VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS INSTITUTAS

Gintautas TAMULEVIČIUS

PAVIENIŲ ŽODŽIŲ ATPAŽINIMO SISTEMŲ KŪRIMAS

Daktaro disertacija

Technologijos mokslai, informatikos inžinerija (07T)

VGTU
LEIDYKLA
TECHNIKA
Vilnius 2008

Disertacija rengta 2003–2008 metais Matematikos ir informatikos institute.

Mokslinis vadovas

Doc. dr. Antanas Leonas Lipeika (Matematikos ir informatikos institutas, technologijos mokslai, informatikos inžinerija – 07T).

http://leidykla.vgtu.lt
VGTU leidyklos TECHNIKA 1495-M mokslo literatūros knyga

ISBN 978-9955-28-279-2

Gintautas TAMULEVIČIUS

PAVIENIŲ ŽODŽIŲ ATPAŽINIMO SISTEMŲ KŪRIMAS

Daktaro disertacija
Technologijos mokslai, informatikos inžinerija (07T)

2008 05 10. 8,5 sp. l. Tiražas 20 egz.
Vilniaus Gedimino technikos universiteto
leidykla „Technika“, Saulėtekio al. 11, 10223 Vilnius
http://leidykla.vgtu.lt
Spausdino UAB „Baltijos kopija“
Kareivių g. 13B, 09109 Vilnius
http://www.kopija.lt

Reziumė

Disertacijoje nagrinėjamas pavienių žodžių atpažinimo klausimas – apžvelgiama atpažinimo sistemų raida, analizuojamos atpažinimo problemos, lyginami atpažinimo metodai, sprendžiami pavienių žodžių atpažinimo sistemos kūrimo klausimai.

Darbo metu sukurta pavienių žodžių atpažinimo ir segmentavimo sistema KAS (Kalbos Atpažinimas ir Segmentavimas). Žodžiams palyginti panaudotas dinaminio laiko skalės kraipymo metodas.

Sukurtas automatinis žodžio ribų nustatymo metodas. Žodžio ribos aptinkamos kaip kalbos signalo savybių pasikeitimo momentai. Atpažinimo procese panaudojus pasiūlytąjį žodžio ribų nustatymo metodą bei klasterizavimu paremtą mokymą, atpažinimo tikslumas išaugo 10–19 %, o tiesinės prognozės modelio analizė tikslumu beveik nenusileido tiesinės prognozės kepstro analizei.

Sistemoje taip pat realizuotas žodžių segmentavimas į garsus. Tam panaudotas kalbos signalo tiesinės prognozės modelio parametrų pasikeitimo momentų nustatymo principas. Sukurti du žodžių segmentavimo metodai, naudojantys skirtingus pasikeitimo momentų įvertinimo kriterijus: tikėtinumo funkcijos maksimizavimo ir prognozės klaidos minimizavimo. Eksperimentiniai tyrimai parodė, jog pastarasis metodas buvo atsparesnis triukšmui ir padarė beveik ketvirtadaliu mažiau klaidų nei pirmasis – 14,5 %.

Apjungus segmentavimo ir atpažinimo procedūras, realizuotas pavienių žodžių atpažinimas garsais. Atpažinimo garsais eksperimento metu gauti tokie rezultatai: pilnai atpažinta 5,4 % žodžių, 21,6 % žodžių buvo suklysta 1 garsu. Vidutiniškai kiekvienam žodžiui buvo atpažinta 57,8 % garsų, padaryta 2,5 sukeitimo ir 0,1 ištrynimo klaidų. Rezultatams pritaikius elementarų lingvistinį apdorojimą, teisingo atpažinimo lygis šoktelėjo iki 15,3 %.

Žodžių atpažinimo garsais idėjos realizacija neparodė gerų rezultatų, tačiau jie leidžia formuoti tiriamojo darbo kryptį – žodžių atpažinimą atkarpomis. Dirbant šia kryptimi reikėtų nagrinėti atpažinimo vienetų optimalumo, atpažinimo rezultatų lingvistinio apdorojimo klausimus, alternatyvius segmentavimo metodus, požymių sistemas.

Abstract

This dissertation analyzes isolated word recognition – surveys evolution of recognition systems, reviews recognition difficulties and problems, compares different recognition methods, deals with implementation of isolated word recognition system.

An isolated word recognition and segmentation system KAS was developed. System uses dynamic time warping method for word recognition.

Original word endpoints detection method is implemented in recognition system. Endpoints are detected as change moments of the linear prediction model of the speech signal. Endpoint detection method and clustering based learning were proposed for isolated word recognition. It induced 10 – 19 % recognition accuracy increase.

The word segmentation into phones method was proposed and implemented also. As in case of endpoint detection, segmentation is based on detection of change moments of the speech signal model. Two segmentation methods are implemented. The first one uses maximum likelihood criterion, the second is based on minimal prediction error estimate. Experimentally minimal prediction error method demonstrated higher performance and robustness to environmental noise – its error level was 14,5 % and was lower at a quarter than maximum likelihood method level.

Isolated word recognition in phones was implemented joining segmentation and recognition processes. Results of isolated word recognition in phones were following: 5,4 % of words were recognized without errors, 21,6 % with one error (one phone was misrecognized). Average amount of recognized phones in word was 57,8 % with 2,5 substitution error and 0,1 deletion error per word. After linguistic processing of recognition results (results were processed with spell-checker) correct recognition level reached 15,3 %.

Accuracy of recognition in phones wasn't high, but there are a lot of problems to solve – selection of recognition unit, lingvistic processing of recognition results, alternative segmentation and signal analysis methods.

Žymėjimai

Simboliai

$x(n)$	diskretinio laiko signalas
$\omega(n)$	diskretinio laiko lango funkcija
$X(\omega)$	tolydusis signalo spektras
$X(k)$	diskretusis signalo spektras
$C(k)$	signalo kepstras
c_k	k-asis kepstro koeficientas
A	tiesinės prognozės modelio parametrų rinkinys
p	tiesinės prognozės modelio eilė
a_k	k-asis tiesinės prognozės modelio koeficientas
b	tiesinės prognozės modelio stiprinimo koeficientas
u	pasikeitimo momentų rinkinys
u_i	i-asis pasikeitimo momentas
$\hat{u}$	pasikeitimo momentų įverčių rinkinys
$\hat{u}_i$	i-ojo pasikeitimo momento įvertis
$d(r, z)$	atstumas tarp vektorių r ir z
$\mathcal{D}_{RZ}$	atstumas tarp pavyzdžių vektorių sekų R ir Z

Santrumpos

DFT	diskrečioji Furjė transformacija
DLK	dinaminis laiko skalės kraipymas
DP	dinaminis programavimas
ES	energijos slenksčio metodas
GFT	greitoji Furjė transformacija

MK	prognozės klaidos minimizavimo metodas
NIR	neribota impulsinė charakteristika
PMM	paslėptieji Markovo modeliai
RIR	ribota impulsinė charakteristika
ST	santykis signalas-triukšmas
TF	tikėtinumo funkcijos maksimizavimo metodas
TPM	tiesinės prognozės modelis
TPMK	tiesinės prognozės modelio kepstras
$\overline{\text{TPMK}}$	tiesinės prognozės modelio kepstras su vidurkio atėmimu

Turinys

1

Įvadas

1.1. Tiriamoji problema

Darbe nagrinėjami pavienių žodžių atpažinimo klausimai – žodžio ribų nustatymas, etalonų kūrimas, jų įtaka atpažinimo tikslumui.

1.2. Temos aktualumas

Šiuolaikiniame pasaulyje vis aktualesniu tampa automatinio kalbos atpažinimo klausimas – kuriama vis daugiau diktavimo, balso operatorių, balsu valdomos paieškos ir navigacijos sistemų, o jų kūrimu užsiima stambiausi informacinių technologijų produktų gamintojai. Atpažinimo sistemų kūrimas reikalauja didelių laiko ir žmogiškojo darbo resursų, kas sąlygoja didelę tokių produktų savikainą. Todėl komerciniai produktai, kuriuose realizuojamas kalbos atpažinimas, kuriami tik didelėms rinkoms, t. y. plačiai paplitusioms kalboms. Tuo tarpu kalbos, vartojamos nedidelėse srityse, lieka be dėmesio. Toks gamintojų atsiribojimas nuo komerciškai nepatrauklių kalbų lemia tų kalbų atpažinimo tyrimų nykimą – tyrimai apsiriboja didžiosioms kalboms sukurtų metodų ir technologijų pritaikymu ir modifikavimu. Įvertinus tai, kad kalbos skiriasi savo fonetinėmis savybėmis, gramatika, tokie tyrimai anksčiau ar vėliau gali tapti neperspektyvūs, ir mokslo bei praktinės realizacijos požiūriu beverčiai.

Aukščiau pateikti teiginiai pilnai atitinka situaciją lietuvių kalboje. Komerciniu požiūriu lietuvių kalbos vartojimo sritis yra per maža, kad atkreiptų didžiųjų

gamintojų dėmesį, todėl lietuvių kalbos atpažinimo klausimus tenka spręsti patiems. Nors ir pasiūlyta keletas originalių sprendimų lietuvių kalbai atpažinti, tačiau patys tyrimai nevyksta labai aktyviai ir tai lemia kalbos atpažinimo atskirtį nuo šiuolaikinių komunikacijos ir informacijos technologijų.

Siekiant sumažinti atskirtį, padidinti praktinę kalbos atpažinimo reikšmę ir įtaką šiuolaikinėms technologijoms, būtina kurti technines ir metodines lietuvių kalbos atpažinimo tyrimo priemones, plėsti atpažinimo klausimų tyrimus, bandyti pritaikyti atpažinimo sprendimus praktiniams uždaviniams. Tuomet pabrėždami lietuvių kalbos prigimtį ir turtingumą, galėsime akcentuoti, kad viena iš seniausių kalbų savo technologijomis niekuo nenusileidžia didžiosioms pasaulio kalboms.

1.3. Darbo tikslas ir uždaviniai

Pagrindinis šio darbo tikslas yra pasiūlyti sprendimus, kurie leistų padidinti pavienių žodžių atpažinimo sistemos tikslumą bei efektyvumą nemodifikuojant naudojamo atpažinimo ir signalo analizės metodų. Siekiant tikslo buvo sprendžiami šie uždaviniai:

1. Pasiūlyti sprendimą žodžio ribų nustatymo stabilumui ir atsparumui triukšmui didinti.
2. Pasiūlyti žodžių etalonų sudarymo metodą, didinantį atpažinimo tikslumą ir efektyvumą.
3. Pasiūlyti žodžių segmentavimo į garsus metodą. Išnagrinėti žodžio garsų atpažinimo galimybę.
4. Realizuoti pasiūlytuosius metodus. Eksperimentiškai įvertinti žodžių ribų nustatymo ir etalonų kūrimo įtaką atpažinimo tikslumui, segmentavimo metodų tikslumą.

1.4. Darbo metodai ir priemonės

Teorinei analizei, praktinei realizacijai ir tyrimams panaudotos matematinės analizės, skaitmeninio signalų apdorojimo, atpažinimo teorijos žinios. Atpažinimo sistema realizuota C++ kalba, naudojant *Microsoft Visual Studio 6.0* programavimo aplinką.

1.5. Mokslinis naujumas

Disertacijoje pasiūlyta keletas sprendimų, didinančių pavienių žodžių atpažinimo tikslumą ir efektyvumą. Sukurtas automatinio žodžio ribų nustatymo metodas.

Metodas pasižymi atsparumu triukšmui, didesniu nei energijos slenksčio metodas, ir leidžia sumažinti atpažinimo klaidų, kylančių dėl klaidingų žodžio ribų, kiekį. Etalonams kurti pasiūlytas klasterizavimas, minimizuojantis vidutinį atstumą iki klasterių centrų. Sprendimo išskirtinumas – atstumai skaičiuojami naudojant dinaminį laiko skalės kraipymą. Toks etalonų kūrimas leidžia padidinti atpažinimo tikslumą su mažesniu etalonų kiekiu nei tiesioginis kūrimas.

Sukurti du žodžių segmentavimo į garsus metodai, grindžiami tikėtinumo funkcijos maksimizavimu ir prognozės klaidos minimizavimu. Abiejuose metoduose garsų ribos aptinkamos kaip kalbos signalo tiesinės prognozės modelio parametrų pasikeitimo momentai. Panaudojus prognozės klaidos minimizavimo metodą suformuluota ir realizuota žodžių atpažinimo garsais idėja. Žodis atpažįstamas 2 etapais: segmentuojamas į garsus, pastaruosius atpažįstant. Toks atpažinimas leido supaprastinti palyginimo procesą ir sumažinti reikalingų etalonų kiekį. Be to, idėja leidžia formuoti tolimesnio darbo kryptis: tobulinti segmentavimo metodą, taikyti alternatyvius kalbos signalo analizės ir klasifikavimo metodus.

1.6. Ginamieji teiginiai

1. Automatinis žodžio ribų nustatymo metodas leidžia sumažinti atpažinimo klaidų, kylančių dėl neteisingai nustatytų ribų, lygį.

2. Klasterizavimu pagrįstas etalonų kūrimas leidžia padidinti atpažinimo sistemos tikslumą su mažesniu etalonų skaičiumi.

3. Sukurtieji kalbos signalo segmentavimo metodai leidžia žodžių garsų ribas signale aptikti kaip signalo tiesinės prognozės modelio parametrų pasikeitimo momentus.

4. Atpažinimo sistemos eksperimentinio tyrimo rezultatai patvirtina pasiūlytųjų sprendimų ir metodų efektyvumą.

1.7. Praktinis taikymas

Disertacijoje suformuluoti pasiūlymai ir sukurti metodai buvo pritaikyti šiose praktinėse realizacijose:

- Žodžio ribų nustatymo ir klasterizavimu pagrįstas etalonų kūrimo metodas buvo panaudoti pavienių žodžių ir frazių atpažinimo sistemoje *Atpazinimas*, skirtoje atpažinimo procesui vizualizuoti ir analizuoti. Sistema įtraukta į 2000–2006 m. programos „Lietuvių kalba informacinėje visuomenėje“ automatinio lietuvių šnekos atpažinimo tiriamuosius darbus. Be to, pristatyta mokslo, inovacijų ir aukštųjų technologijų parodoje „Mokslas 2004“. Be to,

sistema kaip atpažinimo modulis naudota VGTU, VPU ir VDU bakalaurų ir magistrų baigiamuose darbuose, kaip techninė priemonė VGTU, VPU ir VDU laboratoriniuose darbuose.

- Sukurtieji segmentavimo metodai buvo panaudoti žodžių segmentavimo sistemoje *Segmentacija*, skirtoje segmentavimo procesui vizualizuoti ir analizuoti. Sistema įtraukta į 2000–2006 m. programos „Lietuvių kalba informacinėje visuomenėje“ automatinio lietuvių šnekos atpažinimo tiriamuosius darbus.
- Naudojant žodžio ribų nustatymo metodą, klasterizavimu pagrįstą mokymą ir dinaminį laiko skalės kraipymą sukurta sistema, leidžianti balsu valdyti interneto naršyklę, atverti interneto puslapius ir paleisti programas kompiuteryje. Sistema pristatyta informacinių technologijų parodoje „Infobalt 2007“.

1.8. Pranešimai konferencijose

Tarpiniai disertacijos darbo rezultatai buvo pristatyti šiose mokslo konferencijose:

1. Tarptautinėje konferencijoje „Elektronika“ 2004 ir 2005 m., Vilniuje.
2. Lietuvos matematikų draugijos XLIV konferencijoje 2003 m., Vilniuje.
3. Tarptautinėje konferencijoje „Human Language Technologies – The Baltic perspective“ 2004 m., Rygoje (Latvija).
4. Jaunųjų mokslininkų konferencijoje „Lietuva be mokslo – Lietuva be ateities“ 2004 m., Vilniuje.
5. Konferencijoje „Informacinės technologijos 2007“ 2007 m., Kaune.

1.9. Disertacijos struktūra ir turinys

Disertaciją sudaro: 6 skyriai, literatūros ir autoriaus publikacijų sąrašai bei vienas priedas. Disertacijos aiškinamąjį raštą sudaro 124 teksto puslapiai su 35 iliustracijomis ir 14 lentelių. Literatūros sąraše – 121 šaltinis.

Pirmajame skyriuje pristatoma darbo tema, darbo tikslas ir uždaviniai, ginamieji teiginiai bei darbo mokslinis naujumas.

Antrajame skyriuje pateikiama kalbos atpažinimo sistemos struktūra, nagrinėjami kalbos atpažinimo klausimai ir problemos. Apžvelgiama sistemų raida, darbai užsienyje ir Lietuvoje.

Trečiajame skyriuje nagrinėjami kalbos atpažinimo sistemų elementai – signalo analizės ir klasifikacijos metodai, jų privalumai ir trūkumai.

Ketvirtasis skyrius skirtas pavienių žodžių atpažinimo sistemos realizacijai. Sukuriami žodžio ribų nustatymo ir žodžių segmentavimo metodai, nagrinėjami sudarytųjų algoritmų efektyvumai. Etalonams kurti pritaikomas klasterizavimo principas. Apjungus segmentavimo ir pavienių žodžių atpažinimo metodus, suformuluojama žodžių atpažinimo garsais idėja, išdėstomos prielaidos idėjai realizuoti.

Penktajame skyriuje pateikiami sukurtosios pavienių žodžių sistemos eksperimentinio tyrimo rezultatai. Eksperimentais tirtas sukurtųjų žodžio ribų nustatymo, segmentavimo metodų darbingumas, mokymo įtaka atpažinimo tikslumui. Atliktas preliminarus žodžių atpažinimo garsais tyrimas. Suformuluojamos galimos atpažinimo garsais vystymo kryptys.

Šeštajame skyriuje suformuluojamos darbo išvados ir įvardijami ateities darbai vystant žodžių atpažinimą garsais.

Priede pateikiamas sistemai tirti naudotas žodynas.

2

Kalbos atpažinimo sistemos

Visa žmogaus techninė veikla yra nukreipta į įrenginių ir sistemų, imituojančių, pakeičiančių žmogaus fizinius ir protinius sugebėjimus, kūrimą. Naudodamas sukurtąsias priemones, žmogus stengiasi automatizuoti monotoniškus veiksmus, padidinti procesų efektyvumą. Tuo tikslu yra sukurtos priemonės atlikti konkrečias užduotis, judėti, matyti aplinką, fiksuoti ir įsiminti informaciją, priimti sprendimus ir pan. Natūralu, jog analogiškos sistemos pradėtos kurti ir kalbai – atpažinti bei sintezuoti. Kalbos atpažinimo sistemos leistų automatizuoti informacijos įvedimą ir vertimą į kitas kalbas, realizuoti balso sąsajas sistemose.

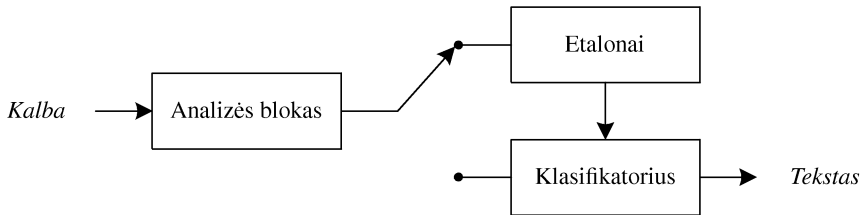
Nors priimta, kad kalbos atpažinimas vystomas jau 50 metų, pirmieji du dešimtmečiai nepasižymėjo didele technologijų vystymosi sparta. To priežastis – tuometinės skaičiavimo technikos ribotumas. Ir tik per paskutinius 20–30 metų, ištobulėjus skaičiavimo technikai, kalbos technologijos pradėjo vystytis itin sparčiai – buvo realizuojami ir tobulinami algoritmai, kuriami hibridiniai metodai, atliekami eksperimentai, kaupiami garsynai, kuriamos praktinės kalbos atpažinimo realizacijos realiems uždaviniams. Nepaisant visų dedamų pastangų ir lėšų, dabartinės atpažinimo sistemos nėra tobulos. Komercinių atpažinimo sistemų gamintojų deklaruojamas 98–99 % atpažinimo tikslumas yra pasiekiamas atskiriems kalbėtojams laboratorinėmis sąlygomis. Praktikoje atpažinimo tikslumas priklauso nuo kalbančiojo fizinės, emocinės, psichologinės būsenos, kalbėjimo manieros, naudojamos įrangos, vartojamos ir gimtosios kalbų ir netgi nedidelis šių faktorių pokytis sukelia atpažinimo tikslumo svyravimus. Kitas svarbus momentas – naujų idėjų nebuvimas. Pastarųjų dešimtmečių darbo rezultatas – tie patys, prieš 30–40 metų sukurti metodai, su daugybe patobulinimų ir parametrų kombinacijų, hibridiniai metodai,

apjungiantys tuos pačius klasikinius metodus.

