. PAGRINDINIAI NUSTATYMO METODAI IR SKIRIAMIEJI BRUOŽAI</b> .....	<b>12</b>
1.1. Balsių ir priebalsių diferenciacija. ....	12
1.1.1. Fonemų klasifikavimo požymiai .....	12
1.2. Pagrindinis tonas.....	15
1.3. Harmoninė garso bangos struktūra .....	16
1.4. Pagrindinio tono nustatymo algoritmai .....	17
1.4.1. Pagrindinio tono radimas autokoreliacijos būdu. ....	17
1.4.2. Pagrindinio tono radimas taikant kepstrą. ....	18
1.4.3. Furjė transformacijų kombinacinis metodas. ....	20
1.4.4. Pagrindinio tono nustatymo metodų klasifikacija.....	21
1.5. Balsių atpažinimo algoritmai.....	21
1.5.1. Balsių atpažinimas iš šnekėtojo lūpų .....	22
1.5.2. Melino transformacija balsių atpažinimui.....	24
1.5.3. Balsių atpažinimas naudojant Bajeso analizę.....	25
1.6. Balsių išskyrimas iš kitų garsų šnekamosios kalbos signale .....	26
1.7. Balsių nustatymo metodai .....	27
1.7.1. Garsumo funkcija .....	28
1.7.2. Formantės .....	29
1.7.3. Formančių išskyrimas .....	29
1.8. Signalų filtravimas.....	34
1.8.1. Juostos praleidimo filtras .....	35
1.8.2. Juostos uždraudimo filtras.....	36
1.8.3. Aukšto ir žemo praleidimo filtrai .....	37
1.9. Apibendrinimas .....	37
<b>2. BALSŲ SKIRIAMOJO POŽYMIO METODO MODELIS</b> .....	<b>39</b>
2.1. Suminės harmonikų energijos metodo modelis.....	39
2.1.1. Garso paruošimas analizei.....	40
2.1.2. Pagrindinio tono metodo parinkimas .....	41

2.1.3.	Skaitmeniniai filtrai.....	43
2.1.4.	Filtro eilės parinkimas.....	48
2.1.5.	Signalų energija.....	48
2.2.	„LTDIGITS“ garsynas.....	49
2.3.	Tyrimo rezultatų apdorojimas.....	51
<b>3.</b>	<b>BALSIŲ SKIRIAMOJO POŽYMIO METODO ĮVERTINIMO TYRIMAS.....</b>	<b>52</b>
3.1.	Eksperimentinio tyrimo aplinka.....	52
3.1.1.	Programinės įrangos architektūra.....	53
3.1.2.	Tyrimo aplinkos pritaikymas vartotojo poreikiams.....	54
3.2.	Pagrindinis eksperimentas.....	55
3.2.1.	Eksperimento eiga.....	56
3.2.2.	Tyrimo rezultatai.....	57
3.3.	Papildomi tyrimai.....	61
3.4.	Pagrindinio ir papildomų tyrimų apibendrinimas.....	63
3.5.	Algoritmo greičio tyrimas.....	64
	<b>IŠVADOS IR PASIŪLYMAI.....</b>	<b>66</b>
	<b>LITERATŪROS SĄRAŠAS.....</b>	<b>68</b>
	<b>PRIEDAI.....</b>	<b>72</b>

## PAVEIKSLŲ SĄRAŠAS

<b>1 pav.</b> Lietuvių kalbos garsų klasifikacija .....	13
<b>2 pav.</b> LTDIGITS fonemų hierarchinė struktūra .....	14
<b>3 pav.</b> Fonemų skirstymo į stambias grupes metodika .....	14
<b>4 pav.</b> Harmoninė garso bangos struktūra .....	16
<b>5 pav.</b> Autokoreliacijos algoritmo veiksmų seka.....	18
<b>6 pav.</b> Autokoreliacijos metodas.....	18
<b>7 pav.</b> Kepstro algoritmas .....	19
<b>8 pav.</b> Kepstras.....	19
<b>9 pav.</b> „Sylvian Marchand“ pasiūlytas Furjė transformacijų kombinacinis metodas .....	20
<b>10 pav.</b> Pagrindinio tono nustatymo metodų klasifikacinė sistema.....	21
<b>11 pav.</b> Balsių atpažinimo metodų klasifikacija .....	22
<b>12 pav.</b> Balsių atpažinimo iš šnekėtojo lūpų metodo modelis.....	23
<b>13 pav.</b> Lūpos aukščio ir pločio nustatymas .....	23
<b>14 pav.</b> Melino transformacija pagrįstas balsių atpažinimo metodas .....	25
<b>15 pav.</b> Balsių atpažinimo metodo naudojant Bajeso analizę algoritmo modelis .....	26
<b>16 pav.</b> ASA sistemos struktūra.....	26
<b>17 pav.</b> Spektro glodinimas .....	30
<b>18 pav.</b> a) standartinė FFT spektrograma; b) LPC spektrograma.....	31
<b>19 pav.</b> a) Dalinių filtrų metodas; b) LPC metodas .....	32
<b>20 pav.</b> Spektras suskaidytas į regionus pagal dinaminio programavimo algoritmą.....	33
<b>21 pav.</b> Formančių radimo metodų klasifikacinė sistema.....	34
<b>22 pav.</b> Juostos praleidimo filtras .....	35
<b>23 pav.</b> Juostos praleidimo filtro realizavimo schema.....	36
<b>24 pav.</b> Juostos uždraudimo filtras .....	36
<b>25 pav.</b> Aukšto ir žemo praleidimo filtrai.....	37
<b>26 pav.</b> Balsių nustatymo metodų optimalumas .....	38
<b>27 pav.</b> Suminės harmonikų energijos metodo modelis .....	39
<b>28 pav.</b> Signalų langavimas .....	40
<b>29 pav.</b> Garso centro iškirpimas .....	41
<b>30 pav.</b> Filtruojami dažnių intervalai .....	43
<b>31 pav.</b> Filtravimo proceso veiksmų seka.....	44
<b>32 pav.</b> 6 pakopų Batervorto filtro perdavimas ir dažninės charakteristikos.....	45
<b>33 pav.</b> Skaitmeninių filtrų palyginimas.....	46

<b>34 pav.</b> FIR filtras .....	47
<b>35 pav.</b> „LTDigits“ žodžių reikšmių failo .ort turinio fragmentas.....	50
<b>36 pav.</b> „LTDigits“ žodžių ribų failo .ort turinio fragmentas .....	50
<b>37 pav.</b> „LTDigits“ fonemų ribų turinio fragmentas .....	50
<b>38 pav.</b> Programinės įrangos panaudos atvejų modelis .....	52
<b>39 pav.</b> Tyrimo programinės įrangos architektūra.....	53
<b>40 pav.</b> „MATLAB“ kodo intarpai į tyrimo aplinką .....	54
<b>41 pav.</b> Pagrindinio tyrimo eiga.....	56
<b>42 pav.</b> Tiriama kriterijaus reikšmių pasiskirstymas skirtingiems balsiams .....	59
<b>43 pav.</b> Tiriama kriterijaus reikšmių pasiskirstymas skirtingiems balsiams .....	61
<b>44 pav.</b> Papildomo tyrimo rezultatai moteriškiems garsams .....	62
<b>45 pav.</b> Papildomo tyrimo rezultatai vyriškiems garsams .....	63

## LENTELIŲ SĄRAŠAS

<b>1 lentelė.</b> Pagrindinio tono nustatymo metodų tyrimo rezultatai.....	42
<b>2 lentelė.</b> Pagrindinio tono nustatymo metodų su apibrėžtais dažnio intervalais tyrimo rezultatai...	42
<b>3 lentelė.</b> Diktorių žodžių sekos.....	49
<b>4 lentelė.</b> Tyrimo aplinkos parametrai .....	54
<b>5 lentelė.</b> Pagrindinio eksperimento konfigūracija .....	55
<b>6 lentelė.</b> Pagrindinio eksperimento rezultatai moteriškiems garsams .....	57
<b>7 lentelė.</b> Pagrindinio eksperimento rezultatai vyriškiems garsams .....	59
<b>8 lentelė.</b> Papildomo eksperimento konfigūracija.....	62
<b>9 lentelė.</b> Eksperimento ir papildomų tyrimų rezultatų apibendrinimo matrica moteriškiems garsams .....	64
<b>10 lentelė.</b> Eksperimento ir papildomų tyrimų rezultatų apibendrinimo matrica vyriškiems garsams .....	64
<b>11 lentelė.</b> Greičio tyrimo rezultatų apibendrinimo matrica moteriškiems garsams .....	65
<b>12 lentelė.</b> Greičio tyrimo rezultatų apibendrinimo matrica vyriškiems garsams .....	65

## SANTRUMPŲ SĄRAŠAS

ASA – automatinis šnekos atpažinimas

FFT – greitoji Furjė transformacija

IFFT – greitoji inversinė Furjė transformacija

LPC – tiesinės prognozės kodavimo algoritmas

TIMIT – šnekos garsynas anglų kalbai

LTDIGITS – šnekos garsynas lietuvių kalbai sukurtas Vilniaus universiteto Kauno humanitariniame fakultete.



Aurelijus, Morkevicius. (2007). *An Algorithm for Consonants and Vowels Differentiation in Speech Signal*. MBA Graduation Paper. Kaunas: Vilnius University, Kaunas Faculty of Humanities, Department of Informatics. 67 p.

## SUMMARY

New algorithm for consonants and vowels differentiation is revealed in this paper. This significantly fast algorithm calculates a feature based on the sum of energies of a filtered signal. This feature allows determining whether a vowel or a consonant is an exploratory. An experiment is done to make estimations how reliable the feature is.

**The main aim** of this paper is to realize and evaluate by the experiments consonants and vowels differentiation algorithm based on the calculation of pitch and the sum of energies of a filtered signal.

**The object** of this paper is the algorithm of consonants and vowels differentiation in speech signal.

To realize the main aim there are several tasks to be resolved:

1. describe the definition of the pitch, pitch detection methods and their possible practical use;
2. evaluate mostly used pitch detection methods;
3. analyze existing vowels differentiation and recognition algorithms;
4. analyze mostly used digital filters;
5. according to analysis results implement an algorithm model;
6. realise information system to calculate desired feature;
7. evaluate feature precision and quickness;

For these tasks to be resolved there are several research techniques used.

1. Qualitative techniques: pitch detection methods classification, vowel recognition methods classification, scientific literature analysis and filing and UML for algorithm modeling.
2. Quantitative techniques: performance evaluation and comparison of algorithms using mathematical and statistical evaluation criterions.

The Paper consists of three major parts. First part defines major concepts, existing pitch vowel recognition and differentiation methods, digital filters. The Second part describes an algorithm model. The third part describes major and supplementary experiments for an evaluation of created model.

The Paper consists of 67 pages, 12 tables and 45 pictures and schemes.

## ĮVADAS

**Temos aktualumas ir problematika.** Su kalbos signalų atpažinimu mes susiduriame kas dieną. Girdėdami žmogaus šneką mes gan nesunkiai galime nusakyti jo lytį, amžių ar tapatybę, jei šį žmogų pažystame ar dažnai girdime jo balsą. Vis daugiau naujų šiuolaikinių technologijų taip pat grindžiamos balsų atpažinimu. Tačiau, nepaisant akivaizdžios pažangos, net ir geriausių kalbos atpažinimo sistemų efektyvumas dar negali susilyginti su žmogaus gebėjimais atpažinti ir suvokti kalbos signalus. Taigi būtina ir toliau tirti atpažinimo metodus, siekiant pagerinti kalbos atpažinimo sistemų patikimumą.

Didžiausią neapibrėžtumą garso signalų tyrimams įneša didelis skaičius kalbų, dialektų, žmonių rasių skirtumai, lytis, amžius ir be abejo aplinka, kurioje atliekamas atpažinimas. Visa tai stengiamasi kiek galima minimizuoti, tačiau apie universalią šnekamosios kalbos atpažinimo sistemą kalbėti dar anksti. Nepaisant to, kalbos atpažinimas pastaruoju metu stipriai pažengęs į priekį. Dabar kalba atpažįstama ne tik iš garso signalo, bet ir vaizdinės informacijos, tokios kaip lūpų ar liežuvio judesiai. Kombinuojant vaizdo ir garso signalo atpažinimo metodus galimas labai geras rezultatas. Deja vaizdinė informacija prieinama tik išskirtiniais atvejais, o didžiajai daliai kalbos signalų atpažinimo problemų yra „juodoji dėžė“.

Svarbiausias kalbos atpažinimo etapas - atpažinti balsius. Yra taikoma įvairių metodų, jie tobulinami ir kuriami nauji, kad išgauti kiek galima tikslesnį, greitesnį ir universalesnį metodą. Svarbiausias momentas atpažįstant balsius yra rasti juos skiriantį požymį. Tokie požymiai dažniausiai remiasi trim pagrindiniais garso skiriamaisiais bruožais: pagrindiniu tonu, harmonine struktūra ir formantėmis. Kad atlikti balsių atpažinimą, pirmiausiai reikia juos identifikuoti, t.y. išskirti iš kitų garsų kalbos signale. Balsių identifikacija plačiai taikoma ir kalbos signalų segmentacijai.

Mirjam'as Vester'is remdamasis akustiniais faktoriais klasifikavo garsus. Šiai klasifikacijai atlikti buvo apmokytas neuroninis tinklas. Atliktas tyrimas parodė, kad balsių teisingo išskyrimo procentas – 73 (WESTER, M. 2003). Danil'as Korčerovas atlikdamas rusų kalbos balsių atpažinimą naudojo balsių išskyrimą remdamasis vidutinės signalo energijos koreliacija ir spektro intensyvumu (KOCHAROV, D. A. 2004). Tai tik keli pavyzdžiai kaip yra atskiriami balsiai nuo priebalsių ir kur tai praktiškai pritaikoma.

Yra daug algoritmų balsiams identifikuoti. Priklausomai nuo tiriamos problemos pobūdžio taikomi įvairūs statistiniai, spektrinės ir laikinės garso charakteristikos algoritmai. Tiesa, daugelis algoritmų yra paremti spektrine signalo analize, o tai stipriai padidina algoritmo imlumą kompiuterio resursams, o tuo pačiu ir skaičiavimo laiką.

Greitas ir tikslus balsių išskyrimo požymis galėtų stipriai palengvinti ir pagreitinti balsių, o tuo pačiu ir kalbos atpažinimą.

**Darbo objektas:** balsių nustatymo algoritmas.

**Darbo tikslas:** realizuoti ir eksperimentais patikrinti balsių nustatymo požymį pagrįstą pagrindinio tono ir harmonikų dažnių energijos suma.

**Hipotezė:** tiriamas balsių nustatymo požymis gali būti taikomas Lietuvos kalbos balsiams atskirti nuo priebalsių.

**Darbo uždaviniai:**

1. apibūdinti pagrindinio tono sąvoką, jo paieškos ypatumus, paieškos poreikį ir galimą praktinį taikymą;
2. išanalizuoti ir ištirti eksperimentu plačiausiai taikomus pagrindinio tono nustatymo algoritmus;
3. išanalizuoti esamus balsių nustatymo algoritmus, naudojamus balsių identifikavimo požymius;
4. išanalizuoti plačiausiai taikomus skaitmeninius filtrus;
5. iš analizės ir eksperimentinių rezultatų sudaryti tiriamo požymio algoritmo modelį;
6. realizuoti informacinę sistemą tiriamam požymiui apskaičiuoti;
7. atlikti pagrindinį eksperimentą tiriamo požymio tikslumo įvertinimui;
8. atlikti papildomus eksperimentus tiriamo požymio rezultatų patikslinimui.

**Tyrimo metodai:**

1. kokybiniai tyrimo metodai: pagrindinio tono paieškos metodų kalbos signaluose klasifikacija, balsių atpažinimo metodų klasifikacija, mokslinės literatūros analizė ir sisteminimas bei UML modeliavimo kalba algoritmų modeliavimui;
2. kiekybiniai tyrimo metodai: algoritmo konfigūracijų našumo palyginimas pagal įvairius matematinius ir statistinius vertinimo kriterijus.

Darbe remtasi užsienio ir Lietuvos autorių mokslinėmis publikacijomis, knygomis, keletu interaktyvių šaltinių. Išanalizuoti panašaus pobūdžio ir tematikos darbai atlikti darbo vadavo mokslinėje grupėje, plačiai apžvelgti pasaulio mokslininkų pasiekimai balsių atpažinimo ir balsių nustatymo srityje. Viso rašant darbą remtasi 36 literatūros šaltiniais.

**Darbo struktūra.** Darbas susideda iš trijų pagrindinių dalių: pirmoje dalyje apibūdinamos pagrindinės sąvokos, aprašomi jau esami balsių atpažinimo ir nustatymo metodai, bei garso signalo filtravimo metodai; antroje – sudaromas metodo modelis, apibūdinami skaičiavimams naudojami algoritmai; trečiojoje – aprašomas pagrindinis ir papildomi eksperimentai, jų rezultatai, eksperimentinė aplinka.

Darbą sudaro 67 puslapiai, 45 paveikslai, 12 lentelių ir 8 formulės.

Darbo rezultatai paskelbti „Informacinės technologijos 2009“ konferencijos pranešimų medžiagoje. Darbo metu atliktas pagrindinio tono nustatymo metodų tikslumo tyrimas paskelbtas „Informacinės technologijos 2008“ konferencijos pranešimų medžiagoje.

# 1. PAGRINDINIAI NUSTATYMO METODAI IR SKIRIAMIEJI BRUOŽAI

Šioje darbo dalyje aprašomi balsių ir priebalsių skiriamieji bruožai, Lietuvių kalbos fonemų klasifikavimo būdai, aiškinama pagrindinio tono prasmė, jo nustatymo metodai, analizuojami balsių atpažinimo ir identifikavimo algoritmai. Pateikiamos pagrindinio tono, balsių atpažinimo metodų, formančių radimo metodų klasifikacijos. Taip pat Aprašomi skaitmeniniai filtrai, analizuojami jų tipai ir dizaino kategorijos.

Remiantis šiame skyriuje atlikta analize, sekančiame darbo skyriuje realizuojamas balsių nustatymo algoritmo modelis.

## 1.1. Balsių ir priebalsių diferenciacija.

Balsiai ir priebalsiai yra skirtingos kilmės garsai. Tariant balsius, oro srovė virpina suartėjusias balso stygas, bet ryklės ir burnos ertmėje nesutinka tikrų kliūčių ir išeina pro burną palyginti laisvai. Todėl balsiai vadinami nekliūtiniais garsais (PAKERYS, A. 1995).

Artikuluojant priebalsius, kurioje nors burnos vietoje susidaro kliūtis: aklina uždaruma arba bent ankštuma (plyšys). Pro tą kliūtį balsinga ar nebalsinga oro srovė turi skverbtis su didesne jėga, negu tariant balsius. Pavyzdžiui, tardami priebalsį p, iš pradžių lūpas sučiaupiame ir užtveriname kelią oro srovei, o po to šią uždarumą susprogdiname; tardami s, tarp priešakinės liežuvio dalies ir beveik sukąstų dantų paliekame siaurą plyšelį, pro kurį skverbiasi oras. Dėl šių priežasčių priebalsiai vadinami kliūtiniais garsais (PAKERYS, A. 1995).

Savitas priebalsių poklasis – pusbalsiai – pagal artikuliacijos būdą užima tarpinę padėtį tarp balsių ir kitų priebalsių (PAKERYS, A. 1995).

Žmonės beveik be klaidų gali atskirti balsius nuo priebalsių, tačiau automatizuotas balsių atskyrimas nuo priebalsių nėra itin paprastas uždavinys. Triukšmas, dvigarsiai, pusbalsiai – tai kliūtys, dėl kurių balsių atskyrimas nuo priebalsių tampa sudėtingu uždaviniu.

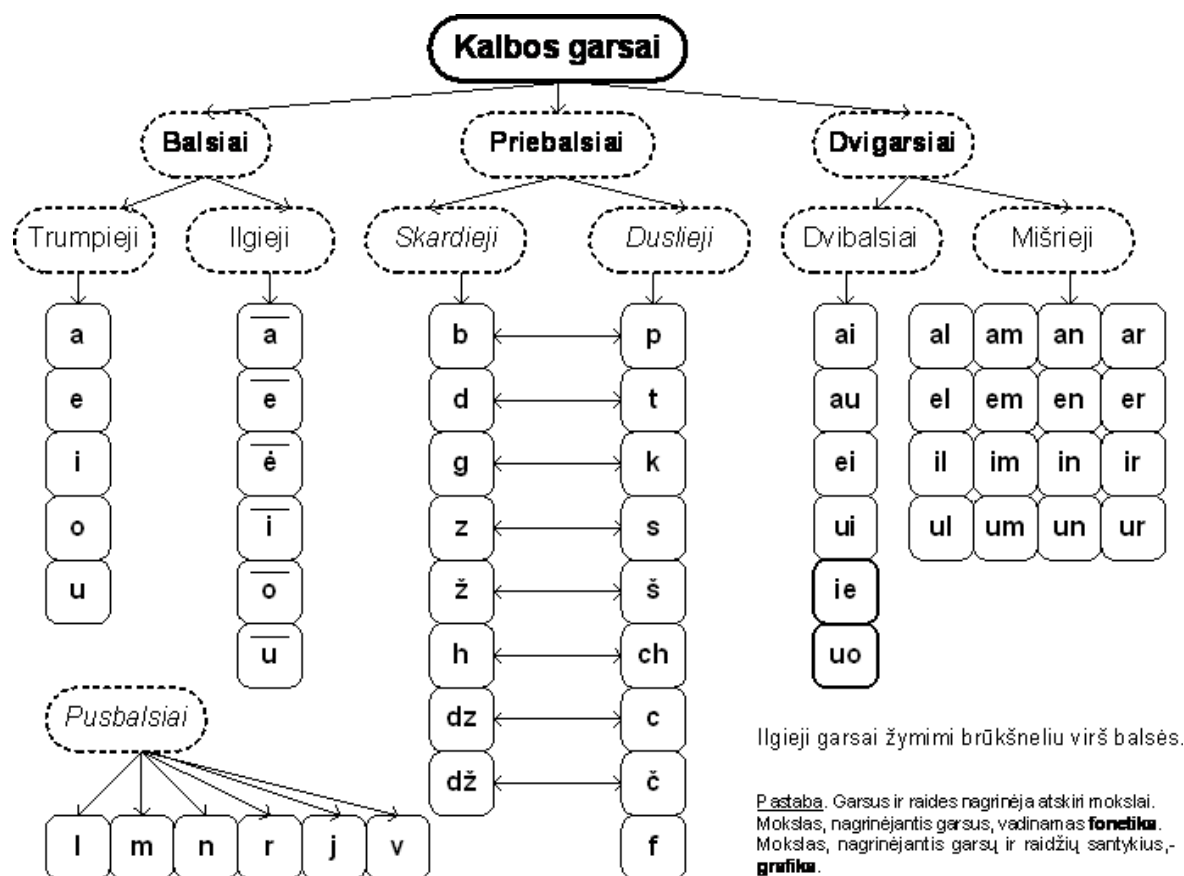
### 1.1.1. Fonemų klasifikavimo požymiai

Lietuvių kalboje fonemos, o tiksliau priebalsiai, klasifikuojamos dvejopai. Pirmasis klasifikavimo būdas – artikuliacinė klasifikacija. Artikuliacinė klasifikacija interpretuoja garsus kaip priežastinę kalbos padargų veiklą. Pagal artikuliacines savybes priebalsiai klasifikuojami į sprogstamuosius ir pučiamuosius. Taip pat iškiriami pusbalsiai, kuriuos dar priimta klasifikuoti į paprastuosius ir balsinguosius sonantus (RAGUOTIENĖ, G. 1998).

Akustinė fonemų klasifikacija remiasi kalbos padargų sukeltais virpesiais. Artikuluojant priebalsius oro srovė skverbiasi pro tam tikras kliūtis ir sukelia neperiodinius virpesius, kuriuos

suvokiame kaip triukšmą (trenksmą, ūžesį, šlamesį, švilpesį). Virpesių pobūdis ir kiekis nėra vienodas. Duslieji priebalsiai susideda tik iš neperiodinių virpesių, o skardieji priebalsiai – iš pagrindinio tono ir neperiodinių virpesių. Pusbalsiai yra gana toningi garsai su nedideliu triukšmo priedu (URBONAVIČIUS, A).

Bendru atveju, abstrakti fonemų klasifikacija tiek paremta akustinėmis fonemų savybėmis, tiek artikuliacinėmis, pateikiama 1 paveiksle. Detali priebalsių artikuliacinė klasifikacija pateikiama darbo prieduose, 5 priede.



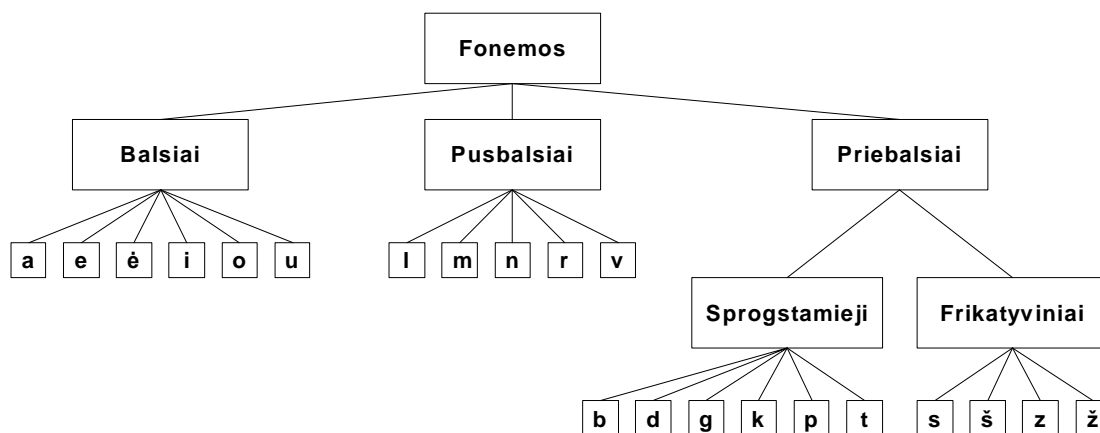
Šaltinis: URBONAVIČIUS, Alfredas. *Fonetikos sąvoka*. Prieiga per internetą:

<http://ualgiman.dtiltas.lt/kalbos%20garsai.html>

### 1 pav. Lietuvių kalbos garsų klasifikacija

Dauguma automatinio kalbos atpažinimo technologijų remiasi pavyzdžiais arba šablonais grįstais metodais (*angl. pattern recognition*). Pagrindinis šių metodų principas yra tai, kad pradžioje yra paruošiami nagrinėjamų kalbos vienetų etalonai, kurie atpažinimo metu palyginami su atpažįstamu matavimų vektoriumi. Fonemų atpažinimas - tai uždavinys, kurio tikslas priimti sprendimą kokiam fonetiniam elementui atitinka analizuojamas kalbos signalo fragmentas. Paprasčiausias pavyzdžiais grįsto fonemų klasifikavimo algoritmas – kai nagrinėjamas kalbos signalo fragmentas yra lyginamas su kiekviena etalonine reikšme ir sprendimo priėmimo etape analizuojamam fragmentui yra priskiriamas geriausias atitikimo rezultatus parodžiusio etalono simbolis (DRIAUNYS, Kęstutis; RUDŽIONIS, Vytautas; ŽVINYS, Pranas. 2005).

Fonetikos teorija lietuviškų fonemų aibę interpretuoja kaip fonetinę medžio tipo hierarchiją, kur fonemos atitinka “lapus” ir jungiamos į tam tikras grupes (balsiai, priebalsiai). Supaprastintas ir adaptuotas LTDIGITS fonemų aibei tokio medžio pavyzdys pateikiamas 2 paveiksle. Šie fonetiniai medžiai vadinami dendrogramomis (DRIAUNYS, Kęstutis; RUDŽIONIS, Vytautas; ŽVINYS, Pranas. 2005).



Šaltinis: DRIAUNYS, Kęstutis; RUDŽIONIS, Vytautas; ŽVINYS, Pranas. (2005). *Hierarchine fonemų struktūra grindžiamo LTDIGITS fonemų klasifikavimo tyrimas*. 2005. p. 1.

## 2 pav. LTDIGITS fonemų hierarchinė struktūra

Fonemų skirstymo į stambias grupes metodika ir požymiai išsamiai aptarti D. Childers monografijoje (CHILDERS, D. G. 2000). Autorius siūlo klasifikuoti balsius atliekant tam tikrą seką veiksmų, kuri plačiau detalizuojama 3 paveiksle.



Šaltinis: sukurta autoriaus pagal CHILDERS. D. G. *Speech processing and synthesis toolboxes*. 2000. – 484 p.

## 3 pav. Fonemų skirstymo į stambias grupes metodika

Kaip matome paveiksle, balsių nustatymo etapas išskirtas raudona spalva. Tai etapas, kuris bus realizuotas šio darbo rezultatu – balsių nustatymo algoritmu.

## 1.2. Pagrindinis tonas

Svarbi sąvoka, tariant balsius, balso stygų sukelti periodiniai oro virpesiai — vadinamasis pagrindinis tonas. Jis sužadina ryklės, burnos (kartais ir nosies) ertmių rezonansą. Rezonansiniai tonai ir lemia balsių kokybę — tembrą. Taigi balsiai yra toningiausi, sonoringiausi (skambiausi) garsai. Pagrindinis tonas gali būti nustatomas tiek iš laikinės, tiek ir iš dažninės kalbos signalo reprezentacijos (URBONAVIČIUS, A).

Artikuluojant priebalsius oro srovė skverbiasi pro tam tikras kliūtis ir sukelia neperiodinius virpesius, kuriuos suvokiame kaip triukšmą (trenksmą, ūžesį, šlamesį, švilpesį). Virpesių pobūdis ir kiekis nėra vienodas. Duslieji priebalsiai susideda tik iš neperiodinių virpesių, o skardieji priebalsiai — iš pagrindinio tono ir neperiodinių virpesių. Pusbalsiai yra gana toningi garsai su nedideliu triukšmo priedu (URBONAVIČIUS, A).

Pagrindinio tono ( $f_0$ ) paieška nuo skaitmeninių signalų tyrimų pradžios iki šių dienų yra aktuali mokslinių tyrimų sritis. Pagrindinis tonas apibūdinamas ir vadinamas labai įvairiai. Tiksliausias apibūdinimas, iš kurio nesunku išskirti dar vieną pagrindinio tono pavadinimą, tai - fundamentalusis dažnis. Fundamentalusis dažnis dažnai apibrėžiamas kaip garso bangos žemiausios pasikartojančio signalo komponentės dažnis. Tačiau  $f_0$  nebūtinai turi būti žemiausio pasikartojančio komponento dažnis. Periodinė garso banga turi daug tarpusavyje harmoniškai susijusių komponentų, o tai turėtų reikšti, kad visos šios komponentės yra susijusios su žemiausia, kurios dažnis dažniausiai yra fundamentalusis dažnis.

Sukurta daugybė pagrindinio tono nustatymo metodų, tačiau nei vieno iš jų negalima vadinti visapusiškai universaliu metodu, t.y. metodu pateikiančiu geriausius rezultatus iš bet kokio garso signalo. Skirtingi metodai, priklausomai nuo garso signalo, pasižymi skirtingais rezultatais. Vieni geriau nustato pagrindinį toną šnekamojoje kalboje, kiti muzikos instrumentų garsuose. Kiekvienas metodas daugiau ar mažiau, tačiau skirtingai reaguoja į signalo triukšmą ir kitus faktorius, tokius kaip šnekėtojo dialektą, lytį, muzikinio instrumento kokybę ir t.t.

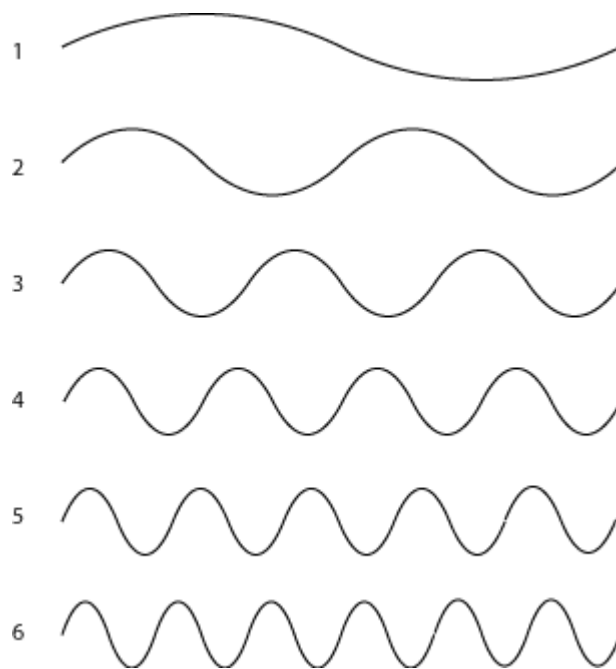
Dauguma šiuolaikinių kalbos signalo atpažinimo sistemų ignoruoja pagrindinio tono nustatymą, o jų kūrėjai traktuoja pagrindinį toną kaip nereikalingą garso signalo parametras. Tam priežasčių iš tikrųjų yra: fonemų įvardijimas nesunkiai atliekamas be pagrindinio tono nustatymo; pagrindinio tono nėra šnabždant. Tačiau nepaisant išvardintų priežasčių pagrindinis tonas yra naudingas dėl perduodamos semantinės informacijos, kuri be abejojimo yra aukščiau fonetinio ir leksinio lygio. Kalbos signalų atpažinimui tai yra itin svarbu siekiant įvardinti tam tikrus signalo požymius kaip pavyzdžiui diktoriaus lytį. Būtent todėl aukščiausio lygio kalbos atpažinimo sistemos turėtų

remtis pagrindiniu tonu, o tuo pačiu ir jo nustatymu, pasirinkus tinkamiausią pagrindinio tono radimo metodą.

### 1.3. Harmoninė garso bangos struktūra

Dauguma garsų pasižymi harmonine struktūra. Garso harmoninė struktūra – tai eilė harmonikų, dar kitaip vadinamų daliniais tonais. Pirmoji harmonika – žemiausio dažnio dalinis tonas – yra pagrindinis tonas (KASPARAITIS, P. 2007).

Kaip jau minėta anksčiau, pagrindinis tonas yra mažiausia pasikartojanti garso bangos komponentė. Žinant pagrindinį toną nesunku suskaičiuoti visas garso bangos harmonikas padauginant pagrindinį toną iš harmonikos skaičiaus pridėjus vieneta. Pavyzdžiui 1 harmonika apskaičiuojama pagrindinį toną padauginus iš dviejų, antroji – iš trijų, ir t.t. (PIERCE. John R. 2001). 4 paveiksle pateikta paprasčiausia harmoninė garso struktūra išgaunama daugelio muzikos instrumentų.



Šaltinis: KASPARAITIS, P. *Kompiuterinė lingvistika*. 2007. Prieiga per internetą:  
<<http://www.mif.vu.lt/~pijus/CL/cl.htm>>.