Šiame skyriuje nagrinėsime kalbos atpažinimo sistemas. Suformulavę atpažinimo sistemos apibrėžimą ir pateikę supaprastintą jos struktūrą, atpažinimo sistemas suklasifikuosime pagal įvairius atpažinimo parametrus. Kalbos atpažinimo uždavinio specifikai išryškinti panagrinėsime sunkumus ir problemas, kylančias atpažįstant kalbą, bei galimus jų sprendimus. Atpažinimo metodų ir sistemų raidai nušviesti apžvelgsime užsienyje ir Lietuvoje atliktus darbus, išryškindami mūsų požiūriu svarbiausias idėjas, sprendimus ir metodus.

2.1. Kalbos atpažinimo sistema

Kalbos atpažinimo sistema – programinė arba aparatinė įranga, sugebanti pateiktąjį kalbos signalą sutapdinti su tekstu. Bendru atveju tekstas gali būti tiek galutinis atpažinimo rezultatas (pvz., kaip duomenys komandai, informacijai įvesti), tiek tarpinis (pvz., kaip duomenys lingvistiniam apdorojimui). 2.1 paveiksle pateikta elementarios atpažinimo sistemos, veikiančios akustiniame lygmenyje (lingvistinis apdorojimas nenagrinėjamas), struktūra.



2.1 pav. *Atpažinimo sistemos struktūra*

Kalbos signalo analizės tikslas – išskirti požymius – signalo charakteristikas, atspindinčias lingvistinį signalo turinį. Tokių požymių naudojimas leidžia sumažinti nagrinėjamų duomenų kiekį bei padidinti jų diskriminantinę galią (sugebėjimą atvaizduoti fonetinius skirtumus). Signalo analizę sudaro keletas etapų: signalo skaidymas į persidengiančius kadrus (kai kuriais atvejais signalas prieš skaidymą filtruojamas), dauginimas iš lango funkcijos, spektrinė analizė, reikiamų požymių išskyrimas. Analizės metu gautoji požymių vektorių seka panaudojama sistemai apmokyti arba žodžiui atpažinti. Apmokymo atveju požymiai ir jų atstovaujamo žodžio fonetinė transkripcija išsaugomi žodyno atmintyje kaip etaloniniai duomenys, kuriais remiantis bus atliekamas atpažinimas – toks procesas vadinamas etalono sukūrimu (bendru atveju duomenims taikoma apmokymo procedūra, kurios paskirtis – transformuoti analizės rezultatus į reikiamą struktūrą). Vykstant atpažinimo procesui išskirtieji požymiai klasifikuojami pagal jų atitikimą etalonams. Etalonas, kurio klasei buvo priskirtas nagrinėjamasis žodis, pateikiamas

kaip atpažinimo rezultatas (paprastai pateikiama etalono fonetinė transkripcija). Žodžių panašumo vertinimą, klasifikacijos kriterijų nulemia sistemoje realizuotas atpažinimo metodas.

Pirmąja kalbos atpažinimo sistema galima laikyti praėjusio amžiaus trečiajame dešimtmetyje pardavinėtą žaislą, pavadintą Radio Rex [14]. Iš celiuloido pagamintas šuo, ištarus vardą „Rex“, „iššokdavo“ iš savo būdos. Veikimas buvo paremtas žaislo pagrinde įtaisyto šunto jautrumu 500 Hz akustinei energijai (kuria pasižymėjo žodžio „Rex“ balsė). Šia energija paveikus šuntus, srovė, maitinanti pagrinde įtaisytą magnetą, nutrūkdavo ir šuo, veikiamas suspaustos spyruoklės, „išbėgdavo“ iš būdos. Savaime suprantama, kad Radio Rex reaguodavo ir į kitus žodžius (ar netgi atsitiktinius garsus), kurių spektre buvo pakankamo lygio 500 Hz dažnio dedamoji. Ir nors Radio Rex buvo labai paprasta sistema, jame buvo įgyvendintas šiuolaikinėse sistemose naudojamas atpažinimo principas: etaloninio garso savybės išsaugojamos ir lyginamos su nagrinėjamuoju garsu (šiuo atveju etaloninio garso savybes atitiko tam tikram dažniui jautrus šuntas). Tačiau Radio Rex buvo būdingas ir trūkumas, neišvengiamas kalbos atpažinimo sistemose – nesugebėjimas atmesti žodyne neesančius žodžius (arba tiesiog nesugebėjimas atskirti panašius garsus).

2.2. Kalbos atpažinimo sistemų tipai

Atpažinimo sistemų darbo režimas išreiškiamas keletu parametrų, nusakančių atpažinimo uždavinio sudėtingumą, turimų etalonų aibę ir sistemos priklausomybę nuo vartotojų. Skirtingose sistemų realizacijose ir jų taikymo srityse tie parametrai skiriasi.

Galima išskirti tris pagrindinius atpažinimo sistemų darbo režimo parametrus, kurių pagalba klasifikuojamos sistemos:

- **Atpažinimo objektas**. Pagal tai kokiame lingvistiniame lygmenyje vyksta kalbos atpažinimas, sistemos gali būti skirstomos į du tipus: pavienių žodžių atpažinimo ir ištisinės kalbos atpažinimo. Pirmuoju atveju nagrinėjamieji vienetai yra pavieniai žodžiai, atskirti tyla. Sistemos, nagrinėjančios kalbą be tylos intarpų, vadinamos ištisinės kalbos atpažintuvais. Ištisinė kalba vėlgi gali būti labai įvairi. Tai gali būti skaitymas balsu, sklandi kalba, spontaniška kalba ir pan. Toks kalbėjimas tarpusavyje skiriasi vartojamu žodynu, sklandumu, daromomis pauzėmis, tempo ir balso kitimu, beprasmių garsų (pvz., „mhm“) kiekiu. Todėl greta atskirų žodžių ir ištisinės kalbos atpažinimo sistemų išskiriamas spontaniškos šnekos atpažinimo sistemų tipas.
- **Žodynas**. Žodynu vadiname visumą skirtingų lingvistinių vienetų, kuriais yra apmokyta sistema (sukurti etalonai). Žodyno dydis yra pats svarbiausias ir dažniausiai naudojamas atpažinimo sistemos parametras, kadangi atpaži-

nimo klaida dažniausiai įvyksta bandant atpažinti žodyne neesantį pavyzdį. Pagal žodyno dydį kalbos atpažinimo sistemos skirstomos į mažo (šimtų etalonų eilės), vidutinio (tūkstančių eilės), didelio (dešimčių tūkstančių eilės) ir labai didelio (šimtų tūkstančių ir daugiau) žodyno sistemas.

- **Priklausomybė nuo kalbėtojo**. Pagal jautrumą kalbančiajam atpažinimo sistemos gali būti skirstomos į priklausomas nuo kalbančiojo ir nepriklausomas. Priklausomoje sistemoje gali būti nagrinėjami tik sistemą apmokiusiojo asmens ištarti žodžiai. Norint tokią sistemą panaudoti kito asmens kalbai atpažinti, tenka sistemą mokyti iš naujo, kito asmens ištartais žodžiais. Nepriklausoma nuo kalbančiojo sistema nereikalauja individualaus apmokymo ir skirtingų asmenų kalbai atpažinti naudoja tą patį etalonų rinkinį. Galima išskirti ir adaptyviąsias atpažinimo sistemas, kuriose kiekvienas naujas vartotojas privalo apmokyti sistemą tik pateikdamas savo balso pavyzdžius (individualaus etalonų rinkinio kurti nereikia). Šiuo požiūriu adaptyviosios atpažinimo sistemos yra tarpinės tarp priklausomų ir nepriklausomų nuo kalbančiojo sistemų.

Greta šių parametrų, atpažinimo sistemoms klasifikuoti taip pat naudojami tokie parametrai, kaip sistemoje realizuotas atpažinimo metodas, signalo analizės metodas, ar netgi atpažįstama kalba (pvz., anglų, lietuvių).

2.3. Sistemų raida

Nėra duomenų, kada buvo iškelta idėja realizuoti mašininį kalbos atpažinimą ar suformuluoti pradiniai teoriniai teiginiai. Darbų pradžia laikomas šeštasis dešimtmetis. Būtent tada pradėtos kurti fonemų, garsų ir žodžių atpažinimo sistemos. Paprastai atpažintuvai naudojo pavyzdžių palyginimo principą, kuomet nagrinėjamasis signalas lyginamas su etalonais ir panašiausias iš jų pateikiamas kaip atpažinimo rezultatas. Kaip požymiai buvo naudojamos formantės [16], juostinių filtrų požymiai [73, 88], laikinės spektrų savybės [88]. Taip pat buvo naudojami akustiniai-fonetiniai atpažinimo metodai [43], kuriuose nagrinėjamas signalas suskaidomas i segmentus su pastoviomis pasirinktomis charakteristikomis, jiems priskiriant atitinkamą tekstą. Fonetiniams vienetams charakterizuoti buvo parenkami požymiai, atspindintys akustines vienetų savybes – nosinumas, frikatyvumas, formančių išsidėstymas, garso vokalizuotumas, atitinkamų spektro juostų energija, pagrindinio tono dažnis [43, 88]. Išskirtiniai šio dešimtmečio techniniai sprendimai – atpažintuvo apmokymo [16] ir kalbos modelio panaudojimas [20].

Septintasis dešimtmetis pasižymėjo ypač svarbiais darbais kalbos technologijoms. 1965 m. buvo pristatytas greitosios Furjė transformacijos (GFT) algoritmas [12], po keleto metų A. V. Oppenheim su kolegomis kalbos signalui apdoroti pritaikė kepstro analizę [74]. Septintojo dešimtmečio pabaigoje – aštunto pradžioje kalbos signalo analizei pasiūlytas tiesinės prognozės modelis (TPM), tuo metu

sėkmingai naudotas kalbos signalui koduoti [3]. Šios trys signalo spektro skaičiavimo metodikos pakeitė iki tol dominavusį spektrinės analizės metodą – juostinius filtrus ir turėjo (bei tebeturi) didelę reikšmę kalbos atpažinime. Septintojo dešimtmečio pabaigoje T. K. Vinciuk pasiūlė kalbai atpažinti naudoti dinaminį programavimą [112]. Pritaikymo idėja buvo panaudoti dinaminį programavimą laiko skalei normuoti, taip išsprendžiant iki tol buvusią nevienodos lyginamų kalbos vienetų trukmės problemą. Beje, po poros metų tokią pačią idėją nepriklausomai nuo T. K. Vinciuk pasiūlė ir H. Sakoe bei S. Chiba [100]. Dinaminiu programavimu paremtas kalbos atpažinimo metodas – dinaminis laiko skalės kraipymas (DLK) – ilgą laiką buvo pagrindinis naudojamas metodas. Kaip ir šeštajame dešimtmetyje, dominavo akustiniai-fonetiniai, pavyzdžių palyginimo metodai, naudojantys juostinių filtrų požymius, spektrines kalbos savybes (pagrindinį toną, spektro gaubiamąsias ir pan.), koreliacinį panašumo įvertinimo būdą [21, 30, 92, 105]. Be minėtųjų metodų, kalbai atpažinti buvo bandoma panaudoti ir statistinius metodus [106].

1969 m. J. R. Pierce atvirame laiške [85] kritikavo to meto tyrėjų požiūrį į kalbos atpažinimą ir naudotus metodus. Jis teigė, jog žmogiškasis kalbos atpažinimo mechanizmas yra kur kas sudėtingesnis nei kalbos pavyzdžių akustinių parametrų palyginimas. Ir jau aštuntojo dešimtmečio pradžioje suformuluota „kalbos supratimo“ idėja. Buvo teigiama, jog kalbos atpažinimo klausimas, lyginant pavienius ištisinius žodžius, su laiku sudėtingės ir anksčiau ar vėliau taps sunkiai išsprendžiamu uždaviniu, todėl greta akustinio apdorojimo buvo akcentuojama lingvistinio apdorojimo būtinybė. 1971 m. JAV inicijuota penkerių metų tyrimų programa ARPA, kurios tikslas buvo sukurti nepriklausomą nuo kalbančiojo atpažinimo sistemą, atpažinimui naudojančią lingvistinį apdorojimą [56]. Programos metu sukurtose atpažinimo sistemose buvo naudojamos sintaksės ir semantikos žinios, fonetikos taisyklės [56, 60, 118]. Panašių idėjų buvo laikomasi ir kitose atpažinimo sistemose [93]. Dešimtmečio viduryje F. Itakura pasiūlė atstumo skaičiavimo metodiką tiesinės prognozės modelio parametrams (ir taip pritaikė tiesinį prognozės modelį kalbos atpažinimui) [46], o F. Jelinek ir J. K. Baker su kolegomis kalbai atpažinti pritaikė paslėptuosius Markovo modelius [6, 47], duodami pradžią statistiniam metodui, sėkmingai taikomam ir šiandien. Bendras aštuntojo dešimtmečio tyrimų bruožas – susidomėjimas nuo kalbančiojo nepriklausančiu atpažinimu ir sujungtos kalbos atpažinimu. Taip pat reikėtų paminėti, jog būtent šiame dešimtmetyje buvo sukurta pirmoji komercinė kalbos atpažinimo sistema, skirta komandoms ir duomenims įvesti [41].

Devintajam dešimtmečiui būdingas susidomėjimas ištisinės kalbos atpažinimu ir statistinio paslėptųjų Markovo modelių (PMM) metodo išpopuliarėjimas. PMM pradėti taikyti tiek fonemoms, tiek pavieniams žodžiams, tiek ištisinei kalbai atpažinti [48, 59, 83, 84, 103, 116] ir kartu su dinaminiu laiko skalės kraipymo metodu tapo plačiausiai naudojamais metodais kalbai atpažinti (per sekančius dešimtmečius PMM galutinai įsitvirtins kaip pagrindinis metodas kalbai atpažin-

ti). Dešimtmečio pabaigoje sukurta paslėptuosius Markovo modelius naudojanti atpažinimo modeliavimo sistema HTK [120], labiausiai paplitusi priemonė statistiniam kalbos atpažinimui modeliuoti ir sėkmingai taikoma iki šiol. Greta minėtųjų kalbos atpažinimo metodų pradėti taikyti neuronų tinklai [9, 31, 115] (nors neuronų tinklai buvo pasiūlyti jau šeštajame dešimtmetyje, jie neprigijo dėl praktinės realizacijos sunkumų). Laikui bėgant bandyta apjungti neuronų tinklus su PMM (ar DLK), tačiau ypatingų rezultatų nepasiekta. Nagrinėjant atpažinimo atsparumą kalbėtojui ir triukšmams buvo pasiūlytos naujos požymių sistemos: melų skalės kepstro koeficientai [17], tiesinės suvokimo prognozės koeficientai [38], dinaminiai kepstro požymiai [28], taip pat kalbos signalo analizės metodai, besiremiantys žmogaus klausos modeliu [109]. Devintojo dešimtmečio viduryje pradėta rinkta pirmoji akustinių duomenų bazė TIMIT [32] (taip buvo išspręsta problema dėl nevienodų duomenų naudojimo atpažinimo sistemose).

Paskutiniajame XX a. ir pirmajame XXI a. dešimtmečiuose pagrindinis dėmesys sutelkiamas ties kalbos atpažinimo atsparumo bei ištisinės spontaniškos kalbos atpažinimo klausimais. Atpažinimo tikslumui didinti pasiūloma naudoti tarties modelius [94], dialogo modelius [29], akustinio ir kalbos modelių adaptavimą kalbančiajam, netiesinius kalbos signalo apdorojimo metodus [24]. Spontaniškos kalbos pertekliškumui kompensuoti, atpažinimo procese siūloma naudoti automatinį kalbos apibendrinimą [42]. Greta naudojamų sintaksinių, semantinių, gramatinių žinių į kalbos atpažinimą įjungiamas kalbos „skaitymas“. Kalbos „skaitymo“ idėja – fiksuoti kalbančiojo veido išraišką ir akustinę kalbos signalo analizę papildyti veido išraiškos analizės rezultatais (pvz., kaip tariamo garso požymį panaudoti lūpų padėties parametrus). Prieš keletą metų pradėti kurti žymėjimo kalbų (HTML, XHTML) išplėtimai VoiceXML ir SALT. Šie sprendimai perkelia kalbos atpažinimą į objektinį lygmenį ir leidžia atpažinimo modulius įjungti į vartotojo sąsajas, visiškai nesigilinant į patį atpažinimo procesą. Tai turėtų paskatinti kalbos technologijų diegimą į komunikacijos sistemas, palengvinti vartotojų sąsajų, naudojančių kalbos technologijas, interaktyvių balso sistemų kūrimą.

2.4. Darbai Lietuvoje

Lietuvoje kalbos atpažinimo darbai pradėti aštuntojo dešimtmečio pabaigoje – devintojo pradžioje, taigi atsiliekant nuo pasaulinių darbų pora dešimtmečių.

Pirmosiose kalbos atpažinimo sistemose buvo naudojamas dinaminio laiko skalės kraipymo metodas, realizuotas naudojant dinaminį programavimą. Kaip požymiai buvo naudojami juostinių filtrų požymiai [80, 81], vėliau pereita prie tiesinės prognozės modelio [55, 63], tiesinės prognozės kepstro koeficientų [A5]. Dinaminis laiko skalės kraipymo metodas ypač populiarus buvo paskutiniaisiais praėjusio amžiaus dešimtmečiais ir taikomas atskirai tariamų žodžių sistemose dar ir šiomis dienomis. Devintojo dešimtmečio viduryje kalbai atpažinti pritaikyti paslėptieji Markovo modeliai [54]. Tuo metu pasaulyje pripažintas kaip efektyviau-

sias ir plačiausiai vartotas atpažinimo metodas, Lietuvoje PMM įsitvirtino tiktai dešimtojo dešimtmečio pabaigoje. Signalui analizuoti buvo naudojama tiesinės prognozės, kepstrinė analizės [58], tačiau didžiausias atpažinimo tikslumas pasiektas naudojant melų skalės kepstro požymių sistemą [51]. Jau šio amžiaus pradžioje kalbai atpažinti pritaikyti dirbtiniai neuronų tinklai [25], kurie taip ir neprigijo kalbos atpažinime. Bandyta kurti ir hibridines atpažinimo sistemas jungiant PMM su DLK [55], su neuronų tinklais [26], tačiau ypatingų rezultatų nepasiekta. Greta tradicinių metodų pasiūlyta ir keletas originalių sprendimų: dichotominis klasifikatorius (atliekantis dvinarį dalijimą) [22], dinaminio laiko skalės kraipymo modifikacija – projekcijų metodas [96], žemos ir aukštos eilių nulio kirtimų požymiai (nulio kirtimų parametras, skaičiuojamas įvairios eilės skirtuminiams signalams) [50], fonemų klasifikacija naudojant suderintąją diskriminantinę analizę [98]. Kaupiant medžiagą eksperimentiniams tyrimams ir siekiant juos standartizuoti prieš keletą metų pradėti kurti lietuvių kalbos garsynai [91, 97]. Pastaruoju metu susidomėta kalbos „skaitymo“ pritaikymu lietuvių kalbos atpažinime [7, 52]. Tikimasi, kad kalbančiojo veido analizės panaudojimas leis padidinti atpažinimo tikslumą.

Šiuo metu kalbos atpažinimo klausimai sprendžiami Matematikos ir informatikos institute (Vilnius), Vytauto didžiojo universitete (Kaunas) bei Kauno technologijos universitete. Pagrindinis dėmesys skiriamas ištisinės kalbos atpažinimui taikant PMM (kuriami kalbos modeliai, atliekami eksperimentai), kalbos duomenų bazėms (kaupiamos pavienių žodžių ir ištisinės kalbos duomenų bazės) bei bazių kaupimo automatizavimui. Ir nors kalbos atpažinimo Lietuvoje pažanga akivaizdi, atotrūkis nuo išsivysčiusių valstybių pasiekimų, nors ir mažėjantis, visgi išlieka.

2.5. Atpažinimo problemos

Nors kalbos atpažinimas vystomas jau penkis dešimtmečius, sukurta nemažai atpažinimo metodų, realizuota žmogaus veikloje taikomų atpažinimo sistemų, vis dar kyla sunkumų realizuojant visiškai nuo kalbančiojo nepriklausomas, triukšmui atsparias, tikslias atpažinimo sistemas. Visus šiuos sunkumus galima įvardinti keturiomis problemomis.

Pirmoji problema – tai kalbos signalo kintamumas, reiškiantis, kad neįmanoma realizuoti dviejų visiškai vienodų, to paties lingvistinio vieneto pavyzdžių. Du to paties žodžio ištarimai tarpusavyje skirsis tempu, energijos lygiu, kitomis laikinėmis ar spektrinėmis savybėmis. Išskiriami du kalbos kintamumo tipai – vidinis kintamumas ir išorinis. Vidinis kintamumas pasireiškia to paties kalbančiojo kalbos nepastovumu. Viena iš šio nepastovumo priežasčių – kalbėjimo maniera. Kalbantysis savo mintis gali išreikšti pakeltu tonu ar net rėkdamas, šnabždėdamas, bandydamas paslėpti savo akcentą, ir t. t. Be to, įtakos turi ir subjektyvūs veiksniai – kalbančiojo laikysena, nuotaika, sveikatos būklė, amžius, pokalbio tematika. Dėl šių priežasčių netgi to paties asmens, vienas paskui kitą ištarti žodžiai tarpusavyje akustiškai skirsis ir tie skirtumai ilgėjant laikotarpiui tarp ištarimų di-

dės. Taip pat reikėtų pažymėti, jog vidinis kintamumas labiau pasireiškia ištisinėje kalboje, kadangi prie išvardintųjų vidinio kintamumo priežasčių prisideda natūralios kalbos savybės (koartikuliacija, įsiterpiantys beprasmiai garsai, kalbos tempo varijavimas ir pan.). Kur kas didesni akustiniai skirtumai atsiranda dėl išorinio kalbos kintamumo tarp skirtingų kalbančiųjų. Kiekvienas kalbantysis pasižymi individualiomis balso trakto fiziologinėmis ir akustinėmis savybėmis, kalbėjimo maniera, todėl neįmanoma rasti dviejų žmonių, kurių sugeneruoti kalbos signalai būtų identiški. Ypač akustiniai skirtumai išryškėja tarp skirtingų lyčių, skirtingo amžiaus kalbančiųjų. Kalbos kintamumo problema gali būti sprendžiama dviem būdais. Pirmasis – kalbančiojo adaptacija. Šios procedūros metu konkretaus kalbančiojo ištarti žodžiai panaudojami akustiniams skirtumams tarp to kalbančiojo ir žodyno modelių įvertinti. Antrasis būdas – kalbančiajam atsparios požymių sistemos naudojimas. Tokie požymiai idealiu atveju turėtų atspindėti tik fonetinį kalbos signalo turinį ir visiškai nevertinti akustinių kalbančiojo savybių.

Antroji problema – natūralios kalbos savybės. Natūraliai kalbai būdingi reiškiniai, kuriuos žmogus suvokia ir apdoroja net jų nepastebėdamas, tuo tarpu realizuojant automatines atpažinimo sistemas tie reiškiniai sukelia sunkumų. Visų pirma tai koartikuliacija – gretimų garsų susiliejimas, susidarantis dėl tolydaus balso trakto persiformavimo sekančiam garsui generuoti. Susilieję garsai tampa sunkiai atskiriami ar netgi įgyja visiškai kito fonetinio vieneto skambesį (pvz., žodį „čia“ girdime kaip „če“ ir teisingai mums jį parašyti padeda tik gramatikos žinios). Natūraliai kalbai taip pat būdingi nelingvistiniai garsai (pvz., abejojimo garsas „mmmm“, kostelėjimas), kurie gali užpildyti pauzes, įsiterpti į žodį ar netgi nutraukti jį. Žmogaus suvokimo sistema šiuos garsus lengvai išskiria kaip nelingvistinius, tuo tarpu atpažinimo sistema juos gali suprasti kaip žodį ar jo dalį (ypač jei tą rodo pvz., akustinės analizės rezultatai). Kai kuriais atvejais gali būti aktualus ribų tarp žodžių ištisinėje kalboje nebuvimo klausimas. Pavyzdžiui žodžių išskyrimo iš kalbos atveju, tikslus žodžio ar jų junginio ribų nustatymas turi lemiamą reikšmę atpažinimo rezultatų tikslumui ir netgi nedidelis netikslumas nustatant ribas gali baigtis atpažinimo klaida. Visos šios problemos turėtų būti sprendžiamos lingvistiniame lygmenyje – naudojami kalbos modeliai, taikomos papildomos gramatikos, prozodikos, semantikos, pragmatikos žinios, t. y. greta akustinio kalbos signalo apdorojimo atsiranda lingvistinio apdorojimo poreikis. Be to, natūralioje kalboje galima išskirti ir aukštesnio lygio signalo kitimus – kalbėjimo tempo, intonacijos kitimus, kuriuose taip pat yra informacijos apie akustines signalo savybes, todėl be akustinių segmentų reikėtų nagrinėti ir aukštesnio lygio segmentus, naudoti dinamines požymių sistemas.

Atpažinimo sistemų žodynai yra dar vienas atpažinimo problemų šaltinis. Dideli žodynai yra painūs – juose yra daug akustiškai panašių pavyzdžių. Ir tas painumas auga kartu su žodyno dydžiu – kuo didesnis žodynas, tuo daugiau akustiškai panašių pavyzdžių, tuo didesnė atpažinimo klaidos tikimybė (10000 žodžių žodyne bus daugiau panašiai skambančių žodžių nei 100 žodžių žodyne). Kai kurie ty-

rėjai teigia, jog kalbos atpažinimo uždavinio sunkumas auga logaritmiškai didėjant žodyno dydžiui [18]. Vienas iš galimų šios problemos sprendimų būdų – kontekstinio (t. y. skirto konkrečiai dalykinei sričiai) žodyno naudojimas. Tačiau toks problemos sprendimas kartu yra ir atpažinimo sistemos apribojimas (jos darbingumas tampa priklausomas nuo konteksto). Kitas galimas problemos sprendimo būdas – didele diskriminantine galia pasižyminčių požymių naudojimas. Dar sunkiau sprendžiama žodyne neesančių žodžių problema. Bet kuri sistema anksčiau ar vėliau susiduria su žodyne neesančiu žodžiu. Tokiu atveju galimi du sprendimai – atmesti žodį kaip neatpažintą arba įtraukti į sistemos žodyną. Antrasis sprendimas sukelia dar aibę sunkiai atsakomų klausimų – kaip garantuoti, kad neatpažintasis pavyzdys yra lingvistiškai prasmingas, kaip sugeneruoti reikalingą transkripciją, kaip atskirti pavyzdį nuo pašalinių triukšmų ir pan. Kol kas nėra pasiūlyta efektyvios procedūros šiems klausimams spręsti, todėl dažnai neatpažintasis pavyzdys tiesiog ignoruojamas.

Ketvirtoji problema – signalo akustinės ir sklidimo aplinkos įtaka signalui. Bet kuris signalo generavimo, sklidimo, priėmimo etape esantis triukšmas gali įtakoti ir būtinai įtakos signalą. Triukšmo šaltiniais gali būti pats kalbantysis (iškvėpimo triukšmas, kalbos padargų mechaniniai triukšmai), sklidimo aplinkos (foninis aplinkos triukšmas, aidas), įvedimo įrenginys (mikrofono elektriniai triukšmai, netiesiniai iškraipymai), perdavimo kanalas (atspindžiai, kanalo netiesiniai iškraipymai), priėmimo įrenginys (elektriniai triukšmai, netiesiniai iškraipymai, kvantavimo triukšmas). Visų šių poveikių rezultatas – užtriukšmintas signalas, lengvai suprantamas žmogui ir kartais visiškai nepriimtinas techninei sistemai. Be to, kiekvienas įrenginys pasižymi savo individualiomis spektrinėmis charakteristikomis (pvz., ribota pralaidumo juosta), kurios taip pat turi įtakos apdorojamam signalui, todėl skirtingų techninių priemonių (skirtingos paskirties, įvairių gamintojų) poveikis signalui yra nevienodas. Sistema apmokyta su vieno tipo mikrofonu (ir puikiai su juo veikianti) gali visiškai prarasti savo savybes pakeitus mikrofoną kitu (šiuo atveju sistemos darbingumas yra priklausomas nuo įrangos). Ši problema turėtų būti sprendžiama ieškant triukšmams atsparių požymių sistemų.

Apibendrinant reiktų pasakyti, kad ne vien dėl išvardintų problemų automatinės atpažinimo sistemos neprilygsta žmogiškajam kalbos suvokimui. Visgi dar nėra perprastas žmogaus kalbos generavimo ir suvokimo mechanizmas, dėl ko ir kyla sunkumai automatizuojant kalbos atpažinimą. Aišku tik tai, kad žmogus verbaliniame bendravime neapsiriboja akustine analize, o panaudoja ir fonetikos, fonologijos, leksikos, sintaksės, prozodikos, semantikos, pragmatikos žinias, pokalbio konteksto duomenis, papildomą informaciją, perduodamą gestais, mimika, laikysena, galbūt net intuiciją ir kitus informacijos šaltinius, kurių technikoje mes negalime realizuoti dėl nežinojimo, sudėtingumo ir savo pačių išankstinių prielaidų.

2.6. Antrojo skyriaus apibendrinimas

- Kalbos technologijų pritaikymo galimybių gausa lemia didelę ir staigiai augančią kalbos atpažinimo technologijų sprendimų poreikį.
- Darbai kalbos atpažinimo srityje Lietuvoje nuo pasaulinių atsilieka maždaug 20–30 metų.
- Kalbos signalo kintamumas, natūralios kalbos savybės (žodyno dydis, koartikuliacijos reiškiniai, netaisyklinga tartis), aplinkos fizinis poveikis signalui daro kalbos atpažinimą sunkiu uždaviniu.
- Nors kalbos atpažinimo klausimas nagrinėjamas jau pusė amžiaus, dar nėra sukurta patikimo, efektyvaus, nepriklausomo nuo kalbėtojo metodo kalbai atpažinti.

3

Kalbos atpažinimo sistemų analizė

Šiame skyriuje smulkiau panagrinėsime atpažinimo sistemos elementus. Skyrių sudaro dvi dalys. Pirmojoje trumpai pristatomas kalbos generavimo modelis bei nagrinėjami kalbos signalo analizės metodai. Antrojoje dalyje nagrinėjamas kalbos signalo klasifikacijos klausimas.

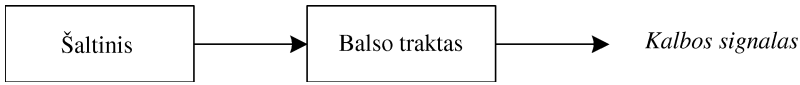
3.1. Kalbos signalas

3.1.1. Kalbos signalo generavimas

Kalbos signalas – akustinis signalas, generuojamas žmogaus kalbos padargais ir skirtas perduoti verbalinę informaciją. Akustinio kalbos signalo generavimo procesą sudaro trys etapai: šaltinio generavimas, artikuliacija ir išspinduliavimas. Generavimo metu iš plaučių išstumiamas oras eina per balso stygas ir per ryklę patenka į burnos ir nosies ertmes, kuriose vyksta artikuliacija. Susidarę akustiniai virpesiai ir oro srautai per lūpas bei nosį išspinduliuojami į aplinką. Struktūriškai kalbos signalo generavimo procesą galima atvaizduoti „šaltinis-sistema“ tipo schema (3.1 pav.).

Šaltinio (plaučių ir balso stygų) generuojamas signalas gali būti dviejų tipų: kvaziperiodinis, charakterizuojamas pagrindiniu tonu (dažniu) arba aperiodinis, t. y. triukšmas. Sugeneruotasis signalas patenka į balso traktą, apimantį ryklę ir burnos ertmę, ir į nosies traktą – nosies ertmes (priklausomai nuo tariamo garso). Balso traktas pasižymi rezonansinėmis savybėmis ir priklausomai nuo jo konfigū-

racijos (liežuvio, lūpų padėties, burnos ertmės tūrio, praeinamumo tarp ryklės ir nosies bei burnos ertmių, dantų sukandimo ir t. t.) tos savybės kinta. Toks balso trakto spektrinių savybių kitimas laike formuoja šaltinio signalo spektrą, taip gaunant skirtingus garsus bei jų sekas – kalbos signalą. Balso trakto konfigūracijos kitimas siekiant išgauti garsus, vadinamas artikuliacija, o balso trakto organų (dar vadinamų artikuliatoriais) judesiai – artikuliaciniais judesiais. Rezonansiniai balso trakto dažniai vadinami formantėmis.

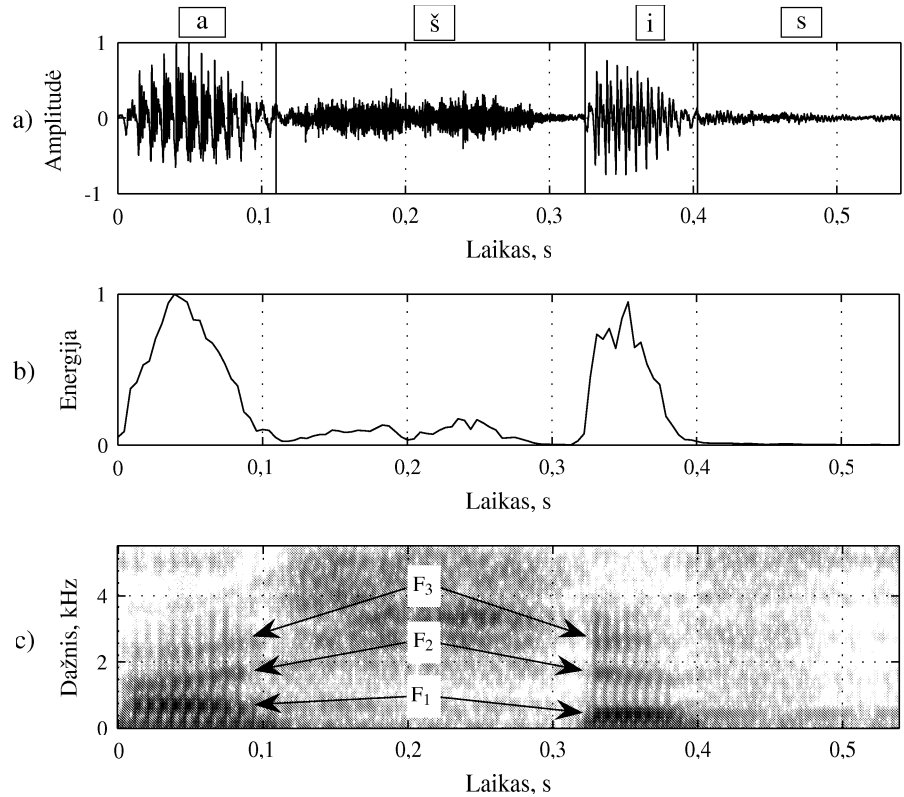


3.1 pav. *Kalbos signalo generavimo schema*

Priklausomai nuo pasirinkto kriterijaus galima įvairi garsų klasifikacija. Akustinėje klasifikacijoje dažniausiai naudojamas šaltinio signalo tipas. Pagal jį garsai skirstomi į vokalizuotus (sužadinantis šaltinio signalas yra kvaziperiodinis) ir nevokalizuotus (šaltinio signalas atsitiktinis), priebalsiai – į skardžiuosius (kvaziperiodinis) ir dusliuosius (atsitiktinis). Artikuliacinėje klasifikacijoje pagal balso trakto uždarumą išskiriami balsiai ir priebalsiai, pagal oro skverbimosi būdus – sprogstamieji, pučiamieji ir sklandieji priebalsiai, pagal minkštumą – minkštieji ir kietieji priebalsiai, pagal lūpų veiklą – lūpiniai ir nelūpiniai balsiai, pagal trukmę – trumpieji ir ilgieji balsiai. Be to, dar galimi tokie artikuliacijos požymiai kaip balsių pakilimas, priebalsių artikuliacijos vieta [75].

3.1.2. Kalbos signalo savybės

Esminė kalbos signalo savybė – nestacionarumas, susidarantis dėl nuolatinio kalbos padargų judėjimo. Kadangi kalbos padargai yra pakankamai inertiški (visgi kalbos generavimas nėra vienintelė ar pagrindinė jų funkcija), tariant žodžius vyksta nuolatinis balso trakto formos kitimas nuo konfigūracijos vienam garsui tarti link sekančio garso konfigūracijos. Todėl kalba gaunasi ne diskreti garsų seka, o nuo vieno garso momentinių savybių iki kito tolydžiai kintantis signalas. 3.2 a) paveiksle pateikta žodžio „ašis“ laiko diagrama su rankiniu būdu išskirtomis garsų ribomis. Paveiksle matome, jog visų garsai savo ribose kinta ir sandūrose gretimi garsai beveik suvienodėja. Diagramoje taip pat matyti skirtumas tarp vokalizuotų („a“, „i“) ir nevokalizuotų („š“ bei „s“) garsų. Pirmųjų diagramoje galima įžvelgti periodinę struktūrą – pagrindinį toną. Pagrindinio tono dažnis yra individuali charakteristika ir priklauso nuo balso stygų parametrų – gali kisti nuo 50 Hz vyrams iki 400–500 Hz moterims ir vaikams. Nevokalizuoti garsai savo struktūra primena triukšmą. Be to, 3.2 b) paveiksle galime pamatyti, kad vokalizuoti garsai pasižymi didesne energija nei nevokalizuoti.



3.2 pav. *Kalbos signalo diagramos: a) laiko; b) energijos; c) spektrograma su nurodytomis pirmosiomis trejomis formantėmis*

Kalbos signalo spektras siekia 7–8 kHz (spektro plotį lemia žadinančiojo signalo spektras). Paprastai kalbos apdorojimo uždaviniams naudojamo signalo spektro plotis siekia 5 kHz, o telefonijoje – 3,4 kHz. Vokalizuotų garsų spektre galima išskirti nuo 2 iki 4 formančių, kurių reikšmės ir tarpusavio padėtis priklauso nuo kalbančiojo asmens, t. y. nuo balso trakto ilgio ir konfigūracijos, kuri kiekvienam garsui yra skirtinga. Paprastai nagrinėjamos pirmosios trys formantės. Pirmosios formantės reikšmė būna 250–900 Hz, antrosios – 400–2400 Hz, trečiosios – 2500–3000 Hz diapazone. Nevokalizuotų garsų spektre ryškesnių formančių išskirti negalima. 3.2 c) paveiksle pateiktoje žodžio spektrogramoje (spektrogramos tamsumas proporcingas spektro intensyvumui) rodyklėmis nurodytos pirmųjų trijų formančių trajektorijos. Iliustracijoje matyti, kad garso ribose formančių reikšmės kinta, t. y. kinta balso trakto konfigūracija. Be to, vokalizuotų garsų energija

susitelkusi žemesniuose dažniuose, tuo tarpu nevokalizuotų garsų energija pasiskirsčiusi daugmaž tolygiai per visą spektro plotį (ypač tai akivaizdu garso „s“ spektre).

3.2. Kalbos signalo analizė

Kalbos signalas yra perteklinis – be lingvistinės informacijos (tai ką kalbantysis nori pasakyti) signale yra informacija apie kalbantįjį, kalbėjimo aplinką, signalo sklidimo kanalą ir t. t. Kalbos atpažinimo atveju mus domina tik lingvistinė informacija, nepriklausomai nuo to, kas ir kokioje aplinkoje kalbėjo. Todėl pirmasis etapas kalbos atpažinimo procese – kalbos signalo analizė (2.1 pav). Galima išskirti 3 pagrindinius analizės tikslus:

- **Duomenų apimties sumažinimas**. Priklausomai nuo skaitmeninio signalo diskretizacijos dažnio, 1 s trukmės įrašą gali sudaryti nuo kelių iki keliolikos tūkstančių atskaitų. Tokio duomenų kiekio apdorojimas gali pareikalauti nemažai laiko, kuris kai kuriais atvejais (pavyzdžiui realaus laiko sistemose) yra labai svarbus. Atlikus analizę paprastai gaunama keletą ar net kelioliką kartų mažesnis informacijos kiekis.

- **Charakteringų duomenų išskyrimas**. Kadangi kalbos atpažinimo tikslas – lingvistinė informacija, vienas iš analizės tikslų – išskirti tas kalbos signalo laikines ar dažnines savybes, kurios labiausiai atspindi fonetinį signalo turinį. Tuo pačiu išskiriami duomenys turi būti invariantiški kalbančiojo ir aplinkos atžvilgiu, t. y. visiškai neturėti informacijos apie kalbantįjį ir aplinką.

- **Diskriminantinės galios didinimas**. Svarbus reikalavimas analizei – išskiriamų duomenų diskriminantinė galia – gebėjimas atskirti skirtingus garsus. Išskiriami duomenys turi būti kuo panašesni sutampantiems garsams ir skirtis skirtingiems garsams.

Kalbos signalas yra nestacionarus – nuolat kinta. Dėl to analizuojant kalbos signalą daroma prielaida, kad tie pokyčiai nėra labai greiti, t. y. tam tikrame laiko intervale šaltinio signalas ir balso trakto forma nekinta arba kinta labai nežymiai – kalbos signalas yra stacionarus. Tuo tikslu signalas skaidomas į 20–40 ms trukmės persidengiančias atkarpas (kadrus) analizės langą perstumiant per 5–10 ms. Kiekviename kadre išskiriamas duomenų vektorius, taip vadinamas požymių vektorius. Analizės kadrų persidengimas garantuoja, kad gautoji požymių vektorių seka atspindės signalo savybių kitimą. Analizės lango tipą (ilgį, formą), išskiriamų požymių vektorių savybes (ilgį, duomenų tipą) nulemia naudojamas analizės metodas.

Tam tikrais atvejais kalbos signalas prieš analizę gali būti apdorojamas – keičiamas diskretizacijos dažnis, filtruojamas ir pan.

Kalbos signalui analizuoti sukurta labai daug metodų. Juose nagrinėjamos kalbos signalo amplitudės reikšmės, trumpalaikės energijos įverčiai, įvairios signalo transformacijos, statistinės charakteristikos, artikuliacinės savybės, netgi taikomi chaoso teorijos elementai. Kiekvienas metodas atspindi savo laikmečio technines galimybes, metodines priemones bei kalbos atpažinimo problemos suvokimą.