#### 4 pav. Harmoninė garso bangos struktūra

Harmoninė garso struktūra laikinėje garso bangos reprezentacijoje nepateikia tiek daug naudingos informacijos kaip dažninėje. Dažnio reprezentacija, dar kitaip vadinama signalo spektru aiškiau išreiškia garso harmoninę struktūrą, garsui pasižymint sudėtinga harmonine struktūra. Harmoninio garso dažninėje reprezentacijoje atsiskleidžia dar vienas itin svarbus garso požymis – formantė. Formančių tyrimai itin aktualūs tiriant kalbos signalus ir analizuojami 1.7.2. skyrelyje.



## 1.4. Pagrindinio tono nustatymo algoritmai

Pagrindinio tono nustatymo metodų yra daug, tačiau šioje srityje plačiau taikoma keletas nusistovėjusių metodų, kurie šiame poskyryje analizuojami detaliau. Analizei parinkti metodai, tai:

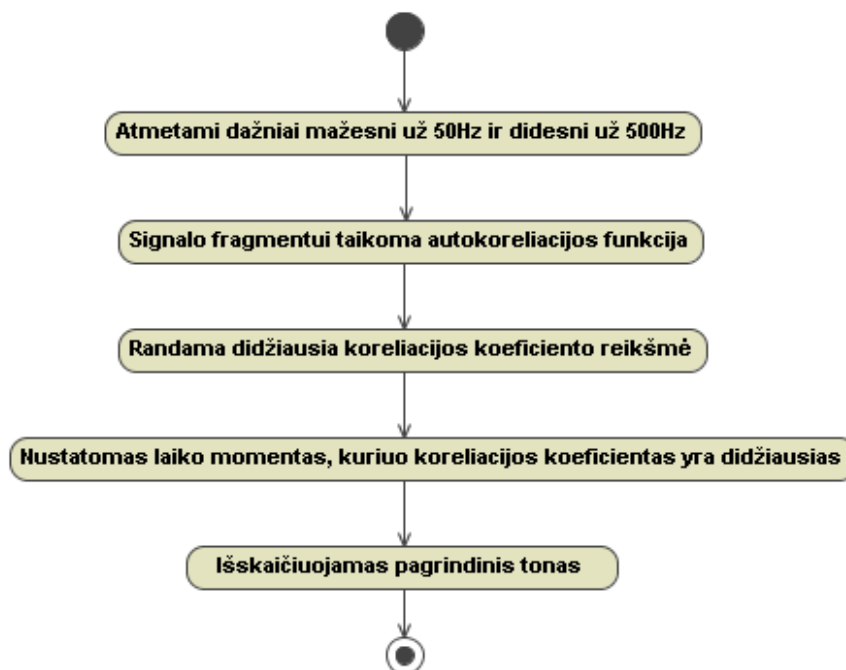
- auto-koreliacinis metodas (laiko domenai);
- kepstras (dažnio domenai);
- mokslininko „Sylvian Marchand“ pasiūlytas Furjė transformacijų kombinacinis metodas.

Tiek autokoreliacija, tiek kepstras – metodai, kurie šiam tikslui taikomi jau daugiau kaip dešimtmetį. Tačiau abu šie metodai yra pakankamai greiti ir efektyvūs, kad išlaikytų pozicijas iki šių dienų. Trečiasis metodas pasirinktas, kad detaliau pažvelgti į šioje srityje bandomas pritaikyti inovacijas. Furjė transformacijų kombinacinis metodas yra vienas naujausių publikuotų metodų skirtų pagrindiniam tonui nustatyti.

### 1.4.1. Pagrindinio tono radimas autokoreliacijos būdu.

Šis būdas vienas seniausių būdų nustatyti pagrindiniam tonui. Jis plačiai taikomas iki šių dienų. Metodo esmė - palyginti dviejų garso signalo laiko domeno intervalų panašumą. Taip pagal intervalo postūmio reikšmę (*ang. lag*) imamas kitas intervalas ir taip iki garso pabaigos. Tiesa kuo tolimesnis intervalas imamas, tuo mažesnis ryšys gaunamas. Šis metodas greitas ir efektyvus išskyrus tam tikrus atvejus, kuriais jis tampa visiškai nenaudingas. Tai dažniausiai nutinka nagrinėjant harmoniškai sudėtingą struktūrą turinčius garsus (PROLOY, Bhattacharya; YUMAO, Lu; YANG, Wu; HUNG\_WEI, Kuo. 2007).

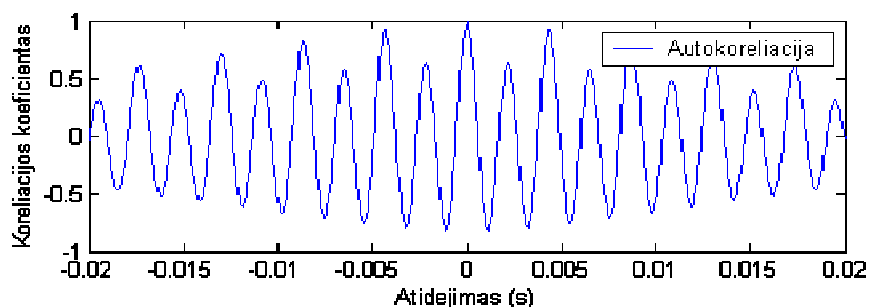
Auto-koreliacinis metodas – tai metodas pagrįstas faktu, jog periodinio signalo (jeigu jis ir ne sinusoidės formos banga) periodai tarpusavyje yra panašūs. Nustatant pagrindinį toną, imamas ne mažesnio ilgio signalas nei signalo fragmentas su dvigubai didesniu ilgiu už ilgiausią signalo periodą (BOERSMA, P. 1993). Signalas filtruojamas eliminuojant dažnius didesnius už 500Hz ir mažesnius už 50Hz, t.y. atmetant dažnius, kuriuose neegzistuoja tikimybė aptikti pagrindinį toną. Turimam signalo fragmentui taikoma autokoreliacijos funkcija. Laiko trukmė, atitinkanti didžiausią šios funkcijos piką, nustatoma kaip pagrindinis tonas (BOERSMA, P. 1993). Pagal ją išskaičiuojamas pagrindinis tonas. Auto koreliacijos algoritmo veiksmų seka pateikiama 5 paveiksle.



Šaltinis: sukurta autoriaus pagal P.Boersma. *Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound*. p. 17.

### 5 pav. Autokoreliacijos algoritmo veiksmų seka

6 paveiksle pateikiamas balsio e, kurį taria moteris autokoreliacijos funkcijos grafikas.



Šaltinis: sukurta autoriaus.

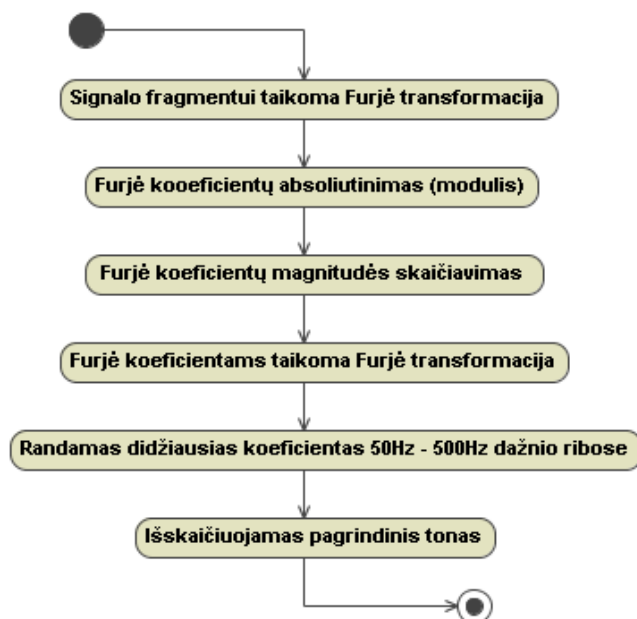
### 6 pav. Autokoreliacijos metodas

Kaip matome autokoreliacijos grafike (6 pav.), didžiausia funkcijos reikšmė 0.9268 yra laiko taške  $\sim 0.0045$ . Tai atitinka dažnį 225.3521Hz. Šis dažnis ir yra pagrindinis signalo dažnis  $f_0$ .

#### 1.4.2. Pagrindinio tono radimas taikant kepstrą.

Šis metodas panašus į aukščiau aprašytą dviejų Furjė transformacijų taikymu, tačiau šiuo atveju antroji transformacija yra inversinė ir taikoma pirmos transformacijos logaritmuotoms reikšmėms. Šis algoritmas taikomas seniai, tačiau jo tikslumas nėra itin didelis (PROLOY, Bhattacharya; YUMAO, Lu; YANG, Wu; HUNG\_WEI, Kuo. 2007).

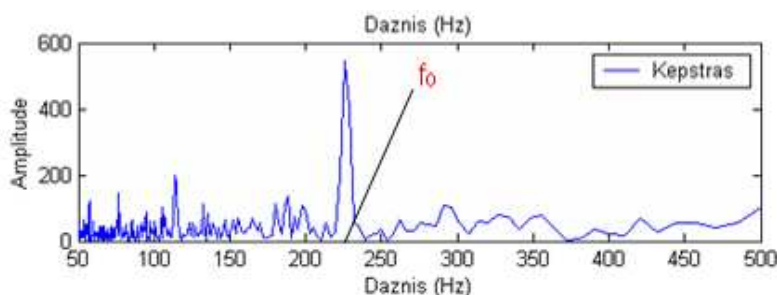
Kepstras - metodas pagrįstas spektru iš spektro. Trumpai apibūdinant, visą algoritmą galima suskaidyti į keletą etapų: Furjė transformacija; modulis; log; Furjė transformacija (7 pav.) (NGUYEN, Danny. 2005).



Šaltinis: sukurta autoriaus pagal NGUYEN, Danny. (2005) *Enhanced Spectrogram Tools for Speech Signals*. Prieiga per internetą: <<http://users.ece.gatech.edu/~danny/ECE%206255%20Project.ppt>>.

**7 pav. Kepstro algoritmas**

Pirmos transformacijos metu gautų koeficientų modulių logaritmai yra antros Furjė transformacijos įreiga. Antros Furjė transformacijos viršūnės dažnis yra pagrindinis signalo dažnis.



Šaltinis: sukurta autoriaus.

**8 pav. Kepstras**

Daugelyje vadovėlių ir mokslinių straipsnių kepstro sąvoka siejama su kiek kitokiu algoritmu. Yra teigiama, jog kepstras – tai inversinė Furjė transformacija pirmos Furjė transformacijos koeficientams. Tačiau iš tikrųjų, t.y. tik dar vienas signalų analizės algoritmas, kuris neturėtų būti vadinamas kepstru (SHIMODAIRA, Hiroshi; NAKAI, Mitsuru. 1992).

8 paveiksle pateikiamas kepstro grafikas tiriamam signalui. Kaip matome, kepstro grafike aiškiai matosi kepstro funkcijos viršūnė ir dažnio taškas, kuris yra pagrindinis tonas. Jo reikšmė šiuo atveju 225.3521Hz, t.y. lygiai tokia pati kaip ir autokoreliacijos metodo taikymo atveju. Svarbu pažymėti, kad funkcijos maksimumas ieškomas tik filtruotame signale, t.y. nuo 50Hz iki 500Hz.

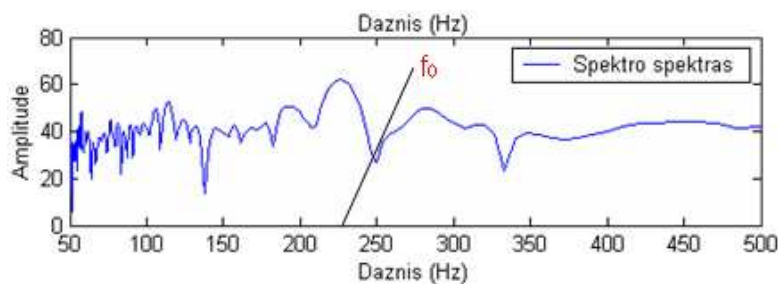
### 1.4.3. Furjė transformacijų kombinacinis metodas.

Šis metodas pagrįstas Furjė transformacijos Furjė transformacija. Kitaip sakant signalui yra pritaikoma Furjė transformacija, o gautoms reikšmėms, kurios vietoje signalo reikšmių tampa FFT įeiga, dar kartelį atliekama transformacija. Pagrindinio tono apskaičiavimas atliekamas taikant 1 formulę, kurioje diskretizavimo (*ang. sampling*) dažnis  $F_s$  dalinamas iš dviejų ir ši išraiška dalinama iš Furjė transformacijos maksimalios reikšmės indekso  $i$  (KUK KIM, Young; LEE, Ki Young; JIN BAE, Myung. 2005).

$$F = \frac{F_s/2}{i_{FT(FT)}} \quad (1)$$

Gautas dydis  $F$  yra garso pagrindinis tonas. Šis metodas yra pakankamai tikslus ir greitas, jei Furjė transformacijai parenkamas 512 taškų skaičius (KUK KIM, Young; LEE, Ki Young; JIN BAE, Myung. 2005). Jis įgalina aptikti pagrindinį toną ne tik kalbos signalų, bet ir muzikos instrumentų bei kitokios prigimties garsams.

„Sylvian Marchand“ pasiūlytas Furjė transformacijų kombinacinis metodas yra labai panašus į keptrą. Šiame algoritme atliekama panaši veiksmų seka, t.y. kombinuojamos dvi Furjė transformacijos. Pirmosios Furjė transformacijos koeficientai išreikšti decibelais yra antrosios Furjė transformacijos įeiga. Antrosios Furjė transformacijos reikšmės taip pat išreiškiamos decibelais ir joms pritaikomas Hamming'o langas (KUK KIM, Young; LEE, Ki Young; JIN BAE, Myung. 2005). Pagrindinis tonas randamas tokiu pačiu principu kaip ir keptro atveju, t.y. ieškant maksimumo dažnio ribose nuo 50Hz iki 500Hz. Paveiksle žemiau (9 pav.) pateikiamas spektro spektras tiriamam signalui.



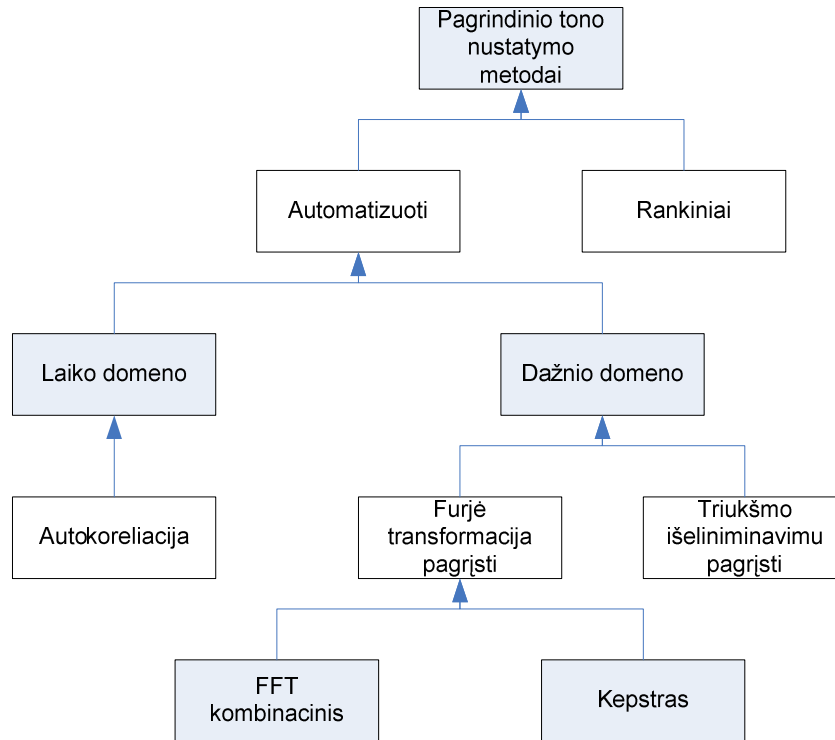
Šaltinis: sukurta autoriaus.

#### 9 pav. „Sylvian Marchand“ pasiūlytas Furjė transformacijų kombinacinis metodas

Kaip matome 9 paveiksle, spektro spektrui galime išvelgti viršūnę, kurios dažnio taškas ir yra pagrindinis tonas. Jo reikšmė šiuo atveju 225.3521Hz - tokia pati kaip ir kitų dviejų metodų atveju. O tai reiškia, jog pagrindinis tonas yra apskaičiuotas teisingai. Tai yra viena iš pagrindinio tono skaičiavimo ypatybių, kad tai yra dydis, kuris, jei teisingai apskaičiuotas, sutampa nepaisant to, kokie algoritmai buvo taikyti pagrindiniam tonui nustatyti.

#### 1.4.4. Pagrindinio tono nustatymo metodų klasifikacija.

Pagrindinio tono radimų algoritmų yra ir kur kas daugiau, tačiau šie metodai yra plačiausiai taikomi dėl savo tikslumo ir nusistovėjimo tarp kitų metodų kaip metodinės priemonės. Bendru atveju visus taikomus metodus galima suklasifikuoti (10 pav.).



Šaltinis: sukurta autoriaus pagal PROLOY, Bhattacharya; YUMAO, Lu; YANG, Wu; HUNG WEI, Kuo. *A Discussion of Pitch and Formant Estimation Methods*.

10 pav. Pagrindinio tono nustatymo metodų klasifikacinė sistema

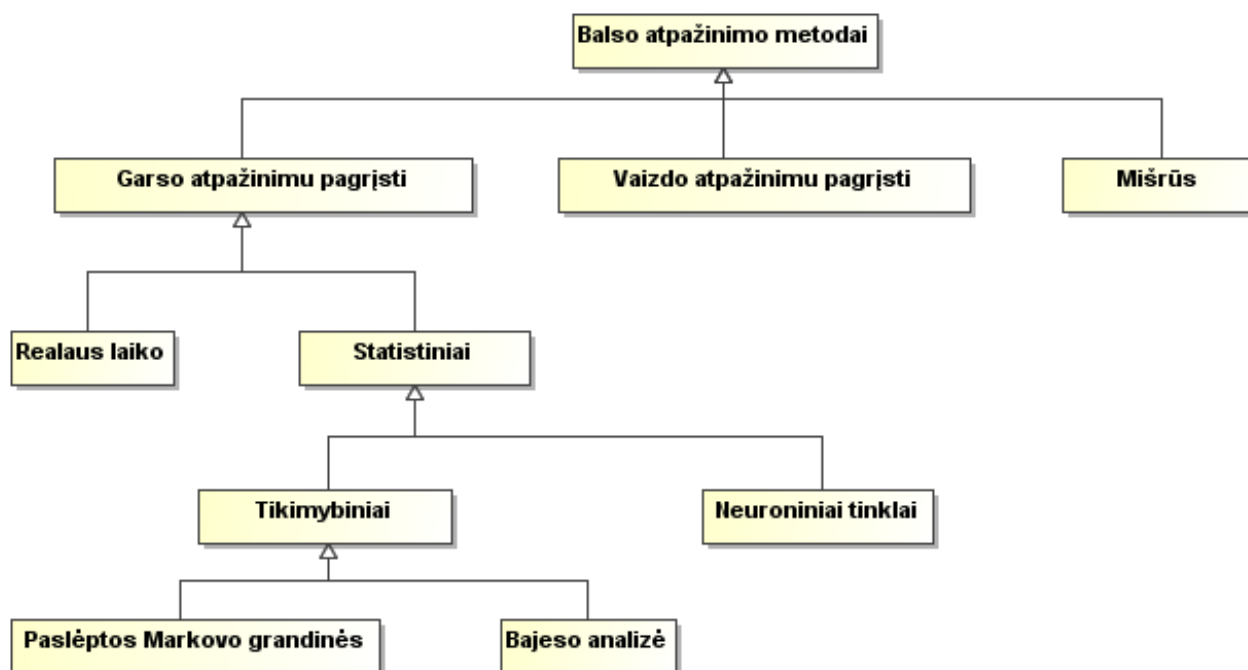
#### 1.5. Balsių atpažinimo algoritmai

Balsių atpažinimas yra viena iš svarbiausių kalbos atpažinimo dalių. Būtent dėl šios priežasties daugelis mokslininkų dirba šioje srityje siekdami atrasti naujus arba efektyvinti jau esamus balsių radimo metodus. O jų yra iš ties nemažai. Balsių atpažinimo metodo realizacijoje svarbiausia yra atrasti požymį, kuris galėtų atskirti balsius nuo kitų garsų. Nuo šio požymio tikslumo priklauso ir balsių atpažinimo metodo sėkmė. Idealus požymis – tai požymis, kuris bet kokioje aplinkoje, bet kokio diktorius šnekamojoje kalboje išskirtų visus balsius, su 100% tikimybe. Deja tokio požymio nėra ir tikėtina, kad nebus. Tai yra todėl, kad skirtingų tautų žmonės labai skirtingai taria tuos pačius garsus, taip pat didelę įtaką turi triukšmai, aplinkiniai garsai ir pan. Tyrimams naudojami laboratorijose įrašyti garsai labai stipriai skiriasi nuo balso atpažinimo praktikoje, kuomet aplinkoje vyrauja daugybė triukšmų.

Pastaruoju metu pradėti taikyti naujo pobūdžio balsių atpažinimo metodai. Jie paremti ne išskiriamu garsu, o šnekėtojo lūpų judesiais. Tai nauja mokslų sritis, kurioje siekiama padidinti kalbos atpažinimo efektyvumą taikant naują metodiką, taip pat kombinuojant naują metodiką su garso atpažinimu iš garso signalo. Šios metodikos trūkumas yra tai, kad norint atpažinti balsius, reikia filmuoti šnekėtoją, o tai tam tikrais atvejais tai yra neįmanoma.

Garso atpažinimu paremtų metodų specifika skiriasi nuo vaizdo atpažinimu pagrįstų. Šiuose methoduose identifikavus tam tikrą požymį sudaroma duomenų bazė, kuriuos duomenimis statistiškai nustatomas ištartas balsis. Tokie metodai dažniausiai remiasi neuroniniais tinklais arba Paslėptomis Markovo grandinėmis.

Apibendrinant galima sudaryti klasifikacinę balsių atpažinimo metodų schemą (11 pav.).



Šaltinis: sukurta autoriaus.

### 11 pav. Balsių atpažinimo metodų klasifikacija

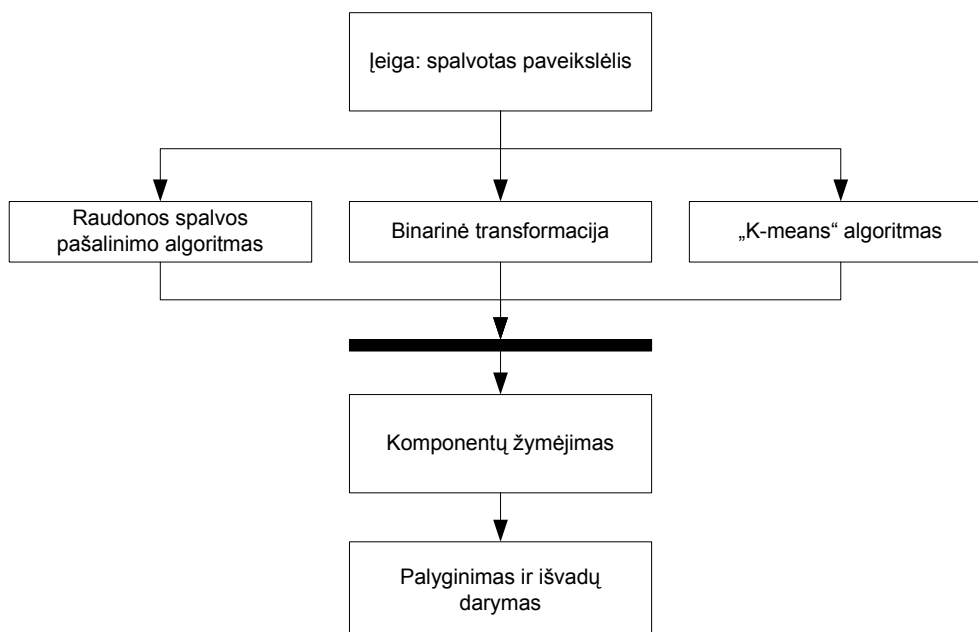
Keletas atpažinimo metodų analizuojami detaliau sekančiuose šio skyriaus poskyriuose.

#### 1.5.1. Balsių atpažinimas iš šnekėtojo lūpų

Pastaruosius keletą metų išplito balsių atpažinimo iš šnekėtojo vaizdo metodai. Analizuojamas metodas atpažįsta balsius remiantis šnekėtojo lūpų judesiais. Judesiai analizuojami ir pagal turimus pavyzdžius nustatomas ištartas balsis (SADEGHI, Vahideh Sadat, YAGHMAIE, Khashayar. 2006).

Analizuojant detaliau, lūpų analizė yra pakankamai sudėtingas dalykas. Lūpų dydžiai (aukštis, plotis), spalva yra labai skirtingi kiekvienam šnekėtojui. Metodas šiuo atveju taiko keletą

algoritmų: raudonos spalvos pašalinimo algoritmą, binarinę transformaciją ir „K-vidurkių“ klasterizavimą (SADEGHI, Vahideh Sadat, YAGHMAIE, Khashayar. 2006).



Šaltinis: sukurta autoriaus pagal Vahideh Sadat Sadeghi, Khashayar Yaghmaie (2006), *Vowel Recognition using Neural Networks* .p. 155. Prieiga per internetą: < [paper.ijcns.org/07\\_book/200612/200612A22.pdf](http://paper.ijcns.org/07_book/200612/200612A22.pdf) >.

### 12 pav. Balsių atpažinimo iš šnekėtojo lūpų metodo modelis

Laikoma, kad žmogaus veidas yra didžiąja dalimi raudonas, t.y. jame dominuoja raudona spalva. Kad atpažinti lūpas bendrame kontekste raudona spalva pašalinama, tuomet išryškėja spalviniai skirtumai tarp lūpų ir kitų veido dalių. Lūpos iškerpamos ir binariškai transformuojamos į juoda balta spalvas (SADEGHI, Vahideh Sadat, YAGHMAIE, Khashayar. 2006). Tuomet pritaikoma klasterizacija ir apskaičiuojamas lūpos aukštis bei plotis šešiose lūpos taškuose, nustatytuose klasterizacijos pagalba (13 pav.).



Šaltinis: SADEGHI, Vahideh Sadat, YAGHMAIE, Khashayar. (2006), *Vowel Recognition using Neural Networks*. p. 156. Prieiga per internetą: < [paper.ijcns.org/07\\_book/200612/200612A22.pdf](http://paper.ijcns.org/07_book/200612/200612A22.pdf) >.

### 13 pav. Lūpos aukščio ir pločio nustatymas

Gauti trys pločiai ir trys ilgiai dviejų neuroninių tinklų pagalba, kurių paslėpto neuronų sluoksnių perdavimo funkcija sigmoidinis tangentas, o aktyvavimo sigmoidinis logaritmas, apdorojami ir nustatomas ištartas balsis. Vienas neuroninis tinklas operuoja tik pločiais, o kitas tik ilgiais. Gauta rezultatų suma sudedama – tai galutinis rezultatas, kuriuo remiantis identifikuojamas balsis (SADEGHI, Vahideh Sadat, YAGHMAIE, Khashayar. 2006).

Šio metodo efektyvumas anot mokslininkų yra 70 % tikslumas atpažįstant balsius (SADEGHI, Vahideh Sadat, YAGHMAIE, Khashayar. 2006). Tai palyginus geras rezultatas, įvertinant tai, kad šis algoritmas remiasi tik vizualine informacija ir visiškai nesiremia garso signalo tyrimu.

### **1.5.2. Melino transformacija balsių atpažinimui**

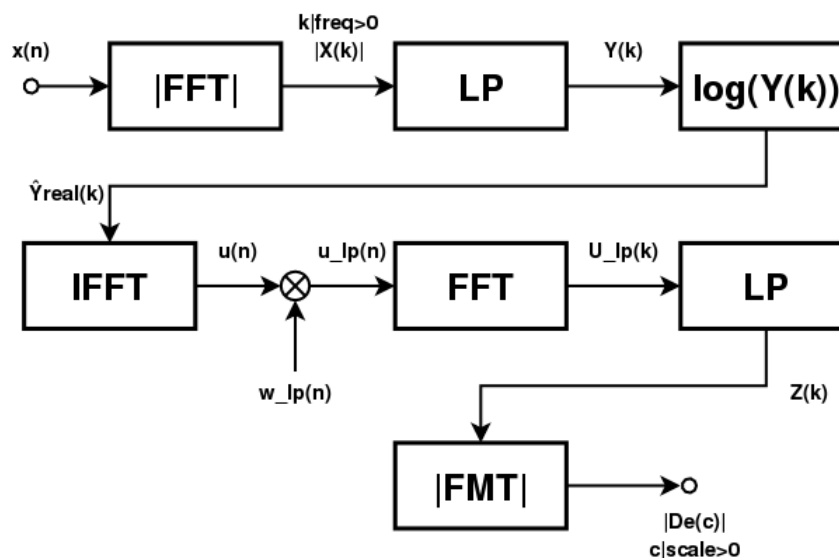
Šis balsių atpažinimo metodas pagrįstas hipoteze, kad tas pats balsis, nesvarbu kieno ir kokioje aplinkoje jis išstartas, visada turės panašią spektro kreivę su skirtingu suspaudimo (*ang. compession*) faktoriumi. Šis algoritmas yra realaus laiko algoritmas, t.y. algoritmas nepagrįstas statistiniais duomenimis, neuroniniais tinklais ar paslėptomis Markovo grandinėmis. Tiesa šioje vietoje yra vienas minusas – algoritmas sugeba skirti balsius, bet nesugeba pasakyti kuris balsis yra kuris (DE SENA, Antonio; ROCCHESSO, Davide. 2005). Be abejo tai tik laiko klausimas, kada algoritmas bus papildytas statistine analize ir sugebės įvardinti balsius.

Algoritmas susideda iš šių žingsnių (DE SENA, Antonio; ROCCHESSO, Davide. 2005):

1. Furjė transformacija;
2. Kepstras;
3. Amplitudė logaritmuojama;
4. Inversinė Furjė transformacija;
5. Furjė transformacija;
6. Amplitudė logaritmuojama;
7. Melino transformacija.

Gauta kreivė laikoma skiriamuoju tam tikro balsio požymiu, nepaisant kas ištarė balsį ar kokiomis sąlygomis jį ištarė (DE SENA, Antonio; ROCCHESSO, Davide. 2005). Grafiškai algoritmo seka atvaizduojama 14 paveiksle.





Šaltinis: DE SENA, Antonio; ROCCHESO, Davide (2005), *A study on using the mellin transform for vowel recognition*. p. 3. Prieiga per internetą: <[smc.afim-asso.org/smc05/papers/AntonioDeSena/DesRocSMC05.pdf](http://smc.afim-asso.org/smc05/papers/AntonioDeSena/DesRocSMC05.pdf)> .

#### 14 pav. Melino transformacija pagrįstas balsių atpažinimo metodas

Šis metodas ganėtinai naujas ir vis dar esantis mokslinių tyrimų stadijoje. Jis išskirtinis tuo, kad tai realaus laiko algoritmas, bet statistinių duomenų sugebančias identifikuoti atskirus balsius (DE SENA, Antonio; ROCCHESO, Davide. 2005). Apie metodo tikslumą autoriai duomenų nepateikia.

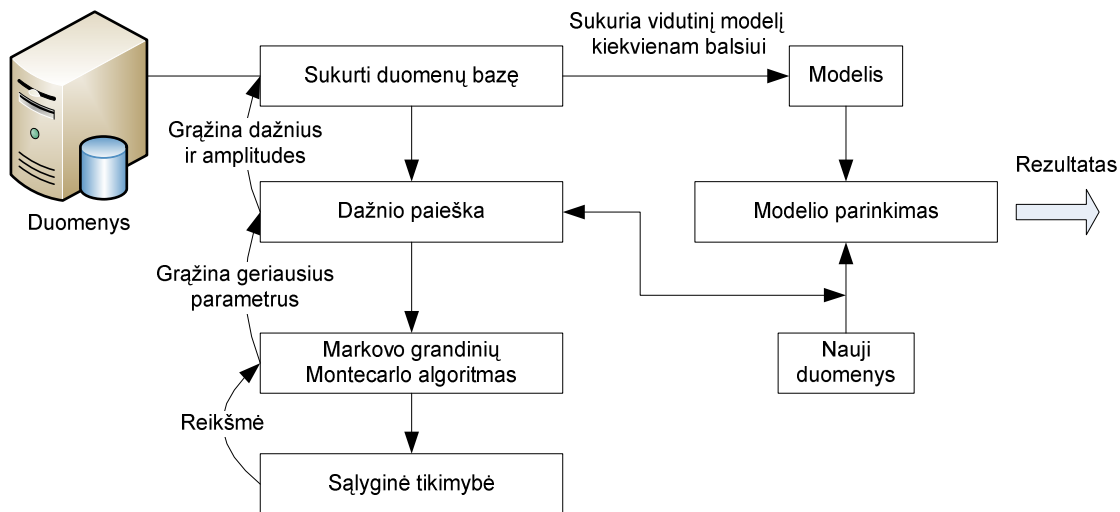
#### 1.5.3. Balsių atpažinimas naudojant Bajeso analize

Tai statistinis tikimybinis metodas. Jis pagrįstas požymiu, kurį sudaro devynių sinusoidžių suma. Šios sinusoidės, tai trys formantės ir šešios harmonikos. Ši suma apskaičiuojama pagal 2 formulę (BIRNSTIEL, Tilman. 2006).

$$s(t) = \sum_{i=1}^9 A_i \sin(2\pi f_i(t + \delta_i)) \quad (2)$$

Formulėje A – amplitudė, f – formančių ir harmonikų dažniai, o  $\delta$  – fazės pokytis.

Šio parametro pagrindu sudaromi balsių modeliai ir lyginami su tiriamo balsio modeliu. Naudojant Bajeso teoremą skaičiuojama tikimybė, kad tiriamas balsis gali būti laikomas tam tikru balsiu (BIRNSTIEL, Tilman. 2006). Metodo algoritmas pateikiamas 15 paveiksle.



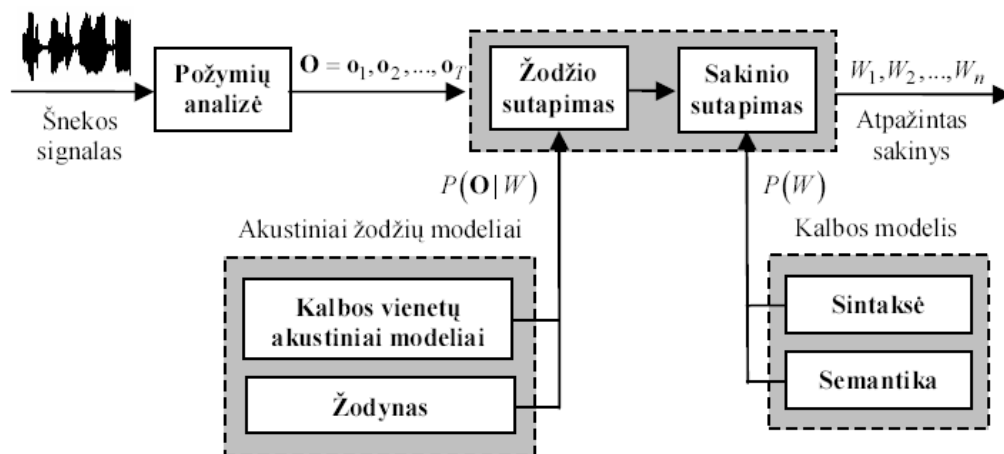
Šaltinis: sukurta autoriaus pagal BIRNSTIEL, Tilman; (2006) *Vowel Recognition using Bayesian Analysis*. p. 2. Prieiga per internetą: <[www.til-birnstiel.de/vortraege/Vowel.pdf](http://www.til-birnstiel.de/vortraege/Vowel.pdf)>.

### 15 pav. Balsių atpažinimo metodo naudojant Bajeso analizę algoritmo modelis

Atlikus tikslumo tyrimą su trejomis balsių a, i, o aibėmis (po tris balsius, tad viso 9 balsiai) prieita išvados, kad algoritmas yra tinkamas naudoti. Tik vienas netikslumas buvo užfiksuotas tyrimo metu su a balsiu. Nepaisant to metodas yra ganėtinai lėtas ir ateityje numatomas optimizuoti (BIRNSTIEL, Tilman. 2006).

## 1.6. Balsių išskyrimas iš kitų garsų šnekamosios kalbos signalė

Balsių išskyrimas (toliau nustatymas) šnekamosios kalbos signalė yra vienas pirmųjų ir vienas svarbiausių etapų. Balsių nustatymo metu siekiama kuo tiksliau identifikuoti balsius, t.y. atskirti juos nuo priebalsių. Apibendrintame šnekamosios kalbos atpažinimo medelyje, balsių nustatymas yra pirmasis etapas – požymių analizė. Šis etapas yra paruošiamasis etapas, kurio metu priklausomai nuo taikomo kalbos atpažinimo metodo, atliekami veiksmai, padedantis suklasifikuoti garsus į tam tikras grupes.



Šaltinis: LAURINČIUKAITĖ, Sigita (2008). *Lietuvių šnekos atpažinimo akustinis modeliavimas*. p. 38.

### 16 pav. ASA sistemos struktūra

Rabiner, Juang (1993m.) pateiktame ASA sistemos apibendrintame modelyje, pateiktame 0, galime matyti visus kalbos atpažinimo etapus ir balsių nustatymo algoritmo vietą tarp jų.

Požymiu analizės modulio funkcija – šnekos signalą parametrizuoti į požymių vektorių seką išlaikant atpažinimui svarbią informaciją apie ištariamus garsus. Išskirti požymiai turi pasižymėti tam tikromis savybėmis (LAURINČIUKAITĖ, Sigita. 2008):

- geru panašių kalbos garsų diskriminatyvumu;
- patogumu tolesniam naudojimui;
- statistinėmis savybėmis, invariantiškomis kalbėtojų ir terpių įvairovei,
- panašumu į žmogaus klausos aparato naudojamus požymius.

Skirtingų požymių tipų naudojimas implikuoja faktą, kad požymių tipo, pasižyminčio visomis išvardintomis savybėmis, nėra (LAURINČIUKAITĖ, Sigita. 2008).

## 1.7. Balsių nustatymo metodai

Balsiams identifikuoti kalbos signale yra keletas populiarių technikų. Plačiausiai taikomi spektrinės analizės metodai ieškant formančių. Formantės leidžia ne tik nustatyti balsių buvimo vietas, bet ir gan tiksliai juos atpažinti. Kitas paplitęs būdas – garsumo funkcija. Tai mokslininko D. Childers pasiūlytas metodas, kuris pasižymi tuo, kad spektrinė analizė nėra taikoma šio metodo apskaičiavimui.

Šių metodų tikslumą sunku įvertinti skaitine išraiška vienareikšmiškai, nei vienas jų praktiškai netaikytas Lietuvių kalbos balsiams nustatyti. Vis tik galima daryti tam tikras išvadas iš užsienio mokslininkų atliktų darbų. Pavyzdžiui arabų kalbos balsių nustatymui taikant formantes, nustatytas ~90% tikslumas (IQBAL, H. Rizwan; AWAIS, M.Muhammad; MASUD Shahid; SHAMAIL, Shafay. 2008) Garsumo funkcija D. Childers publikacijoje atskiriant balsius anglų kalbos garsus įvertinta 85%. Tačiau tai tik keletas bandymų nustatyti balsius. Balsių nustatymo tikslumas priklauso ne tik nuo kalbos, bet ir nuo aplinkybių, tokių kaip diktorius lytis, amžius, dialektas, triukšmas signale ir kt.

Balsių nustatymo algoritmas, kaip pagalbinis algoritmas balsių atpažinimui, klasifikacijai ar segmentacijai turi pasižymėti ne tik tikslumu, tačiau ir greitumu. Kitais žodžiais tariant, balsių nustatymo algoritmas turi būti greitas ir tikslus, kad būtų optimalus taikyti praktikoje.

### 1.7.1. Garsumo funkcija

Nustatant daugumą klasifikavimo požymių, D. Childers naudoja garsumo funkcija (*angl. volume function*)  $V(i)$  (CHILDERS, D. G. 2000). Indeksas  $i$  žymi intervalo numerį, kuriame apskaičiuota funkcija. Garsumo funkcija randama iš lygties (3 formulė) (BALBONAS, D; DAUNYS, G. 2005).

$$V(i) = \frac{1}{N_i} \sqrt{\sum_{m=A}^B \left| H_i \left( e^{j\pi \frac{m}{M}} \right) \right|^2}, \quad (3)$$

Čia  $N_i$  – atskaitų skaičius  $i$  intervale,  $A$  – juostinio filtro apatinę ribą atitinkančio dažnio indeksas,  $B$  – viršutinę ribą atitinkančio dažnio indeksas,  $M$  – maksimalus dažnio indeksas.

Intervalas yra laikomas balsingu, kai jis yra skardus ir žemųjų dažnių garsumo funkcija gerokai didesnė už aukštųjų dažnių garsumo funkciją. Balsingųjų garsų grupei priklauso: balsiai, pusbalsiai, nosiniai priebalsiai, sprogstamųjų priebalsių pradžia. Balsingumui nustatyti naudojami du juostiniai filtrai (žemųjų ir aukštųjų dažnių srityse) ir surandami filtrų išėjimo signalų garsumai. Juostinio filtro žemųjų dažnių srityje pralaidumo juosta yra 98 Hz - 898 Hz, o juostinio filtro aukštesniųjų dažnių srityje – 3691Hz – 5000 Hz (CHILDERS, D. G. 2000). Pirmosios reikšmės santykis su antrąja rodo balsingumo santykį  $R(i)$  (*angl. sonorant ratio*). Čia  $i$  – intervalo indeksas. Balsingumo santykis  $R(i)$  yra lyginamas su slenkstiniu skardumo lygiu  $T_{son}$ . Šitaip nustatomas balsingumo įvertis  $SS(i)$  (*angl. sonorant score*). Jei  $R(i) \geq T_{son}$  tai intervalas yra balsingas ir  $SS(i) = 1$ , priešingu atveju nebalsingas ir  $SS(i) = 0$  (BALBONAS, D; DAUNYS, G. 2005).

Balsiai nustatomi panašiai kaip ir balsingumas (CHILDERS, D. G. 2000). Keičiasi tik filtrų pralaidumo juostos. Jos atitinkamai yra 20 Hz – 996 Hz ir 1016 Hz – 5000 Hz. Tokiu pat principu apskaičiuojamas ir balsio santykis  $VWL(i)$  (*angl. vowel ratio*). Balsės įvertis  $VWLS(i)$  (*angl. vowel score*) yra apskaičiuojamas pagal antrą formulę.

$$VWLS(i) = \left\{ \begin{array}{ll} 0, & \text{jei } VWL(i) \geq T_v \\ 1, & \text{jei } VWL(i) \leq T_a \\ \frac{T_v - VWL(i)}{T_v - T_a}, & \text{jei } T_a \leq VWL(i) \leq T_v \end{array} \right\} \quad (4)$$

Dalies klaidų galima išvengti atlikus korekciją – intervalus laikyti neturinčiais balsių, kai intervalų trukmė yra trumpesnė nei 15 ms (BALBONAS, D; DAUNYS, G. 2005)

Panašiu principu randami pusbalsiai, skardieji, duslieji ir frikatyviniai priebalsiai. Tai vienas iš būdų nustatyti balsius šnekamosios kalbos signale.

### 1.7.2. Formantės

Balsių (iš dalies ir pusbalsių) spektrogramose išryškėja būdingieji dažniai, vadinami formantėmis. Svarbiausios yra dvi pirmosios formantės, sutrumpintai žymimos F1 ir F2. Pirmoji formantė (F1) yra 250 - 900 Hz, o antroji formantė (F2) – 400 - 2400 Hz diapazone (URBANAVIČIUS, Alfredas). Balsiui galima aptikti iki šešių formančių, tačiau užtenka pirmųjų dviejų, kad galima būtų jį identifikuoti. Dėl šios priežasties tyrimuose dažniausiai analizuojamos dvi pirmosios formantės ir tik retais atvejais daugiau formančių.

Balso traktui esant natūralioje būsenoje, formantės yra išsidėsčiusios kas 1000 Hz pradedant nuo 500 Hz. Taigi  $F_1 = 500$  Hz,  $F_2 = 1500$  Hz,  $F_3 = 2500$  Hz ir t.t. Formantės kinta balso traktui keičiant formą, t.y. tariant skirtingas fonemas. Suaugusiam vyrui, pirmoji formantė kinta intervale nuo 300 Hz iki maždaug 1000 Hz (KASPARAITIS, P. 2007).

Pagal pateiktą formančių dažnių statistinį diapazoną akivaizdu, kad antroji formantė prasideda nuo ~400 Hz, o pirmoji baigiasi ties ~900 Hz riba. Tad ieškant pirmosios formantės susiduriame su pavojumi, kad gali būti atrasta antroji ir atvirkščiai. Tai vienas didžiausių trukdžių formančių sėkmingam atpažinimui.

Formantė – tai vienas pagrindinių ir dažniausiai naudojamų parametru balsių atpažinimo ir nustatymo požymiams apskaičiuoti.

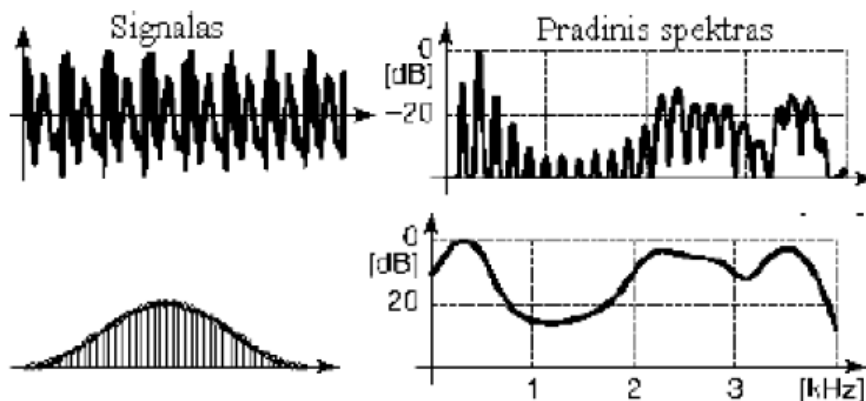
### 1.7.3. Formančių išskyrimas

Formantės tiek rankiniu, tiek automatizuotu būdu išskiriamos iš signalo dažninio spektro. Šis spektras sudaromas transformuojant signalą iš laiko domeno į dažninį domeną.

Formančių išskyrimas iš spektro reikalauja papildomo apdorojimo. Balso stygoms išskiriant pagrindinį toną, jis periodiškai kartojasi spektre ir pasižymi aukštomis amplitudėmis, tarp kurių atstumas yra lygus tono periodo ilgiui. Tokie spektro maksimumai trukdo teisingai įvertinti formantes. Šiai problemai išspręsti naudojamas spektro glodinimas. Glodinimas atliekamas keturiais būdais (KASPARAITIS, P. 2007):

1. tiesiogiai glodinant spektrą;
2. parenkant signalo nagrinėjimo intervalo ilgį mažesnę už tono intervalo ilgį;
3. panaudojant keprą;
4. skaičiuojant spektrą iš tiesinės prognozės koeficientų.

Tiesioginis spektro glodinimas atliekamas panaudojant glodinimo funkciją. Funkcija parenkama taip, kad spektre išsiskirtų formantės, o visi kiti spektriniai maksimumai būtų neutralizuoti. Spektro glodinimo pavyzdys pateiktas 17 paveiksle.



Šaltinis: KASPARAITIS, P. *Kompiuterinė lingvistika*. Prieiga per internetą <<http://www.mif.vu.lt/~pijus/CL/cl.htm>>.

### 17 pav. Spektro glodinimas

17 paveikslo viršuje pateiktas pagrindinis spektras, o apačioje – glodinimo funkcija ir glodintas spektras.

Spektro glodinimas yra pats paprasčiausias, tačiau ne itin efektyvus formančių radimo būdas. Praktikoje taikomi kur kas sudėtingesni ir tikslesni metodai, kuriuose spektras skaitmeninių filtrų pagalba dalinamas į dalis, o spektriniai maksimumai ieškomi kiekvienoje dalyje atskirai. Tačiau šiuo atveju iškyla problema, kaip nustatyti filtruojamas dažnių ribas, jei statistiniai formančių dažniai persidengia. Šis klausimas pastaruoju metu sprendžiamas mokslininkų, siekiant surasti efektyvų būdą spektro skaidymui į regionus. Dauguma tokių tyrimų atliekami su populiariausiomis pasaulio kalbomis, neįvertinant tokių kalbų kaip lietuvių, kuri yra turtinga balsiais ir pusbalsiais.

#### 1.7.3.1. Formančių išskyrimo metodai

Formančių išskyrimo metodai yra pagrįsti sudėtingais matematiniais algoritmais ir susideda iš dviejų pagrindinių algoritmų:

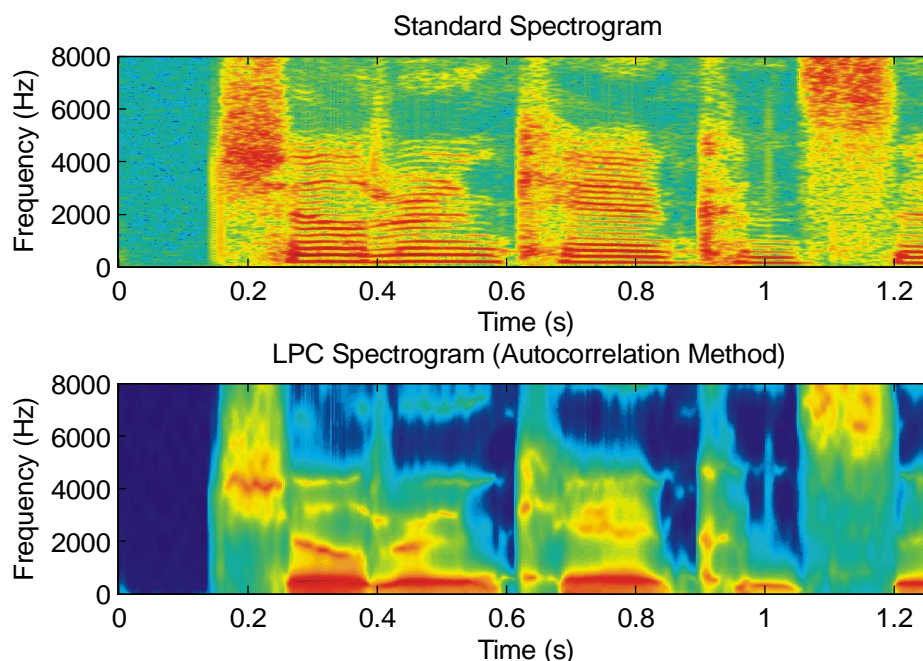
- spektro dalinimo į regionus algoritmo;
- formančių nustatymo algoritmo.

Tiek pirmuoju tiek antruoju atveju taikomų algoritmų yra nemažai. Tiesa vieno, įsigalėjusio kaip rekomenduotino standarto, nėra.

#### 1.7.3.2. LPC

LPC (*ang. Linear prediction coding*) - tai vienas pirmųjų algoritmų pradėtų taikyti formančių paieškai. Nepaisant to jis pasižymi aukštu efektyvumu ir yra plačiai taikomas kalbos signalų analizei. Šis algoritmas pagrįstas tiesiniu spėjimo metodu taikant koreliacijos koeficientus.

Dažniausiai LPC koeficientai apskaičiuojami iš FFT (*ang. Fast Furier transformation*) reikšmių (NGUYEN, Danny. 2005). LPC algoritmo vizualizacija – LPC spektrograma (18 pav.).



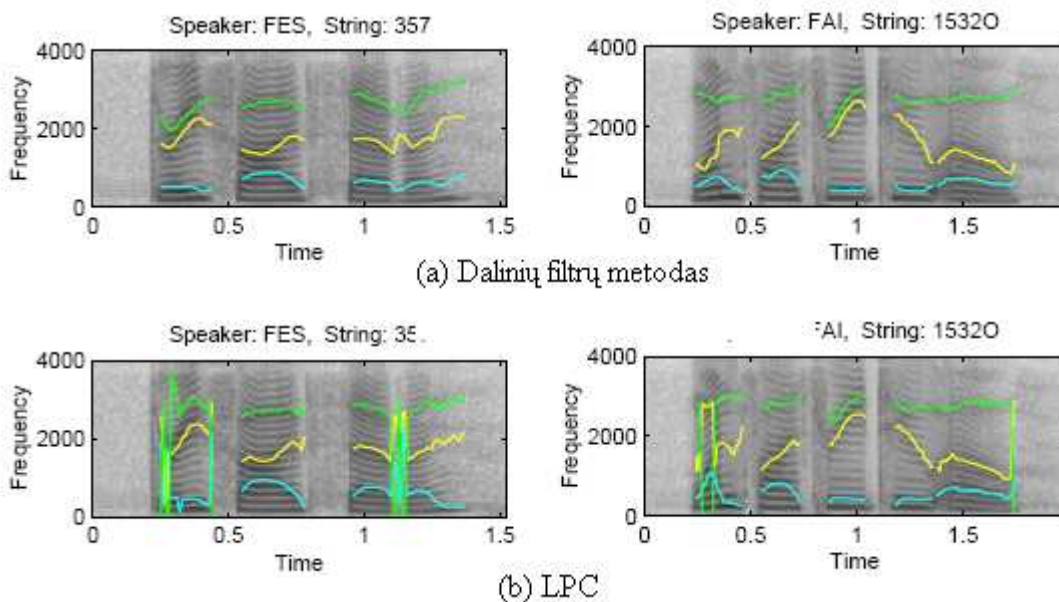
Šaltinis: NGUYEN, Danny. (2005). *Enhanced Spectrogram Tools for Speech Signals*. p.3.

### 18 pav. a) standartinė FFT spektrograma; b) LPC spektrograma

18 paveikslo b dalyje matome aiškiai išsiskiriančias energijos linijas, kurios ir yra vadinamos formantėmis. Šis algoritmas taikomas kaip mokomoji kalbos signalų apdorojimo priemonė, dėl santykinai ne itin sunkaus realizavimo ir ganėtinai didelio tikslumo.

#### 1.7.3.3. Dalinių filtrų metodas

Dalinių filtrų metodas (*ang. Particle filters*) – formančių išskyrimo metodas pasiūlytas 2003m. Jis pagrįstas Monte Carlo simuliacija. Šis metodas analizuoja nespaltotą kalbos signalo spektrogramą ir išskiria formančių linijas iš jos. Esminis šio metodo skirtumas nuo kitų formančių išrinkimų metodų, yra iš esmės ne garso signalo analizė, o spektrogramos, kaip paveikslėlio, analizė. Šis metodas plačiai taikomas medicinoje tiek 2D, tiek 3D paveikslėlių analizei. Tačiau pastaruoju metu jis pritaikomas ir įvairiose kitose srityse, kur tam tikrais atvejais jis pateikia itin tikslius rezultatus. Ne išimtis ir formančių nustatymas spektrogramoje (SHI, Yu; CHANG, Eric. 2003). 19 paveiksle pateikiamos šiuo metodu ir LPC metodu nustatytos formančių trajektorijos.



Šaltinis: SHI, Yu; CHANG, Eric; (2003) *Spectrogram-based format tracking via particle filters*. p. 4.

### 19 pav. a) Dalinių filtrų metodas; b) LPC metodas

Kaip matome 19 paveiksle, naudojant LPC metodą automatizuotam formančių radimui susiduriame su tam tikrais triukšmais, t.y. lokaliais maksimumais, kurie iškraipo formančių trajektorijas. Tuo tarpu naudojant dalinius filtrus, formančių linijos aiškios, t.y. visiškai be triukšmo.

#### 1.7.3.4. Formančių sekimas naudojantis fonetine informacija

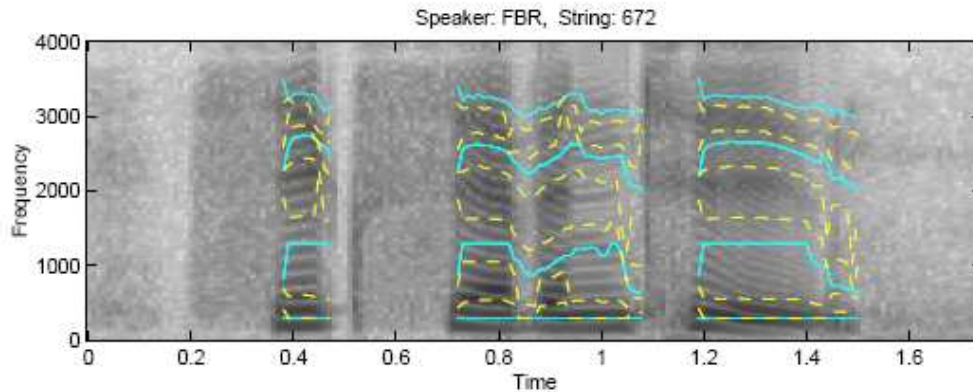
Tai dar vienas būdas išskirti kalbos signalo formantes. Šis būdas bene plačiausiai taikomas, tačiau reikalauja papildomos statistinės informacijos. Metodas plačiausiai taikomas universitetuose, kurie turi sukaupę didelius kiekius kalbos signalų.

Analizuojamas signalas pirmiausiai segmentuojamas į atskiras fonemas. Tai gali būti atliekama realiu laiku automatiškai arba gali būti naudojami susegmentuoti garso įrašai. Žinant kokia fonema yra kokioje vietoje nesudėtinga identifikuoti dažnius, kuriuose galima tikėtis tam tikros fonemos formantės. Tokiu būdu minimizuojama formančių persidengimo klaidos tikimybė. Neabejotinas šio metodo trūkumas – apsiribojimas vienos kalbos tyrimais (LEE, Minkyu; VAN SANTEN Jan; MOBIUS, Bernd; OLIVE, Joseph. 1999). Kitaip sakant rezultatai yra priklausomi nuo surinktų kalbos įrašų kokybės ir regiono, kuriame surinkimas buvo atliktas. Tad šį metodą galima vadinti lokaliu, tinkančiu tik tam tikrai sričiai ir kalbai.



### 1.7.3.5. Dinaminis programavimas

Tai nieko bendro su kompiuteriniu programavimu neturintis terminas, apibūdinantis matematinį algoritmą. Šis algoritmas taikomas jau ne kaip formančių paieškos algoritmas, bet kaip spektro dalinimo į regionus algoritmas.



Šaltinis: SHI, Yu; CHANG, Eric; (2003) *Spectrogram-based format tracking via particle filters*. p. 4.

### 20 pav. Spektras suskaidytas į regionus pagal dinaminio programavimo algoritmą

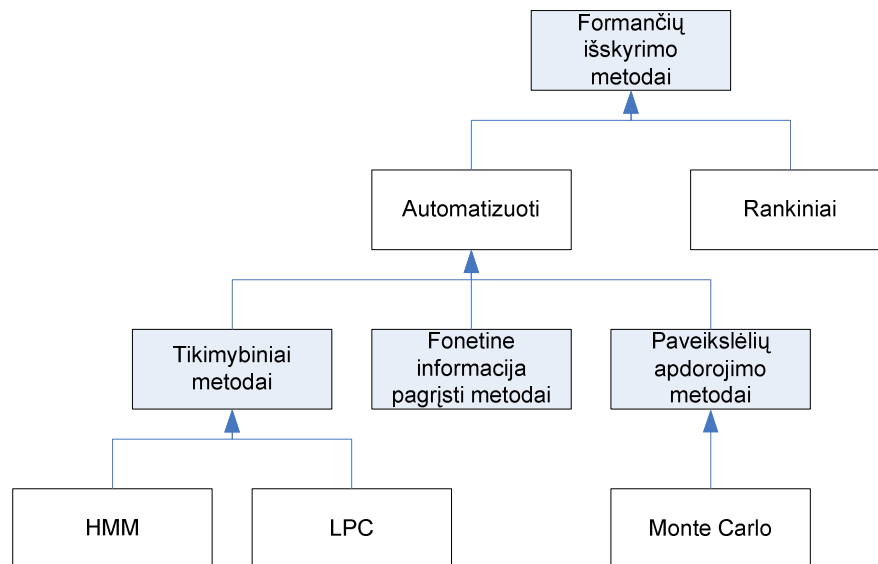
Dinaminis programavimas pagrįstas problemos dalinimu į subproblemas, kurių sprendimas yra išsimenamas. Galimas ir atvirkštinis algoritmas, t.y. didelės problemos sprendimas pagal subproblemas. Bet kokių atveju esminis principas yra teisingo sprendimo išsimenimas ir perpanaudojimas esant tam tikram reikalui (WAGNER, David B 1995). Kalbos signaluose šis algoritmas taikomas su tikslu apibrėžti dažnių intervalus, kuriuose galime tikėtis formantės. 20 paveiksle pateiktas šio algoritmo rezultatas, t.y. spektrograma išskaidyta į tam tikrus formančių regionus.

### 1.7.3.6. Paslėptieji Markovo modeliai

Šis algoritmas – alternatyva dinaminiam programavimui, pasižyminti tikslesniais rezultatais bei taikytina ne tik dažnių regionų išskyrimui. Markovo modeliai pasižymi apsimokymo savybe, t.y. jog syki jau aptikta situacija gali būti identifikuota pagal tai, kaip ji buvo identifikuota ankščiau (THIBAUT, Francois. 2003). Šis algoritmas žinomas kaip plačiai taikomas statistinis tikimybių teorijos algoritmas kalbos signalų atpažinime.

### 1.7.3.7. Formančių nustatymo metodų klasifikacija.

Paminėti formančių aptikimo metodai, vieni plačiausiai taikomų šiuo metu. Apžvelgus keletą modelių, galima sudaryti jų klasifikacinę sistemą pagal formančių radimo specifinius požymius. Ši schema pateikta 21 paveiksle.



Šaltinis: Sukurta autoriaus.

### 21 pav. Formančių radimo metodų klasifikacinė sistema

Straipsnyje klasifikuojami tik automatizuoti formančių išskyrimo metodai, kadangi rankiniai metodai yra neaktualūs darbe keliamam tikslui pasiekti.

## 1.8. Signalo filtravimas

Skaitmeninis filtras – tai sistema, kuri atlieka matematinę operaciją su signalu, kad sumažinti arba išaukštinti tam tikrus signalo aspektus.

Skaitmeninis filtras daugeliu atveju susideda iš trijų pagrindinių dalių: analoginio signalo konverterio į skaitmeninį signalą, procesoriaus ir skaitmeninio signalo konverterio į analoginį signalą. Skaitmeninis procesorius – tai filtro sudedamoji dalis, kuri atlieka matematinės operacijas.

Skaitmeninio filtro našumas yra ganėtinai prastas lyginant jį su analoginiu filtru, tačiau jo galimybės yra stipriai didesnės. Dėl prasto našumo, signalo išeiga laike stipriai vėluoja palyginus su signalo įeiga į filtrą (SMITH, Steven W. 1997).

Skaitmeninis filtras charakterizuojamas perkėlimo funkcija. Yra daug matematinių išraiškų, kurios gali būti taikomos skaitmeniniuose filtruose. Perkėlimo funkcija dažniausiai nusako filtro dizainą. Filto realizacija, tiksliau jo parinkimas, yra labai komplikuota tema ir šių dienų signalų

moksle yra vienas pagrindinių tyrimo šaltinių. Tuo tarpu filtro apskaičiavimas su šiuolaikinėmis techninėmis galimybėmis nėra sudėtingas.

Skaitmeniniai filtrai yra skirstomi į dvi pagrindines dizaino kategorijas (LYONS, R. G. 1997):

- Rekursiniai filtrai (IIR);
- Nerekursiniai filtrai (FIR)

Rekursiniai filtrai yra nestabilūs filtrai, kurie gali turėti ir nulius ir polių  $z$  plokštumoje. Nerekursiniai filtrai yra stabilūs ir  $z$  plokštumoje gali turėti tik nulius. Realizacijų yra kur kas daugiau: vienos jų taikomos vieno pobūdžio filtrams realizuoti, kitos kito pobūdžio filtrams realizuoti. Kai kurie, plačiau naudojami, filtrų dizainai aptariami šio skyriaus poskyriuose.

Pagal pobūdį skaitmeniniai filtrai skirstomi į keturias pagrindines kategorijas (LYONS, R. G. 1997):

- Juostos praleidimo (*ang. Band Pass*);
- Juostos uždraudimo (*ang. Band Reject*);
- Aukšto praleidimo (*ang. Highpass*);
- Žemo praleidimo (*ang. Lowpass*);

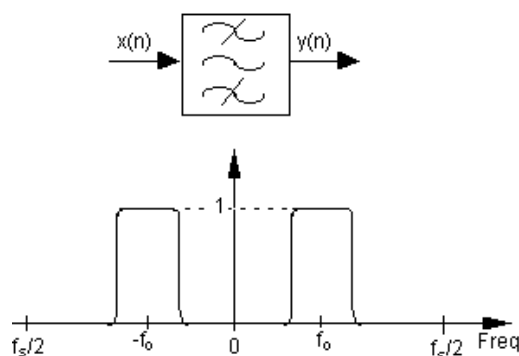
Taikomos ir kitokio tipo skaitmeninių filtrų kategorijos, jų mišrūs dariniai.

### 1.8.1. Juostos praleidimo filtras

Šis filtras, tiksliau jo pagrindinė idėja, yra praleisti tam tikrų dažnių juostą, o kitus (aukštesnius ir žemesnius) dažnius atmesti.

Praktikoje nėra idealaus tokio tipo filtro dizaino. Filtras nevysiškai tiksliai atskiria nenorimus praleisti dažnius. Atliekant tokio tipo signalo filtravimą, priklausomai nuo pasirinkto filtro dizaino ir jo laipsnio siekiama gauti kuo tikslesnius rezultatus (LYONS, R. G. 1997).

Juostos praleidimo filtro iliustracija pateikta 22 paveiksle.

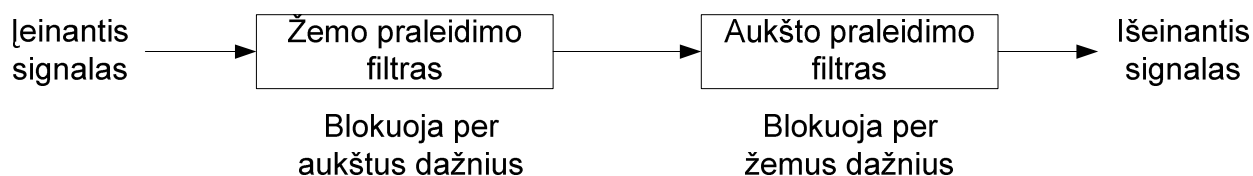


Šaltinis: Lyons, R. G. *Understanding Digital signal processing*. p. 494-505. Prieiga per internetą:

<http://www.dspguru.com/info/terms/filtterm/index2.htm>

22 pav. Juostos praleidimo filtras

Juostos praleidimo filtras lengviausiai realizuojamas kombinuojant žemo ir aukšto praleidimo filtrus (23 pav.).



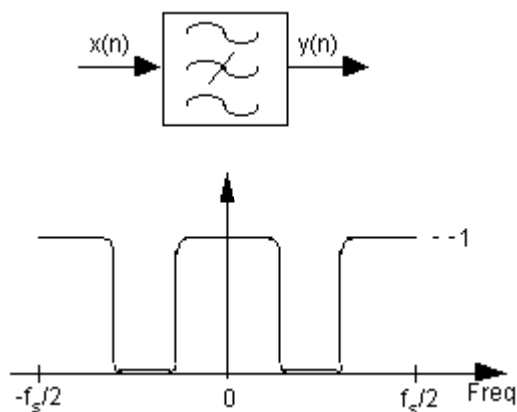
Šaltinis: sukurta autoriaus.

### 23 pav. Juostos praleidimo filtro realizavimo schema

Plačiau taikomi Beselio (*ang. Bessel*), Batervorto (*ang. Butterworth*), Eliptiniai (*ang. Elliptic*) ir nerekursiniai filtrai (FIR).

### 1.8.2. Juostos uždraudimo filtras

Juostos uždraudimo filtras priešingai nei juostos praleidimo filtras, nepraleidžia tam tikros dažnių juostos, o praleidžia aukštesnius ir žemesnius dažnius. Idealus juostos uždraudimo filtras dar vadinamas Nočo filtru (*ang. notch filter*) (LYONS, R. G. 1997). Tokio filtro pavyzdys pateikiamas 24 paveiksle.



Šaltinis: Lyons, R. G. *Understanding Digital signal processing*. p. 494-505. Prieiga per internetą:

<<http://www.dspguru.com/info/terms/filterterm/index2.htm>>

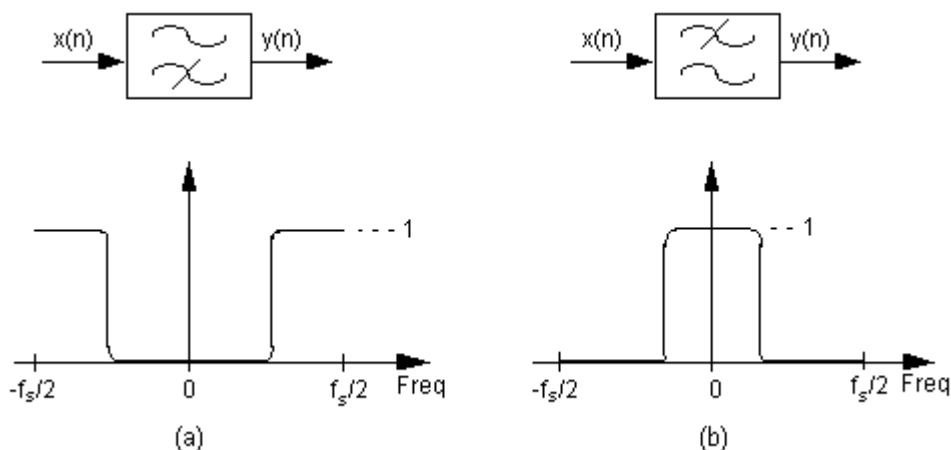
### 24 pav. Juostos uždraudimo filtras

Nočo filtrui realizuoti naudojami rekursiniai filtrai (IIR).