Galima įvairi analizės metodikų klasifikacija. Metodikos požiūriu kalbos analizės metodai gali būti skirstomi į parametrinius ir neparametrinius. Parametriniuose metoduose kalbos signalas modeliuojamas modeliu, o analizės tikslas yra to modelio parametrų nustatymas. Prie parametrinių analizės metodų priskiriama tiesinės prognozės modelis, spektrinių porų požymiai, atspindžio koeficientai, tiesinės suvokimo prognozės modelis. Neparametrinė analizė vykdoma transformuojant kalbos signalo reikšmes. Jiems priskiriama spektrinė ir kepstrinė analizės, koreliacinė (autokoreliacinė) analizė, nulio kirtimų analizė, juostinių filtrų metodas. Pagal pradines analizės prielaidas galima išskirti tris grupes metodų: metodus taikomus signalui ir metodus, pagrįstus kalbos signalo generavimo bei kalbos suvokimo modeliais. Pirmuose metoduose daroma prielaida, jog signale yra visa reikalinga informacija lingvistiniam turiniui nustatyti ir požymiai paprastai išskiriami iš paties signalo arba iš jo darinio, išryškinančio tam tikras savybes. Su signalu dirbama spektro, kepstro, koreliacinėje analizėse. Kalbos generavimo modeliai taikomi darant prielaidą, kad kalbos signalą galima apibrėžti signalą generuojančios sistemos modelio parametrais. Labiausiai paplitęs iš šių metodų grupės yra tiesinės prognozės modelis. Klausos modeliu grindžiamuose metoduose kalbos signalui taikomas toks apdorojimas, koks manoma, yra atliekamas žmogaus klausos sistemoje.

Sekančiuose skyreliuose panagrinėsime požymių sistemas, naudojamas kalbos atpažinime, trumpai apžvelgsime galimą požymių sistemų vystymosi kryptį. Kadangi šiame darbe apsiribojama akustiniu kalbos signalo apdorojimu, nagrinėsime tik akustinės analizės metodus: spektrio analizę, tiesinę prognozę ir kepstro analizę.

3.2.1. Spektro analizė

Ilgą laiką spektro analizė buvo plačiausiai naudojamas kalbos signalo analizės metodas. Šio metodo ilgaamžiškumą lėmė galimybė realizuoti jį tiek programinėmis, tiek aparatinėmis priemonėmis (kas ir buvo daroma, kol programinė realizacija buvo sunkiai įmanoma dėl ribotų skaičiavimo technikos galimybių).

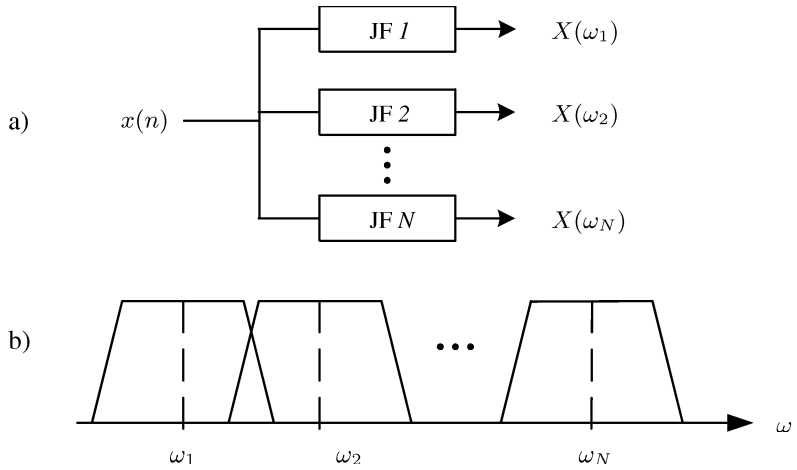
Spektro analizės tikslas – trumpalaikio signalo spektro įvertinimas (tą sąlygoja jau minėta prielaida apie trumpos signalo atkarpos stacionarumą). Trumpalaikio diskretinio signalo $x(n)$ spektras apibrėžiamas

$$X_m(\omega) = \sum_{n=-\infty}^{\infty} x(n)w(m-n)e^{-j\omega n} = |X_m(\omega)|e^{j\theta_m(\omega)}. \qquad (3.1)$$

Šią trumpalaikio signalo spektro išraišką galima interpretuoti dvejopai. Viena vertus, tai yra diskretinio signalo $x(n)$, padauginto iš lango funkcijos $w(m-n)$, Furjė transformacija laiko momentu m, t. y. trumpalaikio signalo spektrą galime gauti skaičiuodami signalo atkarpos Furjė transformaciją. Kita vertus, jei tarsime, kad $w(n)$ yra skaitmeninio filtro impulsinė reakcija, o dažnis ω yra fiksuotas, (3.1) išraiškoje gausime kompozicijos formulę, išreiškiančią išėjimo signalą $X_m(\omega)$, gautą į filtrą su impulsine charakteristika $w(n)$ padavus signalą $x(n)e^{-j\omega n}$. Šiuo atveju trumpalaikio signalo spektrą gauname signalą praleidę pro tiesinį filtrą. Remiantis šiomis interpretacijomis, signalo spektrui įvertinti galimi du būdai: juostinių filtrų blokas ir Furjė transformacija.

3.2.1.1. Juostiniai filtrai

Juostinių filtrų bloko analizėje signalas įvedamas į lygiagrečiai sujungtus juostinius filtrus (3.3 a) pav.). Filtrai tarpusavyje skiriasi centriniais dažniais, o jų pralaidumo juostos tarpusavyje persidengia apimdamos visą signalo dažnių juostą (3.3 b) pav.).



3.3 pav. *Juostinių filtrų bloko analizė: a) supaprastinta filtrų bloko struktūra; b) analizės modelis*

Tuomet i-ojo filtro išėjime laiko momentu m gauname trumpalaikio signalo spektrą $X_m(\omega_i)$ dažnių diapazone su centriniu dažniu ω_i, o visų filtrų išėjimų suma duoda $X_m(\omega)$ – trumpalaikio signalo spektrą laiko momentu m [102]

$$X_m(\omega) = \sum_{i=1}^{N} X_m(\omega_i), \tag{3.2}$$

čia N – filtrų skaičius.

Bendru atveju juostinių filtrų blokas sudaromas iš vienodo pločio pralaidumo filtrų, vienodai nutolusių tiesinėje dažnių ašyje. Kalbos atpažinimo sistemose, remiantis prielaida, kad žmogaus garso suvokimas yra netiesinio pobūdžio, dažniau naudojamos netiesinės dažnių skalės. Paprasčiausias tokios skalės pavyzdys – logaritminė dažnių skalė. Naudojant logaritminę dažnių skalę filtrai išdėstomi taip, kad jų centriniai dažniai būna vienodai nutolę logaritminėje dažnių ašyje, o i-ojo filtro pralaidumo juostos plotis apskaičiuojamas i-1-ojo filtro juostos plotį padauginus iš konstantos (paprastai didesnės už vienetą). Plačiausiai naudoti oktavos ir 1/3 oktavos filtrai (atstumai tarp filtrų centrinių dažnių 1 ir 1/3 oktavos atitinkamai) [88]. Kitas netiesinės dažnių skalės atvejis – skalės dalijimas į kritines juostas. Kritinė juosta – tai dažnių juosta, kurioje pasireiškia maskavimo efektas, t. y. tono slopinimas didesnės amplitudės tonu (už kritinės juostos ribų esantys tonai jokios įtakos neturi). Eksperimentais nustatyta, kad didėjant dažniui, kritinių juostų plotis auga [121]. Kritinių juostų skalei išreikšti įvestas matavimo vienetas – barkas (1 barkas apima vieną kritinę juostą). Remiantis eksperimentų rezultatais buvo pasiūlyta analitinė barkų skalės išraiška

$$f_{\text{bark}} = 13 \arctan(0,76f) + 3,5 \arctan\left(\frac{f}{7,5}\right)^2, \tag{3.3}$$

čia f – linijinis dažnis, išreikštas kHz.

Nagrinėjant įvairių dažnių suvokimą pasiūlyta melų skalė, susiejanti garso ir jo suvokimo dažnius. Melai nurodo suvokiamo garso dažnį, o 1 melas prilygintas 1/1000 suvoktojo 1 kHz garso dažnio. Dažnis melais išreiškiamas

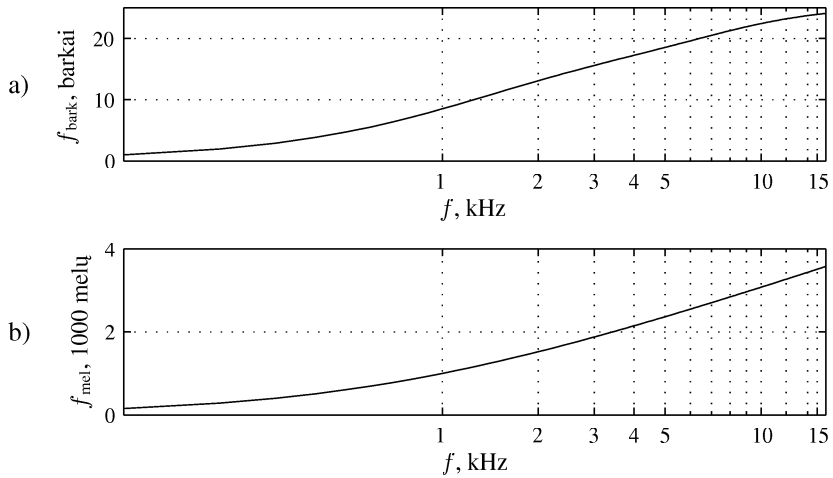
$$f_{\text{mel}} = 1127 \ln(1 + f/0,7), \tag{3.4}$$

čia f – linijinis dažnis, išreikštas kHz.

3.4 paveiksle matome, kad barkų ir melų skalės yra artimos tiesinėms linijinio dažnio logaritmo atžvilgiu, taigi skirtingais eksperimentais gauti rezultatai patvirtina teiginį apie žmogiškosios klausos logaritminį pobūdį. Beje, negalima nepastebėti ir tarpusavio skalių panašumo. Filtrų blokui realizuoti gali būti panaudoti tiek ribotos (RIR), tiek neribotos impulsinės charakteristikos (NIR) filtrai. NIR filtrai yra paprasčiau realizuojami, tačiau RIR filtrai pasižymi tiesine fazės charakteristika, be to juose lengviau realizuoti pageidaujamas dažnines savybes.

Pagrindinis juostinių filtrų bloko analizės privalumas – nepriklausomos spektro juostos, kurių pagalba galima apdoroti konkrečias kalbos signalo spektro sritis (pavyzdžiui, pašalinti triukšmo komponentes, įvertinti spektro sričių įtaką atpažinimui). Be to, skaitmeninių filtrų panaudojimas leidžia nesunkiai realizuoti filtrų bloko analizę. Esminis analizės trūkumas – filtrų jautrumas signalo spektrui. Bet koks šaltinio signalo pokytis, bet kurios prigimties triukšmas įtakoja spektrinės analizės rezultatus, o kadangi natūralioje kalboje šaltinio signalas nuolat kinta

(tiek to paties, tiek skirtingų kalbėtojų atveju), o pats kalbos signalas yra veikiamas įvairių triukšmų, šis filtrų trūkumas apsunkina šio juostinių filtrų analizės panaudojimą kalbos atpažinimo uždaviniuose.



3.4 pav. *Netiesinės dažnių skalės: a) barkų; b) melų*

3.2.1.2. Furjė spektras

Furjė transformacija atvaizduoja signalą dažnių srityje, t. y. išreiškia signalo amplitudę ir fazę kaip dažnio funkcijas. Skaitmeniniams signalams apdoroti naudojama diskrečioji Furjė transformacija (DFT), leidžianti gauti signalo baigtinės trukmės spektro diskrečias reikšmes. Bendru atveju tiesioginė signalo $x(n)$ DFT (dar vadinama analizės lygtimi) išreiškiama

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}, \quad \text{kai } k = 0, 1, \ldots, N-1,$$

čia $X(k)$ – k-oji signalo $x(n)$ spektro imtis, N – signalo trukmė atskaitomis.

Atvirkštinė DFT (arba sintezės lygtis)

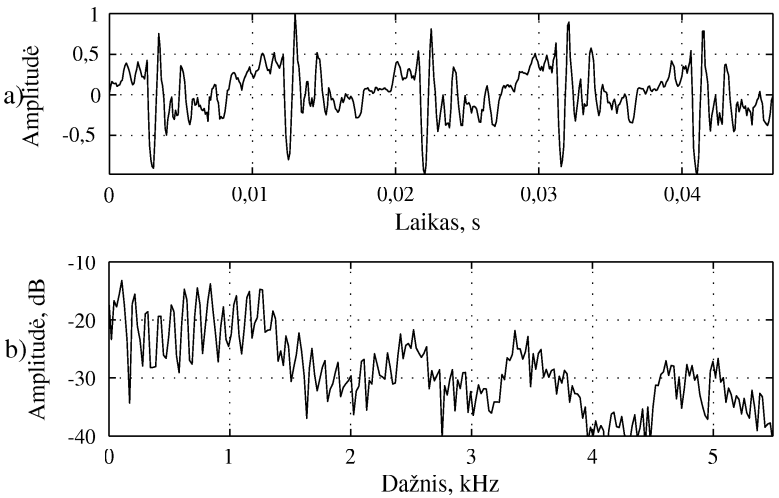
$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi kn/N}, \quad \text{kai } n = 0, 1, \ldots, N-1.$$

Spektras yra kompleksinė dažnio funkcija, kurios modulis apibrėžia amplitudžių spektrą, o argumentas – fazės spektrą. Nagrinėjant kalbos signalą daroma

prielaida, kad fazės spektre lingvistinė informacija neatspindima, todėl paprastai nagrinėjamas tik amplitudžių spektras (3.5 pav.). Kalbos signalo analizės tikslas yra trumpalaikio signalo spektras, todėl skaičiuojant DFT, signalą tenka dauginti iš lango funkcijos

$$X(k) = \sum_{n=0}^{N-1} x(n)w(m-n)e^{-j2\pi kn/N}, \quad \text{kai } k = 0, 1, \ldots, N-1,$$

čia $w(n)$ – lango funkcija.



3.5 pav. *Garso „a“: a) laiko diagrama; b) spektras, gautas atlikus DFT (naudojant Hamming langą)*

Lango funkcijos paskirtis – išskirti signalo atkarpą ir sumažinti signalo trūkių atkarpų galuose įtaką spektrui. Be to, naudojamo lango ilgis lemia spektro savybes – spektro dažnių skiriamoji geba atvirkščiai proporcinga lango ilgiui. Signalų analizei pasiūlyta dešimtys lango funkcijų bei jų modifikacijų, besiskiriančių savo forma, spektrinėmis savybėmis (įvairiais pralaidumo ir slopinimo juostų parametrais) [35, 72]. Ir nors teigiama, jog nė viena lango funkcija negali garantuoti visų parametrų optimalių reikšmių vienu metu [1,34], kalbos atpažinimo srityje dėl realizacijos paprastumo ir pakankamo efektyvumo dažniausiai naudojamas Hamming langas [90].

Tiesioginės DFT operacijų skaičius yra proporcingas N^2 (N – nagrinėjamos signalo atkarpos ilgis). Ši savybė kurį laiką buvo kliūtis Furjė transformacijos realizacijai skaičiavimo technikoje. 7-tajame dešimtmetyje J. W. Cooley ir J. W. Tukey [12] pasiūlė optimizuotą DFT algoritmą – greitąją Furjė transformaciją (GFT).

GFT esmė – DFT taikymas duomenų matricoms, sudarytoms iš suskaidytos signalo sekos, taip išvengiant pasikartojančių operacijų. Toks veiksmų organizavimas leido sumažinti operacijų skaičių iki $Nlog_2N$ eilės. Ir nors literatūroje [11,37] teigiama, jog optimizuoto algoritmo idėja ir pagrindai buvo suformuluoti daug anksčiau, būtent Cooley ir Tukey publikacija tapo lūžio tašku skaitmeninio signalo apdorojime ir Furjė transformacijos taikyme kalbai analizuoti.

Pagrindinis DFT privalumas – diskreti transformacijos prigimtis, kas daro ją labai patogia skaitmeninių signalų analizei. Tačiau anot literatūros [32], DFT algoritmo skaičiavimo operacijų skaičius yra tos pačios eilės kaip analizuojant signalą juostinių filtrų bloku. Įvertinus tai, kad tiek Furjė transformacijos, tiek filtrų bloko panaudojimo tikslas – signalo spektro išskyrimas, galima teigti, jog Furjė transformacija ir juostinių filtrų blokas iš esmės dubliuoja vienas kitą. Tikrasis Furjė transformacijos pranašumas prieš juostinius filtrus išryškėja, naudojant greitąją tranformacijos versiją. Šiuo atveju gauname labai spartų ir aiškiai algoritmizuotą metodą signalo spektrui išskirti.

3.2.2. Tiesinės prognozės modelis

3.2.2.1. Tiesinės prognozės modelio analizė

Skirtingai nei aukščiau nagrinėti analizės metodai, tiesinės prognozės modeliu modeliuojamas pats signalas, o ne jo spektras.

Tiesinės prognozės modelio pagrindas yra kalbos signalo generavimo modelis „šaltinis-sistema“ (3.1 pav.). Šiame modelyje esantis balso traktas gali būti modeliuojamas tiesiniu filtru su laike kintančiais parametrais. Tokio filtro sistemos funkciją bendru atveju sudaro sistemos poliai ir nuliai. Kalbos signalo atveju naudojama visų polių sistemos funkcija (dėl to tiesinės prognozės modelis dar vadinamas autoregresijos modeliu) [3]

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}, \tag{3.5}$$

čia $H(z)$ – balso trakto sistemos funkcija, G – balso trakto stiprinimo koeficientas, a_k – filtro koeficientai, p – sistemos eilė.

Įvertinę į balso traktą patenkantį šaltinio signalą ir atlikę atvirkštinę z transformaciją iš išraiškos (3.5) gauname

$$x(n) = \sum_{k=1}^{p} a_k x(n-k) + G \cdot v(n), \tag{3.6}$$

čia $v(n)$ – šaltinio signalas.

Kadangi paprastai šaltinio signalas $v(n)$ nežinomas, signalo atskaitos įvertis (prognozė) išreiškiama kaip p ankstesnių atskaitų suma

$$\hat{x}(n) = \sum_{k=1}^{p} a_k x(n-k). \tag{3.7}$$

Koeficientai a_k vadinami tiesinės prognozės (arba tiesiog prognozės) koeficientais. Su prognoze susijusi klaida, vadinama prognozės klaida (arba klaidos signalas), apibrėžiama

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{k=1}^{p} a_k x(n-k). \tag{3.8}$$

Palyginę (3.6) ir (3.8) matome, jog prognozės klaidą galime laikyti sužadinimo signalo įverčiu.

TPM analizės tikslas – rasti kažkuria prasme optimalius prognozės koeficientų a_k įverčius. Tradiciškai naudojamas minimalios kvadratinės klaidos (o taip pat ir minimalios vidutinės kvadratinės klaidos) kriterijus. Įrodyta [65], kad minimali kvadratinė klaida atitinka minimalų signalo ir jo modelio spektrų tarpusavio iškraipymą [32].

Kvadratinė prognozės klaida apibrėžiama

$$E = \sum_{n} e^2(n) = \sum_{n} \left(x(n) - \sum_{k=1}^{p} a_k x(n-k) \right)^2 \tag{3.9}$$

čia E – kvadratinė prognozės klaida, apskaičiuota signalo atkarpai (atkarpos dydžio kol kas nenagrinėjame).

Minimizuodami kvadratinę klaidą, skaičiuojame (3.9) išvestines pagal visus prognozės koeficientus a_k ir jas prilyginame nuliui. Atlikę pertvarkymus, gauname p tiesinių lygčių (Yule-Walker lygčių) sistemą

$$\sum_{m=1}^{p} a_m \varphi(m,k) = \sum_{n} \varphi(0,k), \quad \text{kai } 1 \leqslant k \leqslant p, \tag{3.10}$$

čia

$$\varphi(m,k) = \sum_{n} x(n-m) \cdot x(n-k). \tag{3.11}$$

Lygčių sistema (3.10) gali būti išspręsta naudojant standartinius matematinius metodus. Dažniausiai naudojami du sprendimo metodai: kovariacinis ir koreliacinis. Metodai tarpusavyje skiriasi kvadratinės klaidos minimizavimo intervalu (3.10).