### 1.8.3. Aukšto ir žemo praleidimo filtrai

Aukšto praleidimo filtras praleidžia aukštus dažnius, bet nepraleidžia žemų dažnių. Žemo praleidimo filtras yra atvirkščio veikimo – praleidžia žemus dažnius, o nepraleidžia aukštų. Šie filtrų tipai yra labiausiai paplitę skaitmeninėje signalų analizėje ir gali būti realizuoti dideliu skaičiumi matematinių metodų. Šių filtrų dėka galima realizuoti juostos praleidimo filtrus ir juostos uždraudimo filtrus.

Aukšto ir žemo praleidimo filtrai pateikiami 25 paveiksle.



Šaltinis: Lyons, R. G. *Understanding Digital signal processing*. p. 494-505. Prieiga per internetą:

<http://www.dspguru.com/info/terms/filtterm/index2.htm>

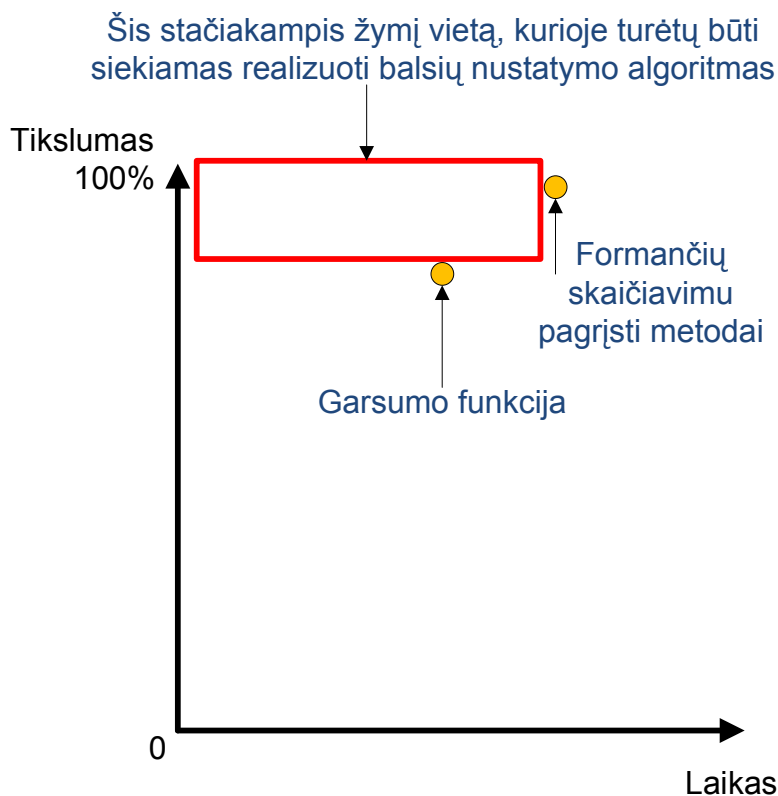
### 25 pav. Aukšto ir žemo praleidimo filtrai

Labiausiai paplitusios realizacijos, tai: Čebiševo, (*ang. Chebyshev*), Batervorto (*ang. Butterworth*), Eliptiniai (*ang. Elliptic*), nerekursiniai filtrai (FIR), rekursiniai filtrai (IIR) ir daug kitų.

## 1.9. Apibendrinimas

Atlikus Lietuvių kalbos fonemų klasifikavimo analizę, nustatyta, kad Lietuvių kalbos balsiai gali būti analizuojami akustiniu arba artikuliaciniu požiūriais. Nuo pasirinkto požiūrio priklauso ir fonemų klasifikacija. Bendru atveju išskiriami fonemų tipai: balsiai, pusbalsiai, skardieji ir duslieji priebalsiai.

Išanalizavus taikomus balsių nustatymo metodus, jų pritaikymo sritis, pagrindinio tono nustatymo ir skaitmeninių filtrų pritaikymo galimybes, galima sudaryti, kaip manoma, greitesnio ir ne mažiau tikslaus balsių nustatymo algoritmo modelį, kurio pagrindu būtų realizuotas optimalus algoritmas balsių ir priebalsių skiriamajam požymiui apskaičiuoti.



Šaltinis: sukurta autoriaus.

### 26 pav. Balsių nustatymo metodų optimalumas

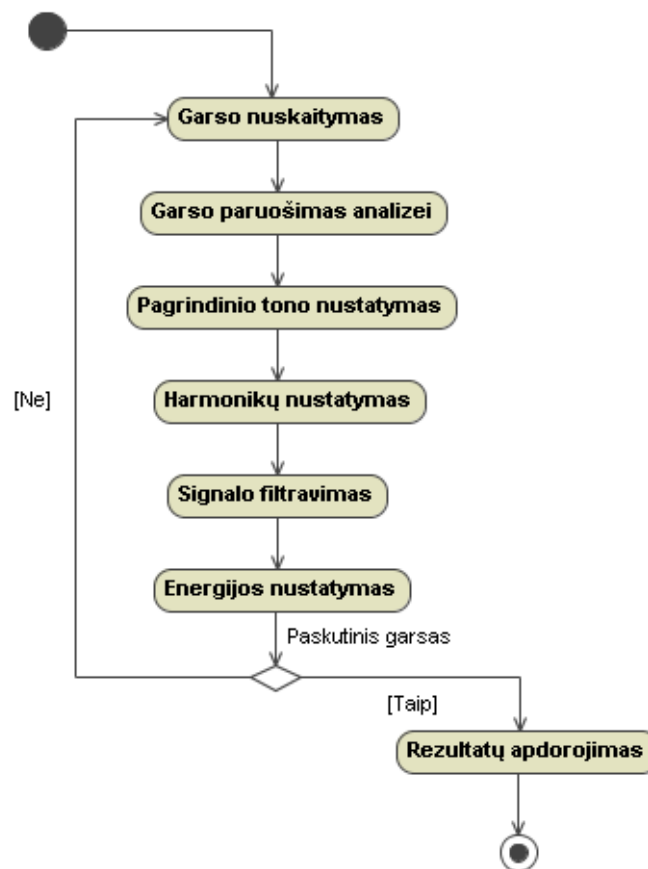
26 paveiksle matome kokią padėtį užima analizuoti balsių nustatymo metodai optimalumo erdvėje. Tikslumo atžvilgiu, kaip matome, taikant spektrinės analizės metodus pasiekiamas itin aukštas tikslumas. Tačiau nukenčiama laiko atžvilgiu. Kad sukurtas balsių nustatymo algoritmas būtų optimaliausias už kitus praktikoje taikomus algoritmus, jis turi patekti į raudona spalva išskirtą sritį. Tik tuo atveju bus galima teigti, kad iškelta darbo hipotezė yra teisinga.

## 2. BALSŪ SKIRIAMOJO POŽYMIO METODO MODELIS

Darbe siekiama patikrinti garso signalo požymį, paremtą signalo harmonine struktūra ir energijų tam tikruose filtruotuose regionuose suma. Tikėtina, kad šio požymio pagalba galima identifikuoti šnekamosios kalbos balsius. Kad apskaičiuoti ir statistiškai įvertinti šio požymio tikslumą ir praktinio panaudojimo galimybes turi būti atlikti tai įrodantys eksperimentai. Eksperimentams atlikti reikalinga sudaryti detalų požymio radimo metodo modelį: įvertinti ir parinkti pagrindinio tono radimo metodą; aprašyti kalbos signalų parengimo analizei algoritmus; identifikuoti papildomus skaičiavimus ir kalbos signalų šaltinį.

### 2.1. Suminės harmonikų energijos metodo modelis

Siekiant nustatyti ieškomą požymį, svarbiausia - pagrindinio tono radimas. Tačiau prieš apskaičiuojant pagrindinį toną, būtina identifikuoti šnekamosios kalbos įrašų šaltinį, kuris turėtų papildomos semantinės informacijos apie garsų ribines (pradžios ir pabaigos) reikšmes tiriamame signale. Taip pat skaičiuojant pagrindinį toną, reikia parinkti garso parengimo algoritmą garso išgryninimui ir skaidymui į smulkesnes dalis - langavimui.



Šaltinis: sukurta autoriaus.

27 pav. Suminės harmonikų energijos metodo modelis

Apskaičiavus pagrindinį toną, turi būti suskaičiuojamos harmonikos. Požymio įvertinimui numatoma naudoti pirmąsias aštuonias harmonikas.

Nustačius garso harmoninę struktūrą, numatoma signalą filtruoti artimuose harmonikai dažniuose ir skaičiuoti filtruoto signalo energiją. Signalu filtravimui būtina parinkti tinkamus filtrus ir filtro eilę. Gautų energijų suma – ieškomas požymis.

Metodo modelis pateikiamas 27 paveiksle.

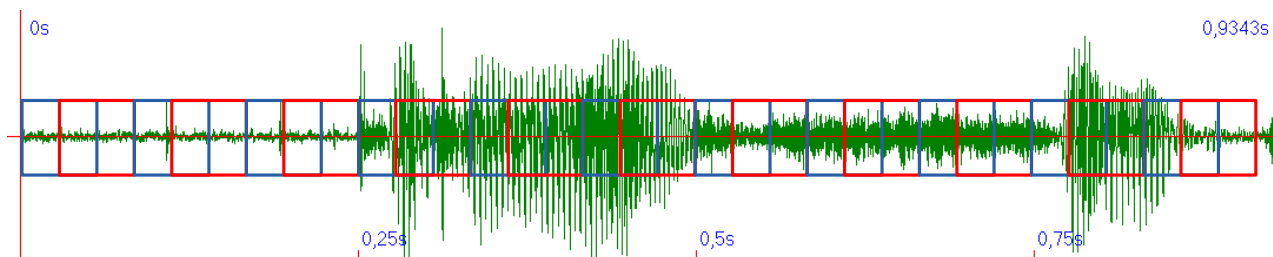
### 2.1.1. Garso paruošimas analizei

Siekiant gauti kuo tikslesnius analizės rezultatus, analizuojamus garsus reikia vienaip ar kitaip apdoroti, siekiant išgauti kuo grynesnį ir skaidresnį garso. Tai yra ypač svarbu skaičiuojant pagrindinį toną, kurio rezultatus labai stipriai įtakoja aplinkiniai garsai.

Šiuo atveju taikytini algoritmai, tai:

- Garso fragmento langavimas;
- Garso centro iškirpimas.

Garso fragmento langavimas tai signalo dalinimas į langus, kurių kiekvienas analizuojamas atskirai, o kaip vieningas rezultatas pateikiama vidurkinė atskirų langų tam tikro parametro reikšmė.



Šaltinis: sukurta autoriaus.

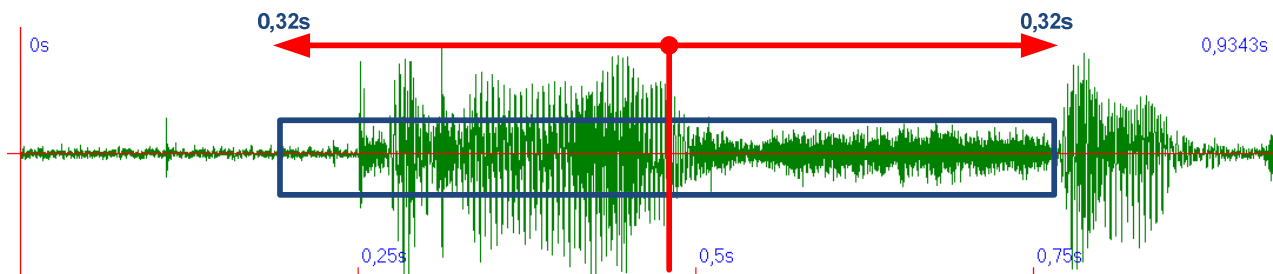
#### 28 pav. Signalu langavimas

Signalu langai, siekiant didesnio skaičiuojamo parametro reikšmės tikslumo, perdengiami, t.y. sekantis langas iškerpamas nuo prieš tai buvusio lango vidurio ir taip toliau. Galimi įvairūs langavimo variantai ir langų intervalai. Intervalai daugiausiai renkami atsižvelgiant į lyginį milisekundžių skaičių taip pat į Furjė transformacijos įėjimo taškų skaičių, jei garso analizei numatoma naudoti Furjė transformaciją. Tokiu atveju garso langas turi būti ne mažiau 512 diskretų. Dažniausiai imamas lango dydis 800 diskretų – 50ms.

Garso centro iškirpimas greitesnis, tačiau tam tikrais atvejais mažiau tikslumo teikiantis metodas. Šiuo atveju randama garso centrinė diskreta ir iškerpamas garso fragmentas 25ms prieš ir 25ms už nustatytos centrinės diskretos. Intervalas gali būti ir kitoks. Šiuo atveju, tikslumas gali būti stipriai sumažintas, jei garso ribos nustatytos klaidingai.



Aprašomo metodo atveju pasirinkta iškirpti 64ms garso fragmentą 32ms į vieną ir 32ms į kitą pusę nuo centrinės diskretos. Įvertinus diskretomis, šis skaičius, 16000Hz diskretizacijos garsams, yra 512 diskretų į vieną ir 511 diskretų į kitą pusę (plius centrinė diskreta). Šis intervalas pasirinktas dėl greitajai Furjė transformacijai reikalingo 1024 taškų skaičiaus.



Šaltinis: sukurta autoriaus.

### 29 pav. Garso centro iškirpimas

Centro iškirpimo metodas pasirinktas įvertinus, kad langavimo metodo taikymo atveju, įvertinus garsų ilgį, garsas suskaidomas į tris – keturis garso intervalus, kurių pirmasis ir paskutinis intervalas tam tikrais atvejais stipriai iškraipo matavimus, o viduriniai intervalai pateikia beveik identiškus rezultatus su centro iškirpimo metodu.

#### 2.1.2. Pagrindinio tono metodo parinkimas

Naudojant pirmoje darbo dalyje aprašytus automatinius pagrindinio tono nustatymo metodus (Autokoreliacija, Kepstrą ir Furjė kombinaciją), atliktas lyties nustatymo tyrimas, naudojantis „LTDIGITS“ garsyno surinktais šnekamosios kalbos įrašais. Tyrimas atliktas su 50 diktorių vyrų ir 50 diktorių moterų įrašais. Pagrindinis tonas skaičiuotas kiekvienam balsiui, visais trim metodais. Kiekvienam žodžiui išvestas jo balsių pagrindinio tono vidurkis. Taip daryta todėl, kad tam tikrais atvejais dėl ne itin tiksliai sužymėtų raidžių ribų, galimi gan dažni pagrindinio tono iškreipimai.

Tyrimais paremta pagrindinio tono riba tarp vyriškų ir moteriškų balsų yra 160Hz (LAWRENCE, M.D.; ANNE, A. 2004). Taip pat teigiama, jog vyriško pagrindinio tono intervalas turėtų svyruoti nuo 100Hz iki 150Hz, o moters – nuo 170Hz iki 220Hz (LAWRENCE, M.D.; ANNE, A. 1998). Šiais teiginiais remiantis atlikti du tyrimai:

1. Pagrindinio tono nustatymo klaida yra laikoma pažeista 160Hz riba;
2. Pagrindinio tono nustatymo klaida yra laikoma išėjimas iš ribų: vyrams intervalas 100Hz – 150Hz; moterims – 170Hz – 220Hz.

Remiantis gautais rezultatais palyginamas pagrindinio tono nustatymo metodų tikslumas kiekvienu atveju, skaičiuojant kiekvieno metodo klaidas pagal aprašytus kriterijus.

### 2.1.2.1. Rezultatai

Tyrimai atlikti „MATLAB“ programinės įrangos aplinkoje.

Pirmasis tyrimas atliktas su pagrindinio tono riba 160Hz, skiriant diktorių vyrą ir diktorių moterį. Jei nustatytas pagrindinis tonas daugiau už 160Hz, o tirtas vyro įrašas – fiksuojama klaida, atitinkamai, tik atvirkščiai - moterų atžvilgiu.

Viso ištirta 4800 žodžių. Pirmojo tyrimo rezultatai pateikiami 1 lentelėje.

1 lentelė

#### Pagrindinio tono nustatymo metodų tyrimo rezultatai

Metodas	Klaidų skaičius ir procentas
Autokoreliacija	122 klaidos 2.54%
Kepstras	134 klaidos 2.79%
Furjė transformacijų kombinacinis metodas	189 klaidos 3.94%

Šaltinis: sudaryta autoriaus.

Antrasis tyrimas atliktas su griežtais dažnio intervalais. Vyro pagrindinio tono intervalas 100Hz – 150Hz, moters 170Hz – 220Hz. Išėjus iš šių intervalų fiksuojama metodo klaida. Antrojo tyrimo rezultatai pateikiami 2 lentelėje.

2 lentelė

#### Pagrindinio tono nustatymo metodų su apibrėžtais dažnio intervalais tyrimo rezultatai

Metodas	Klaidų skaičius ir procentas
Autokoreliacija	1149 klaidos 23.94%
Kepstras	1049 klaidos 21.85%
Furjė transformacijų kombinacinis metodas	850 klaidos 17.70%

Šaltinis: sudaryta autoriaus.

Atliktų tyrimų rezultatus galima vertinti įvairiai. Remiantis 160Hz riba aptinkama apie 10 kartų mažiau klaidų nei remiantis pagrindinio tono intervalais. Tai didele dalimi sąlygoja vyriško pagrindinio tono nukrypimai į žemesnius dažnius ir moteriško tono nukrypimai į aukštesnius, tačiau negalima atmesti ir intervalų netinkamumo turimiems garso įrašams tirti. Pagrindinio tono skirtumus tarp lietuvių ir kitataučių, šiuo atveju amerikiečių, kur nustatyti skiriamieji diktorių lyties pagrindinio tono intervalai, gali lemti daugybė veiksnių, kaip aplinkos, kalbos ir kt. skirtumai (MORKEVIČIUS, Aurelijus. 2008).

Dar vienas įdomus dalykas, kad pagrindinio tono nustatymų naudotų tyrimuose rezultatai yra labai skirtingi tarp pirmojo ir antrojo tyrimų. Pirmame tyrime akivaizdu, kad autokoreliacijos metodas suklydo mažiausiai kartų, tačiau antrajame tyrime, šis metodas suklydo daugiausiai kartų. Apibendrinant visus tris pagrindinio tono nustatymo metodus ir abiejų tyrimų rezultatus, galima teigti, jog Kepstras – stabiliausias metodas (MORKEVIČIUS, Aurelijus. 2008).

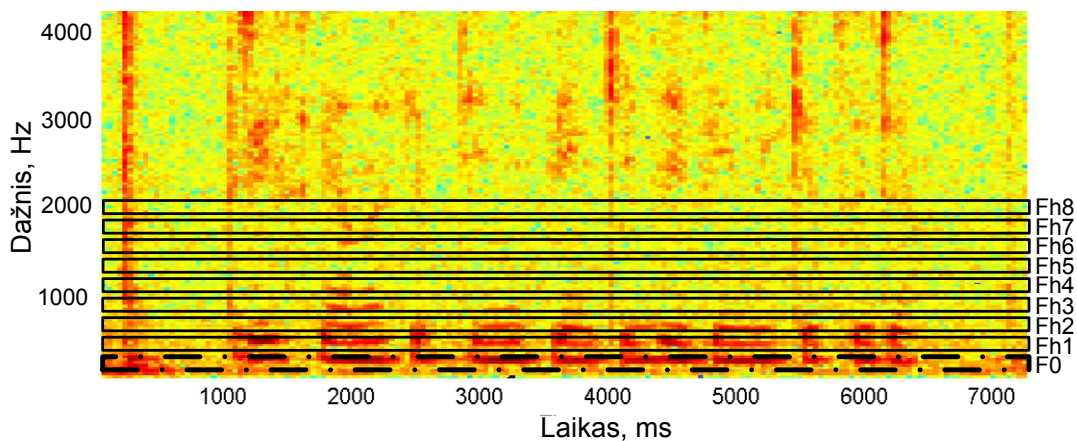
### 2.1.2.2. Apibendrinimas

Atlikus diktorių lyties nustatymo tyrimą remiantis pagrindinio tono nustatymu, parodyta, kad pagrindinis tonas yra tinkamas parametras diktoriaus lyčiai nustatyti (MORKEVIČIUS, Aurelijus. 2008):

- Nustatyta, kad pagrindinio tono nustatymo klaida svyruoja nuo 2,54% iki 3,94% taikant ribinę 160Hz pagrindinio tono reikšmę ir nuo 17,70% iki 23,94% taikant griežtus pagrindinio tono intervalus diktoriaus lyčiai įvardinti.
- Siekiant efektyvinti pagrindinio tono nustatymo metodų efektyvumą, reikia atlikti daugiau tyrimų, su tikslu nustatyti kiek galima tikslesnius vyrų ir moterų pagrindinio tono dažnio intervalus.
- Atlikti tyrimai leidžia daryti išvadą, kad iš tirtų pagrindinio tono nustatymo metodų stabiliausias savo tikslumu – Kepstras. Tad šis metodas bus naudojamas pagrindiniam šio darbo eksperimentui atlikti. Neatmetama tikimybė naudoti ir auto koreliacijos metodą ir palyginti jų parinkimo įtaką galutinio eksperimento rezultatams.

### 2.1.3. Skaitmeniniai filtrai

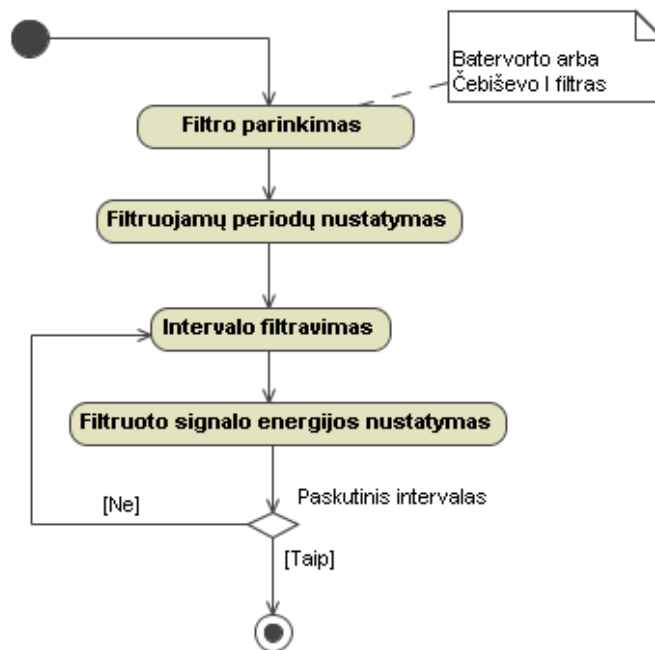
Identifikavus garso pagrindinį toną ir harmonikas, pagal metodo modelį, garsas turi būti filtruojamas. Filtravimas vyksta dažniuose, kur rastas pagrindinis tonas ir harmonikos (30 pav.).



Šaltinis: sukurta autoriaus.

#### 30 pav. Filtruojami dažnių intervalai

Turėdami pagrindinį toną ar vieną iš aštuonių harmonikų turime apskaičiuoti filtruojamus intervalus. Pasirinkto intervalo dydis gali būti svarbus rezultatams, tačiau, kad nustatyti koks intervalas geriausiai tinka norimiems rezultatams gauti reikia atlikti tyrimus bandymu ir klaidų keliu, jau turint metodo realizaciją. Šiuo metu imamas intervalas yra  $[F \cdot 0,75; F \cdot 1,25]$ .



Šaltinis: sukurta autoriaus.

### 31 pav. Filtravimo proceso veiksmų seka

Akivaizdu, kad esamai problemai spręsti reikia naudoti juostos praleidimo filtrą (*ang. band-pass*). Kur kas didesnė problema parinkti filtro realizaciją. Tai itin sunkus uždavinys, kuriam išspręsti reikia atskiro mokslinio tiriamojo darbo. Tad filtrai pasirinkti atsižvelgiant į jų realizaciją su „Matlab“ programiniu paketu, išraiškos sudėtingumą ir paplitimą. Parinkti trys filtrai:

- Batervorto (*ang. Butterworth*);
- Čebiševo (*ang. Chebyshev*);
- FIR (*ang. Finite Impulse Response*).

Kiekvienas jų aprašomas detalčiau sekančiuose šio skyriaus poskyriuose.

Filtravimo procesas aprašomame modelyje gali būti atvaizduotas veiksmų seka, kuri pateikta 31 paveiksle.

#### 2.1.3.1. Batervorto filtras

Batervorto filtras sukurtas taip, kad turėtų kiek galima matematiškai platų dažnio atsaką praleidimo juostoje. Jis realizuojamas aproksimuojant idealų filtrą su maksimaliai plokščia amplitudine dažnine charakteristika dažnių pralaidumo juostoje ir tam tikru charakteristikos kritimu už jos ribų. Fazinė dažninė charakteristika šiek tiek netiesinė (BARANIUK, Richard. 2005). Pagal savo architektūrą filtras dar vadinamas maksimaliai plačios amplitudės filtru.

Batervorto filtro amplitudės charakteristikos iki ribinio dažnio turi palyginti ilgą horizontalios dalies atkarpą ir pastovų nuolydį už ribinio dažnio. Pereinamajai charakteristikai

būdingi gana greitai nuslopstantys svyravimai. Kuo aukštesnės eilės filtras, tuo didesnė svyravimų amplitudė (PETRAUSKAS, Vytautas).

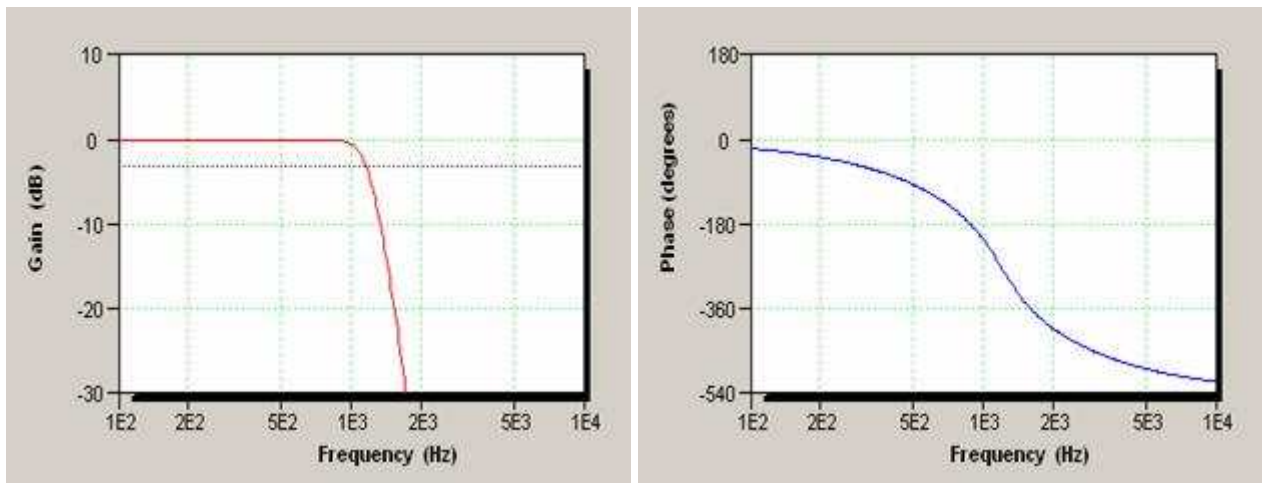
Kaip ir kitų filtrų, tipinis prototipas yra žemo praleidimo filtras, kurį modifikavus gaunamas aukšto praleidimo filtras, o juos kombinuojant juostos praleidimo ir uždraudimo filtrai.

Žemo praleidimo n pakopos filtro perdavimas  $G(\omega)$  apskaičiuojamas naudojant perdavimo (*ang. transfer*) funkciją, kurios formulė pateikta žemiau.

$$G^2(\omega) = |H(j\omega)|^2 = \frac{G_0^2}{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}} \quad (5)$$

Formulėje: n – filtro pakopa,  $\omega_c$ - iškirpimo dažnis (*ang. cutoff frequency*),  $G_0$  – perdavimas, kai dažnis lygus 0.

Nesunku pastebėti, kad n artėjant į begalybę, perdavimas tampa kvadratine funkcija ir dažniai žemesni už  $\omega_c$  praleidžiami su perdavimu  $G_0$ , o dažniai aukštesni nei  $\omega_c$  - nepraleidžiami. Mažesnės pakopos realizacijose iškirpimas yra mažiau smailus.



Šaltinis: PETRAUSKAS, Vytautas . *Žemų dažnių ketvirtos eilės Beselio filtro tyrimas*. Prieiga per internetą:

<<http://ifko.ktu.lt/~vypet/T120B012/filtrai1.htm>>

### 32 pav. 6 pakopų Batervorto filtro perdavimas ir dažninės charakteristikos

Paveiksle (32 pav.) pateikta 6 pakopų Žemo praleidimo Batervorto filtro realizacija.

### 2.1.3.2. Čebiševo filtras

Čebiševo filtro amplitudinė dažninė charakteristika turi normuojamą nelygiškumą dažnių pralaidumo zonoje ir staigų jo kritimą už pralaidumo dažnio ribos. Filtro perdavimo charakteristika netiesinė (PETRAUSKAS, Vytautas).

Čebiševo filtro amplitudės charakteristika pralaidumo juostoje yra banguota, o tuoj už ribinio dažnio staigiai krinta žemyn. Pereinamosios charakteristikos svyravimų amplitudė didesnė ir jie užgęsta ne taip sparčiai kaip Batervorto filtras (PETRAUSKAS, Vytautas).

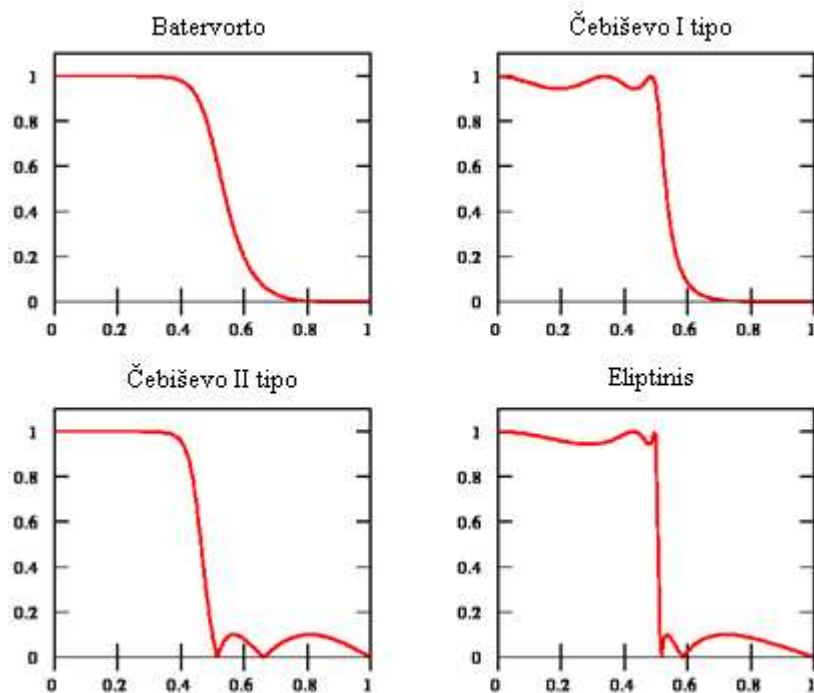
Tipinis prototipas yra žemo praleidimo filtras, tačiau šio filtro pagalba galima realizuoti bet kokius filtrų tipus taip pat ir tyrimui reikalingą juostos praleidimo filtro tipą.

Galimos dviejų tipų Čebiševo filtro realizacijos. Toliau analizuosime tik pirmo tipo Čebiševo filtrą, kurio perdavimo funkcija apskaičiuojama pagal 6 formulę.

$$G_n(\omega) = |H_n(j\omega)| = \frac{1}{\sqrt{1 + \epsilon^2 T_n^2\left(\frac{\omega}{\omega_0}\right)}} \quad (6)$$

Formulėje  $\epsilon$  – pulsavimo faktorius (*ang. ripple factor*),  $\omega_c$ - iškirpimo dažnis (*ang. cutoff frequency*),  $T_n()$  – n-tosios eilės Čebiševo polinomas.

Žemo praleidimo 6 eilės Čebiševo I ir II tipo filtrai kartu su Batervorto filtru ir Eliptiniu filtru pateikiami 33 paveiksle.



Šaltinis: DANIELS, Richard. (1974). *Approximation Methods for Electronic Filter Design*. New York: McGraw-Hill. ISBN 200-07-015308-6.

33 pav. Skaitmeninių filtrų palyginimas

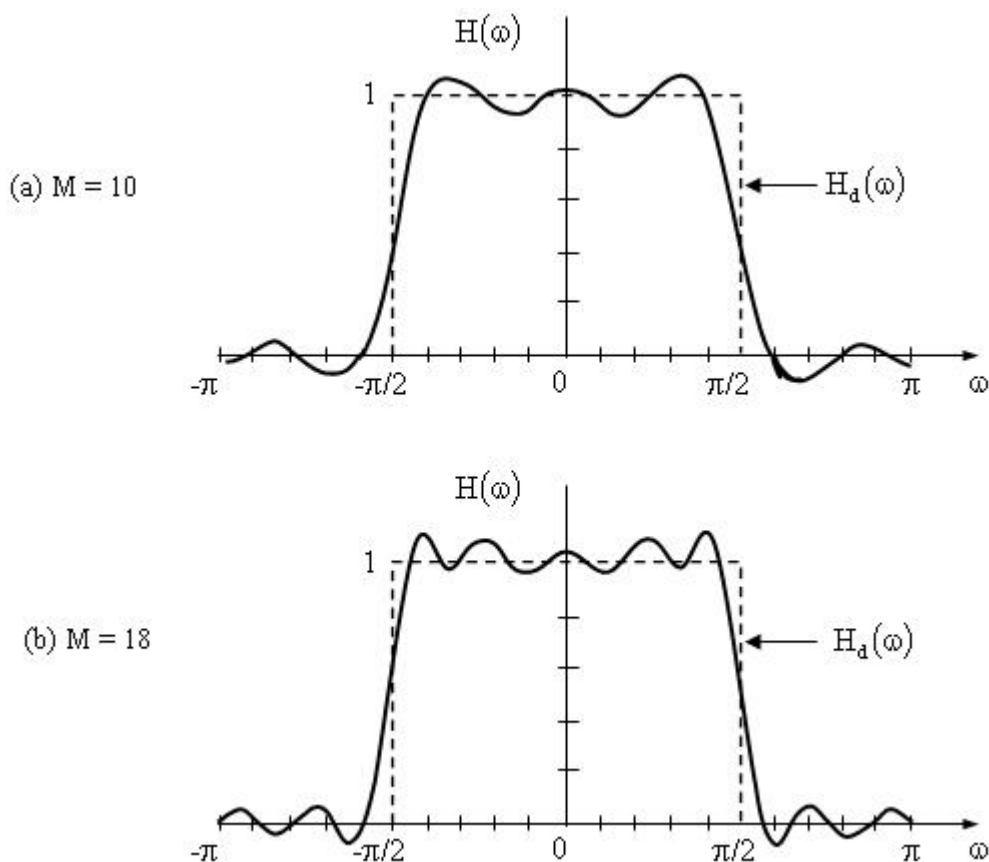
Paveikslėlyje (33 pav.) matome, kad Čebiševo filtrai yra smailesni už Batervorto filtrą, tačiau jie yra ne tokie smailūs kaip Eliptinis filtras. Taip pat akivaizdu, kad Eliptinis filtras labiausiai banguoja ir praleidžiamuose, ir nepraleidžiamuose dažniuose, ko negalima pasakyti apie Batervorto filtrą.