Kovariaciniame metode [3] kvadratinė klaida skaičiuojama visam signalui

$$E = \sum_{n=0}^{N-1} e^2(n). \tag{3.12}$$

čia N – signalo ilgis.
Tuomet lygčių sistema (3.10) išskleidžiama

$$\begin{pmatrix} \varphi(1,1) & \varphi(1,2) & \cdots & \varphi(1,p) \\ \varphi(2,1) & \varphi(2,2) & \cdots & \varphi(2,p) \\ \cdots & \cdots & \cdots & \cdots \\ \varphi(p,1) & \varphi(p,2) & \cdots & \varphi(p,p) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_p \end{pmatrix} = \begin{pmatrix} \varphi(0,1) \\ \varphi(0,2) \\ \cdots \\ \varphi(0,p) \end{pmatrix} \tag{3.13}$$

arba sutrumpinta forma

$$\Phi \boldsymbol{a} = \boldsymbol{\varphi}. \tag{3.14}$$

Matrica Φ gaunasi simetrinė pagrindinės įstrižainės atžvilgiu. Dėl šios savybės, būdingos kovariacinei matricai, kovariacinis metodas ir įgijo savo pavadinimą. Gautajai lygčių sistemai spręsti naudojami dekompozicijos metodas – matrica Φ išskaidoma į kanoninės formos matricų sandaugą, kurių pagalba gaunamos rekurentinės išraiškos prognozės koeficientų vektoriui gauti. Kalbos apdorojime dažniausiai naudojama Cholesky dekompozicija.

Autokoreliaciniame metode [69] kvadratinė klaida vertinama baigtinės trukmės signalo atkarpoje, išskiriamoje naudojant lango funkciją

$$x_m(n) = \begin{cases} x(m+n)w(n), & \text{kai } 0 \leqslant n \leqslant L-1; \\ 0, & \text{kitur,} \end{cases} \tag{3.15}$$

čia $w(n)$ – lango funkcija, L – lango ilgis atskaitomis.

Tuomet kvadratinė klaida laiko intervale $0 \leqslant n \leqslant L+p-1$ yra nelygi nuliui ir išreiškiama

$$E_m = \sum_{n=0}^{L+p-1} e^2(n). \tag{3.16}$$

Įvertinus nagrinėjamąjį intervalą, vietoj lygties narių (3.11) gauname signalo autokoreliacijos funkcijos išraiškas, o lygčių sistema (3.10) įgyja formą

$$R\boldsymbol{a} = \boldsymbol{r}, \tag{3.17}$$

čia R – signalo autokoreliacijos matrica, r – signalo autokoreliacijos matrica-stulpelis.

Matrica R yra Toeplitz tipo – simetrinė, su vienodais elementais pagrindinėje ir jai lygiagrečiose įstrižainėse. Be standartinių matematinių metodų lygčių siste-

mai (3.17) spręsti pasiūlyta keletas rekursyvių procedūrų: Levinsono, Robinsono, Durbino (dar vadinamas Levinsono-Durbino) [65, 69]. Dažniausiai dėl savo efektyvumo naudojamas Levinsono-Durbino algoritmas

1. Inicializacija
$$E^{(0)} = r(0). \tag{3.18}$$

2. Iteracija, kai $1 \leqslant i \leqslant p$
$$k_i = \left(r(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} r(i-j) \right) / E^{(i-1)}; \tag{3.19}$$
$$a_i^{(i)} = k_i; \tag{3.20}$$
$$a_j^i = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)}, \quad \text{kai } 1 \leqslant j \leqslant i; \tag{3.21}$$
$$E^{(i)} = \left(1 - k_i^2\right) E^{(i-1)}. \tag{3.22}$$

3. Užbaigimas
$$a_j = a_j^{(p)}, \quad \text{kai } 1 \leqslant j \leqslant p. \tag{3.23}$$

Išraiškose: E – signalo energija, lygi signalo autokoreliacijai su nuliniu intervalu $r(0)$, k_i – i-osios eilės modelio atspindžio (dar vadinamas dalinės koreliacijos – PARCOR) koeficientas.

Kaip matome, skaičiuojant p-osios eilės modelio koeficientus, tenka apskaičiuoti visus žemesnės eilės modelius.

Pagrindinis reikalavimas modeliuojant signalą – modelio stabilumas. Naudojant kovariacinį metodą, gaunamo modelio stabilumas nėra garantuotas, tačiau jei analizei naudojamas pakankamas skaičius signalo atskaitų, gaunamas modelis praktiškai visada būna stabilus [90]. Autokoreliacinis metodo naudojimas garantuoja, jog gautasis modelis bus stabilus. Tačiau jų stabilumui didelę reikšmę turi skaičiavimo tikslumas, kuriam esant nepakankamo lygio koeficientai gali gautis nestabilūs. Įrodyta, kad nestabilumų tikimybę sumažina išankstinis filtravimas aukštų dažnių filtru (atliekamas prieš signalo analizę), sumažinantis kalbos signale dominuojančių žemų dažnių įtaką skaičiuojamiems parametrams. Tradiciškai naudojamas pirmos eilės ribotos impulsinės charakteristikos filtras, kurio sistemos funkcija
$$H(z) = 1 - \alpha z^{-1}, \tag{3.24}$$
α – filtro koeficientas, paprastai prilyginamas 0,95.

Teigiama, jog filtravimas taip pat leidžia sumažinti klaidos signalo svyravimus, gaunamus tiesinės prognozės analizei naudojant autokoreliacinį metodą [87].
Kita vertus, autokoreliaciniame metode naudojama lango funkcija mažina modelio

tikslumą. Taigi šiuo požiūriu kovariacinis metodas, kuriame langa funkcija nenaudojama, yra pranašesnis. Įvertinus tai, kad reikalinga tiesinės prognozės modelio eilė nepriklauso nuo modelio parametrų skaičiavimo metodo (teigiama, kad abiem atvejais 10–12 eilė yra pakankama [3,65]), galima teigti, jog pasirinkimas tarp autokoreliacinio ir kovariacinio metodų iš esmės yra pasirinkimas tarp stabilumo ir tikslumo. Praktikoje pirmenybė dažniausiai teikiama stabilumui – šį teiginį patvirtina autokoreliacinio metodo paplitimas tiesinės prognozės analizėje.

Be nagrinėtųjų buvo pasiūlyta ir daugiau tiesinės prognozės metodų, modelio vertinimų kriterijų, skaičiavimo algoritmų. Signalų apdorojimo teorijoje šalia autokoreliacinio ir kovariacinio metodų dažnai nagrinėjami pynučių metodai. Juose visų polių filtras modeliuojamas pynučių filtru, o prognozės klaida minimizuojama randant optimalius filtro atspindžio koeficientus [8, 67]. Esminis šių metodų grupės bruožas – atspindžio koeficientai gaunami tiesiogiai iš signalo reikšmių (o ne signalo autokoreliacijos reikšmių kaip kad Levinsono-Durbino algoritme), išvengiant prognozės koeficientų skaičiavimo. Pastaruosius apskaičiuoti galima panaudojus išraišką (3.21). Atspindžio koeficientus taip pat galima apskaičiuoti naudojant Schur algoritmą [36], kuris atspindžio koeficientus išreiškia per autokoreliacijos funkcijos reikšmes (nenaudojant prognozės koeficientų). Prognozės charakteringąjį daugianarį pakeitus simetriniu daugianariu, kurio šaknims rasti reikia dvigubai mažiau operacijų, sukurtas suskaidytasis Levinsono algoritmas [19]. Prognozės modeliui vertinti pasiūlyti kriterijai, be prognozės klaidos vertinantys modelio eilę, nagrinėjamos realizacijos ilgį [2, 79, 95]. Tokių kriterijų panaudojimas leidžia gauti modelio parametrų rinkinius, minimizuojančius ne tik prognozės klaidą, bet ir modelio eilę, parametrams įvertinti reikalingą realizacijos ilgį.

3.2.2.2. Išvestiniai tiesinės prognozės parametrai

Pagrindinis tiesinės prognozės modelio analizės rezultatas – p modelio koeficientų, kurie su papildomais parametrais (dažniausiai su nuliniu modelio koeficientu $a_0 = 1$ ir modelio stiprinimo koeficientu) naudojami parametrų vektoriams formuoti. Modelio koeficientai taip pat gali būti transformuoti į kitą parametrų rinkinį, pasižymintį jam būdingomis savybėmis. Trumpai apžvelgsime parametrus, gaunamus tiesiogiai iš tiesinės prognozės koeficientų.

Daugianario šaknys. Daugianario šaknys yra z reikšmės, su kuriomis sistemos (3.5) daugianaris yra lygus nuliui

$$A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} = \prod_{i=1}^{p} (1 - z_i z^{-1}),$$

čia z_i – bendru atveju kompleksinės i-osios daugianario šaknies rodiklinė forma

$$z_i = r_i e^{j\omega_i},$$

čia r_i – šaknies modulis, ω_i – šaknies argumentas (nurodo kampą su teigiama

realiąja ašimi kompleksinėje plokštumoje).

Gaunamos šaknys (prognozės modelio poliai) nurodo modeliuojamos sistemos rezonansinius dažnius, o kalbos signalo atveju – formantes. Šaknies modulis atitinka formantės amplitudę, šaknies argumentas – kampinį formantės dažnį. Nagrinėjamos tik kompleksinės šaknys su teigiamais argumentais (kas atitinka neneigiamo ir nenulinio dažnio formantes). Daugianario šaknys naudotos tiek kalbos sintezėje [3], tiek analizėje [62]. Pagrindinis daugianario šaknų privalumas – nesudėtingas formančių nustatymas.

Spektro poros. Spektro porų transformacijos esmė – tiesinės prognozės daugianario $A(z)$ išskaidymas į linijinio spektro porų daugianarius ir jų nulių radimas. Daugianariai apibrėžiami

$$P(z) = A(z) + z^{-(p+1)}A(z^{-1});$$
$$Q(z) = A(z) - z^{-(p+1)}A(z^{-1}).$$

Daugianarių nuliai yra ant vienetinio apskritimo, taigi juos galima išreikšti

$$z_i = e^{j\omega_i},$$

čia ω_i vadinami spektro porų dažniais.

Dėl savo atsparumo kvantavimo efektams, spektro porų parametrai ilgą laiką buvo naudojami kalbai koduoti, vėliau pritaikyti kalbai atpažinti [77].

Tiesinės prognozės kepstras. Šie parametrai yra dažniausiai naudojama tiesinės prognozės analizės forma. Signalo kepstru vadiname signalą, kurio spektras yra pradinio signalo Furjė transformacijos (spektro) logaritmas. Praktikoje TPM kepstro (TPMK) koeficientams skaičiuoti yra gautos rekurentinės išraiškos [90]

$$c_0 = \log G; \tag{3.25}$$

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \quad \text{kai } 1 \leqslant m \leqslant p; \tag{3.26}$$

$$c_m = \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \quad \text{kai } m > p; \tag{3.27}$$

čia c_m – m-asis TPM kepstro koeficientas.

Kaip matome (3.26) ir (3.27) išraiškose, esant ribotai tiesinės prognozės analizės eilei, kepstrinės analizės eilė neribota. Pakankama laikoma 12–20 kepstrinės analizės eilė. Reikėtų pažymėti, kad inversinis prognozės koeficientų skaičiavimas iš tiesinės prognozės kepstro koeficientų negarantuoja gaunamo modelio stabilumo, todėl yra vengtinas.

Atlikti atpažinimo eksperimentai parodė, kad TPM kepstro analizė leidžia pasiekti didžiausią atpažinimo tikslumą tarp visų tiesinės prognozės išvestinių para-

metrų grupės [17, 78].

Kalbai atpažinti taip pat bandyta taikyti atspindžio koeficientų (dalinės koreliacijos) analizė [17, 45], tačiau kalbos atpažinime šie parametrai nepaplito.

3.2.2.3. Modifikuoti tiesinės prognozės modeliai

Tiesinės prognozės modelio analizėje visi nagrinėjamo signalo spektro dažniai vertinami vienodai. Tačiau įvertinus tai, kad žmogaus garso suvokimas yra netiesinio pobūdžio, o kalbos signalo didesnė energijos dalis yra sutelkta žemuose dažniuose, kyla klausimas dėl šių faktorių įvertinimo tiesinės prognozės modelyje. Pasiūlyta nemažai tiesinės prognozės modelių, įvertinančių klausos sistemos ir kalbos signalo spektrines savybes. Plačiausiai jų naudojama tiesinė suvokimo prognozė – tiesinės prognozės kepstro analizė, įvertinanti klausos sistemos poveikį signalo spektrui. Šiuo atveju analizei naudojamos ne tiesiogiai iš signalo reikšmių apskaičiuotos autokoreliacijos reikšmės, o gautosios atlikus signalo galios spektro atvirkštinę Furjė transformaciją. Signalo galios spektras formuojamas remiantis apdorojimo klausos sistemoje schema [38]:

1. Netiesinės spektro dažnių skalės formavimas. Signalas filtruojamas 18 barkų skalės juostinių filtrų rinkiniu, apimančiu 0–5 kHz dažnių diapazoną.

2. Spektro glotninimas. Signalas apdorojamas aukštų dažnių filtru, išlyginančiu galios spektrą.

3. Spektro dinaminio diapazono mažinimas. Atliekama signalo amplitudės kompresija. Remiantis dėsniu, nusakančiu ryšį tarp garso intensyvumo ir suvokimo garsumo, paprastai naudojama kubinės šaknies kompresija.

Gautajam galios spektrui atlikus atvirkštinę Furjė transformaciją, gaunamos autokoreliacijos reikšmės, kurios panaudojamos 5 eilės tiesinės prognozės kepstro analizei pagal (3.18)–(3.23) ir (3.25)–(3.27). Teigiama, jog tiesinė suvokimo prognozė leidžia padidinti atpažinimo proceso nepriklausomumą nuo kalbėtojo [40]. Vėliau tiesinei klausos prognozei pritaikytas RASTA apdorojimas (kiekvieno kritinės juostos filtro išėjimas papildomai apdorojamas juostiniais filtrais), siekiant pašalinti lėtus kalbos signalo pokyčius ir padidinti analizės atsparumą triukšmams [39].

Be tiesinės suvokimo prognozės kalbos analizei pasiūlyta selektyvioji tiesinė prognozė, kurios esmė – tiesinę prognozę taikyti tik daliai signalo spektro, taip apsiribojant tik dominančios spektro dalies analize [66]. Iškreiptajame tiesinės prognozės modelyje analizuojamas signalas su „iškraipyta“ spektro dažnių skale, siekiant sumodeliuoti klausos sistemos įtaką signalo spektrui [57]. Prognozuojant signalo atskaitos reikšmę ne tik iš buvusių, bet ir būsimų signalo reikšmių, pasiūlytas dvipusės tiesinės prognozės modelis [15, 119].

Apibendrinsime tiesinės prognozės modelio analizės savybes. Esminis analizės privalumas – sugebėjimas modeliuoti kalbos signalą nevertinant signalo šalti-

nio. Tai pagrindinė tiesinės prognozės modelio panaudojimo kalbos signalui analizuoti priežastis. Kaip privalumus taip pat galima įvardinti metodų, naudojamų modelio parametrams vertinti, standartiškumą (tiek autokoreliacinio, tiek kovariacinio metodų atvejais, modelio parametrams skaičiuoti naudojami standartiniai matematiniai algoritmai) ir neaukštą gaunamo modelio eilę. Kita vertus, tiesinės prognozės modelis pasižymi ir nemenkais trūkumais. Visų pirma, tai visų polių sistemos panaudojimas modelyje. Toks modelis sunkiai įvertina spektrinius nulius, o tai lemia ne visada tikslų spektro įvertinimą. Antra, tiesinės prognozės modelis nesugeba išvengti balso stygų ir balsaskylės įtakos modeliuojamam signalui [49]. Šie trūkumai reiškia, kad atpažinimo sistema, naudojanti tiesinės prognozės modelio analizę, bus neatspari triukšmams ir spektrinių kalbančiojo savybių pokyčiams. Šiuos teiginius patvirtina ir eksperimentinių tyrimų rezultatai [17, 110] [A4, A5].

3.2.3. Kepstro analizė

3.2.3.1. Homomorfinė analizė ir signalo kepstras

Signalo generavimo modelyje „šaltinis-sistema“ (3.1 pav.) šaltinio signalas ir balso trakto impulsinė charakteristika yra susietos kompozicijos operacija

$$x(n) = v(n) * h(n),$$

čia $x(n)$ – kalbos signalas, $v(n)$ – šaltinio signalas, $h(n)$ – balso trakto impulsinė charakteristika.

Jei signalus $v(n)$ ir $h(n)$ atskirsime, t. y. atliksime dekompoziciją, gausime dvi laiko sekas, apibūdinančias šaltinį ir balso traktą.

Siekiant atskirti signalus tiesinės sąveikos atveju (pvz., naudingą signalą atskirti nuo adityvinio triukšmo) paprastai laiko sritį transformuojame į dažnių sritį, kurioje nagrinėjamojo signalo dedamosios būna atskirtos. Netiesinės operacijos atveju (signalų sandaugos ar kompozicijos atveju) naudojama homomorfinė analizė [74] – taikant tam tikrą transformacijos schemą, netiesinė signalų sąveika transformuojama į signalų sumą, kuriai taikomas apdorojimas pagal superpozicijos principą. Dekompozicijos atveju transformacija atliekama logaritmuojant Furjė transformaciją [71]. Furjė transformacija kompoziciją paverčia sandauga, o logaritmas – sandaugą suma

$$X(\omega) = V(\omega) \cdot H(\omega);$$

$$log\left[X(\omega)\right] = log\left[V(\omega) \cdot H(\omega)\right] = log\left[V(\omega)\right] + log\left[H(\omega)\right].$$

Siekiant grįžti į laiko sritį ir išlaikyti operacijos tiesiškumą atliekama atvirkštinė Furjė transformacija

$$C_x(k) = C_v(k) + C_h(k),$$

čia C_x, C_v ir C_h – signalų $x(k)$, $v(k)$ ir $h(k)$ kepstrų k-ieji koeficientai atitinkamai;

k turi laiko dimensiją ir vadinamas žadniu (nuo termino „dažnis“).

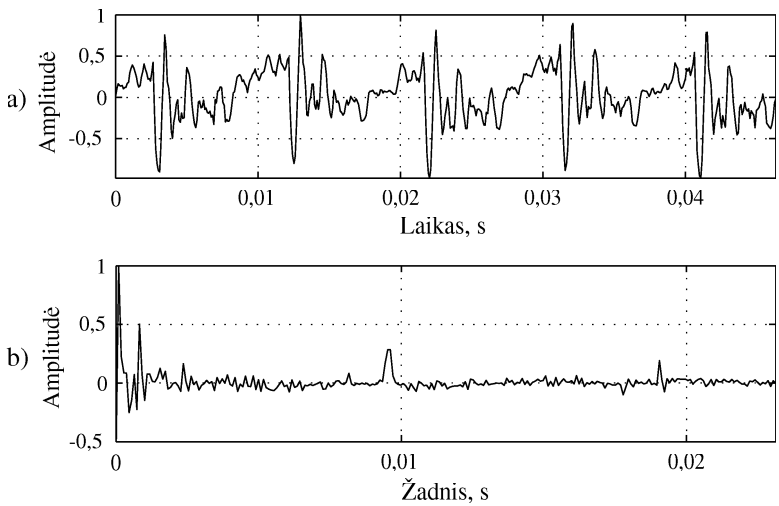
Taigi signalo kepstru vadiname signalo Furjė transformacijos logaritmo atvirkštinę Furjė transformaciją

$$C_x(k) = F^{-1}\{logF\,[x(n)]\}.$$

Bendru atveju kepstras yra gęstanti, lyginė, kompleksinė funkcija, tačiau kalbos apdorojime dažniausiai apsiribojama realaus kepstro analize, nagrinėjant baigtinį skaičių kepstro koeficientų (3.6 pav.)

$$C_x(k) = F^{-1}\{log|F\,[x(n)]\,|\,\},$$

t. y. logaritmuojamas ne kompleksinis spektras, o spektro modulis. Tokiu būdu informacija apie signalo fazę prarandama ir atstatyti signalą iš kepstro neįmanoma.



3.6 pav. *Garso „a“: a) laiko diagrama; b) kepstras*

3.2.3.2. Statinio kepstro analizė

Kalbos signalo spektrui skaičiuoti dažniausiai naudojama klasikinė skaičiavimo schema, paremta kepstro apibrėžimu, t. y. kepstras gaunamas skaičiuojant signalo Furjė transformaciją, ją logaritmuojant ir atliekant atvirkštinę Furjė transformaciją. Kadangi analizės tikslas būna trumpalaikio signalo įvertinimas, signalo

atkarpos išskyrimui tenka naudoti langa funkciją

$$X(k) = \sum_{n=0}^{N-1} x(n)w(m-n)e^{-j2\pi kn/N}, \quad \text{kai } 0 \leqslant k \leqslant N-1;$$

$$X_l(k) = \log|X(k)|, \quad \text{kai } 0 \leqslant k \leqslant N-1;$$

$$c_x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X_l(k)e^{j2\pi kn/N}, \quad \text{kai } 0 \leqslant n \leqslant N-1.$$

Paprastai naudojamas natūralusis arba logaritmas pagrindu 10, bet iš esmės gali būti panaudotas logaritmas bet kuriuo pagrindu. Išraiškose esanti DFT sėkmingai gali būti pakeista greitąja transformacijos versija. Kalbos signalo analizei apsiribojama 10–14 koeficientų skaičiavimu. Beje, reikėtų pažymėti, kad DFT panaudojimas kepstrui skaičiuoti lemia skaičiavimo paklaidą, kuri gali būti sumažinta didinant nagrinėjamos realizacijos ilgį N.

Pasiūlyti ir alternatyvūs metodai kepstrui skaičiuoti: signalo z transformacijos skaidymo daugikliais metodas [108], Hartlio transformacija paremtas algoritmas [107], tačiau tradiciškai naudojamas klasikinis kepstro skaičiavimo metodas logaritmuojant spektrą.

Praktikoje dažnai naudojami netiesinės dažnių skalės koeficientai, gaunami signalo spektrą skaičiuojant netiesinėje dažnių skalė. Dažniausiai naudojama melų skalė, rečiau – barkų. Melų skalės kepstro koeficientai skaičiuojami apibrėžiant L trikampių filtrų, tolygiai išdėstytų melų skalėje, ir taikant kosinusinę transformaciją energijos kiekvieno filtro išėjime logaritmui

$$X(k) = \sum_{n=0}^{N-1} x(n)w(m-n)e^{-j2\pi kn/N}, \quad \text{kai } 0 \leqslant k \leqslant N-1;$$

$$S(l) = \log\left[\sum_{k=0}^{N-1}|X(k)|^2 H_l(k)\right], \text{ kai } 0 \leqslant k \leqslant N-1,\ 1 \leqslant l \leqslant L;$$

$$c_{\text{x MEL}}(n) = \sum_{l=0}^{L-1} S(l)\cos\left(\frac{\pi n}{L}(l-1/2)\right), \quad \text{kai } 0 \leqslant n \leqslant L-1,$$

čia $H_l(k)$ – l-ojo trikampio filtro dažninė amplitudės charakteristika.

Atlikti tyrimai parodė [17], kad melų skalės kepstro koeficientai leidžia pasiekti didžiausią kalbos atpažinimo tikslumą. Dėl šios priežasties melų skalės koeficientų skaičiavimas yra plačiausiai vartojamas kalbos signalo analizės metodas kalbos atpažinime.

3.2.3.3. Dinaminio kepstro analizė

Be statinio kepstro (tiek tiesinės, tiek netiesinės dažnių skalės atveju) dažnai nagrinėjamas skirtuminis arba dinaminis kepstras, taip vadinamas delta-kepstras. Delta-kepstras skaičiuojamas dvejopai. Pirmu atveju skirtuminis kepstras skaičiuojamas kaip skirtumas tarp dviejų kadrų kepstro vektorių

$$\Delta c_m = c_{m+\delta} - c_{m-\delta},$$

čia Δc_m – m-ojo kadro delta-kepstras, $c_{m+\delta}$, $c_{m-\delta}$ – $(m-\delta)$ ir $(m+\delta)$-ojo kadrų kepstrai atitinkamai, δ paprastai prilyginamas 1–2.

Antruoju atveju delta-kepstras skaičiuojamas kaip kepstro išvestinės aproksimacija

$$\Delta c_n(m) = \sum_{k=-n_0}^{k=n_0} k c_{n+k}(m) / \sum_{k=-n_0}^{k=n_0} k^2,$$

čia $\Delta c_n(m)$ – m-asis kepstro laiko momentu n koeficientas.

Nors teigiama, kad antruoju būdu gautas delta-kepstras pasižymi glotnesniu pokyčio įvertinimu [32], praktiniuose uždaviniuose sėkmingai taikomas tiek pirmasis, tiek antrasis delta-kepstro skaičiavimo metodai.

Tačiau reiktų pažymėti, kad pats savaime skirtuminis kepstras turi informaciją tik apie kalbos signalo kepstro dinamiką ir vien jos neužtenka patikimai atpažinti kalbą. Todėl dažnai dinaminio kepstro analizės rezultatai apjungiami su statinio kepstro analizės rezultatais. Greta delta-kepstro analizės kartais atliekama taip vadinama delta-delta-kepstro analizė, t. y. skaičiuojama antroji kepstro išvestinė, rodanti kepstro kitimo pagreitį. Delta-delta-kepstro analizės rezultatai taip pat naudojami tik apjungus su statinės (ir delta-kepstro) analizės rezultatais.

Pagrindinis kepstro analizės privalumas – sugebėjimas atskirti šaltinio signalo nuo balso trakto spektrines savybes. Šiuo požiūriu kepstras yra pranašesnis už tiesinės prognozės modelį. Beje, kaip ir TPM analizės atveju, analizuojant kepstrą gaunamas pakankamai nedidelis parametrų skaičius.

3.3. Atpažinimo metodai

Sekantis etapas po signalo analizės – išskirtųjų požymių klasifikacija. Klasifikacijos metodą sistemoje nulemia naudojamas atpažinimo metodas. Sukurta daugybė metodų kalbai atpažinti, tačiau tik keletas jų pasiteisino kaip verti naudoti atpažinimo sistemose. Mes išskirtume keturias kalbos atpažinimo metodų grupes:

- akustiniai-fonetiniai metodai;
- pavyzdžių palyginimo metodai;
- statistiniai metodai;

- dirbtinio intelekto metodai.

Akustiniuose-fonetiniuose metoduose daroma prielaida, kad kalbos signalą sudaro baigtinės trukmės, akustiškai skirtingi fonetiniai vienetai, išsiskiriantys savitomis laikinėmis ir dažninėmis savybėmis. Išmatavus šias savybes (dažniausiai tai būdavo rezonansiniai dažniai, energijos ir amplitudės lygiai, nulio kirtimų skaičius) ir pritaikius elementarias taisykles (akustinės fonetikos žinias, slenkstines reikšmes, sprendimų medžius) kalbos signalas segmentuojamas ir sužymimas (priskiriama fonetinė transkripcija) taip kalbos atpažinimą realizuojant tiesioginiu signalo dekodavimu į transkripciją. Tačiau segmentavimas buvo ribota ta prasme, kad nebuvo atsižvelgiama į koartikuliacijos reiškinį, garsų kintamumą skirtingose žodžio versijose, o klasifikacija buvo pakankamai primityvi, todėl akustiniai-fonetiniai metodai nepasiteisino ir šiuolaikinėse atpažinimo sistemose yra beveik nenaudojami.

Pagrindinė pavyzdžių palyginimo metodų prielaida – to paties kalbos pavyzdžio versijos turi daugiau akustinių panašumų nei skirtingi pavyzdžiai. Pavyzdžių palyginimo metodikos esmė – nagrinėjamo pavyzdžio palyginimas su visais etaloniniais pavyzdžiais – etalonais. Tuo tikslu akustiniai pavyzdžiai išreiškiami požymių vektorių laiko sekomis, modeliuojančiomis akustinių savybių kitimą laike. Požymių sekos lyginamos tarpusavyje, o jų panašumas (skirtumas) įvertinamas skaitine forma. Remiantis gautais palyginimo rezultatais, priimamas sprendimas, kuris etalonas yra artimiausias nagrinėjamam pavyzdžiui.

Statistiniuose metoduose kalbos signalo kintamumas įvertinamas statistiniais metodais. Daroma prielaida, kad signalas yra atsitiktinis procesas, kurį galima sumodeliuoti – etalonams kurti panaudojami statistiniai modeliai, kurių parametrų įverčiai randami iš apmokymo duomenų aibės. Atpažinimo etape nagrinėjamo kalbos signalo atstumas iki etaloninių modelių įvertinamas naudojant statistinę klasifikaciją – dažniausiai Bejeso taisyklę. Labiausiai paplitęs statistinių metodų grupės atstovas – paslėptieji Markovo modeliai.

Dirbtinio intelekto metoduose stengiamasi imituoti žmogiškąjį kalbos atpažinimą: į atpažinimo procesą įjungiamos papildomos kalbos žinios, atpažinimo sistemoms suteikiamas sugebėjimas adaptuotis ir mokytis. Ankstyvuose metoduose, naudotuose ekspertinėse sistemose, greta akustikos žinių į atpažinimo procesą buvo papildomai įjungiamos akustikos, leksikos, sintaksės, semantikos ir net pragmatikos žinios. Žinioms įdiegti naudotos „viršus-apačia“, „apačia-viršus“, lentos metodai. Kita, stambiausia dirbtinio intelekto metodų grupė – neuronų tinklai. Neuronų tinklais imituojamas žmogaus sugebėjimas mokytis iš gaunamų duomenų, pagal juos koreguojant savo suvokimą.

Sekančiuose poskyriuose panagrinėsime plačiausiai paplitusius kalbos atpažinimo metodus: dinaminį laiko skalės kraipymą, paslėptuosius Markovo modelius ir neuronų tinklus. Atpažinimo sistemose nagrinėjamus kalbos vienetus (žodžių junginį, žodį, skiemenį, garsų junginį ar garsą) vadinsime kalbos pavyzdžiais, siekdami išvengti metodų susiejimo su konkrečiais kalbos vienetais. Taip pat vengsime

tiesioginio atpažinimo metodų pasiekiamo tikslumo palyginimo. Atpažinimo sistemos, naudojančios skirtingus metodus, eksperimentiškai tiriamos skirtingomis sąlygomis – naudojama skirtinga techninė įranga, sistemai apmokyti ir testuoti naudojamos skirtingos kalbos pavyzdžių bazės (ypač tai galioja devintojo ir ankstesnių dešimtmečių sistemoms, kadangi tuo metu dar nebuvo standartizuotų akustinių duomenų bazių), skirtingos akustinės eksperimentų sąlygos. Todėl sistemų atpažinimo tikslumo palyginimas, mūsų nuomone, nebūtų objektyvus.

3.3.1. Dinaminis laiko skalės kraipymas

Dinaminis laiko skalės kraipymo metodas, priklausantis pavyzdžių palyginimo metodų grupei, itin išpopuliarėjo praėjusio amžiaus 7–8-ajame dešimtmečiuose. Metodas dažniausiai buvo taikomas pavieniams žodžiams atpažinti [89, 100, 117], tačiau buvo bandymų pritaikyti žodžių junginiams [70, 99] ir net ištisinei kalbai atpažinti [114]. Nepaisant to, kad pats metodas ir jo algoritmas sukurtas prieš keturis dešimtmečius, jis sėkmingai taikomas ir šiais laikais.

3.3.1.1. Atstumas

Pavyzdžių palyginimas neįmanomas be atstumo – pavyzdžių panašumo mato – sąvokos. Atstumas (dar kartais naudojama „tarpusavio iškraipymo“ sąvoka) tarnauja kaip skaitinis pavyzdžių panašumo įvertinimas, kuriam pritaikius ekstremumo (dažniausiai minimumo) kriterijų, mes priimame sprendimą apie geriausiai atitinkantį etaloną. Atstumas tarp pavyzdžių paprastai išreiškiamas kaip atstumas tarp pavyzdžių vektorių.

Atstumui keliami du reikalavimai: simetrijos ir teigiamo apibrėžtumo. Atstumo simetrija reiškia, kad atstumas nepriklauso nuo lyginamų vektorių poros eiliškumo, o teigiamu apibrėžtumu reikalaujama, kad skirtumas turi būti teigiamas ir baigtinis dydis esant skirtingiems vektoriams ir nulinis esant lygiems vektoriams.

Atstumo tipas priklauso nuo požymių sistemos, kadangi kiekviena sistema reikalauja jos fizikinę interpretaciją atitinkančio atstumo.

Spektrinės analizės atveju atstumu dažniausiai tarnauja spektrų skirtumas, dar vadinamas norma

$$d(r_i, z_i) = \sqrt[n]{\sum_{k=1}^{L} \left(r_i^k - z_i^k\right)^n},$$

čia r_i ir z_i – etalono ir nežinomojo pavyzdžio i-ieji požymių vektoriai, L – nagrinėjamų vektorių eilė, k – vektoriaus elemento indeksas, n – normos eilė. Siekiant garantuoti atstumo simetriją esant nelyginiam laipsnio rodikliui, dažnai nagrinėjamas skirtumo modulis.

Dažniausiai naudojamos 1 ir 2 eilės normos, taip vadinamos Čebyševo ir Euklido normos. Teigiama, jog Euklido normos naudojimas lemia šiek tiek geresnius rezultatus [117].

Tiesiniam prognozės modeliui pasiūlyta keletas atstumų. Vienas jų – tikėtinumų santykio matas, išreiškiamas

$$d(r_i, z_i) = \frac{r_i R r_i^T}{z_i R z_i^T} - 1, \tag{3.28}$$

čia R – nežinomojo pavyzdžio autokoreliacijos matrica, indeksas ^{T} žymi transponavimo operaciją.
Labai panašus į pirmąjį tikėtinumų santykio logaritmo matas

$$d(r_i, z_i) = \log\left(\frac{r_i R r_i^T}{z_i R z_i^T}\right).$$

Abu panašumo matai yra sudėtingesnio, taip vadinamo Itakura-Saito, atstumo daliniai atvejai. Visiems šiems tiesinės prognozės modelių panašumo matams būdingas vienas trūkumas – asimetriškumas, t. y. $d(r_i, z_i) \neq d(z_i, r_i)$. Paprastai ši problema sprendžiama skaičiuojant abu atstumo atvejus ir kaip tikrąjį atstumą nagrinėjant abiejų aritmetinį vidurkį. Pasiūlytas ir simetrinis modelių panašumo matas – COSH atstumas [33], tačiau kalbos atpažinime jis nepaplito.

Kepstrui (tiek Furjė, tiek tiesinės prognozės) taip pat pasiūlyta keletas atstumo skaičiavimo būdų. Paprasčiausias jų išreiškiamas kaip Euklido atstumas tarp vektorių (kartais vartojamas atstumo kvadratas)

$$d(r_i, z_i) = \sqrt{\sum_{k=1}^{L} \left(r_i^k - z_i^k\right)^2}. \tag{3.29}$$

Sudėtingesnėje atstumo formoje kepstro dedamųjų įtakai įvertinti naudojama svorio funkcija

$$d(r_i, z_i) = \sqrt{\sum_{k=1}^{L} \left(g(k)\left(r_i^k - z_i^k\right)\right)^2}.$$

Kaip svorio funkcija siūlyta naudoti kepstro koeficientų indeksus [76], kepstro koeficientų dispersijai atvirkščius koeficientus [111], įvairios formos langus (požymių vektorius dauginant iš lango funkcijos) [49]. Visais atvejais pasvertasis kepstrinis atstumas buvo pranašesnis už įprastą Euklido atstumą, tačiau vėlgi reiktų pažymėti, kad svorio funkcija yra tiktai eksperimentais argumentuojamas sprendimas.

3.3.1.2. Pavyzdžių sutapdinimas

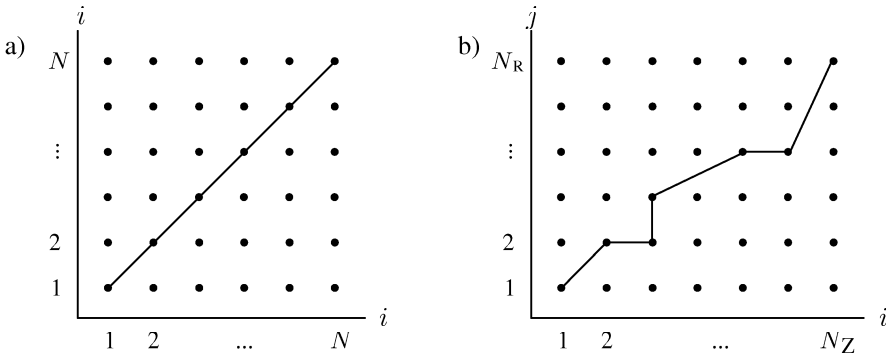
Norėdami įvertinti dviejų pavyzdžių panašumą, turime nagrinėti atstumus tarp atitinkamų požymių vektorių. Tarkime turime etaloną R ir nežinomą pavyzdį Z.

Paprasčiausiu atveju, kai abu pavyzdžiai yra vienodo ilgio, atstumas tarp R ir Z

$$D_{RZ} = \frac{1}{K} \sum_{i=1}^{N} d(r_i, z_i), \tag{3.30}$$

čia r_i ir z_i – etalono ir nežinomojo pavyzdžio i-ieji požymių vektoriai, $d(r_i, z_i)$ – atstumas tarp i-ųjų vektorių, K – normuojantis koeficientas, dažnai prilyginamas nagrinėtų pavyzdžių trukmei.

Grafinė atstumo tarp pavyzdžių skaičiavimo interpretacija pateikta 3.7 a) paveiksle.



3.7 pav. *Laiko skalės kraipymo atvejai: a) tiesinis; b) dinaminis*

Kiekvienas tinklelio taškas atitinka požymių vektorių, kurių indeksą nurodo taško koordinatės, porą. Su kiekvienu tašku sutapdintas dalinis atstumas, nurodantis atstumą tarp atitinkamų požymių vektorių. Vienodo ilgio pavyzdžių atveju panašumas skaičiuojamas taškuose, esančiuose ant plokštumos įstrižainės (paveiksle jie sujungti linija).

Bendru atveju nagrinėjami pavyzdžiai yra skirtingo ilgio, todėl tiesinis laiko skalių sutapdinimas netinka. Šiuo atveju tenka naudoti kraipymo funkciją $\phi(i, j)$, kurios pagalba atstumui skaičiuoti ieškomos artimiausių požymių vektorių poros, darant prielaidą, kad tie požymiai atstovauja vienodus akustinius įvykius. Toks laiko skalės kraipymas leidžia kompensuoti pavyzdžių trukmių skirtumą, o atstumas tarp pavyzdžių išreiškiamas

$$D_{RZ}(\phi) = \frac{1}{K} \sum_{i=1}^{K} d(\phi(i, j)), \tag{3.31}$$

čia $D_{RZ}(\phi)$ – suminis atstumas tarp pavyzdžių, gautas naudojant konkrečią kraipymo funkciją, K – trukmė pavyzdžio, kurio atžvilgiu atliekamas normavimas.

Toks dinaminio laiko skalės kraipymo atvejis pateiktas (3.7 b) pav.), kuriame matyti kreive (dar vadinama trajektorija) sujungtos skaičiuotų dalinių atstumų poros. Trajektoriją lemia pasirinktoji kraipymo funkcija $\phi(i, j)$.

Akivaizdu, kad galimų trajektorijų (o tuo pačiu ir atstumų tarp pavyzdžių) skaičius yra be galo didelis ir priklauso nuo kraipymo funkcijos $\phi(i, j)$ bei nuo nagrinėjamų pavyzdžių trukmių (arba tiesiog taškų tinklelio dydžio). Todėl tenka naudoti trajektorijos, vienareikšmiškai nusakančios atstumą tarp nagrinėjamų pavyzdžių, kriterijų. Atstumų tarp pavyzdžių atveju savaime suprantamas pasirinkimas yra minimalaus atstumo kriterijus

$$D_{RZ} = \min_{\phi} D_{RZ}(\phi). \tag{3.32}$$

Taigi pavyzdžių palyginimas naudojant dinaminį laiko skalės kraipymo metodą yra minimizavimo uždavinys – tenka nustatyti trajektoriją, minimizuojančią atstumą tarp nagrinėjamų pavyzdžių.

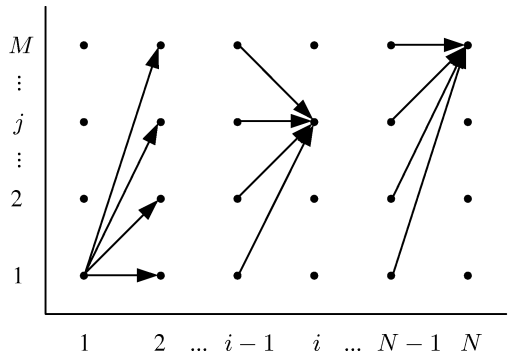
3.3.1.3. Dinaminis programavimas

Tiesiogiai ieškant minimalaus atstumo trajektorijos iškyla problema – didelis galimų trajektorijų skaičius. Jeigu mes nagrinėjame taškų tinklelį, sudarytą iš N stulpelių ir M eilučių, (kas atitinka pavyzdžių su N ir M požymių vektorių nagrinėjimą), galimų trajektorijų skaičius – M^N. Įvertinus tai, kad N ir M, priklausomai nuo pasirinktų analizės parametrų, būna kelių dešimčių eilės bei didesni ir tai, kad tenka skaičiuoti bei saugoti kiekvieną trajektoriją atitinkantį atstumą, pavyzdžių palyginimas tampa sunkiai realizuojamu. Šiam uždaviniui spręsti pasiūlytas dinaminio programavimo principas [100, 112].

Dinaminis programavimas – tai nuoseklaus optimizavimo metodas, daugelio optimizavimo etapų uždaviniuose leidžiantis pasiekti globalaus kriterijaus optimumą. Esminis dinaminio programavimo principas – kiekviename optimizavimo etape priimamas sprendimas garantuoja globalaus kriterijaus optimalumą. Kadangi kiekviename etape yra išsaugoma informacija apie priimtą sprendimą, paskutiniame etape pasiekus globalaus kriterijaus optimumą, galimas priimtų sprendimų sekos atstatymas. Dinaminio programavimo pritaikymas optimalios trajektorijos paieškai tinklelyje reiškia, kad bus nagrinėjamas baigtinis skaičius trajektorijų, iš kurių paskutiniame etape bus išrinkta viena, ir kad bus išsaugota informacija apie optimalios trajektorijos taškus.