### 2.1.3.3. FIR filtras

Baigtinio impulso atsako, dar kitaip vadinamas nerekursiniu, filtras (ang. *Finite Impulse Response*) yra baigtinis, nes  $N$ -tosios eilės filtro impulso atsakas tęsiasi iki  $N+1$  pavyzdžio ir tuomet tampa lygus 0 (PHOUNG, Nguyen Huu. 2008).

FIR filtrai pasižymi stabilumu, yra lengvai realizuojami, nes turi tiesinę fazę, kuri keičiasi proporcingai dažniui. Didžiausias trūkumas, kuriuo pasižymi FIR filtrai, yra sudėtingas skaičiavimas, kuris atitinkamai reikalauja daugiau resursų nei dauguma kitų skaitmeninių filtrų rūšių (PHOUNG, Nguyen Huu. 2008).

Paveikslėlyje žemiau (34 pav.) pateiktos dvi FIR filtro realizacijos: kai filtro eilė lygi dešimt ir kai filtro eilė lygi 18.



Šaltinis: PHOUNG, Nguyen Huu (2008). *FIR Filter Design: The window Design Method*. prieiga per internetą: <http://vocw.edu.vn/content/m11285/latest/> .

34 pav. FIR filtras

Kaip matome paveiksle (34 pav.), 18 eilės filtras yra tik šiek tiek tikslesnis, tačiau stipriai imlesnis skaičiavimams. Didinant filtro eilę, imlumas skaičiavimams didėja, o tikslumo pasiekama vis mažiau.

#### 2.1.4. Filtro eilės parinkimas

Filtro eilės parinkimas – sudėtingas uždavinys. Iš pirmo žvilgsnio dažnai klystama, manant, kad parinkti skaitmeninio filtro eilę yra nesudėtinga, vadovaujantis tuo, kad kuo filtro eilė aukštesnė, tuo filtro perdavimo charakteristika pereinamoje srityje statesnė ir atitinkamai gaunami geresni filtravimo rezultatai. Tačiau daugelyje skaičiavimų yra svarbus našumo faktorius. O kuo didesnė filtro eilė, tuo didesnis filtro koeficientų skaičius tad ilgiau vykdomas filtravimas bei reikalaujama kur kas daugiau resursų skaičiavimams atlikti.

Siekiant sukurti greitą balsių identifikavimo požymį svarbu, kad filtras būtų apskaičiuojamas greitai, bet tuo pačiu pateiktų ir patenkinamus rezultatus.

Filtro eilės pasirinkimas taip pat priklauso nuo naudojamo filtro. Pavyzdžiui FIR filtrai turi būti aukštesnės eilės nei Batervorto filtrai, kurių skaičiavimas stipriai lėtėja didinant filtro eilę.

Filtro eilei parinkti atlikta keletas smulkių tyrimų, kurių metu skaičiuotas filtro skaičiavimo laikas ir klaida. Remiantis gautais rezultatais pasirinktos pradinės filtrų eilės:

- Čebiševo 4 eilės filtras;
- Batervorto 4 eilės filtras.
- FIR 10 eilės filtras.

Tyrimo metu, siekiant optimalių rezultatų, numatoma taikyti įvairios eilės filtrus, tačiau ne mažesnės eilės nei pasirinktos pradinės filtrų eilės.

#### 2.1.5. Signalų energija

Signalų energijai skaičiuoti naudojamas pjūvio energijos skaičiavimo metodas. Jis grindžiamas diskretų kvadratų sumos nustatymu ir apskaičiuojamas pagal 7 formulę.

$$E_N = \sum_{n=-N}^N |x(n)|^2 \quad (7)$$

Energija skaičiuojama filtruotam signalui, kuris prieš tai parengiamas skaičiavimams naudojant centro iškirpimo metodą aprašytą 2.1.1 skyrelyje.

Energija skaičiuojama pagrindinio tono ir pirmų aštuonių harmonikų dažniu remiantis nustatytame intervale filtruotam signalui. Ieškomas požymis gaunamas sudėjus gautas energijų reikšmes.



$$E = E_{F_0} + E_{F_{h1}} + E_{F_{h2}} + \dots + E_{F_{h8}} \quad (8)$$

Požymis skaičiuojamas visiems pasirinkto kalbos signalo balsiams, analizuotai imčiai sudaromi statistiniai matavimai, pagal kuriuos po eksperimento bus daromos išvados apie šio požymio gebėjimą identifikuoti balsius.

## 2.2. „LTDIGITS“ garsynas

Kaip pagrindinis šaltinis kalbos signalų įrašams, kurie bus taikomi eksperimento metu pasitrintas „LTDIGITS“ garsynas. Šis garsynas yra vienas pirmųjų Lietuvoje. Jis pagrįstas hibridiniu duomenų modeliu ir sukurtas remiantis „TIMIT“ garsynų etalonu. Pagrindinė šio garsyno paskirtis – kalbos signalų atpažinimo pritaikymas telekomunikacijose ir internete.

Šiame garsyne diktoriai išskirti į dvi pagrindines grupes: vyrai ir moterys. Tai atsispindi ir hierarchinėje duomenų struktūroje, kuri artima „TIMIT“ struktūrai, tačiau susideda iš mažesnio skaičiaus elementų. Kiekvienas diktorius įrašo po dešimt įrašų, kurių kiekvieną sudaro po šešis žodžius. Pirmi penki įrašai susideda iš skaitmenų nuo nulio iki devynių. Šie skaitmenys diktoriui parenkami atsitiktine tvarka. Sekantys trys įrašai susideda iš lietuviškų valdymo komandų. Devintas įrašas – šeši skiemenys, kur du nosiniai priebalsiai (m, n) eina prieš tris kontrastinguosius balsius (a, u, i). Paskutiniame įrašė šie skiemenys sujungti į rišlią frazę. Diktorių įrašomos žodžių sekos pateikiamos 3 lentelėje

3 lentelė

**Diktorių žodžių sekos**

Sekos Nr.	Sekos turinys					
1-5	trys	penki	keturi	vienas	devyni	nulis
6	pradėti	baigti	sustoti	pauzė	laukti	tęsti
7	pirmyn	atgal	į pradžią	į pabaigą	sekantis	perduoti
8	taip	ne	pagalba	saugoti	start	stop
9	ma	na	mu	nu	mi	ni
10	Mikas	mato	nišoje	mūsų	namo	numerį

Šaltinis: sudaryta autoriaus.

Įrašai įrašomi naudojant 16kHz diskretizacijos dažnį ir vieną garso kanalą bei saugomi .vdi formato failuose. Kiekviena diskreta yra dviejų baitų sveikas skaičius intervale nuo -32768 iki 32767.

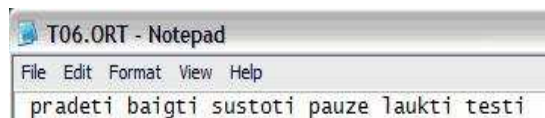
Garsyno struktūra susideda iš tokių elementų:

1. duomenų bazės vardas, pvz.: /LTDIGITS;
2. diktoriaus lyties identifikatorius, pvz.: /LTDIGITS/MALE;
3. diktoriaus identifikatorius, pvz.: /LTDIGITS/F001;
4. failai (.vdi, .ort, .rbb, .phn), pvz.: /LTDIGITS/F001/T01.phn.

Kaip matome, informacija saugoma keturių tipų failuose:

- .vdi – garso failas su antrašte;
- .ort – žodžių reikšmės (turinio fragmentas pateiktas 35 pav.);
- .rbb – žodžių ribos (turinio fragmentas pateiktas 36 pav.);
- .phn – anotacijos fonemų lygyje (turinio fragmentas pateiktas 37 pav.).

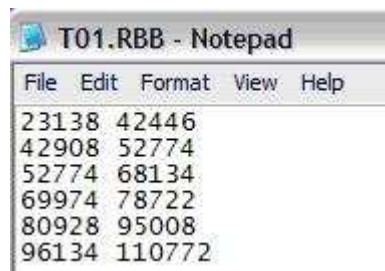
Žodžių ir fonemų ribos saugomos papildomuose tekstiniuose failuose (.rbb, .phn). Tai vieni svarbiausių duomenų kalbos signalų garsyne, tad jie privalo būti korektiški, būtent toks yra vienas iš šiuolaikinio kalbos signalų garsyno reikalavimų. Siekiant užtikrinti šių failų korektiškumą, reikalinga akustinio segmentavimo taikomoji programinė įranga.



Šaltinis: sukurta autoriaus.

### 35 pav. „LTDigits“ žodžių reikšmių failo .ort turinio fragmentas

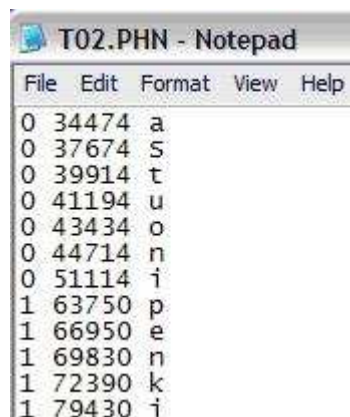
Žodžių reikšmės .ort faile saugomos vienoje eilutėje, atskirtos viena nuo kitos tarpais.



Šaltinis: sukurta autoriaus.

### 36 pav. „LTDigits“ žodžių ribų failo .ort turinio fragmentas

Žodžių ribų faile .rbb žodžio pradžios ir pabaigos baitai rašomi vienoje eilutėje, atskirti tarpu. Šiame faile pateikiamos žodžių ribos baitais nurodo žodžio vietą garso įrašo faile .vdi.



Šaltinis: sukurta autoriaus.

### 37 pav. „LTDigits“ fonemų ribų turinio fragmentas

Fonemų anotacijų faile, skirtingai nei žodžių ribų faile, matome ne tik ribas baitais, bet ir žodžio numerį bei fonemos reikšmę. Vieną duomenų eilutę sudaro: žodžio, kuriam priklauso garsas, eilės numeris, kuris prasideda nuo nulio, garso pabaigos baitai garso įrašo faile .vdi ir fonemos reikšmė. Duomenys vienoje eilutėje atskirti tarpais.

Tokia duomenų struktūra nors ir aiški, tačiau taikomuojų požiūriu nėra ideali. Vis tik garsyno struktūra yra tinkama metodo eksperimento realizacijai.

### **2.3. Tyrimo rezultatų apdorojimas**

Remiantis sudarytu algoritmo modeliu, bus atlikta keletas eksperimentų ieškomo požymio tikslumui patikrinti. Tyrimu metu numatoma nustatyti ribą, kuri turėtų skirti balsius nuo priebalsių ir apskaičiuoti nukrypimų nuo nustatytos ribos skaičių. Nustatytas nukrypimų skaičius leis procentaliai įvertinti metodo tikslumą tiek atliekant tyrimą su vyriškais, tiek su moteriškais garsais.

Siekiant išvengti požymio skaičiavimo nukrypimų, numatomi naudoti 20-ties ir 80-ties percentiliai. Tai reiškia, jog mažiausios ir didžiausios reikšmės bus atmetamos iš tiriamos imties. Tokiu būdu yra atmetamos reikšmės, kurios iškraipo imtį dėl vieno ar kito skaičiavimo netikslumo, pvz. neteisingai apskaičiuoto pagrindinio tono ar neteisingų vieno ar kito garso ribų.

Taip pat numatoma apskaičiuoti kitus statistinius rodiklius, kaip mediana, dispersija, vidurkis ir standartinis nuokrypis.

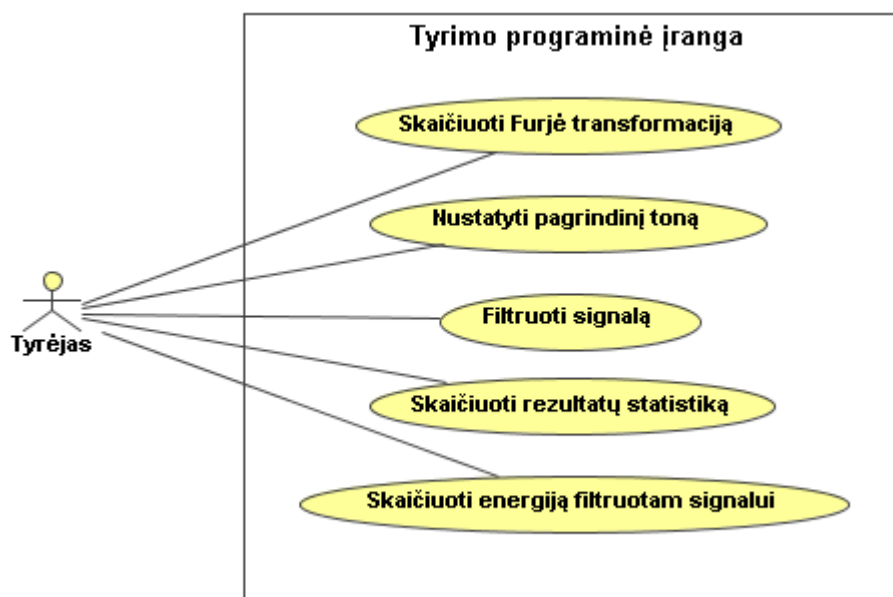
Remiantis gautais rezultatais numatoma pateikti išvadas apie tiriamo algoritmo kokybę ir galimą praktinį pritaikymą.

### 3. BALSŲ SKIRIAMOJO POŽYMIŲ METODO ĮVERTINIMO TYRIMAS

Šioje darbo dalyje aprašoma eksperimentinio tyrimo aplinka, programinės ir techninės priemonės reikalingos eksperimentui realizuoti. Aprašoma pagrindinio eksperimento eiga, pateikiamos išvados. Remiantis pagrindiniu tyrimu atliekami papildomi tyrimai, jų rezultatas lyginamas su pagrindinio tyrimo rezultatu. Taip pat įvertinamas algoritmo optimalumas, ne tik tikslumo, bet ir greitumo atžvilgiu. Šiam įvertinimui pagrįsti atliekamas dar vienas papildomas eksperimentas, kurio rezultatas – optimaliausia sukurto algoritmo konfigūracija.

#### 3.1. Eksperimentinio tyrimo aplinka

Tyrimui atlikti naudojama specialiai tyrimui sukurta programinė įranga. Ši programinė įranga pasižymi visu tyrimui atlikti reikalingu funkcionalumu. Kokias funkcijas tyrėjas gali atlikti su programine įranga, galima matyti žemiau pateiktoje panaudos atvejų diagramoje (38 pav.).



Šaltinis: sukurta autoriaus.

#### 38 pav. Programinės įrangos panaudos atvejų modelis

Tyrėjas naudodamasis sistema, gali sudarinėti garsų sąrašus. Tai parengiamasis tyrimo etapas. Pasirinktiems garsams vartotojas gali nesunkiai apskaičiuoti spektrą, sudaryti spektrogramą. Pagrindinio tono skaičiavimui vartotojas gali pasirinkti Kepstro arba Autokoreliacijos metodus, nesunkiai gali integruoti savo norimą metodą.

Filtravimas taip pat galimas dviem metodais – tai Čebiševo ir Batervorto metodai. Filtru algoritmą tyrėjas taip pat gali nesunkiai pakeisti savo norimu metodu, kurio sistema standartiškai nerealizuoja.

Filtruotam signalui sistema skaičiuoja energiją. Tai standartinis signalo energijos algoritmas. Suskaičiuotas energijas sistema sumuoja ir pateikia vartotojui apskaičiuotą kiekvieno garso pagrindinio tono ir harmonikų energijų sumą. Šis skaičius tai ir yra ieškomas parametras.

Gautiems rezultatams tirti sistemoje realizuoti statistiniai metodai: vidurkis, dispersija, standartinis nuokrypis, taip pat skaičiuojami kiti parametrai: mediana, minimumas, maksimumas, nurodyto procentinio dydžio percentiliai.

Sistema skaičiavimo rezultatus pateikia dvejopai:

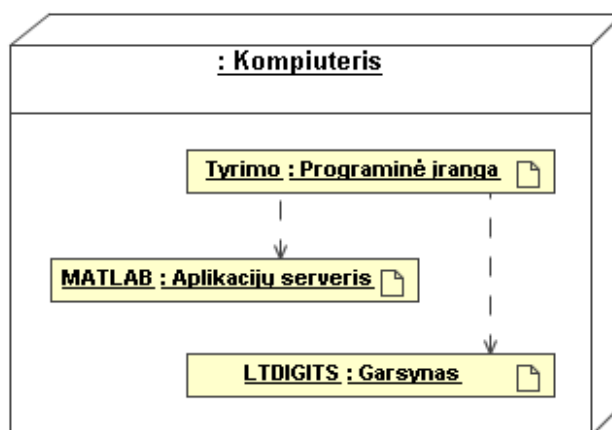
- Detalūs rezultatai – pateikiamas apskaičiuotas rodiklis kiekvienam garsui;
- Suminiai rezultatai – pateikiami apibendrinti rezultatai: pagrindinio tono, harmonikų ir energijos - kiekvienam skirtingam garsui.

Sistemos pateikiami rezultatai toliau analizuojami, remiantis jais daromos tyrimo išvados, atliekami papildomi tyrimai, keičiant tam tikrus laisvai koreguojamus sistemoje parametrus.

### 3.1.1. Programinės įrangos architektūra

Tyrimui naudojama programinė įranga sukurta taip, kad kuo daugiau parametrų ir taikomų algoritmų tyrėjas galėtų naudoti. Pagrindinio tono ir filtravimo metodai gali būti integruojami į sistemą pagal vartotojo poreikius, taip pat dauguma skaitinių reikšmių gali būti laisvai koreguojamos, siekiant išgauti kuo tikslesnius eksperimentinio tyrimo rezultatus.

Sistemos architektūrinis sprendimas pagrįstas C#.NET 3.0 programavimo platforma, tačiau matematiniai skaičiavimai atliekami su „MATLAB“ serveriu. Tai leidžia ne tik naudoti daugumą „MATLAB“ standartinių funkcijų, bet ir nesunkiai leisti tyrėjui koreguoti naudojamą „MATLAB“ kodą. Sistemos architektūra pateikiama 39 paveiksle.



Šaltinis: sukurta autoriaus.

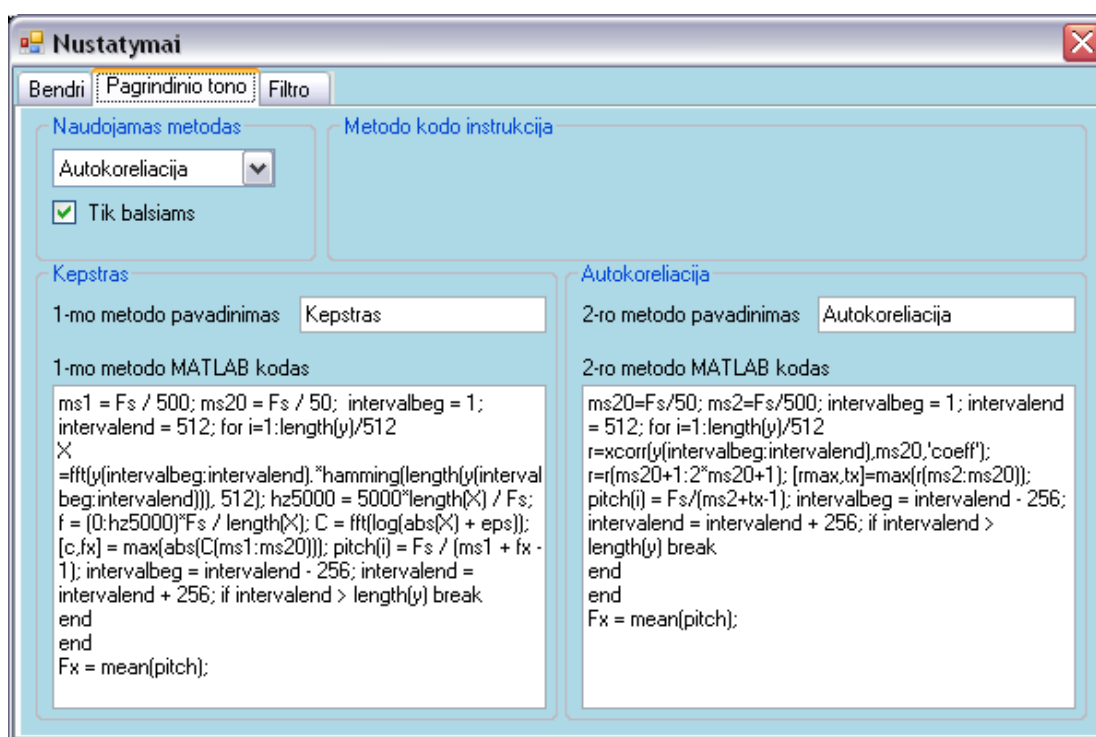
### 39 pav. Tyrimo programinės įrangos architektūra

Kaip matome paveiksle (39 pav.), tyrimo aplinka lokali, t.y. pilnai integruota viename vartotojo kompiuteryje. Programinė įranga skirta dirbti tik su „LTDIGITS“ garsynu.

### 3.1.2. Tyrimo aplinkos pritaikymas vartotojo poreikiams

Tyrimo aplinka tyrėjo atžvilgiu nėra statinė. Ji gali būti dinamiškai modifikuojama keičiant naudojamus algoritmus ir daugybę skaičiavimo ir algoritmų parametrų. Tai suteikia galimybę atlikti įvairius tyrimus, stebėti tyrimų rezultatus keičiant algoritmus arba jų parametrus.

Pats svarbiausias tyrimo aplinkos bruožas yra galimybė įterpti savo norimą „MATLAB“ kodą, taip įsiterpiant į sistemos funkcionalumą ir keičiant jos atliekamus veiksmus. Toks įsikišimas galimas į pagrindinio tono ir filtro algoritmus.



Šaltinis: sukurta autoriaus.

#### 40 pav. „MATLAB“ kodo intarpai į tyrimo aplinką

Visos „MATLAB“ komandos siunčiamos į serverį yra dinamiškai pateikiamos tyrėjui. Tai dar vienas būdas, iškirpus tam tikrą fragmentą, jį modifikuoti ir paleisti jau „MATLAB“ aplinkoje.

Tyrėjas taip pat turi galimybę keisti įvairius tyrimo aplinkos parametrus. Bendras parametrų sąrašas pateikiamas 4 lentelėje.

4 lentelė

#### Tyrimo aplinkos parametrai

Parametras	Aprašymas
Pagrindinio tono normalizavimas	Pagrindinio tono normalizavimas tai garsų nepatenkančių į nustatytą dažnio intervalą atmetimas iš analizės. Atskiri intervalai gali būti nurodyti vyriškiems ir moteriškiems garsams. Pagrindinio tono normalizavimas gali būti ir netaikomas. Taikant, rekomenduojama vyriško intervalo reikšmės parinkti tarp 100 ir 160Hz, o moteriško tarp 170 ir 220Hz

Parametras	Aprašymas
Percentiliai	Tai statistinės analizės parametras, kuriame galima nurodyti dvi procentines reikšmes, dviejų percentilių apskaičiavimui.
Analizė tik balsiams	Pasirinkus šį parametą iš analizės išmetami visi kiti garsai, o paliekami tik balsiai.
Pagrindinio tono metodo parinkimas	Vienu metu sistemoje galima saugoti dviejų pagrindinio tono metodų realizacijas. Tačiau naudojama skaičiavimams vienu metu gali būti tik viena realizacija. Šis parametras susijęs su jos parinkimu.
Filtro metodo parinkimas	Vienu metu sistemoje galima saugoti du filtravimo algoritmus, tačiau skaičiavimams vienu metu gali būti taikomas tik vienas metodas. Šis parametras apsprendžia pasirinktą metodą.
Filtro eilė	Filtro eilė-tai filtro algoritmo parametras parenkamas taikomam filtro algoritmui.

Šaltinis: sudaryta autoriaus.

Galimų parametų visuma apsprendžia vartotojo poreikius konkretaus eksperimento atveju.

### 3.2. Pagrindinis eksperimentas

Pagrindinis eksperimentas yra skirtas sudaryto algoritmo įvertinimui. Eksperimento tikslas atsakyti į pagrindinę darbo hipotezę: ar tiriamas parametras gali būti taikomas balsiams nustatyti. Siekiant patikrinti šią hipotezę bus atliekami tyrimai su vyriškais ir moteriškais garso įrašais atskirai.

Eksperimento metu numatoma ištirti šimtą moteriškų frazių ir šimtą vyriškų frazių, po šešis žodžius kiekvienoje frazėje.

Tyrimo metu numatoma naudoti kepstro pagrindinio tono nustatymo metodą. Filtravimui numatomas naudoti FIR filtras. Keičiant pagrindinio tono nustatymo ir filtravimo algoritmus numatoma atlikti papildomus tyrimus.

Tyrimo konfigūracija pateikiama 5 lentelėje.

5 lentelė

#### Pagrindinio eksperimento konfigūracija

Parametras	Aprašymas	
Pagrindinio tono normalizavimas	Netaikomas	
Pagrindinio tono metodas	Algoritmas	Kepstras
	Tiriamas dažnio intervalas	50Hz-500Hz
Filtrai	Algoritmas	FIR filtras
	Filtro eilė	10
Percentiliai	20 ir 80	

Šaltinis: sudaryta autoriaus.

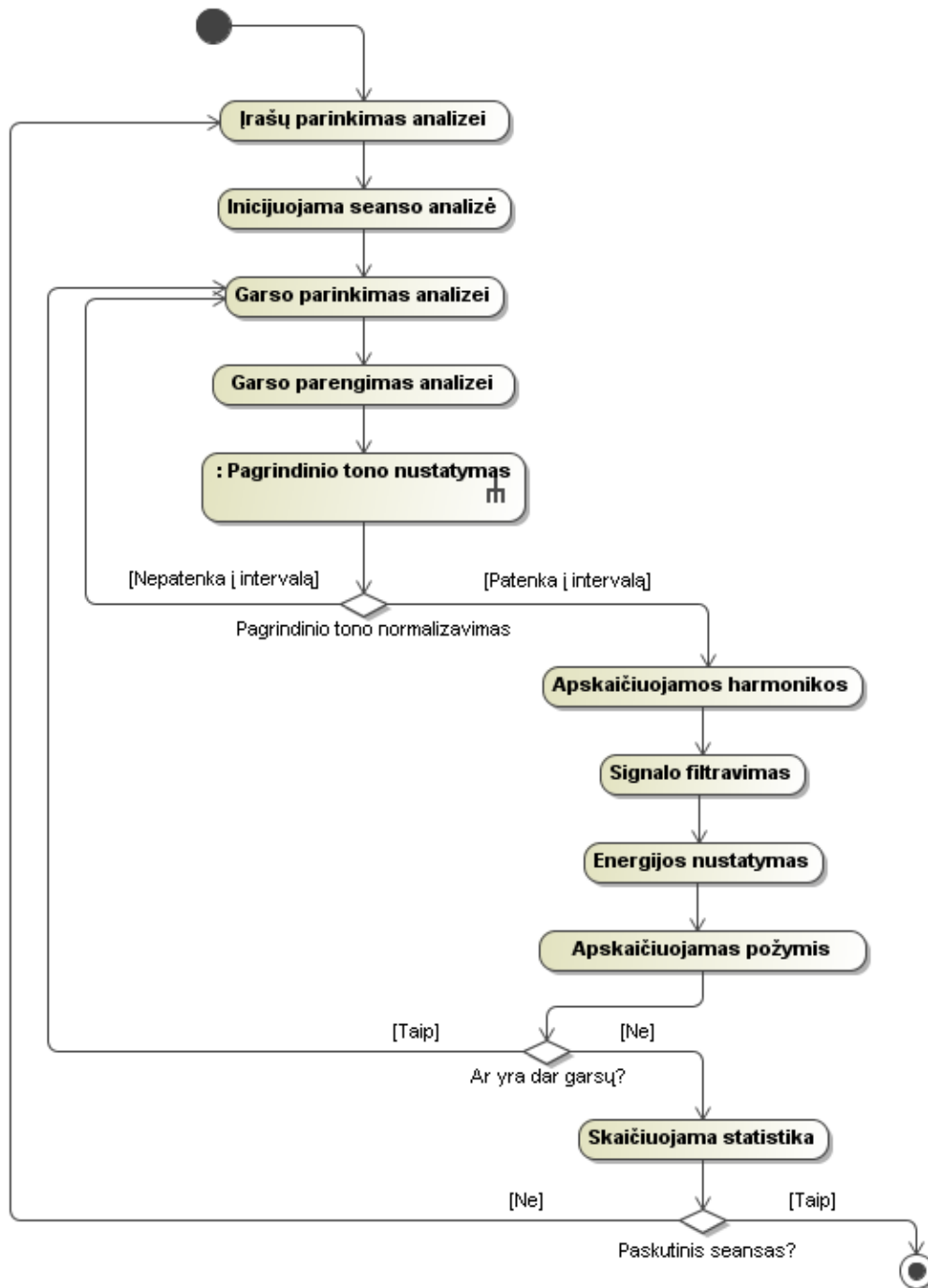
Pagrindiniam tyrimui pasirinkta konfigūracija pagrindžiama antroje darbo dalyje.

### 3.2.1. Eksperimento eiga

Pirmiausiai eksperimentui atrenkami garso įrašai. Atrinkimas atliekamas atsitiktiniu būdu, tačiau pagal tam tikras taisykles. Pagrindinė taisyklė – kad tyrimo metu nebūtų daugiau kaip dviejų to paties diktoriaus garso įrašų.

Tyrimas vykdomas dviem seansais. Pirmasis seansas – moteriškų garso įrašų tyrimas, antrasis – vyriškų.

Pasirinkus garso įrašus nustatoma pasirinkta tyrimo konfigūracija.



Šaltinis: sukurta autoriaus.

41 pav. Pagrindinio tyrimo eiga



Tyrimas atliekamas pagal numatytą metodą, t.y. tokias etapus:

- Kiekvienam balsiui apskaičiuojamas pagrindinis tonas;
- Kiekvienam balsiui apskaičiuojamos aštuonios harmonikos;
- Filtruojami dažnio intervalai aplink pagrindinį toną ir harmonikas;
- Skaičiuojama filtruotų intervalų energija;
- Apskaičiuojama energijų suma vienam garsui;
- Sumuojami rezultatai (pateikiama statistika);
- Tyrimas kartojamas kitos lyties garso įrašams.

Vizualizuotai tyrimo eiga pateikiama 41 paveiksle.

### 3.2.2. Tyrimo rezultatai

Atlikus tyrimą gauti rezultatai moteriškiems garso įrašams pateikiami 6 lentelėje.

6 lentelė

**Pagrindinio eksperimento rezultatai moteriškiems garsams**

Garsas	Garsų skaičius	Pagrindinis tonas	Požymio vidurkis	Dispersija	Standart. nuokrypis	Percentilis (20)	Percentilis (80)	Klaidos
t	303	224,87	0,004056	0,00019	0,013775	0,000418	0,003854	0
r	125	263,9	0,045436	0,006637	0,081468	0,004989	0,055991	56
i	544	223,91	0,140961	0,049651	0,222826	0,022713	0,20866	0
s	234	241,32	0,00291	0,000029	0,005366	0,000419	0,003683	0
d	109	234,72	0,040462	0,002668	0,051651	0,006484	0,066008	69
e	252	202,66	0,249144	0,244883	0,494857	0,031658	0,314099	0
v	69	217,17	0,037316	0,003908	0,062516	0,008918	0,046623	39
n	229	204,99	0,141634	0,048528	0,22029	0,027289	0,200821	229
š	97	275,2	0,036692	0,011782	0,108546	0,001907	0,028252	21
a	251	200,54	0,321892	0,569111	0,754395	0,034154	0,32734	0
p	164	267,15	0,010067	0,000757	0,027517	0,000646	0,010945	0
k	77	241,51	0,003717	0,000024	0,004881	0,000788	0,004742	0
u	186	199,87	0,216746	0,133112	0,364844	0,021776	0,297349	0
o	74	194,68	0,167674	0,146585	0,382864	0,016988	0,201843	0
l	71	214,5	0,0788	0,012809	0,113176	0,013136	0,11658	62
ė	21	204,8	0,349883	0,61079	0,781531	0,055933	0,356937	0
b	35	198,23	0,042518	0,004426	0,066526	0,006547	0,055186	21
g	57	229,53	0,024135	0,001819	0,04265	0,00387	0,033731	9
z	11	199,36	0,017481	0,000512	0,02263	0,001933	0,033374	2
m	11	203,29	0,336765	0,17375	0,416834	0,115348	0,521363	11
ž	10	290,51	0,009595	0,000133	0,011551	0,001534	0,018066	0

Šaltinis: sudaryta autoriaus.

Tyrimo metu viso ištirta 2930 garsų. Iš jų 1328 balsiai ir 1602 priebalsiai. Klasifikuojant toliau, galima pastebėti, kad iš priebalsių 745 sprogstamieji, 352 frikatyviniai ir 505 pusbalsiai.

Atlikus skaičiavimus, gauti rezultatai, kuriais remiantis galime daryti pirmąsias išvadas. Pirmiausiai nesunku pastebėti, kad atsitiktinai parinktuose garso įrašuose, ištirti net 544 i balsiai. Daugiausiai ištirtų priebalsių – t. Mažiausiai ištirta ž, m bei z priebalsių. Šie skaičiai turi tam tikros įtakos požymio tikslumo įvertinimui. Pavyzdžiui m – pusbalsis, o pusbalsių atskyrimas nuo balsių labiausiai komplikuoatas. Jeigu tyrime būtų 500 ar daugiau m pusbalsių, tikėtina, kad tyrimo rezultatai būtų prastesni.