Sudarydami dinaminio programavimo išraiškas optimalios trajektorijos paieškai, nagrinėsime 3.8 paveikslą bei priimsime sąlygą, kad trajektorija turėtų prasidėti taške $(1, 1)$ ir baigtis taške (N, M), čia N ir M – lyginamų pavyzdžių ilgiai.

Su kiekvienu tinklelio tašku susieti du dydžiai: $d(i, j)$ ir $D(i, j)$. $d(i, j)$ nurodo atstumą tarp pavyzdžių i-ojo ir j-ojo vektorių, $D(i, j)$ – suminį atstumą, apskaičiuotą trajektorijos, prasidedančios taške $(1, 1)$ ir pasibaigiančios (i, j), taškuose.



3.8 pav. *Optimalios trajektorijos tinklelyje paieška, naudojant dinaminio programavimo principą*

Visa tai įvertinus, trajektorijos paieškos, naudojant dinaminį programavimą, algoritmas

1. Inicializacija

$$D(1,1) = d(1,1); \tag{3.33}$$
$$P = (1,1). \tag{3.34}$$

2. Rekursija

$$D(i,j) = \min_k [D(i-1,k)] + d(i,j); \tag{3.35}$$
$$P = \arg\min_k [D(i-1,k)]; \tag{3.36}$$
$$\text{kai } i = 2,3,\ldots,N-1; \quad j,k = 1,2,\ldots,M.$$

3. Užbaigimas

$$D(N,M) = \min_k [D(N-1,k)] + d(N,M); \tag{3.37}$$
$$P = \arg\min_k [D(N-1,k)]; \tag{3.38}$$
$$\text{kai } k = 1,2,\ldots M.$$

Optimalios trajektorijos paieška pradedama taške $(1,1)$. Nagrinėjant tinklelio taškus, išrenkama tik viena trajektorija, ateinanti į tašką, taigi nagrinėjamų trajektorijų skaičius sumažėja iki M. Pasiekus tašką (N,M) telieka vienintelė galima trajektorija, kuri ir garantuoja minimalų atstumą tarp lyginamųjų pavyzdžių.

3.3.1.4. Laiko skalės kraipymo apribojimai

Svarbus reikalavimas atliekant kalbos pavyzdžių palyginimą – nuoseklumo laike išlaikymas. Tai reiškia, kad pavyzdžiai turi būti nagrinėjami nuo pradžios iki pabaigos, nuosekliai, lyginant tas pačias pavyzdžio dalis. Nesilaikant šio reikalavimo palyginimas tampa beprasmis. Pavyzdžiui vardą „aras“, nagrinėdami nuo pabaigos, jo neatskirsime nuo „sara“, nenuosekliai nagrinėdami „mada“, galime gauti maksimalų jo panašumą ir su „mada“, ir su „dama“, ir gal net su „adam“. Dinaminiame laiko skalės kraipyme palyginimo nuoseklumas realizuojamas apribojimais ir sąlygomis ieškomai sutapimo trajektorijai. Šių apribojimų visuma iš esmės ir sudaro kraipymo funkciją $\phi(i, j)$, kompensuojančią lyginamų pavyzdžių trukmių skirtumą. Aptarsime šiuos kraipymo trajektorijos apribojimus.

Galo taškų apribojimai. Šiais apribojimais nurodomi trajektorijos pradžios ir pabaigos taškai, t. y. nurodomos lyginamų pavyzdžių ribos. Paprastai prieš lyginant pavyzdžius yra atliekamas pavyzdžių ribų nustatymas, todėl ribos yra žinomos. Tokiu atveju galo taškų apribojimai

$$\phi(1,1) = (1,1); \tag{3.39}$$
$$\phi(N_Z, N_R) = (N, M). \tag{3.40}$$

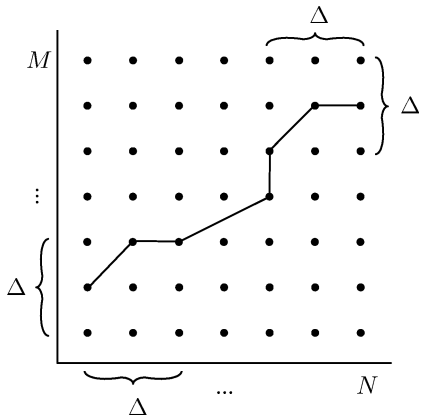
Tai reiškia, kad laiko skalės kraipymo trajektorija turi prasidėti taške $(1, 1)$ ir baigtis taške (N, M). Jei pavyzdžio ribos yra nežinomos arba jos nepatikimos, tenka taikyti šiek tiek laisvesnius reikalavimus

$$(1,1) \leqslant \phi(1,1) \leqslant (1,1) + \Delta; \tag{3.41}$$
$$(N, M) - \Delta \leqslant \phi(N_Z, N_R) \leqslant (N, M). \tag{3.42}$$

Šiuo atveju darome prielaidą, kad nagrinėjami pavyzdžių ribos gali nesutapti, todėl sudarome galimybę palyginimą pradėti nebūtinai pirmaisiais vektoriais ir užbaigti nebūtinai paskutiniaisiais (3.9 pav.).

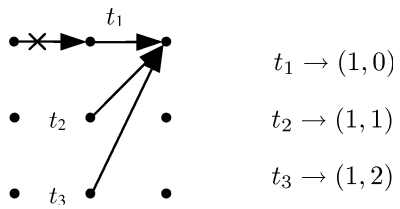
Konkrečiu atveju intervalai Δ galis skirtis pradžios ir pabaigos taškams ar ašims. Literatūroje [89] yra pateiktos konkrečios intervalų išraiškos, tačiau drįstume teigti, kad galo taškų apribojimų sumažinimas yra pakankamai euristinio pobūdžio ir sunkiai formalizuojamas.

Lokalus krypties apribojimas. Idealaus lyginamų pavyzdžių sutapimo atveju, gautoji trajektorija turėtų būti paieškos tinklelio įstrižainės aproksimacija (kadangi dėl skirtingų pavyzdžių trukmių ideali įstrižainė retai kada gaunasi) (3.7 a) pav.). Panašių pavyzdžių atveju trajektorija neturėtų nutolti nuo įstrižainės srities. Trajektorijos nutolimas nuo įstrižainės reiškia, kad nagrinėjami taškai, kurie atstovauja atstumą tarp skirtingose pavyzdžių dalyse esančių požymių vektorių, o toks palyginimas yra beprasmis, kadangi pažeidžiamas palyginimo nuoseklumas. Taigi formuodami galimas trajektorijas turime stengtis užtikrinti jų nenutolimą nuo įstrižainės. Reikiamą trajektorijos kryptį padeda išlaikyti lokalūs krypties apri-



3.9 pav. *Praplėstų galo taškų apribojimų įtaka trajektorijai*

bojimai. Jų formuluotės yra euristinio pobūdžio ir gali būti pateikti įvairiomis formomis. Paprasčiausi yra grafinio pateikimo būdas ir apribojimų pateikimas galimais trajektorijos koordinačių pokyčiais. Pirmajame rodyklėmis nurodoma, iš kurių taškų ir kokia trajektorija gali būti pasiektas nagrinėjamasis taškas. Antruoju būdu yra pateikiamos skaičių poros, kurios nurodo galimus koordinačių pokyčius ėjimo metu, t. y. nurodo galimas judėjimo kryptis. Kaip pavyzdį 3.10 paveiksle pateikiame Itakura pasiūlyto [46] lokalaus krypties apribojimo grafinę formą ir išraišką galimais koordinačių pokyčiais.



3.10 pav. *Itakura lokalus krypties apribojimas*

Lokalius apribojimus galima suskirstyti į dvi grupes: simetrinius ir asimetrinius. Simetriniuose apribojimuose nagrinėjamasis taškas gali būti pasiektas tinklelio įstrižainės atžvilgiu simetriškomis trajektorijomis. Asimetriniuose galimos trajektorijos yra nesimetriškos (3.10 pav.). Teigiama, jog simetriniai apribojimai leidžia pasiekti didesnį atpažinimo tikslumą [100]. Pasiūlyta nemažai krypties apribojimų, tačiau dėl jų euristinės prigimties apribojimus tarpusavyje galima palyginti tik eksperimentais. Apribojimų parinkimas taip pat nėra formalizuotas ir gali būti

argumentuojamas tik eksperimentų rezultatais.

Literatūroje [88] be lokalių krypties apribojimų dar naudojamas trajektorijos monotoniškumo reikalavimas. Juo iš esmės nurodoma, kad didėjant argumentui (nagrinėjamo taško koordinatėms tinklelyje), kraipymo funkcija turi nemažėti, t. y. paieškos kreivė negali būti mažėjanti arba pasukti atgal. Mūsų nuomone, lokalių krypties apribojimų naudojimas garantuoja paieškos trajektorijos monotoniškumą, todėl monotoniškumo reikalavimą galima laikyti lokalaus krypties apribojimo daliniu atveju ir jo neišskirti.

Prie lokalių apribojimų priskirtume ir krypties svorio koeficientus. Jų idėja – skirtingoms perėjimo į nagrinėjamąjį tašką kryptims suteikti skirtingus svorius. Svorio koeficientai į optimalios trajektorijos paiešką (3.33)–(3.38) įtraukiami kaip taškų atstovaujamų vektorių atstumų daugikliai. Taip galima valdyti trajektorijos pobūdį, neleidžiant per daug staigaus trajektorijos kilimo ar leidimosi žemyn. Vėlgi, egzistuojanti daugybė svorio koeficientų rinkinių rodo, kad šių koeficientų parinkimas ir naudojimas yra euristinio pobūdžio sprendimas, patikrinamas tik eksperimentais.

Globalus trajektorijos apribojimas. Aukščiau teigėme, jog tinklelio taškai, nutolę nuo tinklelio įstrižainės, atstovauja skirtingų pavyzdžių dalių požymių vektorių poras. Tokių vektorių palyginimas yra beprasmis, todėl siekiant atmesti tokius taškus, taikomas nagrinėjamų taškų srities apribojimas – globalus trajektorijos apribojimas. Siekdami paprastumo, apribojimo išraišką suskaidysime į dvi dalis – apribojimus iš kairės ir apribojimus iš dešinės

$$1 + (\phi(i) - 1)\,/Q_{\max} \leqslant \phi(j) \leqslant 1 + Q_{\max}\,(\phi(i) - 1)\,; \tag{3.43}$$

$$N_Z + Q_{\max}\,(\phi(i) - N_R) \leqslant \phi(j) \leqslant N_Z + (\phi(i) - N_R)\,/Q_{\max}. \tag{3.44}$$

$Q_{\max}$ – maksimalaus nuokrypio koeficientas, priklausantis nuo pasirinktojo lokalaus apribojimo tipo.

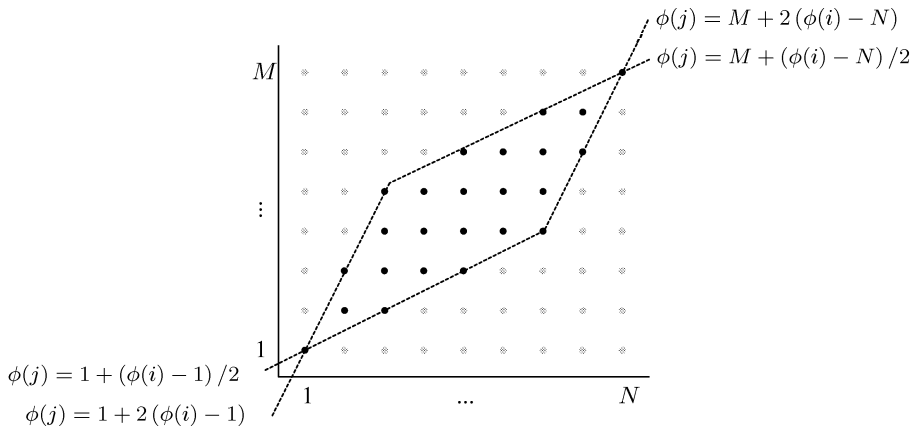
Globalus apribojimas taškų tinklelyje suformuoja lygiagretainio formos sritį, kurioje ir vykdoma optimalios trajektorijos paieška. 3.11 paveiksle pateiktas globalaus apribojimo pavyzdys (nagrinėjami tik juodos spalvos taškai, pilkos – atmetami). Kaip matome, globalus apribojimas atlieka labai svarbią funkciją – sumažina analizuojamų taškų aibę.

Jeigu išraiškos (3.43) kairiąją pusę sulyginsim su (3.44) dešiniąja puse, o (3.43) dešiniąją su (3.44) kairiąja, gausime ribines pavyzdžių ilgių išraiškas

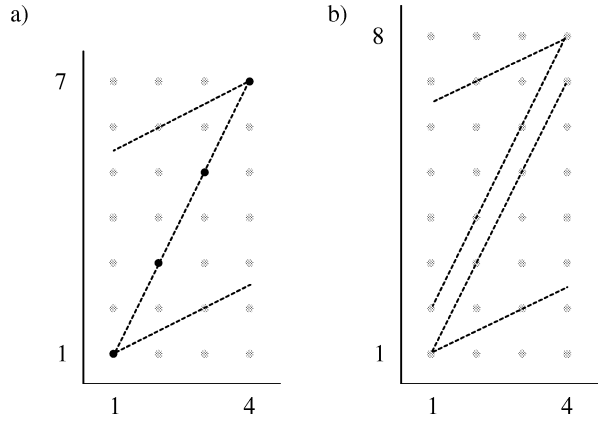
$$N_R = Q_{\max}\,(N_Z - 1) + 1; \tag{3.45}$$

$$N_Z = Q_{\max}\,(N_R - 1) + 1. \tag{3.46}$$

Jeigu nagrinėjamų pavydžių ilgiai tenkina vieną iš šių ribinių sąlygų, įmanoma vienintelė laiko skalės kraipymo trajektorija, sutampanti su tiesinio laiko skalės kraipymo linija (3.12 a) pav.). Jeigu N_Z ir N_R viršija (3.45) ir (3.46) reikšmes, dinaminis laiko skalės kraipymas tampa nebeįmanomas (3.12 b) pav.).



3.11 pav. *Globalus trajektorijos apribojimas, kai* $Q_{\max} = 2$



3.12 pav. *Globalaus apribojimo ribiniai atvejai: a) kai galima vienintelė trajektorija, b) kai jokia trajektorija negalima*

3.3.1.5. Metodo privalumai ir trūkumai

Nagrinėjant dinaminio laiko skalės metodo savybes galima įžvelgti, jog praktiškai visi metodo privalumai ir trūkumai yra nulemti naudojamos atpažinimo metodikos, kai nagrinėjamasis pavyzdys lyginamas su visais žodyno etalonais. Kaip metodo privalumus būtų galima paminėti:

- Algoritmo paprastumas. DLK algoritmas yra pakankamai paprastas, lengvai formuluojamas. Pritaikius dinaminio programavimo principą išvengiama

tiesioginio algoritmo taikymo trūkumo – milžiniško skaičiavimų kiekio.

- DLK metodas apsiriboja akustine pavyzdžių analize, išvengiant gramatikos, sintaksės ir kitų lingvistinės analizės lygių. Toks atsiribojimas reiškia, kad metodas yra nepriklausomas nuo nagrinėjamo lingvistinio vieneto – metodas gali būti pritaikomas tiek atskiriems žodžiams (ar net jų junginiams), tiek skiemenims, tiek garsams atpažinti.

- Nesudėtingas lingvistinio apdorojimo įdiegimas. Pavyzdžių palyginimo proceso pradiniai duomenys – akustinis signalas, rezultatas – su artimiausiu etalonu sutapdinta fonetinė transkripcija, kuri gali būti panaudota kaip duomenys lingvistiniam apdorojimui.

Kita vertus, metodo paprastumas ir orientavimas į akustinį lygmenį lemia trūkumus. Trūkumais įvardintume šias savybes:

- Atpažinimo tikslumo priklausomybė nuo etalonų skaičiaus. Kuo didesnis skaičius etalonų apibrėžtas konkrečiam pavyzdžiui, tuo mažesnė tikimybė, kad tas pavyzdys bus neatpažintas – tuo atpažinimo tikslumas bus aukštesnis.

- Pavyzdžio analizės trukmės priklausomybė nuo pavyzdžio ilgio. Kadangi pavyzdžio analizė atliekama nuosekliai, egzistuoja analizės trukmės priklausomybė nuo pavyzdžio trukmės – kuo ilgesnis pavyzdys, tuo ilgiau vyksta pavyzdžio analizė.

- Atpažinimo proceso trukmės priklausomybė nuo žodyno dydžio. Esminis pavyzdžių palyginimo metodų požymis – nagrinėjamojo pavyzdžio palyginimas su visais žodyno etalonais. Taigi kuo daugiau etalonų sistemos žodyne, tuo atpažinimo procesas ilgesnis.

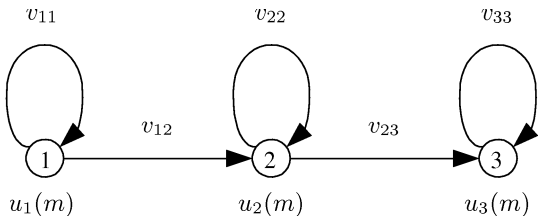
Kaip matome, bandydami spręsti pirmojo trūkumo sukeliamus klausimus, mes iš esmės paaštrinsime trečiojo trūkumo pasekmes. Įvertinus visus trūkumus, galima teigti, jog aukščiau išdėstytas (nemodifikuotas) dinaminis laiko skalės kraipymo metodas dėl savo nuoseklumo ir diskretumo (nagrinėjamas kiekvienas pavyzdys nuo pradžios iki galo) yra netinkamas ištisinei kalbai atpažinti. Sunkumų galima tikėtis ir atpažįstant pavienius žodžius tuo atveju, kai žodynas yra labai didelis.

3.3.2. Paslėptieji Markovo modeliai

Paslėptieji Markovo modeliai kalbai atpažinti buvo pritaikyti 8-ajame dešimtmetyje ir dabar yra vienas plačiausiai naudojamų metodų kalbai atpažinti. Dažniausiai PMM naudojami ištisinei kalbai atpažinti, tačiau neretai naudojami ir pavieniams žodžiams.

3.3.2.1. Statistiniai pavyzdžių modeliai

PMM panaudojimas kalbai atpažinti remiasi prielaida, kad kalbos signalas yra atsitiktinis procesas, kurio parametrus galima nustatyti. Šiame metode kalbos pavyzdžiai modeliuojami paslėptaisiais Markovo modeliais (3.13 pav.).



3.13 pav. *Trijų būsenų Markovo modelis*

Šiuose modeliuose nagrinėjamas dvigubas atsitiktinis procesas, kuriame vyksta du atsitiktiniai procesai: perėjimas iš vienos būsenos į kitą ir stebėjimas būsenoje. Dažniausiai naudojamas pirmos eilės Markovo modeliai, kuriuose perėjimo tikimybė priklauso tik nuo ankstesnės būsenos. Paslėptuoju vadinamas antrasis procesas, kadangi jis vykdomas per pirmąjį procesą ir tiesiogiai nestebimas. Perėjimus nusako tikimybių v_{ij} aibė, m-ojo simbolio b_m stebėjimą būsenoje i – tikimybių $u_i(b_m)$ aibė. Tokį pavyzdžio modelį nusako 5 parametrai: M – stebėjimo simbolių skaičius būsenoje, I – būsenų skaičius, V – perėjimų tikimybių pasiskirstymas, U – stebėjimų būsenoje tikimybių pasiskirstymas ir π – pradinio buvimo būsenose tikimybių pasiskirstymas (paprastumo dėlei tarsime, kad nagrinėjame diskrečiuosius atsitiktinius procesus). Toks modelis gali būti panaudotas kaip stebėjimų sekos $O = (o_1, o_2, \ldots, o_T)$ generatorius:

1. Laiko momentu 1 su tikimybe π_i patenkame į pradinę būseną $t_1 = i$.
2. Būsenoje i su tikimybe $u_i(b_m)$ stebime m-ąjį simbolį $o_1 = u_i(b_m)$.
3. Laiko momentu 2 su tikimybe v_{ij} pereiname į būseną $t_2 = j$.
4. Būsenoje j su tikimybe $u_j(b_m)$ stebime m-ąjį simbolį $o_2 = u_j(b_m)$.
5. Žingsnius 3 ir 4 kartojame visiems $t_1, t_2, \ldots, t_T$, taip gaudami stebėjimų vektorių $O = (o_1, o_2, \ldots, o_T)$.

Trumpumo dėlei modelis išreiškiamas trimis parametrais $\lambda = (V, U, \pi)$.

Beje, reikėtų pasakyti, kad 3.13 paveiksle pateiktoji modelio struktūra (ji vadinama „kairės-dešinės“ arba Bakis modeliu) nėra vienintelė įmanoma. Dažnai naudojamas ergodinis (kai visos būsenos tarpusavyje yra susietos perėjimais), lygiagretusis „kairės-dešinės“ (jame egzistuoja keletas lygiagrečių perėjimų iš kairės į dešinę) modeliai.