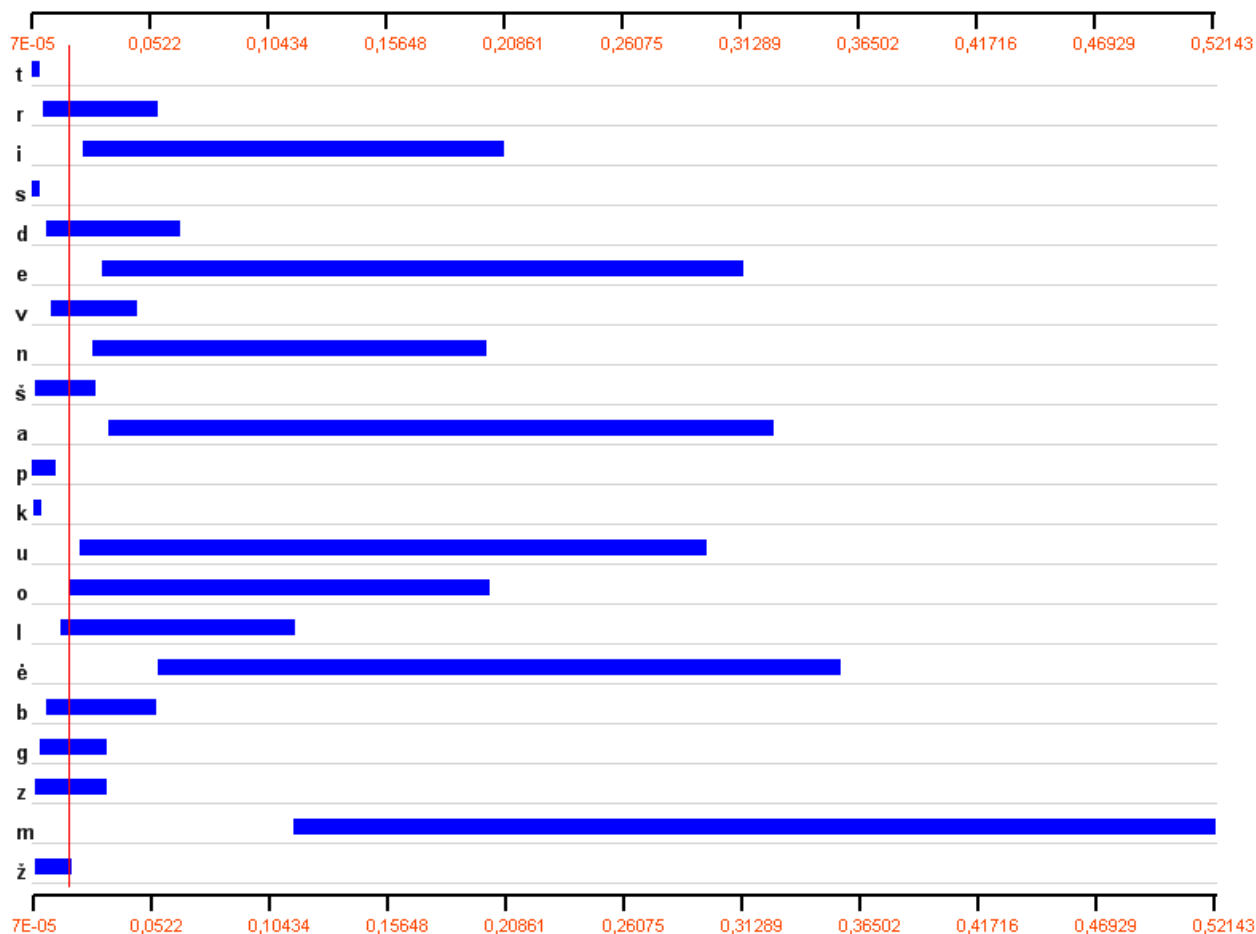
Iš 6 lentelėje pateiktų rezultatų matome, kad didžiausias ieškomo požymio standartinis nuokrypis yra balsių è – 0,78 ir a – 0,75. Tai parodo, kad šie balsiai gali būti tariami su stipresne ir silpnesne energija. Nesunku pastebėti, kad balsių ir pusbalsių ieškomo požymio standartiniai nuokrypiai yra didžiausi, stipriai didesni nei skardiesiems ir dusliesiems priebalsiams. Galime daryti išvadą, jog moterys šiuos priebalsius taria su panašia energija nepriklausomai nuo aplinkybių.

Vienos tyrimų sesijos kontekste nesunku išvelgti ir energijų skirtumus tarp skirtingų garsų, tačiau šis rodiklis nieko neatskleidžia viso tyrimo kontekste. Svarbiausi parametrai, kuriais remiantis daromos pagrindinės išvados yra percentiliai. Percentiliai taikomi atmesti labiausiai išsiskiriančias energijos reikšmes, kurios dėl vienu ar kitų priežasčių gali būti traktuojamos kaip nukrypimas. Apskaičiuoti 20 ir 80 percentiliai sudaro intervalą požymio reikšmių, kurios tikėtina turėtų išskirti balsius iš priebalsių. Siekiant tai patikrinti, kiekvienam balsiui suskaičiuojamas percentilių intervalo vidurkis ir gauti rezultatai sudedami į bendrą intervalų sistemą (42 pav.).

Kaip matome paveikslėlyje (42 pav.), situacija kebli su pusbalsiais, taip pat su kai kuriais sprogstamaisiais priebalsiais. Raudona spalva paveiksle išskirta riba, kur prasideda pirmojo balsio požymio intervalas. Nuo šios ribos balsiai negali būti atskirti nuo priebalsių. Visi priebalsių pasirodymai esantys aukščiau šios ribos traktuojami kaip algoritmo klaidos, kuriomis remiantis daromos išvados apie algoritmo tikslumą.

Paveikslėlyje nesunku pastebėti, kad priebalsiai t, s, p, k ir ž nesikerta su nustatyta riba, o tai reiškia, jog jie gali būti 100% tikslumu atskirti nuo balsių. Pusbalsiai m ir n priešingai – yra visiškai neatskiriami. Tiesa pusbalsį m, konkrečiu atveju, galima dalinai atskirti brėžiant ribą kitoje pusėje, nes kaip matome paveikslėlyje šio pusbalsio požymio reikšmių intervalas didele dalimi viršija artimiausią balsį è.

Požymio tikslumas vertinamas skaičiuojant priebalsių atpažinimo klaidas. Klaidingai atpažintais traktuojami visi priebalsiai, kurie yra už nubrėžtos ribos. Skaičiuojamas procentinis dydis, kuris įvertina kiek aptikta klaidų iš visų tirtų garsų. Šios sesijos atveju atpažinimo tikslumas 82,27%.



Šaltinis: sukurta autoriaus.

#### 42 pav. Tiriamo kriterijaus reikšmių pasiskirstymas skirtingiems balsiams

Gautas tikslumo įvertis yra pakankamai aukštas, tačiau tyrimo rezultatai ne itin geri tiriamų pusbalsių atžvilgiu. Nepaisant to, teigti, kad šis kriterijus negali išskirti pusbalsių dar per anksti. Reikia atlikti papildomus tyrimus keičiant tyrimo konfigūraciją, t.y. atlikti tyrimus su kitais pagrindinio tono ir filtravimo algoritmais. Tikėtina, kad tyrimo rezultatus galima pagerinti taikant didesnę skaičiuojamų garso harmonikų skaičių.

Pagrindinio tyrimo rezultatai vyriškiems garsams pateikiami 7 lentelėje.

7 lentelė

#### Pagrindinio eksperimento rezultatai vyriškiems garsams

Garsas	Garsų skaičius	Pagrindinis tonas	Požymio vidurkis	Dispersija	Standart. nuokrypis	Percentilis (20)	Percentilis (80)	Klaidos
a	262	132,79	1,023161	1,557233	1,247891	0,159804	1,597303	0
š	97	273,75	0,074461	0,016834	0,129745	0,009548	0,098151	0
t	299	224,45	0,013258	0,001036	0,032185	0,000355	0,017458	0
u	196	137,07	0,699071	0,879535	0,937836	0,101091	1,09875	0
o	79	130,28	0,629449	0,671239	0,819292	0,118808	0,930982	0
n	236	135,5	0,315784	0,062366	0,249732	0,089503	0,490373	228
i	538	143,95	0,601371	0,390345	0,624776	0,127003	0,998637	0
d	107	201,38	0,064545	0,004319	0,065722	0,017343	0,093503	0

Garsas	Garsų skaičius	Pagrindinis tonas	Požymio vidurkis	Dispersija	Standart. nuokrypis	Percentilis (20)	Percentilis (80)	Klaidos
r	122	215,91	0,145855	0,025505	0,159704	0,031589	0,228298	61
s	224	262,67	0,025729	0,002236	0,047289	0,00049	0,041175	0
k	86	270,28	0,01608	0,000507	0,022528	0,001126	0,026013	0
e	256	127,43	0,633547	0,709934	0,842576	0,124658	0,911947	0
p	164	261,02	0,020809	0,001633	0,040412	0,000406	0,03029	0
v	73	148,95	0,116167	0,009647	0,098221	0,036088	0,184962	28
l	71	149,23	0,377188	0,11454	0,338437	0,095415	0,671955	68
ė	22	132,68	1,296221	1,784804	1,335966	0,312529	1,807777	0
b	35	186,87	0,09323	0,006681	0,081736	0,033883	0,137332	9
g	57	183,81	0,059238	0,00256	0,050601	0,016913	0,092712	0
z	11	208,74	0,104223	0,017989	0,134123	0,020276	0,148486	2
m	11	118,45	0,45522	0,088431	0,297373	0,15357	0,747141	11
ž	11	277,13	0,130253	0,160206	0,400258	0,017489	0,061734	0

Šaltinis: sudaryta autoriaus.

Vyriškų garo įrašų tyrimo metu ištirta 2957 garsai. Iš jų 1353 balsiai, 513 pusbalsiai, 748 sprogstamieji priebalsiai ir 343 frikatyviniai priebalsiai.

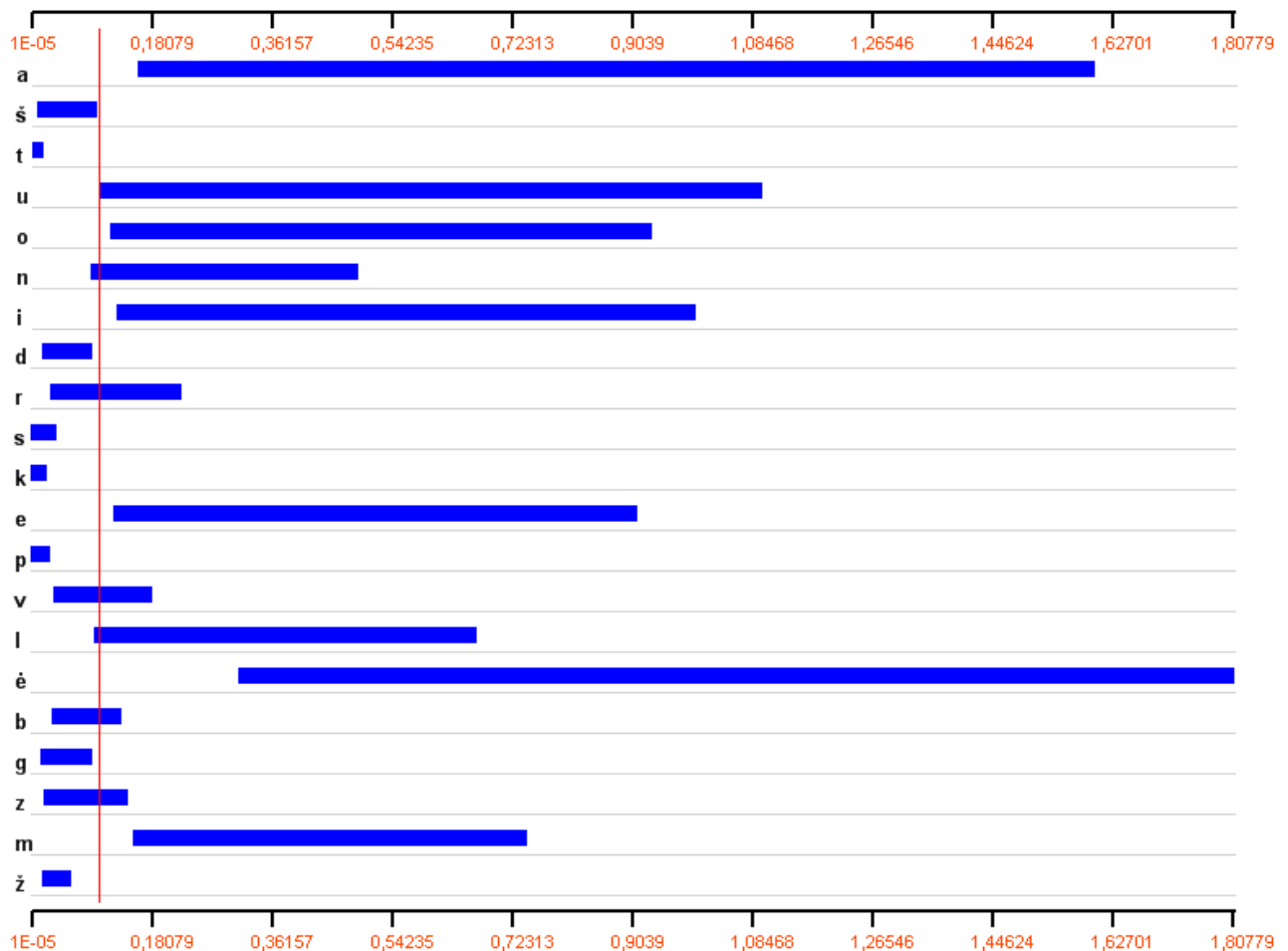
Kaip ir moteriškų garso įrašų tyrimo metu, vyrauja tie patys pagrindiniai garsai, t.y. i, t, e, a.

Didžiausias standartinis nuokrypis vyriškų garsų tyrimo atveju yra garsams ė, a, u. Skirtingai nuo ankstesnio tyrimo rezultatų, pusbalsiai vyriškų garso įrašų tyrime standartinio nuokrypio reikšme taip neišsiskiria iš kitų tipų priebalsių ir stipriau išsiskiria nuo balsių. Galime daryti išvadą, kad tarp skirtingų diktorių vyrų yra mažesni energijos svyravimai tariant pusbalsius ir didesni tariant balsius.

Sudarome bendrą intervalų vizualizacija kaip ir moteriškiems garsams (43 pav.). Paveiksle matome, kad pusbalsių išskyrimas ne mažiau komplikotas kaip ir pirmojo tyrimo rezultatuose. Pusbalsis m negali būti atskirtas visiškai, pusbalsiai n ir l gali būti atskirti su nedidesne kaip 5% tikimybe.

Sprogstamieji priebalsiai d, k, p, g taip pat frikatyviniai priebalsiai s, š, ž gali būti išskirti su 100% tikimybe. Likę garsai, t.y. r, v, b, z gali būti išskirti su ~50% tikimybe. Apskaičiuotas bendras požymio tikslumas – 86,24 procentai. Lyginant su algoritmo tikslumu moteriškų garsų atveju, galime teigti, jog vyrų ištartus balsius algoritmas atskiria tiksliau nei moterų.

Įdomu pastebėti, kad apskaičiuotas požymis vyrų ištartiems garsams yra daug didesnis nei moterų ištartiems. Palyginimui galima pateikti apskaičiuotas balsių atskyrimo ribas. Moteriškų garsų atveju ji - 0,0167877, o vyriškų 0,101091. Matome, jog skirtumas daugiau negu akivaizdus.



Šaltinis: sukurta autoriaus.

### 43 pav. Tiriama kriterijaus reikšmių pasiskirstymas skirtingiems balsiams

Siekiant padidinti požymio efektyvumą atskiriant priebalsius nuo pusbalsių, reikia atlikti papildomus tyrimus keičiant tyrimo konfigūraciją.

Keičiant tyrimo aplinkos konfigūraciją atliekami papildomi tyrimai.

### 3.3. Papildomi tyrimai

Papildomas tyrimas atliekamas keičiant taikoma pagrindinio tono nustatymo metodą iš Kepstro į Autokoreliacijos. Taip pat keičiamas skaitmeninio filtro metodas į ketvirtos eilės Batervorto filtrą. Taikant Auto Koreliaciją pagrindiniam tonui nustatyti, algoritmas bus skaičiuojamas be Furjė transformacijos. Tai turėtų pagreitinti ieškomo požymio apskaičiavimą, tačiau kaip jau įrodyta kito tyrimo metu, gali sumažėti pagrindinio tono, o kartų ir ieškomo požymio tikslumas.

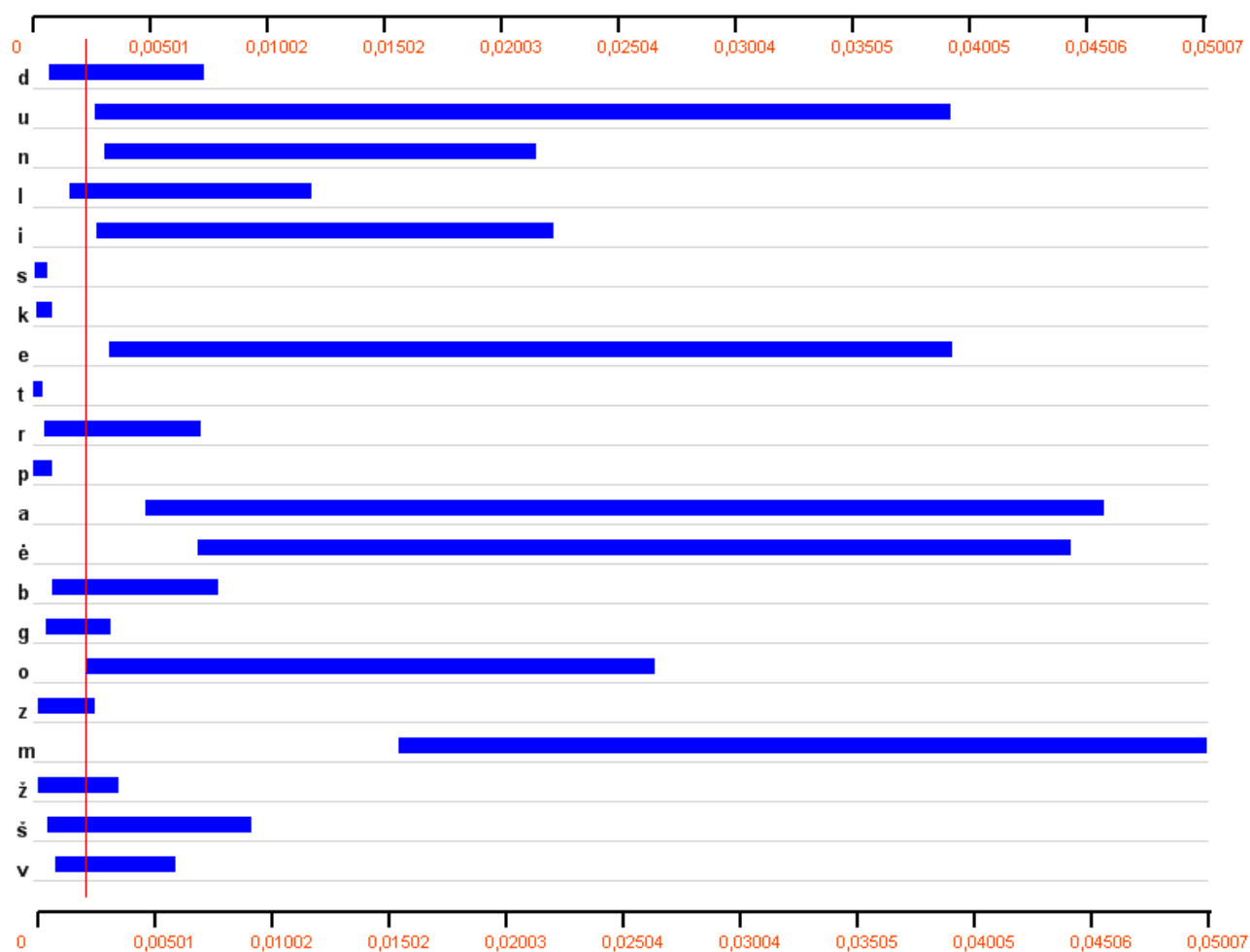
Papildomo tyrimo konfigūracija pateikiama 8 lentelėje.

## Papildomo eksperimento konfigūracija

Parametras	Aprašymas	
Įrašų skaičius	100 diktorių vyrų ir 100 diktorių moterų įrašų. Viso 200 įrašų.	
Pagrindinio tono normalizavimas	Netaikomas	
Pagrindinio tono metodas	Algoritmas	Auto Koreliacija
	Tiriamas dažnio intervalas	50Hz-500Hz
Filtrai	Algoritmas	Batervorto filtras
	Filtro eilė	4
Percentiliai	20-80	

Šaltinis: sudaryta autoriaus.

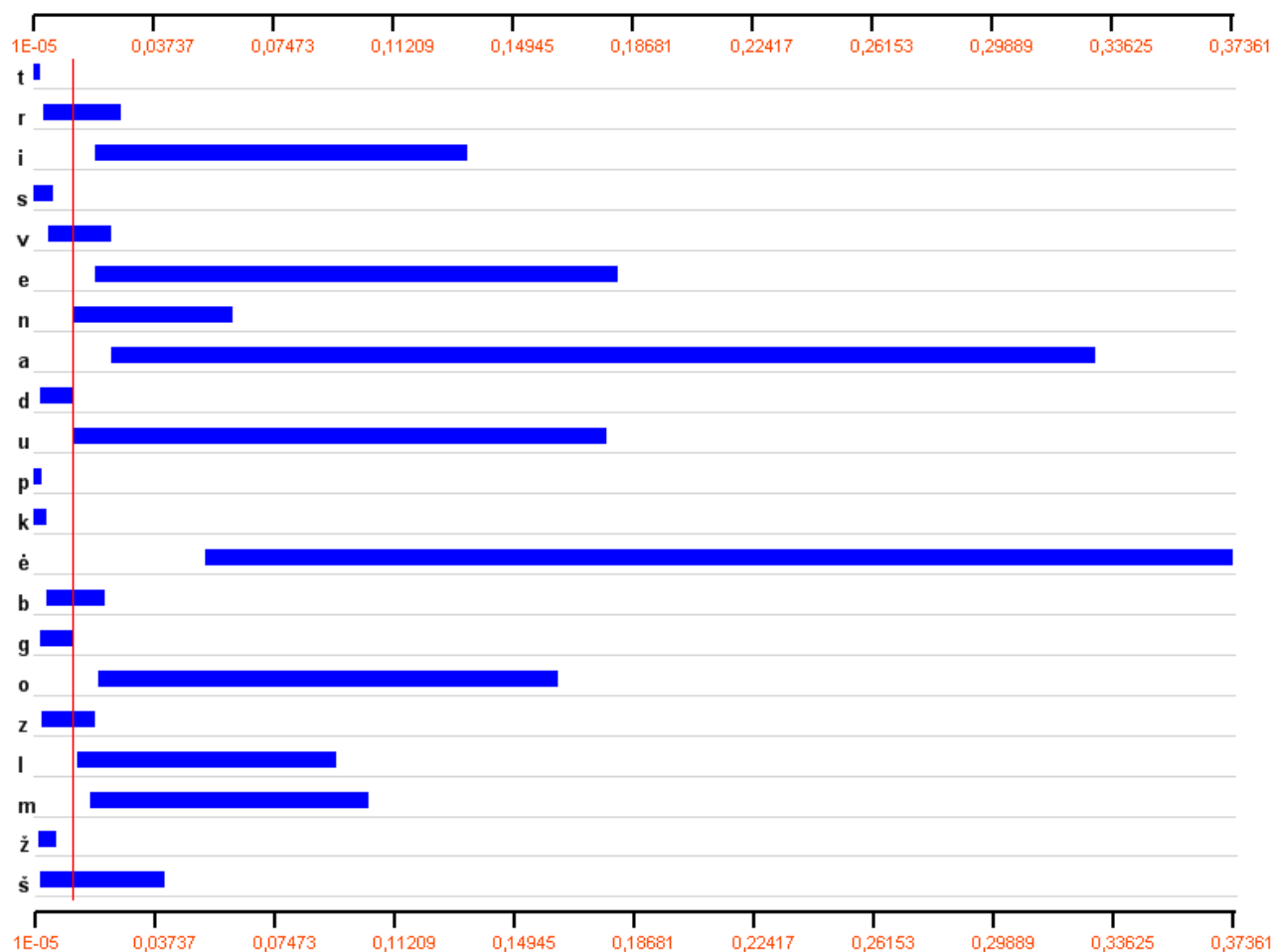
Tyrimo rezultatai pateikiami 44 ir 45 paveiksluose. 44 paveiksle pateikiami moteriškų garsų tyrimo rezultatai, 45 – vyriškų.



Šaltinis: sukurta autoriaus.

## 44 pav. Papildomo tyrimo rezultatai moteriškiems garsams

Kaip matome 44 paveiksle, šiuo atveju rezultatai artimi pagrindinio tyrimo rezultatams. Pusbalsiai m ir n yra visiškai neatskiriami. Likę pusbalsiai atskiriami su labai maža tikimybe. Prie balsiai s, k, t ir p atskiriami 100%. Nustatytas bendras tyrimo tikslumas moteriškiems garsams – 81,36%.



Šaltinis: sukurta autoriaus.

#### 45 pav. Papildomo tyrimo rezultatai vyriškiems garsams

Vyriškų garsų atveju, rezultatai šiek tiek geresni. 45 paveiksle matome, kad priebalsiai t, s, d, p, k, g, ž gali būti atskirti 100%. Kaip ir ankstesnių tyrimų atveju, pusbalsiai m, n ir l negali būti atskirti visiškai. Bendras atpažinimo tikslumas šiuo atveju – 83,66%.

### 3.4. Pagrindinio ir papildomų tyrimų apibendrinimas

Siekiant rasti tiksliausią tyrimo metu taikomų metodų derinį, atlikta dar keletas tyrimų. Šių tyrimų ir ankstesnių tyrimų duomenys pateikiami 9 ir 10 lentelėse. 9 lentelėje pateikiami tyrimo rezultatai moteriškiems garsams, o 10 – vyriškiems.

Tyrimai atliekami kaitaliojant skaitmeninio filtro metodą ir pagrindinio tono nustatymo metodus. Ištirtos visos galimos šių metodų poros.

**Eksperimento ir papildomų tyrimų rezultatų apibendrinimo matrica moteriškiems garsams**

	Kepstras	Auto Koreliacija
<b>FIR filtras</b>	82,27%	81,51%
<b>Čebiševo filtras</b>	82,99%	81,36%
<b>Batervorto filtras</b>	83,96%	82,58%

Šaltinis: sudaryta autoriaus.

Kaip matome 9 lentelėje, moteriškiems garso įrašams geriausias tikslumas pasiektas taikant Batervorto filtrą ir Kepstro pagrindinio tono nustatymo metodą – 83,96%. Taikant auto koreliacijos pagrindinio tono nustatymo metodą, tiksliausias rezultatas su tuo pačiu – Batervorto filtru. Mažiausias tikslumas gautas taikant auto koreliacijos metodą ir FIR filtrą.

Nesunku pastebėti, kad skaičiuojant pagrindinį toną auto koreliacijos metodu, požymio tikslumas yra apytiksliai 1% punktu mažesnis.

**Eksperimento ir papildomų tyrimų rezultatų apibendrinimo matrica vyriškiems garsams**

	Kepstras	Auto Koreliacija
<b>FIR filtras</b>	86,24%	84,81%
<b>Čebiševo filtras</b>	86,29%	83,66%
<b>Batervorto filtras</b>	86,57%	84,91%

Šaltinis: sudaryta autoriaus.

Vyriškiems garsams, kaip matome 10 lentelėje, geriausias tikslumas gaunamas taikant Batervorto filtrą ir kepstro pagrindinio tono nustatymo metodą – 86,57%. Taikant auto koreliacijos pagrindinio tono nustatymo metodą rezultatai prastesni apytiksliai 2% punktais.

Galime daryti išvadą, kad Batervorto filtro ir Kepstro pagrindinio tono radimo metodo derinys yra pats tiksliausias tiek vyriškų, tiek moteriškų garso įrašų atžvilgiu. Taip pat galime teigti, jog vyriškiems garso įrašams požymio tikslumas didesnis ~3% punktais.

Remiantis iki šiol atliktais tyrimais akivaizdu, kad realizuotas požymis negali išskirti balsių nuo pusbalsių, ypačingai pusbalsių m, n ir l. Įvertinus geriausią požymio tikslumo rezultata, iš tyrimo pašalinus pusbalsius, vyriškų garso įrašų atveju gauname 98,90% tikslumą. Moteriškų garsų atveju – 96,49%. Tad galima teigti, jog požymis geba puikiai atskirti balsius nuo sprogstamųjų ir friktyvinių priebalsių, tačiau yra praktiškai nepritaikomas atskirti balsius nuo pusbalsių.

**3.5. Algoritmo greičio tyrimas**

Viena iš esminių tiriamo algoritmo savybių yra algoritmo greitis. Tiesa jis stipriai priklauso nuo taikomos tyrimo konfigūracijos. Labiausiai algoritmo greitį įtakoja pagrindinio tono nustatymo metodas ir skaitmeninio filtro eilė.



Algoritmo greičio tyrimo metu siekiama nustatyti kaip skiriasi algoritmo greitis taikant Kepstro ir Auto koreliacijos pagrindinio tono radimo metodus. Skaitmeninio filtro metodas abiem tiriamais atvejais parinktas vienodas, kaip ir taikoma filtro eilė.

Šio tyrimo metu siekiama parodyti kaip skaičiavimus apsunkina dažninės charakteristikos analizės metodai, konkrečiu atveju - Furjė transformacija, kuri naudojama kepstro algoritmui apskaičiuoti. Šis tyrimas taip pat turėtų atskleisti kiek tiriamas algoritmas gali būti greitesnis už spektrine signalo analize pagrįstus algoritmus tai pačiai problemai spręsti.

Tyrimo rezultatai pateikiami moteriškiems (11 lentelė) ir vyriškiems (12 lentelė) garso įrašams su įvairiomis filtravimo ir pagrindinio tono nustatymo metodų kombinacijomis.

11 lentelė

#### Greičio tyrimo rezultatų apibendrinimo matrica moteriškiems garsams

	Kepstras	Auto Koreliacija
<b>FIR filtras</b>	26s	27s
<b>Čebiševo filtras</b>	38s	38s
<b>Batervorto filtras</b>	48s	44s

Šaltinis: sudaryta autoriaus.

12 lentelė

#### Greičio tyrimo rezultatų apibendrinimo matrica vyriškiems garsams

	Kepstras	Auto Koreliacija
<b>FIR filtras</b>	19s	35s
<b>Čebiševo filtras</b>	35s	54s
<b>Batervorto filtras</b>	28s	43s

Šaltinis: sudaryta autoriaus.

Remiantis gautais greičio tyrimo rezultatais, nesunku pastebėti, kad greičiausia požymio skaičiavimo konfigūracija yra FIR filtras ir kepstro pagrindinio tono radimo metodas. Ši konfigūracija yra greičiausia ir vyriškiems, ir moteriškiems garso įrašams.

Apibendrinus tikslumo ir greitumo tyrimus, galima teigti, kad greičiausia požymio skaičiavimo konfigūracija, nors ir nėra tiksliausia, vis tik geriausiai atitinka balsių nustatymo metodui iškeltus poreikius.

## IŠVADOS IR PASIŪLYMAI

1. Pagrindinis tonas gali būti nustatomas tiek iš laikinės, tiek ir iš dažninės kalbos signalo reprezentacijos. Sukurta daugybė pagrindinio tono nustatymo metodų, tačiau nei vieno iš jų negalima vadinti visapusiškai universaliu metodu, t.y. metodu pateikiančiu geriausius rezultatus iš bet kokio garso signalo.
2. Atlikus lyties nustatymo tyrimą pagal pagrindinį toną, nustatyta, kad pagrindinio tono nustatymo klaida svyruoja nuo 2,54% iki 3,94% taikant ribinę 160Hz pagrindinio tono reikšmę ir nuo 17,70% iki 23,94% taikant griežtus pagrindinio tono intervalus diktoriaus lyčiai įvardinti.
3. Atlikti tyrimai leidžia daryti išvadą, kad iš tirtų pagrindinio tono nustatymo metodų stabiliausias savo tikslumu – kepstras.
4. Balsių nustatymas šnekamosios kalbos signale yra vienas pirmųjų ir vienas svarbiausių etapų, kurio metu siekiama kuo tiksliau identifikuoti balsius, t.y. atskirti juos nuo priebalsių.
5. Balsių radimo metodai gali būti klasifikuojami į garso signalo analize pagrįstus, vaizdo analize pagrįstus ir mišrius metodus. Garso signalu pagrįsti metodai gali būti klasifikuojami į realaus laiko ir statistinius metodus.
6. Skaitmeniniai Filtrai pasirinkti atsižvelgiant į jų realizaciją su „Matlab“ programiniu paketu, išraiškos sudėtingumą ir paplitimą. Parinkti trys filtrai: Batervorto (*ang. Butterworth*); Čebiševo (*ang. Chebyshev*); FIR (*ang. Finite Impulse Response*).
7. Filtro eilei parinkti atlikta keletas smulkių tyrimų, kurių metu skaičiuotas filtro skaičiavimo laikas ir klaida. Remiantis gautais rezultatais pasirinktos pradinės filtrų eilės: Čebiševo 4 eilės filtras; Batervorto 4 eilės filtras; FIR 10 eilės filtras.
8. Sudarytas suminės harmonikų energijos metodo modelis susideda iš trijų pagrindinių etapų: pagrindinio tono apskaičiavimo, signalo filtravimo ir filtruotų signalų energijų sumos apskaičiavimo. Šiems skaičiavimams atlikti turi būti taikomi papildomi algoritmai: signalo paruošimas analizei, harmonikų nustatymas ir filtruojamų intervalų nustatymas
9. Tyrimui atlikti naudojama specialiai tyrimui sukurta programinė įranga. Ši programinė įranga pasižymi visu tyrimui atlikti reikalingu funkcionalumu.
10. Sukurtos sistemos architektūrinis sprendimas pagrįstas C#.NET 3.0 programavimo platform, tačiau matematiniai skaičiavimai atliekami su „MATLAB“ serveriu. Tai leidžia ne tik naudoti daugumą „MATLAB“ standartinių funkcijų, bet ir nesunkiai leisti tyrėjui koreguoti naudojamą „MATLAB“ kodą.

11. Atlikus pagrindinį ir papildomus eksperimentus, nustatyta, kad moteriškiems garso įrašams geriausias tikslumas pasiektas taikant Batervorto filtrą ir kepstro pagrindinio tono nustatymo metodą – 83,96%. Taikant auto koreliacijos pagrindinio tono nustatymo metodą, tiksliausias rezultatas su tuo pačiu – Batervorto filtru. Mažiausias tikslumas gautas taikant auto koreliacijos metodą ir FIR filtrą.
12. Vyriškiems garsams geriausias tikslumas gaunamas taikant Batervorto filtrą ir kepstro pagrindinio tono nustatymo metodą – 86,57%. Taikant auto koreliacijos pagrindinio tono nustatymo metodą rezultatai prastesni apytiksliai 2% punktais.
13. Batervorto filtro ir kepstro pagrindinio tono radimo metodo derinys yra pats tiksliausias tiek vyriškų, tiek moteriškų garso įrašų atžvilgiu.
14. Vyriškiems garso įrašams požymio tikslumas didesnis ~3% punktais.
15. Realizuotas požymis negali išskirti balsių nuo pusbalsių, ypač m, n ir l.
16. Įvertinus geriausią požymio tikslumo rezultata, iš tyrimo pašalinus pusbalsius, vyriškų garso įrašų atveju gauname 98,90% tikslumą. Moteriškų garsų atveju – 96,49%. Tad galima teigti, jog požymis geba puikiai atskirti balsius nuo sprogstamųjų ir friktyvinių priebalsių.
17. Remiantis gautais greičio tyrimo rezultatais, nustatyta, kad greičiausia požymio skaičiavimo konfigūracija yra FIR filtras ir kepstro pagrindinio tono radimo metodas. Ši konfigūracija yra greičiausia ir vyriškiems, ir moteriškiems garso įrašams.
18. Apibendrinus tikslumo ir greitumo tyrimus, galima teigti, kad greičiausia požymio skaičiavimo konfigūracija, nors ir nėra tiksliausia, vis tik geriausiai atitinka balsių nustatymo metodui iškeltus poreikius.
19. Remiantis visais iki šiol atliktais tyrimais galime teigti, kad tiriamas požymis gali būti taikomas balsiams nustatyti.
20. Siekiant padidinti algoritmo optimalumą, būtina atlikti daugiau tyrimų siekiant pagerinti ne tik algoritmo tikslumą, bet ir greitumą.
21. Siekiant didesnio algoritmo tikslumo reikėtų atlikti papildomus tyrimus skaičiavimuose remianti daugiau nei 8 harmonikomis.

## LITERATŪROS SĄRAŠAS

1. BALBONAS, D; DAUNYS, G (2005). *Fonemų klasifikavimas panaudojant garso ir vaizdo informaciją*. Elektronika ir elektrotechnika. 2005. Nr. 5(61). ISSN 1392 – 1215. [interaktyvus]. [žiūrėta 2009m. gegužės 6 d.]. Prieiga per internetą: <<http://www.ee.ktu.lt/journal/2005/5/Balbonas.pdf>>.
2. BIRNSTIEL, Tilman. *Vowel Recognition using Bayesian Analysis*. (2006) State University of New York University at Albany Physics Department, 2006 Sausio 13d. [interaktyvus]. [žiūrėta 2009 m. sausio 6 d.]. Prieiga per internetą: <[www.til-birstiel.de/vortraege/Vowel.pdf](http://www.til-birstiel.de/vortraege/Vowel.pdf)>.
3. BOERSMA. P. *Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound*. Institute of Phonetic Sciences, University of Amsterdam, 1993, p.17.
4. CHEN, Jau-Hung; KAO, Yung-An (2001). *Pitch Marking Based on an Adaptable Filter and a Peak-Valley Estimation Method*. Computational Linguistics and Chinese Language Processing Vol. 6, No. 2, February 2001, p. 1-12 [interaktyvus]. [žiūrėta 2007 m. sausio 4 d.]. Prieiga per internetą <<http://www.aclclp.org.tw/clclp/v6n2/v6n2a3.pdf>>.
5. CHILDERS, D. G. *Speech processing and synthesis toolboxes*. John Wiley & Sons, 2000. – 484 p.
6. CHU, C.Way (2003). *Speech Coding Algorithms*. Published by John Wiley & Sons, Inc., Hoboken, New Jersey. 2003 [interaktyvus]. [žiūrėta 2007 m. gruodžio 28 d.]. ISBN 0-471-37312-5.
7. DE SENA, Antonio; ROCCHESO, Davide (2005). *A study on using the mellin transform for vowel recognition*. Universit`a di Verona Dipartimento di Informatica. [interaktyvus]. [žiūrėta 2009 m. sausio 6 d.]. Prieiga per internetą: <[smc.afim-asso.org/smc05/papers/AntonioDeSena/DesRocSMC05.pdf](http://smc.afim-asso.org/smc05/papers/AntonioDeSena/DesRocSMC05.pdf)>.
8. DRIAUNYS, Kęstutis; RUDŽIONIS, Vytautas; ŽVINYS, Pranas (2005). *Hierarchine fonemų struktūra grindžiamo LTDIGITS fonemų klasifikavimo tyrimas*. Informacinės technologijos. – Kaunas: Technologija, 2005. – p. 283-288.
9. GERHARD, D. *Pitch Extraction and Fundamental Frequency: History and Current Techniques*. Technical Report TR-CS 2003-06, 2003.11, ISBN 0 7731 0455 0, ISSN 0828-3494.
10. GOLD, B; MORGAN, N. *Processing and Perception of Speech and Music*. Speech and Audio Signal Processing John Wiley and Sons, Inc: New York, 2000.

11. IQBAL, H. Rizwan. AWAIS, M. Muhammad. MASUD Shahid. SHAMAIL, Shafay. *On Vowels Segmentation and Identification Using Formant Transitions in Continuous Recitation of Quranic Arabic*. Springer Berlin / Heidelberg. 2008. ISSN 1860-949X. ISBN 978-3-540-79354-0. p.155-162
12. YUK YAN CHENG, Octavian (2003). *Speech Processing Workstation*, The university of auckland Department of Electrical & Electronic Engineering 2003 [interaktyvus]. [žiūrėta 2007 m. gruodžio 29 d.]. Prieiga per internetą <[www.ele.auckland.ac.nz/~p4p\\_2005/archive/reports2003/pdfs/p60\\_hlai015.pdf](http://www.ele.auckland.ac.nz/~p4p_2005/archive/reports2003/pdfs/p60_hlai015.pdf)>.
13. KASPARAITIS, P. *Kompiuterinė lingvistika*. Paskaitų konspektai, VU 2007 m. [interaktyvus]. [žiūrėta 2007 m. lapkričio 17 d.]. Prieiga per internetą <<http://www.mif.vu.lt/~pijus/CL/cl.htm>>.
14. KUK KIM, Young; LEE, Ki Young; JIN BAE, Myung (2005). *On a Pitch Detection Method Using Noise Reduction*, Computational Linguistics and Intelligent Text Processing p. 429-432, Springer Berlin / Heidelberg 2005 [interaktyvus]. [žiūrėta 2008 m. sausio 3 d.]. Prieiga per internetą <<http://www.springerlink.com/index/1T64BXJ044MFTW1W.pdf>>.
15. LAURINČIUKAITĖ, Sigita (2008). *Lietuvių šnekos atpažinimo akustinis modeliavimas*. Daktaro disertacija. Vilniaus Gedimino technikos universitetas, matematikos ir informatikos institutas, Vilnius 2008.
16. LAWRENCE, M.D.; ANNE, A. *Selected References: Speech Therapy and the Transsexual Voice*, 1998.
17. LAWRENCE, M.D.; ANNE, A. *Voice Feminization Surgery: A Critical Overview*, 2004.01.
18. LEE, Minkyu; VAN SANTEN Jan; MOBIUS, Bernd; OLIVE, Joseph. (1999) *Formant Tracking Using Segmental Phonemic Information*, iš EUROSPEECH'99, Budapest, Hungary [interaktyvus]. [žiūrėta 2007 m. lapkričio 18 d.]. Prieiga per internetą <[www.ims.uni-stuttgart.de/~moebius/papers/euro99.pdf](http://www.ims.uni-stuttgart.de/~moebius/papers/euro99.pdf)>.
19. LYONS, R. G. *Understanding Digital signal processing*. Addison Wesley Longman Inc. 1997. p. 494-505. [interaktyvus]. [žiūrėta 2009 m. sausio 8 d.]. Prieiga per internetą: <<http://www.dspguru.com/info/terms/filtterm/index2.htm>>.
20. MARCHAND, Sylvain (2001). *An Efficient Pitch-tracking Algorithm using a Combination of Fourier Transforms*, Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01), Limerick, Ireland, December 6-8, 2001 [interaktyvus]. [žiūrėta 2007 m. gruodžio 30 d.]. Prieiga per internetą <[www.csis.ul.ie/dafx01/proceedings/papers/marchand.pdf](http://www.csis.ul.ie/dafx01/proceedings/papers/marchand.pdf)>.

21. MORKEVIČIUS, Aurelijus (2008). *Diktoriaus lyties nustatymas pagal pagrindinį toną*. Informacinės technologijos 2008 konferencijos pranešimų medžiaga, Kauno technologijos universitetas, 2008, 88-92 p.
22. MORKEVIČIUS, Aurelijus; ŽVINYS, Pranas (2009). *Algoritmas skirtas balsių ir priebalsių diferenciacijai šnekamosios kalbos signale*. Informacinės technologijos 2009 konferencijos pranešimų medžiaga, Vilniaus universitetas, Kauno humanitarinis fakultetas, 2009, 17-20 p.
23. NGUYEN, Danny (2005). *Enhanced Spectrogram Tools for Speech Signals*. School of Electrical and Computer Engineering, Georgia USA. [interaktyvus]. [žiūrėta 2007 m. lapkričio 17 d.]. p. 3. Prieiga per internetą <<http://users.ece.gatech.edu/~danny/ECE%206255%20Project.ppt>>.
24. PAKERYS, A. *Lietuvių bendrinės kalbos fonetika*. Žara, Vilnius, 1995.
25. PETRAUSKAS, Vytautas. *Žemų dažnių ketvirtos eilės Beselio filtro tyrimas*. Kauno Technologijos Universitetas.[interaktyvus].[žiūrėta 2009m. sausio 8 d.]. Prieiga per internetą: <<http://ifko.ktu.lt/~vypet/T120B012/filtras1.htm>>.
26. PHOUNG, Nguyen Huu (2008). *FIR Filter Design: The window Design Method*. Open Course Ware. [interaktyvus].[žiūrėta 2009m. balandžio 6 d.]. Prieiga per internetą <<http://vocw.edu.vn/content/m11285/latest/>>.
27. PIERCE, John R. (2001). *Consonance and Scales*. Perry R. Cook. Music, Cognition, and Computerized Sound. MIT Press. ISBN 9780262531900.
28. PROLOY, Bhattacharya; YUMAO, Lu; YANG, Wu; HUNG WEI, Kuo. *A Discussion of Pitch and Formant Estimation Methods*, University of California, Los Angeles 405 Hilgard Avenue Los Angeles [interaktyvus]. [žiūrėta 2007 m. gruodžio 29 d.].
29. RAGUOTIENĖ, G. *Žodžiai apie kalbą*. Vilnius, 1998.
30. SADEGHI, Vahideh Sadat, YAGHMAIE, Khashayar (2006). *Vowel Recognition using Neural Networks*. IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.12, December 2006. [interaktyvus].[žiūrėta 2009 m. sausio 4 d.]. Prieiga per internetą: <[paper.ijcsns.org/07\\_book/200612/200612A22.pdf](http://paper.ijcsns.org/07_book/200612/200612A22.pdf)>.
31. SHI, Yu; CHANG, Eric (2003). *Spectrogram-based format tracking via particle filters*, iš ICASSP 2003 Hong Kong, China. p I-168-I-171. ISBN: 0-8493-3692-9.
32. SHIMODAIRA, Hiroshi; NAKAI, Mitsuru (1992). *Robust Pitch detection by narrow Band Spectrum Analysis*. Japan Advanced Institute of Science and Technology, Hokuriku 1992 [interaktyvus]. [žiūrėta 2007 m. gruodžio 30 d.]. Prieiga per internetą <[www.cstr.ed.ac.uk/downloads/publications/1992/icslp92.pdf](http://www.cstr.ed.ac.uk/downloads/publications/1992/icslp92.pdf)>.

33. SMITH, Steven W (1997). *Introduction to Digital Filters*. The Scientist and Engineer's Guide to Digital Signal Processing. [interaktyvus]. [žiūrėta 2009m. sausio 6 d.]. Prieiga per internetą: <<http://www.dspguide.com/ch14/5.htm>>.
34. THIBAUT, Francois; (2003) *Formant Trajectory Detection Using Hidden Markov Models*, Sound Processing and Control Lab (SPCL) at McGill University 2003 [interaktyvus]. [žiūrėta 2007 m. lapkričio 13 d.]. Prieiga per internetą <[www.music.mcgill.ca/musictech/spcl/formants\\_tracking\\_using\\_hidden\\_markov\\_models](http://www.music.mcgill.ca/musictech/spcl/formants_tracking_using_hidden_markov_models)>.
35. URBANA VIČIUS, Alfredas. *Fonetikos sąvoka*. Gerbkime žodį [interaktyvus]. [žiūrėta 2007 m. lapkričio 16 d.]. Prieiga per internetą <[http://ualgiman.dtiltas.lt/fonetikos\\_savoka.html](http://ualgiman.dtiltas.lt/fonetikos_savoka.html)>.
36. WAGNER, David B (1995). *Dynamic Programming*. An introductory article on dynamic programming in Mathematica. [interaktyvus]. [žiūrėta 2007 m. lapkričio 18 d.]. Prieiga per internetą <<http://citeseer.ist.psu.edu/268391.html>>

## PRIEDAI

<b>1 PRIEDAS.</b> XIII tarpuniversitetinės magistrantų ir doktorantų konferencijos pranešimų medžiaga: Diktoriaus lyties nustatymas pagal pagrindinį toną .....	73
<b>2 PRIEDAS.</b> 14-osios tarpuniversitetinės magistrantų ir doktorantų konferencijos pranešimų medžiaga: Algoritmas skirtas balsių ir priebalsių diferenciacijai šnekamosios kalbos signale .....	78
<b>3 PRIEDAS.</b> Pagrindinio tyrimo papildomi suminiai duomenys.....	83
<b>4 PRIEDAS.</b> Tyrimo aplinkos informacinės sistemos rezultatų langas .....	85
<b>5 PRIEDAS.</b> Artikuliaciniai priebalsių požymiai. ....	86
<b>6 PRIEDAS.</b> Papildomų tyrimų rezultatai .....	87
<b>7 PRIEDAS.</b> Dirbtinių neuroninių tinklų taikymas diktoriaus lyčiai nustatyti.....	95

























## Pagrindinio tyrimo papildomi suminiai duomenys

## Vyriškiems garsams

Balsis	F0 - pagrindinis tonas (vidurkis)	FH1 - pirma harmonika (vidurkis)	FH2 - antra harmonika (vidurkis)	FH3 - trečia harmonika (vidurkis)
a	132,79	265,59	398,38	531,17
š	273,75	547,5	821,24	1.094,99
t	224,45	448,89	673,34	897,79
u	137,07	274,14	411,22	548,29
o	130,28	260,56	390,84	521,12
n	135,5	271,01	406,51	542,02
i	143,95	287,9	431,85	575,8
d	201,38	402,76	604,14	805,52
r	215,91	431,82	647,74	863,65
s	262,67	525,34	788,01	1.050,67
k	270,28	540,57	810,85	1.081,13
e	127,43	254,85	382,28	509,7
p	261,02	522,04	783,06	1.044,08
v	148,95	297,9	446,86	595,81
l	149,23	298,47	447,7	596,93
ė	132,68	265,36	398,04	530,72
b	186,87	373,75	560,62	747,5
g	183,81	367,61	551,42	735,23
z	208,74	417,47	626,21	834,94
m	118,45	236,89	355,34	473,79
ž	277,13	554,25	831,38	1.108,51

Šaltinis: sudaryta autoriaus.

## Moteriškiems garsams

Balsis	F0 - pagrindinis tonas (vidurkis)	FH1 - pirma harmonika (vidurkis)	FH2 - antra harmonika (vidurkis)	FH3 - trečia harmonika (vidurkis)
t	224,87	449,74	674,61	899,48
r	263,9	527,8	791,69	1.055,59
i	223,91	447,82	671,73	895,64
s	241,32	482,64	723,95	965,27
d	234,72	469,44	704,15	938,87
e	202,66	405,32	607,98	810,64
v	217,17	434,34	651,51	868,69
n	204,99	409,99	614,98	819,98
š	275,2	550,39	825,59	1.100,79
a	200,54	401,08	601,62	802,16
p	267,15	534,29	801,44	1.068,59
k	241,51	483,02	724,54	966,05
u	199,87	399,75	599,62	799,5
o	194,68	389,37	584,05	778,74
l	214,5	429	643,5	858
ė	204,8	409,6	614,4	819,21
b	198,23	396,47	594,7	792,93
g	229,53	459,07	688,6	918,13
z	199,36	398,73	598,09	797,46
m	203,29	406,57	609,86	813,14
ž	290,51	581,01	871,52	1.162,02

Šaltinis: sudaryta autoriaus

## Tyrimo aplinkos informacinės sistemos rezultatų langas

Failas	Žodis	Garsas	Pagrindinis tonas	1-ma harmonika	2-ra harmonika	3-čia harmonika	Energija
T07.WAV	pirmyn	i	231,88	463,77	695,65	927,54	0,042381
T07.WAV	pirmyn	i	206,15	412,31	618,46	824,61	0,045301
T07.WAV	atgal	a	220,70	441,40	662,10	882,80	0,028229
T07.WAV	atgal	a	206,37	412,73	619,10	825,47	0,076789
T07.WAV	ipradzia	i	210,53	421,05	631,58	842,11	0,002278
T07.WAV	ipradzia	a	179,58	359,15	538,73	718,30	0,036216
T07.WAV	ipradzia	e	299,01	598,02	897,02	1.196,03	0,012408
T07.WAV	ipabaiga	i	205,13	410,26	615,38	820,51	0,002536
T07.WAV	ipabaiga	a	150,98	301,96	452,94	603,92	0,012601
T07.WAV	ipabaiga	a	195,18	390,36	585,54	780,72	0,011299
T07.WAV	ipabaiga	i	238,81	477,61	716,42	955,22	0,007753
T07.WAV	ipabaiga	a	235,53	471,06	706,59	942,12	0,035321
T07.WAV	sekantis	e	379,06	758,12	1.137,19	1.516,25	0,016448
T07.WAV	sekantis	a	212,63	425,26	637,90	850,53	0,008420
T07.WAV	sekantis	i	211,10	422,20	633,31	844,41	0,009614
T07.WAV	perduoti	e	137,08	274,16	411,24	548,32	0,009998
T07.WAV	perduoti	u	165,35	330,69	496,04	661,38	0,004475
T07.WAV	perduoti	o	192,88	385,77	578,65	771,53	0,002755

Garsas	Pagrindinis tonas (avg)	1-ma harmonika (avg)	2-ra harmonika (avg)	3-čia harmonika (avg)	Energija (avg)	Dispersija	Standartinis nuokrypis	Mediana	Minimumas	Maksimumas	Percentilis (20)	Percentilis (80)
i	237,48	474,96	712,44	949,92	0,048793	0,001319	0,036318	0,043625	0,000840	0,152967	0,016314	0,076509
a	224,63	449,27	673,90	898,53	0,058736	0,001769	0,042064	0,051029	0,008420	0,151288	0,018660	0,090835
e	267,74	535,47	803,21	1.070,94	0,077028	0,006524	0,080771	0,057316	0,009998	0,346970	0,016272	0,081801
u	243,84	487,68	731,52	975,36	0,072248	0,003849	0,062041	0,064890	0,003824	0,224540	0,015460	0,104273
o	209,34	418,68	628,03	837,37	0,033097	0,001014	0,031841	0,015437	0,002755	0,078332	0,006505	0,070256
y	221,77	443,54	665,30	887,07	0,147137	0,004793	0,069231	0,147137	0,098183	0,196091	0,098183	0,196091

## Artikuliaciniai priebalsių požymiai

Oro skverbimosi būdas	Artikuliacijos vieta		Balso stygų veikla	Patališkumas (minkštumas)	Priebalsis		
	aktyvusis kalbos padargas	pasyvusis kalbos padargas					
Sprogstamieji	lūpiniai	lūpiniai	duslieji	kietasis	<i>p</i>		
				minkštasis	<i>p'</i>		
			skardieji	kietasis	<i>b</i>		
				minkštasis	<i>b'</i>		
	liežuvio priešakiniai	dantiniai	duslieji	kietasis	<i>t</i>		
				minkštasis	<i>t'</i>		
			skardieji	kietasis	<i>d</i>		
				minkštasis	<i>d'</i>		
	liežuvio užpakaliniai	gomurio užpakaliniai	duslusis	kietasis	<i>k</i>		
			skardusis	kietasis	<i>g</i>		
	liežuvio viduriniai	gomurio viduriniai	duslusis	minkštasis	<i>k'</i>		
			skardusis	minkštasis	<i>g'</i>		
Pučiamieji	liežuvio priešakiniai	dantiniai	duslieji	kietasis	<i>s</i>		
				minkštasis	<i>s'</i>		
			skardieji	kietasis	<i>z</i>		
				minkštasis	<i>z'</i>		
		alveoliniai	duslieji	kietasis	<i>š</i>		
				minkštasis	<i>š'</i>		
			skardieji	kietasis	<i>ž</i>		
				minkštasis	<i>ž'</i>		
	liežuvio užpakaliniai	gomurio užpakaliniai	duslusis	kietasis	<i>x</i>		
			skardusis	kietasis	<i>h</i>		
	liežuvio viduriniai	gomurio viduriniai	duslusis	minkštasis	<i>x'</i>		
			skardusis	minkštasis	<i>h'</i>		
	lūpiniai	dantiniai	duslieji	kietasis	<i>f</i>		
				minkštasis	<i>f'</i>		
Sklandieji	paprastieji	lūpiniai	dantiniai	skardieji	kietasis	<i>v</i>	
					minkštasis	<i>v'</i>	
		liežuvio vidurinis	gomurio vidurinis		skardusis	minkštasis	<i>j'</i>
						nosiniai	lūpiniai
	minkštasis	<i>m'</i>					
		liežuvio priešakiniai	dantiniai	skardieji	kietasis	<i>n</i>	
					minkštasis	<i>n'</i>	
	šoniniai	liežuvio priešakiniai	dantinis	skardusis	kietasis	<i>l</i>	
			alveolinis	skardusis	minkštasis	<i>l'</i>	
	virpamieji	liežuvio priešakiniai	alveoliniai	skardusis	kietasis	<i>r</i>	
				skardusis	minkštasis	<i>r'</i>	

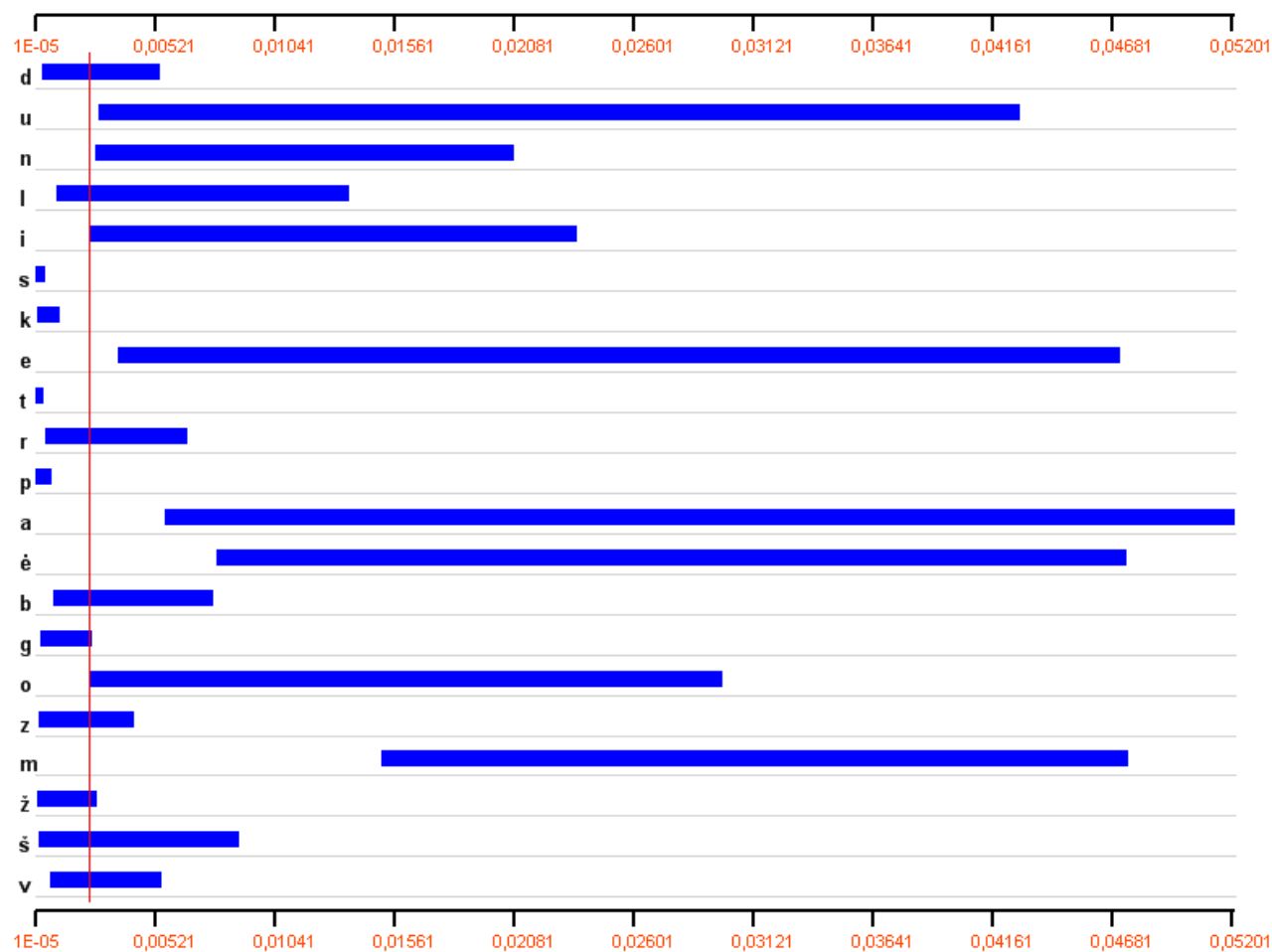
Šaltinis: Šaltinis: URBANAVIČIUS, Alfredas. *Fonetikos sąvoka*. Prieiga per internetą:

<[http://ualgiman.dtiltas.lt/artikuliaciniai\\_priebalsiu\\_pozymiai.html](http://ualgiman.dtiltas.lt/artikuliaciniai_priebalsiu_pozymiai.html)>

Pastaba. Apostrofo ženklas (') čia žymi suminkštintą (patalizuotą) priebalsį.

## Papildomų tyrimų rezultatai

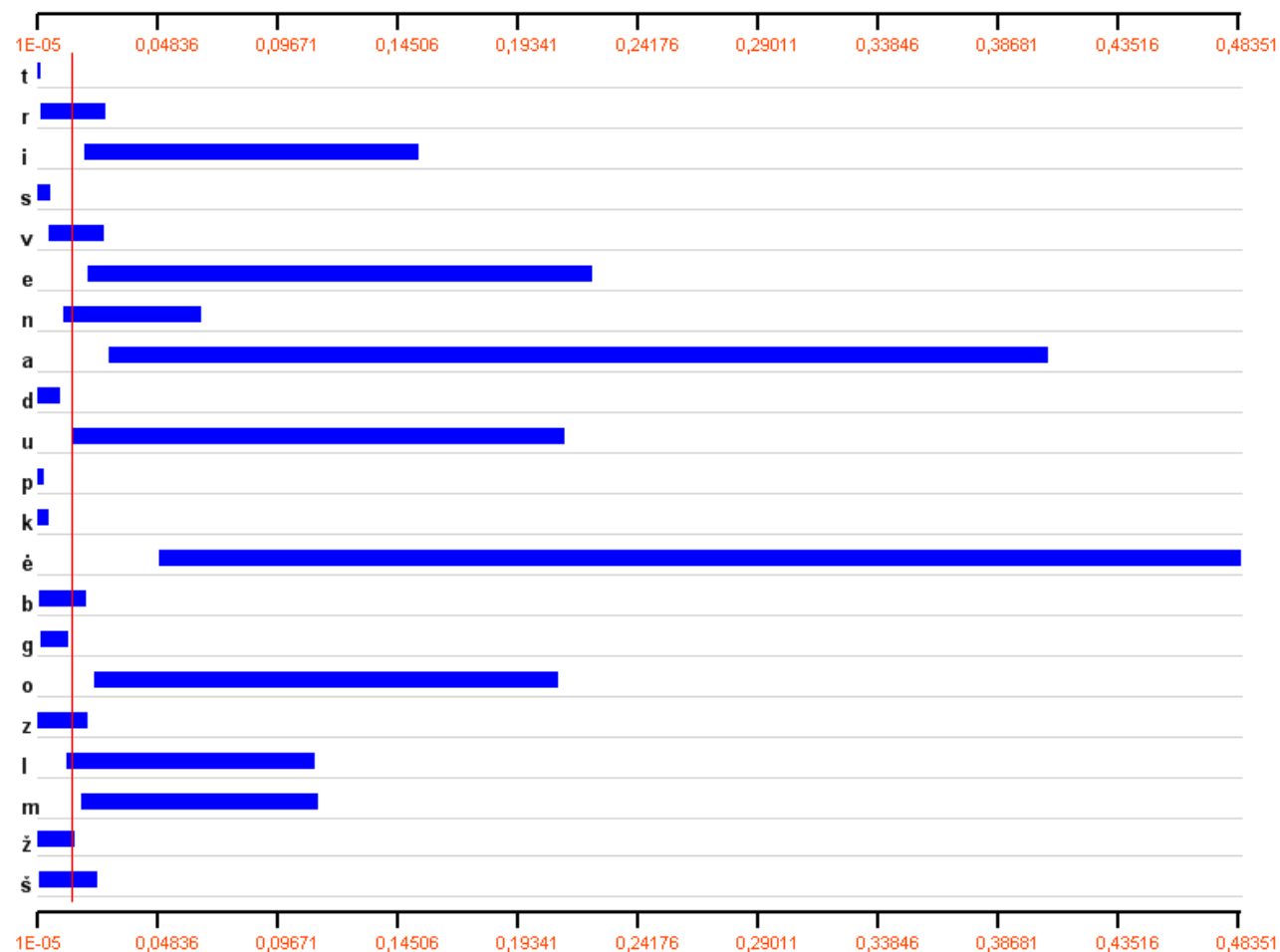
Tyrimas, taikant Batervorto filtrą ir kepstro pagrindinio tono nustatymo metodą moteriškiems garso įrašams.



Tikslumas - 83,96%

## 6 PRIEDO TĘSINYS

Tyrimas, taikant Batervorto filtrą ir kepstro pagrindinio tono nustatymo metodą vyriškiems garso įrašams.

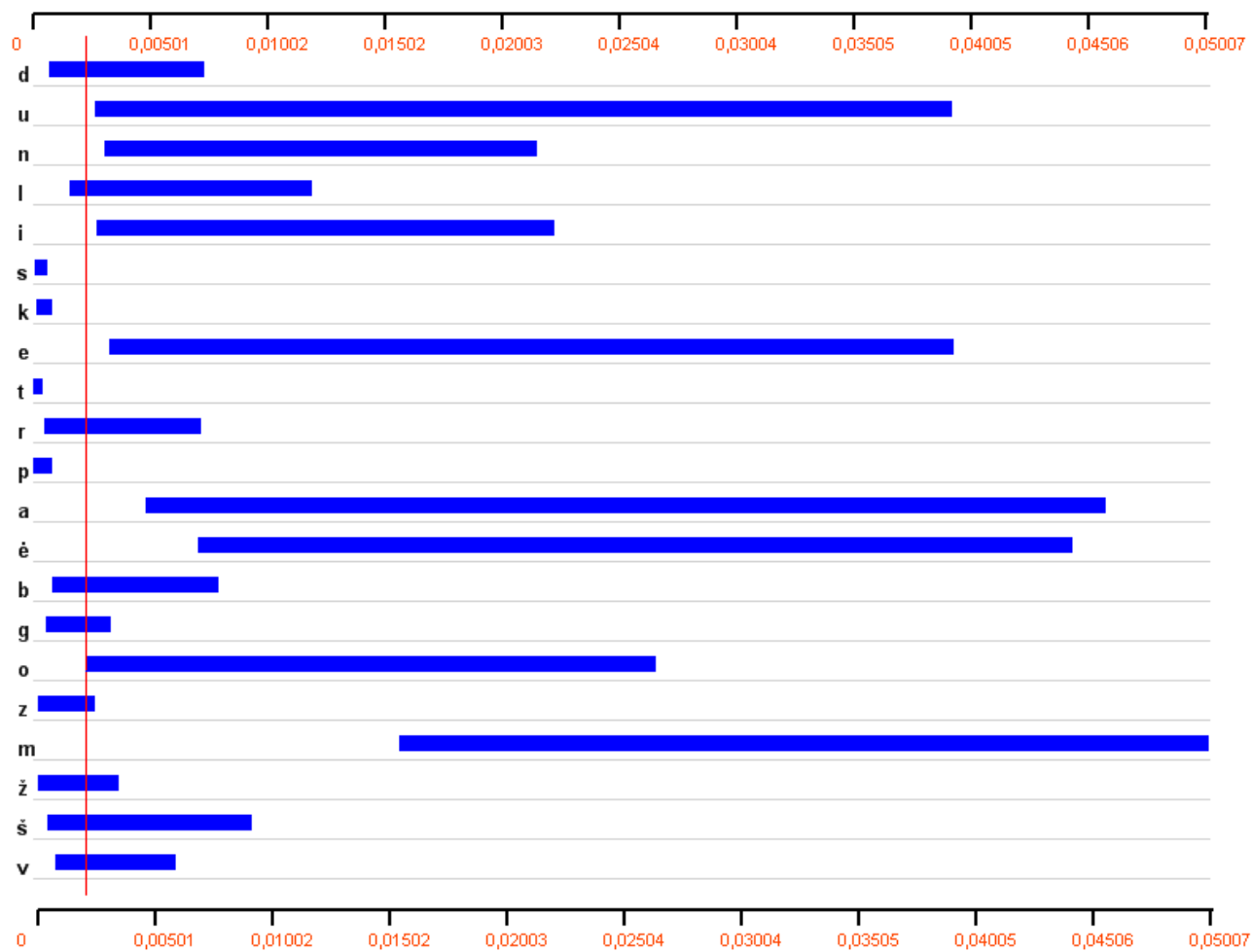


Tikslumas - 86,57%



## 6 PRIEDO TĘSINYS

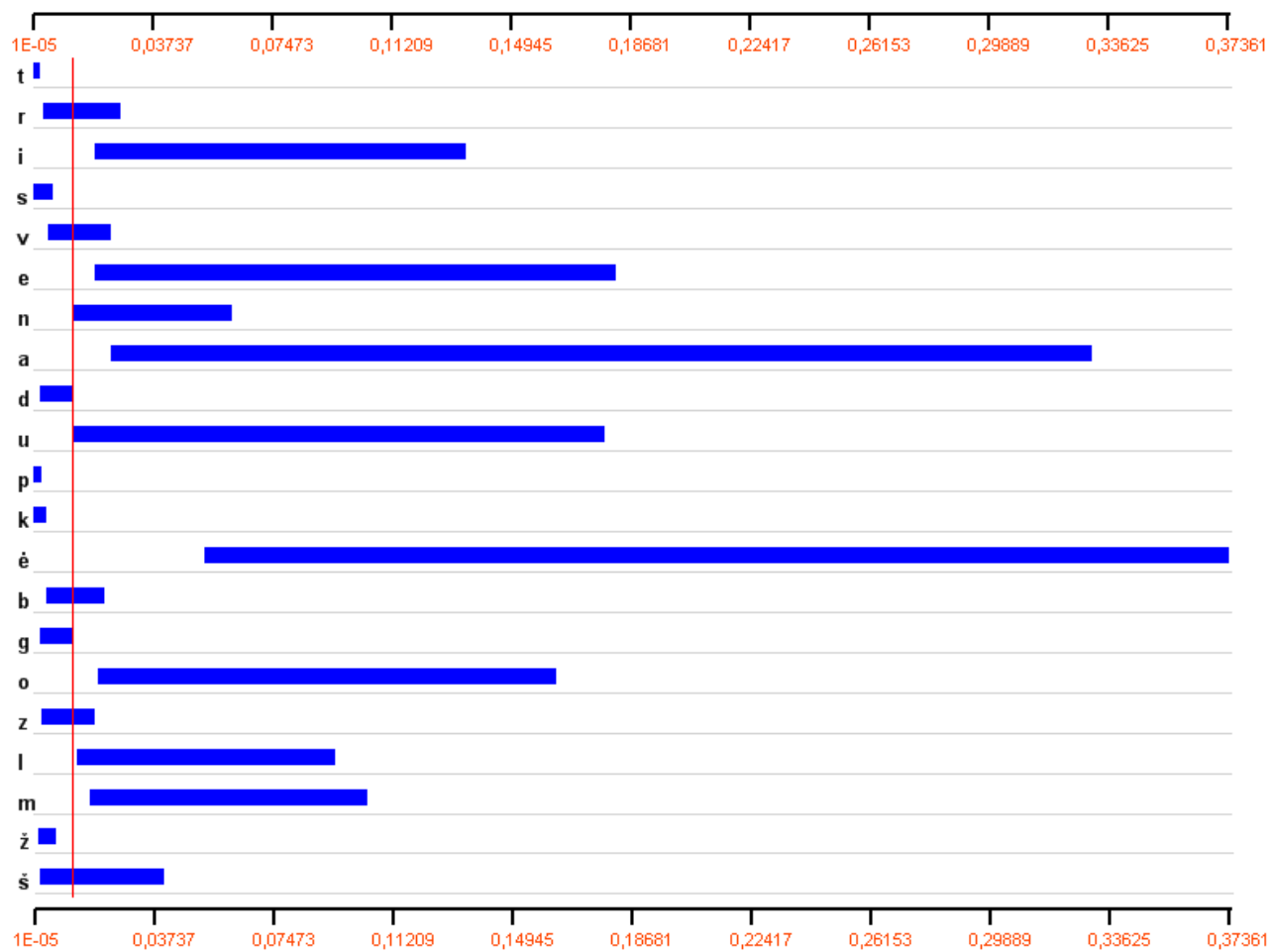
Tyrimas, taikant Čebiševo filtrą ir autokoreliacijos pagrindinio tono nustatymo metodą moteriškiems garso įrašams.



Tikslumas - 81,36%

## 6 PRIEDO TĘSINYS

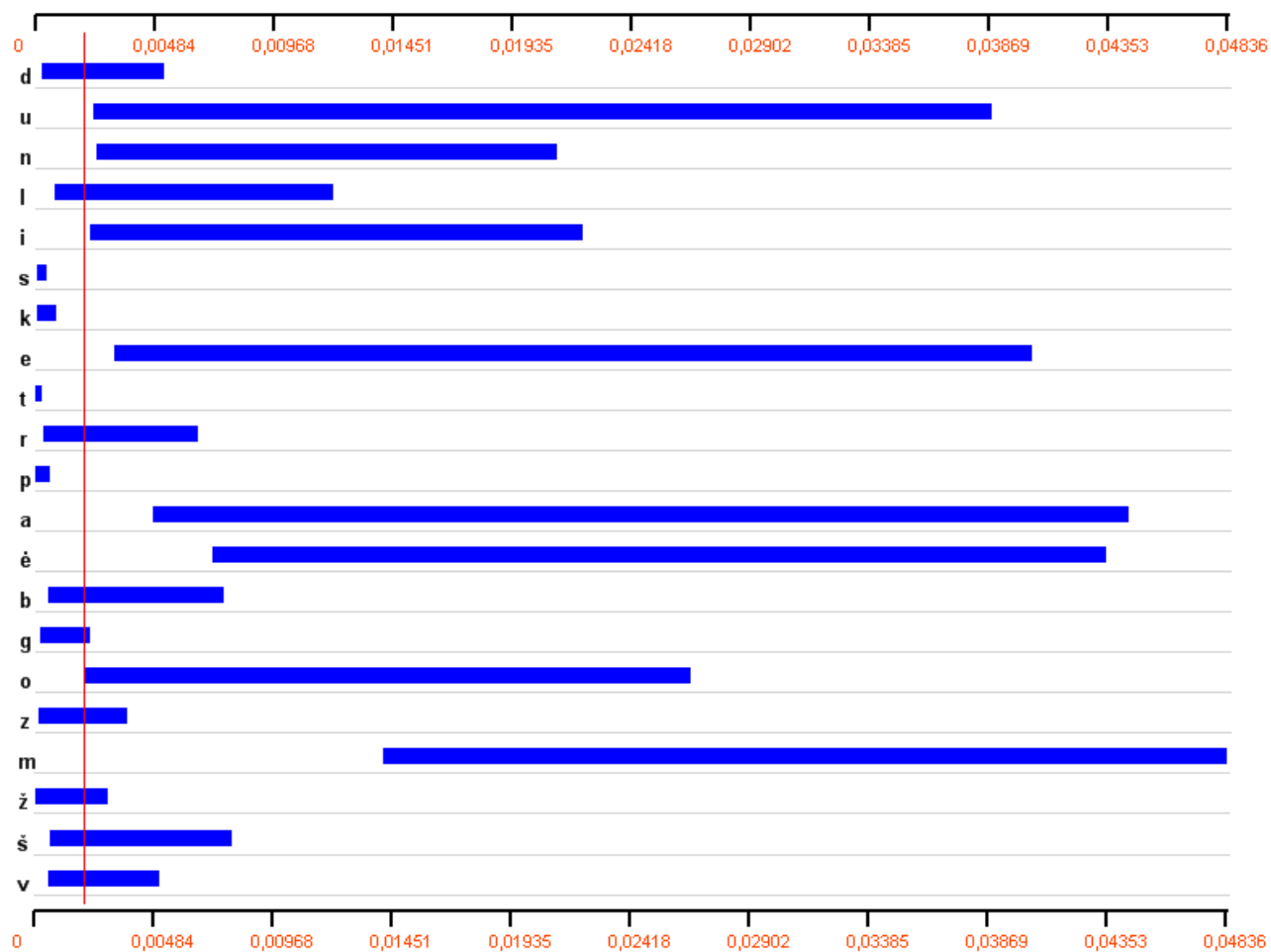
Tyrimas, taikant Čebiševo filtrą ir autokoreliacijos pagrindinio tono nustatymo metodą vyriškiems garso įrašams.



Tikslumas - 83,66%

## 6 PRIEDO TĘSINYS

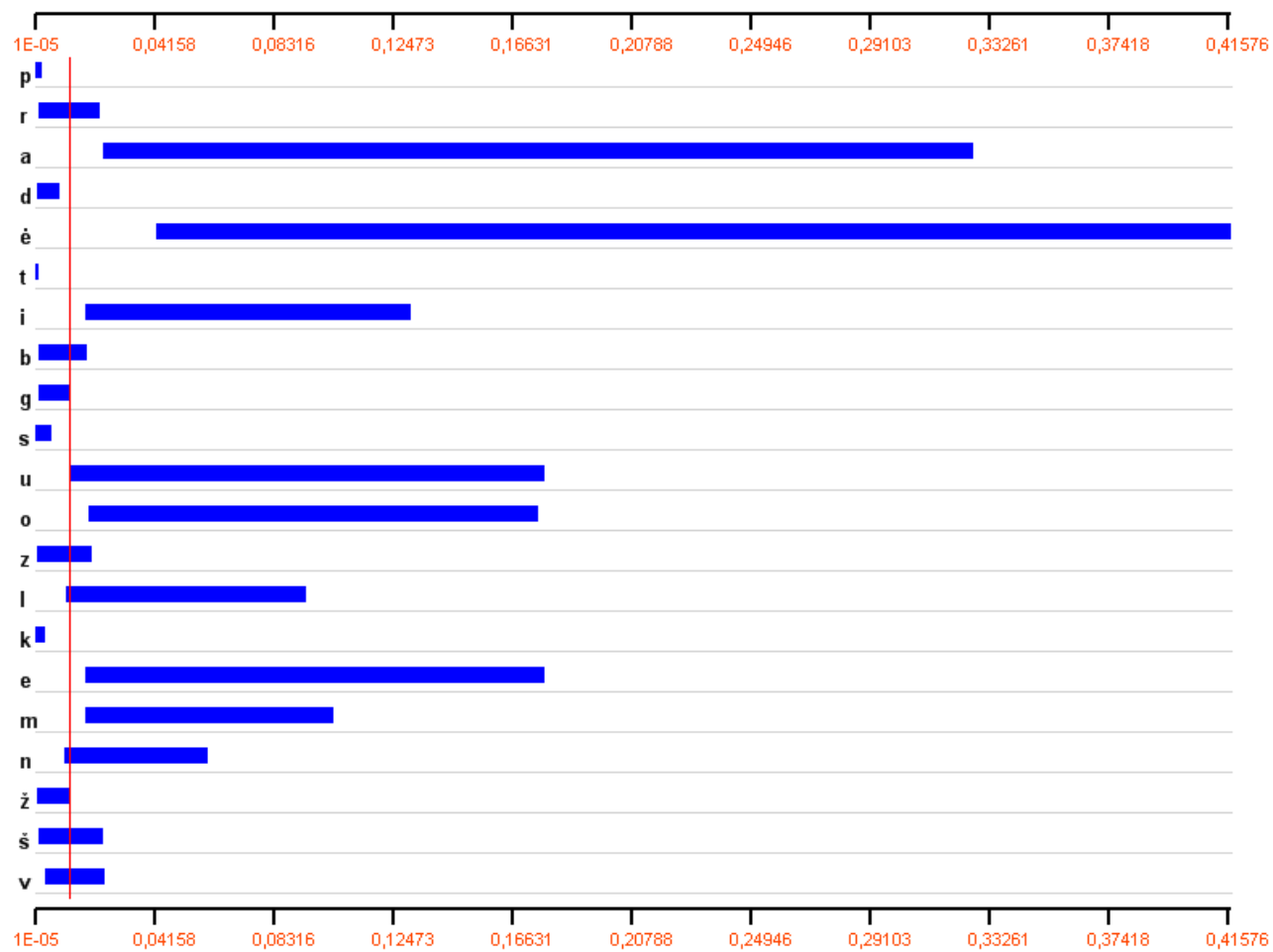
Tyrimas, taikant Čebiševo filtrą ir kepstro pagrindinio tono nustatymo metodą moteriškiems garso įrašams.



Tikslumas - 82,99%

## 6 PRIEDO TĘSINYS

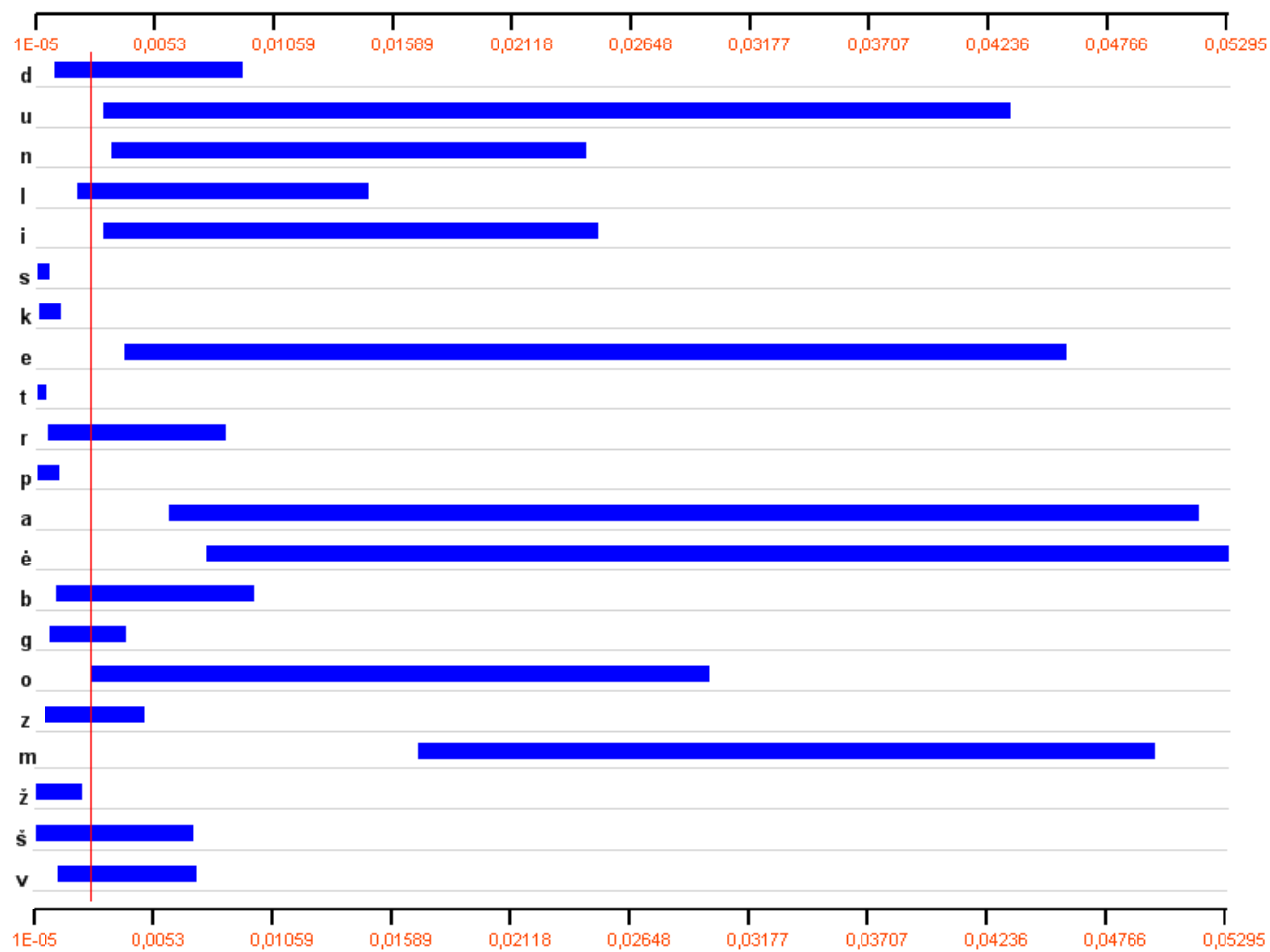
Tyrimas, taikant Čebiševo filtrą ir kepstro pagrindinio tono nustatymo metodą vyriškiems garso įrašams.



Tikslumas - 86,29%

## 6 PRIEDO TĘSINYS

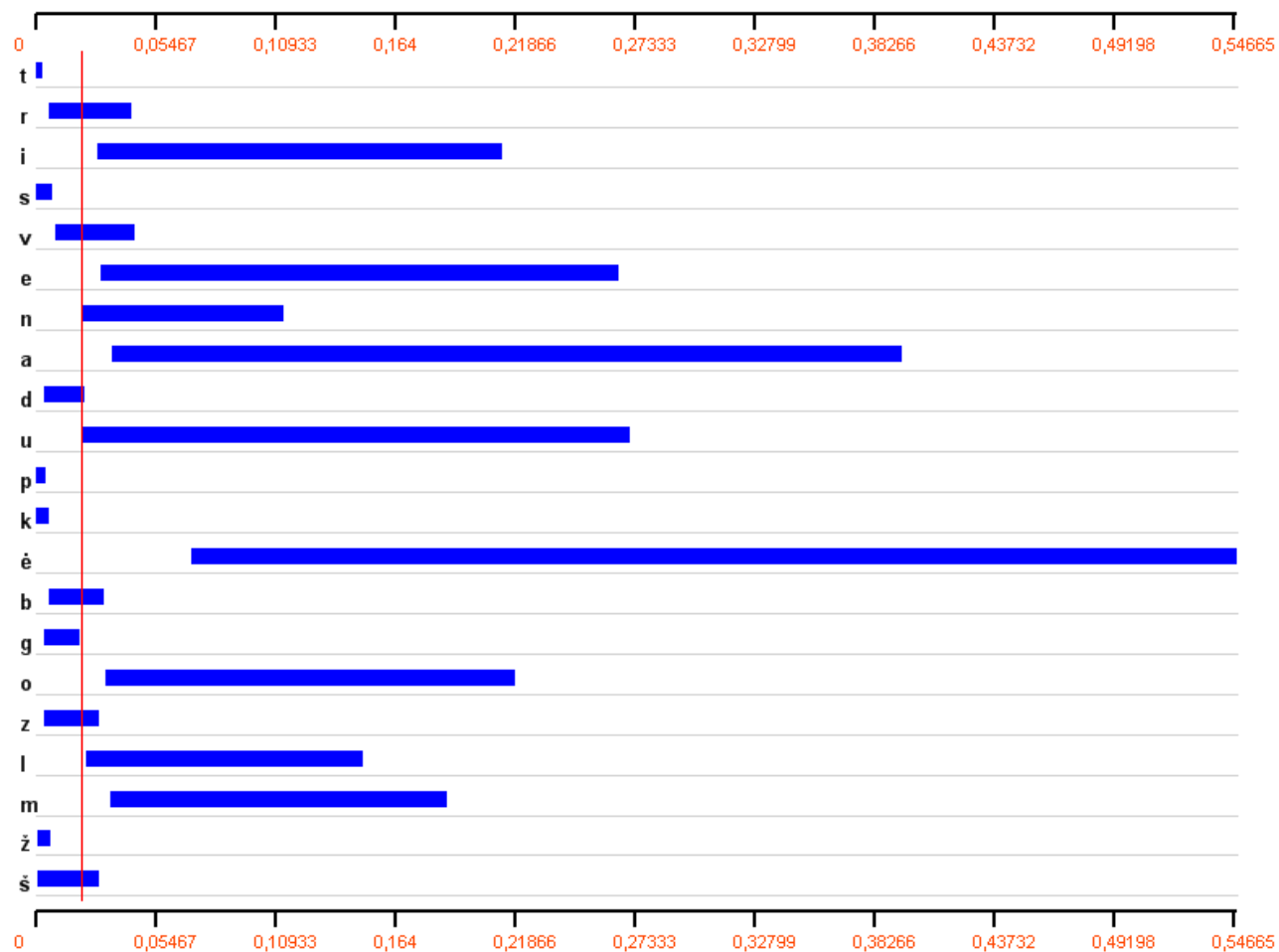
Tyrimas, taikant FIR filtrą ir autokoreliacijos pagrindinio tono nustatymo metodą moteriškiems garso įrašams.



Tikslumas - 81,51%

## 6 PRIEDO TĘSINYS

Tyrimas, taikant FIR filtrą ir autokoreliacijos pagrindinio tono nustatymo metodą vyriškiems garso įrašams.



Tikslumas - 84,81%

### Dirbtinių neuroninių tinklų taikymas diktoriaus lyčiai nustatyti

**Tyrimo tikslas:** Ištirti, kurio lietuvių kalbos balsio (a, e, i, u, o) pagrindinio tono pagalba, galima tiksliausiai atspėti diktoriaus lytį dirbtinio neuroninio tinklo pagalba.

**Duomenų šaltinis:** Duomenys tinklui rinkti eksperimentiniu būdu, skaičiuojant garso pagrindinį toną Kepstro metodu. Šis metodas pasirinktas atlikus pagrindinio tono metodų palyginamąją analizę ir eksperimentinį palyginamąjį tyrimą, kuris aprašytas moksliniame straipsnyje (Aurelijus Morkevičius, *Diktoriaus Lyties nustatymas pagal pagrindinį toną*, Informacinės technologijos XIII, KTU 2008.05.09). Pagrindinis tonas skaičiuotas balsiams (a, e, i, u, o) po tris kartus, t.y. balsio pradžios, balsio vidurio ir balsio pabaigos. Balsio pradžios ir pabaigos pagrindinis tonas nėra švarus. Jis veikiamas šalia esančių priebalsių arba kitų balsių, tad dažnai atvejais neatspindi tikrojo balsio pagrindinio tono. Tuo tarpu vidutinės balsio dalies pagrindinis tonas yra švariausias. Pagrindinis tonas skaičiuotas 50 diktorių vyrų ir 50 diktorių moterų žodžiams (Garso įrašai iš „LTDIGITS“ garsyno). Pirmo eksperimento atveju jis skaičiuotas tik a balsiams (971 balsių), antro – i balsiams (1747), trečio – e balsiams (1118), ketvirto – u balsiams (802), penkto – o balsiams (324).

Duomenys imami iš failų (output\_A.txt, output\_E.txt, output\_I.txt, output\_U.txt, output\_O.txt)

**Sprendimas:** Sprendimui pasirinktas MATLAB „NNSYSID20“ paketas. Modeliuojamas tiesinio sklidimo neuroninis tinklas su trim įėjimais (balsio pradžios pagrindinis tonas, balsio vidurio pagrindinis tonas, balsio pabaigos pagrindinis tonas) ir vienu išėjimu. Išėjimo reikšmės, tai 0 arba 1. Nulis - jei diktorius moteris, 1 – jei vyras. Vyra ir moterį skiria pagrindinio tono riba 160Hz (M.D.Lawrence, A.Anne, *Selected References: Speech Therapy and the Transsexual Voice*, 1998). Žemiau šios ribos vyriškas tonas, aukščiau – moteriškas.

Sprendimo eiga:

1. Duomenys iš duomenų failų nuskaitomi į MATLAB aplinką.
2. Apibrėžiamas dirbtinis neuroninis tinklas;
3. Apmokomas dirbtinis neuroninis tinklas (su 50% duomenų);
4. Tikrinamos tinklo generalizacijos savybės (su kita puse duomenų);
5. Atliekamas bendras palyginimas, su kuriuo balsiu tinklas daro mažiausią paklaidą, o su kuriuo didžiausią.

1 lentelė

#### Eksperimento su balsiu a tinklo apmokymo rezultatai

Paslėptų neuronų skaičius	Iteracijų skaičius	NSSE
3	2000	0.28048

Paslėptų neuronų skaičius	Iteracijų skaičius	NSSE
5	2000	0.24018
7	2000	0.22509
9	100	0.21278
20	1000	0.14868
40	1000	0.10082
80	1000	0.047007

Šaltinis: Sudaryta autoriaus.

Remiantis 1 lentelės duomenimis, tolesniems tyrimams nuspręsta pasirinkti 20 paslėptų neuronų tinklą, kad išvengtų ilgų skaičiavimų, bet tuo pačiu ir visiškai neprarasti tikslumo. Reikia pridurti, kad tyrimo tikslas ne gauti kuo tikslesnius rezultatus, o patikrinti, kuris balsis geriausiai tinkamas neuroniniam tinklui, diktoriaus lyties nustatymui.

### Apibendrinimas

Atliktų eksperimentų rezultatai pateikiami 2 lentelėje.

2 lentelė

#### Eksperimentų rezultatai

Balsis	Mokymų klaida (NSSE)	Generalizavimo klaida (NSSE)
a	0.14868	0.29069
e	0.14589	0.28921
i	0.1913	0.29051
u	0.16227	0.26553
o	0.09071	0.56686

Šaltinis: Sudaryta autoriaus.

Kaip matome 2 lentelėje, rezultatai labai panašūs. Galbūt šiek tiek daugiau išsiskiria o baltio rezultatai. Taip yra todėl, kad o baltio eksperimento atveju duomenų skaičius buvo apytiksliai trigubai mažesnis negu kitų balsių atveju. Vis tik, vertinant tiksliausius generalizavimo rezultatus, reikėtų išskirti balsio u eksperimentą. Šio eksperimento atveju generalizavimo klaida mažiausia, nepaisant to, kad mokymo klaida didžiausia. Be abejo tai gali sąlygoti simuliacijos duomenų aibę, kuri gali būti panaši į apmokymo duomenų aibę.

Daryti konkrečią išvadą apie tai, kurio baltio pagrindinis tonas geriausiai atspindi diktoriaus lytį yra pakankamai sudėtinga. Pagal atliktus tyrimus, neanalizuojant įvairių galimų šalutinių veiksnių, tiksliausiai diktoriaus lytį atspindi balsis u.



## Pagrindinio tono duomenų klasterizavimo tyrimas

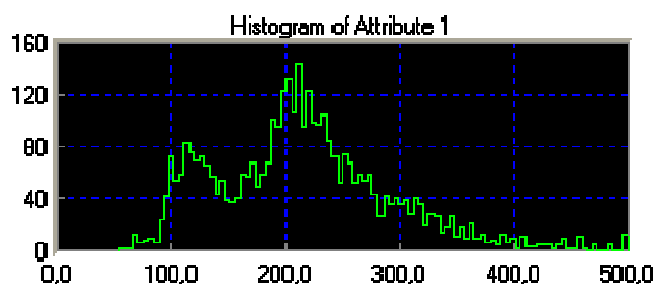
Papildomai prie pagrindinio tyrimo, atliekamas pagrindinio tono duomenų klasterizacijos tyrimas naudojantis „SOMine“ programiniu paketu. Šiuo atveju tyrimui naudojami visi balsiai, t.y. neišskiriant nei vieno jų, su tikslu patikrinti į kelis klasterius suskils duomenų aibė ir ar išryškės moteriškų ir vyriškų pagrindinio tono dažnių ribos. Tai leis įvertinti kokio tikslumo yra pagrindinio tono nustatymo metodas, taikytas pagrindiniame eksperimente.

Tyrimui surinkti duomenys iš 50 diktorių vyrų ir 50 diktorių moterų balsių a, e, i, u, o. Duomenys suskaidyti į tris aibes: pradžios pagrindinis tonas, vidurio pagrindinis tonas, pabaigos vidurio tonas.

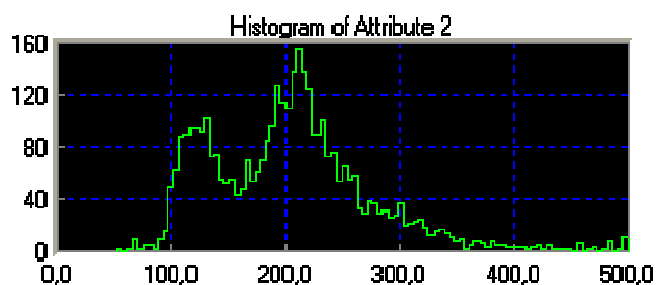
### Tyrimo eiga:

- Duomenys įkeliami į „SOMain“ aplinką;
- Sugeneruojamas klasterių žemėlapis.
- Dirbtinai keičiamas segmentų skaičius, siekiant gauti norimus rezultatus.

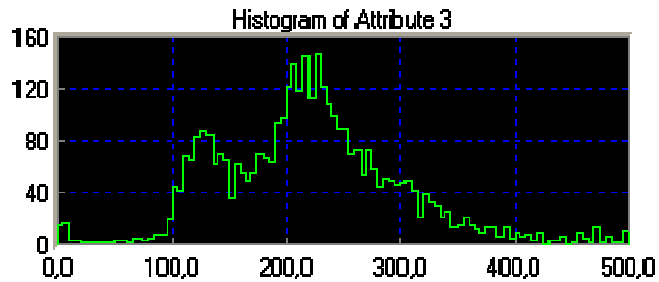
Duomenų aibės pasiskirstymas:



### 1. Pradžios pagrindinio tono pasiskirstymas aibėje



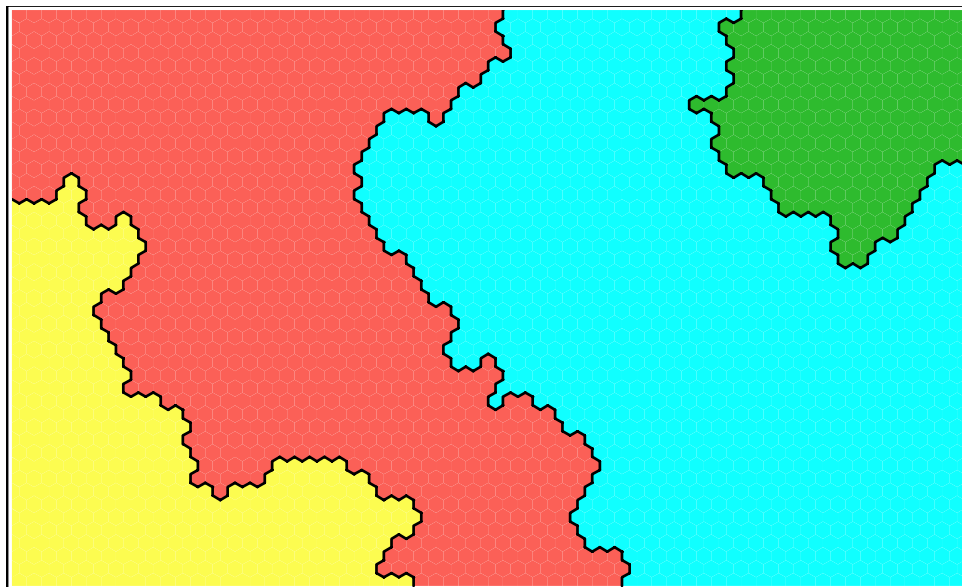
### 2. Vidurio pagrindinio tono pasiskirstymas aibėje



### 3. Pabaigos pagrindinio tono pasiskirstymas aibėje

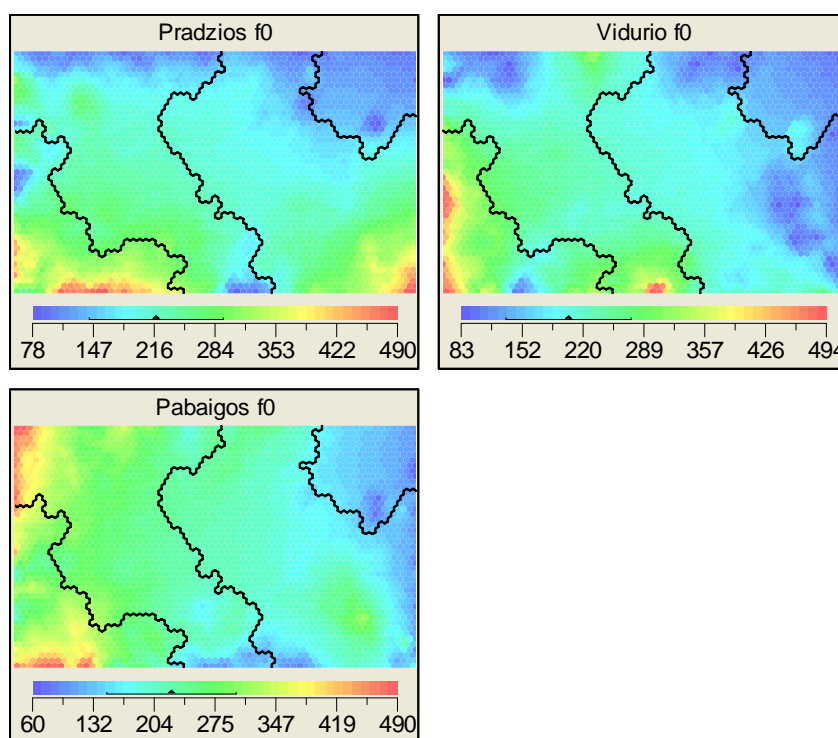
Pagal pateiktą duomenų pasiskirstymą nesunku išskirti dvi viršūnes, t.y. vieną vyriško pagrindinio tono dažnio ribose, kitą - moteriško pagrindinio tono dažnio ribose.

Pagal nutylėjimą, „SOMain“ sistema duomenis suskirstė į keturis segmentus (4.pav.).



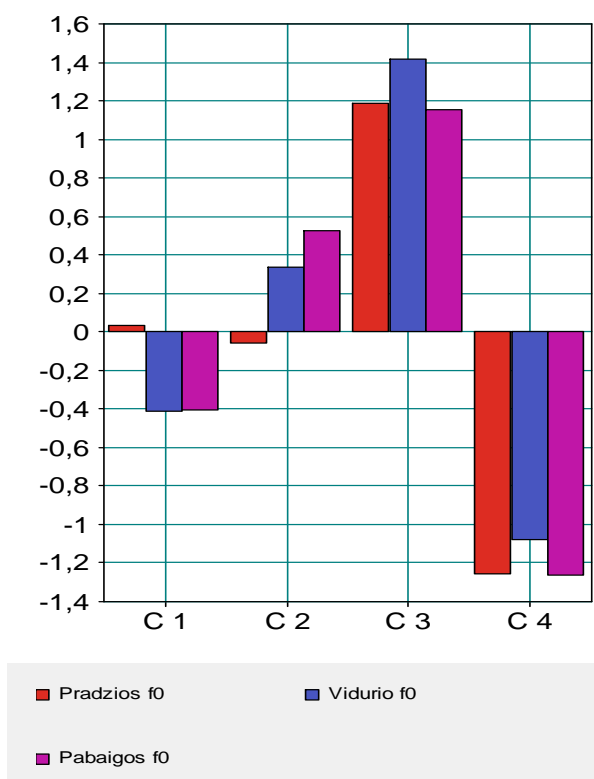
### 4. Duomenų klasterizacijos rezultatas

Kaip matome 4 paveiksle, duomenų klasterizacija puikiai atspindi pagrindinio tono pasiskirstymą. Pirmasis segmentas (geltona spalva) ir ketvirtasis segmentas (žalia spalva) yra metodo skaičiavimo nuokrypiai nuo vyriško ir moteriško pagrindinio tono. Antrasis segmentas (raudona spalva) yra vyriško pagrindinio tono duomenų sandauga, o trečiasis (mėlyna spalva) – moteriškojo. Čia reikėtų pastebėti, kad segmentai nėra itin tiksliai pasiskirstę pagal dažnio skalę. Vyriškas pagrindinis tonas nukrypsta į žemesnius dažnius, o moteriškas – į aukštesnius. Pati riba taip pat nėra itin tiksli. Tai matosi detalesniuose pasiskirstymo pagal duomenų požymius žemėlapiuose (5 pav.).



### 5. Klasterių žemėlapiai

Taip pat galime paanalizuoti klasterių atsiradimą lėmusių požymių įtakos lygį pagal pateiktą įtakos diagramą (6.pav.).



### 6. Požymių įtaka klasterių atsiradimui

## 7 PRIEDO TĘSINYS

Kaip matome įtakos diagramoje (6. Pav.) pirmojo segmento atsiradimą įtakuoja pradžios pagrindinis tonas. Taip ir turėtų būti, nes būtent pradžios pagrindinis tonas yra labiausiai veikiamas prieš tai einančio garso. Antrąjį segmentą labiausiai įtakuoja pabaigos pagrindinis tonas. Tai ganėtinai įdomus rezultatas. Trečiojo segmento atsiradimą įtakuoja vidurio pagrindinis tonas, o ketvirtojo pradžios ir pabaigos tonai. Ketvirtasis segmentas kaip ir pirmasis yra sąlygojamas nukrypimų nuo moteriškų ir vyriškų pagrindinio tono dažniu, tad dėsninga, kad būtent šį segmentą labiausiai įtakuoja ne vidurio pagrindinis tonas, o pradžios ir pabaigos pagrindiniai tonai.

### **Apibendrinimas**

Duomenų klasterizacija nors ir atspindi pagrindinio tono pasiskirstymą, reikėtų pastebėti, kad segmentai nėra itin tiksliai pasiskirstę pagal dažnio skalę. Vyriškas pagrindinis tonas nukrypsta į žemesnius dažnius, o moteriškas – į aukštesnius. Pati riba tarp vyriško pagrindinio tono ir moteriškojo taip pat nėra itin tiksli.

Duomenų klasterizacijos nukrypimus daugiausiai sąlygoja pradžios ir pabaigos pagrindinio tono klaidos, kadangi tiek balsio pradžia, tiek balsio pabaiga yra įtakojama gretimų garsų.