3.3.2.2. Pavyzdžių klasifikacija

Tarkime turime V kalbos pavyzdžių, kuriems atpažinti norime pritaikyti PMM metodą. Pirmasis žingsnis – žodyno sukūrimas. Kiekvienam iš V pavyzdžių sukuriame modelį λ. Modelio būsenų ir stebėjimų būsenoje skaičių nulemia pasirinktasis modelio tipas, modeliuojamas kalbos pavyzdys, priimtos pradinės prielaidos. Modelio tikimybiniai parametrai V, U ir π nustatomi taikant įvertinimo procedūras iš apmokymui pateiktųjų pavyzdžių (vienam etalonui sukurti reikia keleto pavyzdžio versijų).

Nagrinėjant nežinomąjį kalbos pavyzdį, mes atliekame signalo analizę, taip gaudami stebėjimų seką O (kiekvienas stebėjimas gali atitikti, pavyzdžiui požymių vektorių). Atpažintuoju pavyzdžiu paskelbiamas etaloninis pavyzdys, kurio modelis geriausiai atitinka nagrinėjamąją stebėjimų seką. Modelio atitikimą stebėjimų sekai įvertinant tikėtinumu, kad nagrinėjamoji stebėjimų seka yra sugeneruota modelio, atpažintuoju pavyzdžiu skelbiamas etalonas

$$Z = \arg \max_{1<k<V} P\left(O|\lambda_k\right).$$

Pritaikant PMM kalbai atpažinti tenka spręsti tris klausimus:

- Tikėtinumo $P\left(O|\lambda\right)$ skaičiavimas.
- Optimalios būsenų sekos parinkimas esant užduotai stebėjimų sekai O ir modeliui λ.
- Modelio parametrų, maksimizuojančių tikėtinumą $P\left(O|\lambda\right)$, parinkimas.

Tiesiogiai skaičiuodami tikėtinumo $P\left(O|\lambda\right)$ reikšmę, susidurtume su milžiniškomis skaičiavimo apimtimis – nagrinėjant T ilgio stebėjimų sekas ir N būsenų modelius, skaičiavimų kiekis būtų proporcingas $2TN^T$. Šis skaičius netgi esant mažiems T ir N gaunasi milžiniškas. Problemai spręsti pasiūlyti rekurentiniai skaičiavimo būdai – ėjimo į priekį, ėjimo atgal metodai [86]. Juose apibrėžiami tarpiniai kintamieji – daliniai tikėtinumai, kurių pagalba organizuojamas iteratyvus tikėtinumo skaičiavimas, kiekvieną dabartinę tarpinio kintamojo reikšmę išreiškiant per buvusiąją jo reikšmę. Naudojant šiuos metodus skaičiavimų kiekis, reikalingas $P\left(O|\lambda\right)$ apskaičiuoti, sumažėja iki TN^2.

Stebėjimų sekos ir modelio atžvilgiu optimalios būsenų sekos parinkimas yra kur kas sunkesnis uždavinys negu tikėtinumo skaičiavimas. Visų pirma, yra keletas galimų sekos optimalumo kriterijų. Antra, tiesioginis visų galimų sekų perrinkimas ir optimalios išrinkimas reiškia milžinišką skaičiavimų kiekį. Dažniausiai naudojamas tikėtiniausios sekos kriterijus, maksimizuojantis sekos tikėtinumą esant užduotai stebėjimų sekai ir modeliui. Maksimalaus tikėtinumo sekai nustatyti pasiūlytas nuoseklaus optimizavimo Viterbi algoritmas [113], realizuojamas dinaminiu programavimu. Kaip efektyvesnė alternatyva pasiūlytas modifikuotas Viterbi algoritmas, reikalaujantis mažesnio skaičiavimo operacijų skaičiaus [82].

Trečiasis PMM taikymo klausimas yra sunkiausiai sprendžiamas, kadangi neįmanoma analitiškai gauti modelio λ parametrų, maksimizuojančių stebėjimų sekos tikėtinumą $P(O|\lambda)$. Lokalus tikėtinumo maksimumas gali būti pasiektas modelio parametrams įvertinti naudojant Baum-Welch metodą, kuris iš esmės yra dalinis matematinės vilties maksimizavimo atvejis [86].

Pabaigai reiktų pasakyti, jog diskrečiaisiais Markovo modeliais sunku modeliuoti kalbos modelius, kadangi kalbos signalo analizės rezultatas būna tolydžiosios erdvės vektoriai, kas sunkiai suderinama su diskrečiais stebėjimais nagrinėtajame modelyje. Todėl realiose atpažinimo sistemose dažniausiai naudojami tolydieji Markovo modeliai, kuriuose stebėjimai modeliuojami tolydžiomis tikimybinio tankio funkcijomis, dažniausiai pasvertąja keleto Gauso tankio funkcijų suma.

3.3.2.3. Paslėptųjų Markovo modelių metodo modifikacijos

Kaip atpažinimo vienetas PMM metoduose naudoti įvairūs vienetai. Paprasčiausias atvejis yra pavienių žodžių atpažinimas. Šiuo atveju tenka kurti modelį kiekvienam žodžiui atskirai, taigi iš esmės neišvengiama trūkumo, būdingo pavyzdžių palyginimo metodikai – atpažinimo proceso trukmės priklausomybės nuo žodyno dydžio. Todėl pasiūlyti smulkesni atpažinimo vienetai, kurių skaičius paprastai yra ribotas – fonemos [59, 103] ir dvigarsiai [10], pusiau-skiemenys ir skiemenys [27]. Greta šių kalbos vienetų pasiūlyti ir dirbtiniai savo prigimtimi vienetai – trigarsiai, įvertinantys ir nagrinėjamosios fonemos aplinkines fonemas [104], klasterizacijos priemonėmis gaunami akustiniai segmentai – fenonai [5], su modelių būsenomis sutapdinami segmentai – senonai [44], konkrečios kalbos savybes modeliuojantys fonetiniai vienetai [13].

Kaip alternatyva maksimalaus tikėtinumo kriterijui vertinant modelio parametrus pasiūlytas maksimalios abipusės informacijos [4], minimalios diskriminantinės informacijos kriterijai [23]. Taikant pirmąjį maksimizuojama tikimybė, kad stebėjimų seka buvo sugeneruota nagrinėjamojo modelio ir tuo pačiu minimizuojama tikimybė, kad seką sugeneravo kiti modeliai. Šiuo atveju parametrams vertinti naudojami gradientiniai metodai. Tačiau šie metodai turi rimtų trūkumų – negarantuojamos parametrų konvergencija bei teigiamos parametrų reikšmės. Taikant antrąjį kriterijų ieškomi modelio parametrai, minimizuojantys stebėjimų sekos ir modelio tarpusavio entropiją, parametrus vertinant Baum-Welch metodu. Tačiau ir šiuo atveju modelio parametrų konvergencija negarantuojama.

Paslėptiesiems Markovo modeliams taikyti ir kiti sprendimai. Iš jų galima paminėti pusiau tolydžiųjų, antros eilės ir autoregresinių Markovo modelių taikymą kalbai atpažinti, buvimo būsenoje trukmės modeliavimą, stebėjimo negeneruojantys nuliniai perėjimai, modelio adaptavimo kalbančiajam procedūras.

3.3.2.4. Metodo privalumai ir trūkumai

Statistinis kalbos atpažinimo uždavinio sprendimas PMM metode lemia ne tik privalumus, dėl kurių jis tapo plačiausiai pripažintu ir vartojamu metodu ištisinei

kalbai atpažinti, bet ir trūkumus. Privalumais įvardintume šias metodo savybes:

- Statistinė analizė leidžia sudaryt daug bendresnį kalbos pavyzdžio modelį – etaloną, todėl pasiekiamas didesnis atpažinimo tikslumas ir nepriklausomybė nuo kalbėtojo.

- Kalbos pavyzdžių modeliavimas susietomis būsenomis leidžia nesunkiai į akustinį apdorojimą įjungti lingvistinio apdorojimo elementus. Be to, paslėptieji Markovo modeliai kalbos modelių forma gali tarnauti ir kaip lingvistinio apdorojimo metodas.

Metodo trūkumai:

- Sudarant etaloninių kalbos pavyzdžių modelius, dažnai į juos įtraukiama papildoma lingvistinė informacija. Tokie modeliai tampa priklausomais nuo tikslinės kalbos ir skirtingoms kalboms jie skiriasi (pvz., modeliai anglų ir lietuvių kalboms atpažinti skirsis). Todėl šiuo atveju pritaikant atpažinimo sistemą kitai kalbai, apmokymo kitos kalbos pavyzdžiais neužteks – teks kurti naujus modelius.

- Statistinė metodo prigimtis reikalauja nemažo apmokymo duomenų kiekio. Jeigu pavyzdžių palyginimo atveju netgi vieno kalbos pavyzdžio užtenka etalonui sukurti, tai PMM atveju minimalus reikalingas duomenų kiekis – bent kelios dešimtys kalbos pavyzdžio versijų.

- Egzistuoja pavyzdžių atpažinimo, naudojant PMM, tikslumo priklausomybė nuo apmokymo duomenų kiekio. Viena vertus, jei jis yra nepakankamas, galimas netikslus modelių parametrų įvertinimas. Kita vertus, per didelis duomenų kiekis gresia permokymu, kai modelių parametrai atspindi ne pavyzdžių klasės, o apmokymo pavyzdžių savybes. Abiem atvejais prarandamas atpažinimo tikslumas.

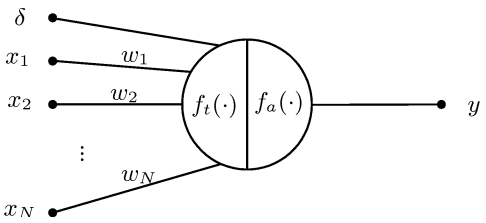
3.3.3. Dirbtiniai neuronų tinklai

Dirbtiniai neuronų tinklai, kalbai atpažinti pradėti taikyti praėjusio amžiaus devintajame dešimtmetyje, yra jauniausias iš nagrinėjamų metodų. Iš pradžių neuronų tinklai taikyti fonemoms [68, 115], skiemenims [101] atpažinti, vėliau atskiriems žodžiams [9, 25, 31] ir ištisinei kalbai [53]. Tačiau reiktų pasakyti, jog neuronų tinklai nepasiteisino kaip savarankiškas atpažinimo metodas (ypač ištisinės kalbos atpažinime) ir dažnai naudojami kartu su paslėptaisiais Markovo modeliais.

Nagrinėdami neuronų tinklų taikymą kalbai atpažinti apsiribosime perceptronu ir jo pagrindu sudarytais tinklais, trumpai paminėdami alternatyvius neuronų tinklų tipus.

3.3.3.1. Neurono modelis

Pagrindinis visų dirbtinių neuronų tinklų elementas – neuronas – supaprastintas biologinio neurono modelis, sudarytas iš branduolio su įėjimo ir išėjimo taškais (3.14 pav.).



3.14 pav. *Neurono modelis*

Neurono branduolį sudaro dvi funkcijos: tinklo ir aktyvuojanti. Tinklo funkcija nulemia įėjimo duomenų apdorojimą (dažniausiai tai būna pasvertųjų įėjimų sumavimas). Aktyvuojanti funkcija formuoja neurono išėjimo signalą, tiesiškai arba netiesiškai transformuodama įėjimų signalų darinį (dažniausiai naudojama sigmoidinė, rečiau slenksčio funkcija [64]). Tokio neurono modelio išėjimo signalas apibrėžiamas

$$y = f_a\left(\sum_{i=1}^{N} w_i x_i - \delta\right),$$

čia x_i – i-asis neurono įėjimas, w_i – įėjimo svorio koeficientas, δ – nuolatinė dedamoji, veikianti kaip įėjimo slenkstis, f_a – aktyvuojanti funkcija, y – neurono išėjimas.

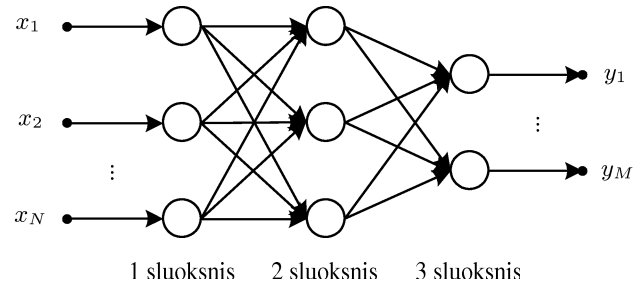
Taigi neurono reakciją į įėjimo signalą nulemia svorio koeficientai w_i, bei tinklo ir aktyvuojanti funkcijos.

3.3.3.2. Neuronų tinklai

Jungdami atskirus neuronus į masyvus gauname taip vadinamus neuronų tinklus (tinklą gali sudaryti ir vienas neuronas). Tinklai apibūdinami jų topologija, t. y. neuronų sujungimo schema. Neuronų tinklai dažniausiai pateikiami orientuotais grafais, kur mazgai atitinka atskirus neuronus, lankai nurodo duomenų sklidimo kryptį.

Daugiasluoksniame perceptrone neuronai yra organizuoti tarpusavyje sujungtomis grupėmis, taip vadinamais sluoksniais (3.15 pav.).

Sluoksniai, kurių išėjimo tiesiogiai nestebime, vadinami paslėptaisiais. 3.15 paveiksle pateiktoji struktūra turi du paslėptuosius sluoksnius – pirmą ir antrą. Paprastai vieno sluoksnio neuronai turi tą pačią aktyvuojančią funkciją, tačiau bendru atveju ji gali skirtis.



3.15 pav. *Daugiasluoksnio perceptrono su dviem paslėptaisiais sluoksniais struktūra*

Be perceptrono, objektams klasifikuoti nagrinėjami ir kitų tipų tinklai [88]: rekurentiniai tinklai, savaime besiformuojantys (Kohonen) tinklai, vėlinantys neuronų tinklai.

Viena iš svarbiausių neuronų tinklų savybių – sugebėjimas mokytis, t. y. keisti struktūrą pagal gaunamus duomenis. Mokymo procesas vyksta koreguojant svorio koeficientus w_i. Mokymas gali būti su mokytoju arba be mokytojo. Mokymo su mokytoju (taikomo perceptronui) metu tinklui pateikiama mokymo aibė, sudaryta iš įėjimo duomenų ir laukiamo išėjimo, o korekcija vykdoma pagal gaunamų rezultatų atitikimą pateiktiesiems. Mokymui be mokytojo (naudojamam Kohoneno tinklams) pateikiami tik įėjimo duomenys, o svorio koeficientų korekcija vyksta siekiant panašiems įėjimo signalams gauti atitinkamo panašumo išėjimo signalus. Sukurta nemažai tinklų apmokymo procedūrų, tačiau daugumai jų būdingas nedidelis sprendimo konvergavimo greitis.

3.3.3.3. Kalbos atpažinimas naudojant neuronų tinklus

Neuronų tinklų sugebėjimas atvaizduoti N įėjimų į M išėjimų daro juos tinkamais klasifikacijai. Kaip įėjimo signalus pateikę kalbos pavyzdžio analizės duomenis (požymių vektorius ar jų elementus) galime atlikti kalbos pavyzdžių klasifikaciją. Tinklams mokinti tokiu atveju tenka naudoti segmentuotą ir žymėtą kalbos signalą. Segmentavimo lygmenį lemia nagrinėjamas kalbos pavyzdys – atpažįstant fonemas signale turi būti nurodytos fonemų ribos ir jas atitinkančios klasės žymė, pavienių žodžių atveju – žodžių ribos.

Kalbai atpažinti dažniausiai naudotas daugiasluoksnis perceptronas. Nedidelių žodynų atveju atpažinimas neuronų tinklais tikslumu beveik prilygo paslėptiesiems Markovo modeliams. Neigiama perceptrono savybė (beje, būdinga daugeliui neuronų tinklų tipų) – pakankamai sudėtingas apmokymo etapas. Tinklui mokinti reikia turėti segmentuotus ir žymėtus kalbos pavyzdžius, o pats mokymas yra pakankamai ilgas procesas.

Dėl statiškos prigimties, daugiasluoksnių perceptronų panaudojimo galimybės kalbai atpažinti yra pakankamai ribotos. Rekurentiniuose tinkluose galimi atgaliniai ryšiai tarp išėjimo ir įėjimo bei ryšiai tarp to paties sluoksnio neuronų. Dalinai rekurentiniuose (tarpiniuose tarp daugiasluoksnių perceptronų ir rekurentinių) tinkluose yra tik tam tikri grįžtamieji ryšiai (pavyzdžiui, grįžtamasis ryšys su ankstesniu sluoksniu). Tokie tinklai pasižymi dinaminėmis savybėmis – išėjimo signalas priklauso nuo buvusių įėjimo reikšmių, bei nepriklauso nuo to, ar yra įėjime signalas nagrinėjamu momentu. Dėl šių savybių rekurentiniai tinklai gali būti panaudoti laike kintančiam kalbos signalui analizuoti ir atpažinti. Tačiau jų panaudojimą apsunkina kur kas sudėtingesnis nei perceptronų tinklo apmokymas.

Vėlinančiuose neuronų tinkluose [115] kiekvienas įėjimo signalas (tiek tinklo išorėje, tiek viduje) perleidžiamas per keletą vėlinimų elementų, taip nagrinėjamąjį vektorių susiejant su ankstesniais įėjimais. Toks susiejimas leidžia įvertinti laikinius ryšius tarp skirtingų įėjimų ir padidina klasifikacijos tikslumą.

Paslėpto valdymo neuronų tinkluose [61] neurono išėjimo signalas priklauso ir nuo patenkančių duomenų ir nuo papildomai į neuroną įvedamo valdymo signalo. Šis valdymo signalas suteikia galimybę valdyti neuronų išėjimus ir taip realizuoti apdorojimo kitimą laike, suteikiant tinklui dinaminių savybių.

3.3.3.4. Metodo privalumai ir trūkumai

Neuronų tinklo, kaip kalbos atpažinimo metodo, privalumais įvardintume šias savybes:

- Neuronų tinklų sugebėjimas mokytis, t. y. adaptuotis įvertinus gautus naujus duomenis. Ši savybė puikiai tinka kalbos atpažinimo uždaviniams, kuriuose visada susiduriama su žodyne neesančiu žodžiu, su nauju kalbėtoju – aplinkybėmis, prie kurių sistema turėtų adaptuotis.

- Neuronų tinklų skaičiavimų lygiagretumas ir netiesiškumas. Pakankami didelis neuronų tinklas gali aproksimuoti bet kurią netiesišką sistemą, reikalaujančią didelių skaičiavimo pajėgumų – ši savybė labai patraukli kalbos signalo apdorojime.

Kita vertus, neuronų tinklų sugebėjimas mokytis, skaičiavimų lygiagretumas ir netiesiškumas nulemia ir trūkumus, kuriais įvardintume:

- Literatūroje neaptikta metodikos, kurios pagalba būtų galima nustatyti, kokia tinklo topologija, kokios tinklo charakteristikos (įėjimo ir išėjimų skaičius, neuronų skaičius sluoksnyje, sluoksnių skaičius, pradinės svorio koeficientų reikšmės, ryšiai tarp neuronų) tinkamiausios konkrečiam atpažinimo uždaviniui.

- Tarp neuronų tinklo topologijos, apmokymo duomenų aibės dydžio, apmokymo laiko ir klasifikacijos tikslumo egzistuoja tarpusavio ryšys. Pavyz-

džiui, didėjant tinklui, didėja ir reikalingas apmokymo duomenų kiekis, ilgėja mokymo laikas, auga klasifikacijos tikslumas. Tačiau metodikos, konkrečiai apibrėžiančios minėtus ryšius ir formalizuojančios sprendimus, neaptikta.

- Griežtas ryšys tarp analizės duomenų eilės, klasių skaičiaus ir tinklo topologijos. Pasirinktos analizės eilė turi būti lygi neuronų tinklo įėjimų skaičiui, o išskiriamų klasių skaičius – išėjimų skaičiui. Kadangi analizės eilės arba klasių skaičiaus pakeitimas reiškia naujo neuronų tinklo sukūrimą, dažnai sprendimai apie duomenis priimami remiantis tinklo topologija arba atvirkščiai – topologija sudaroma remiantis duomenų savybėmis, t. y. optimizuojama tik dalis atpažinimo sistemos parametrų.

Šie neuronų tinklų trūkumai yra sunkiai sprendžiami ir paprastai tam taikomi euristiniai metodai.

3.4. Trečiojo skyriaus apibendrinimas

- Kalbos signalas yra perteklinis. Todėl kalbos signalui analizuoti naudojami požymiai – analizei svarbias savybes atspindintys duomenys. Šiuo metu kalbos atpažinime signalui analizuoti plačiausiai naudojama melų skalės kepstro analizė.

- Nors skirtingi atpažinimo metodai remiasi skirtingomis prielaidomis, naudoja skirtingus klasifikacijos kriterijus, nė vienas jų neturi akivaizdaus pranašumo prieš kitus atpažinimo tikslumo, efektyvumo požiūriu.

- Kiekvienas atpažinimo metodas turi savų trūkumų ir pranašumų, todėl manytume jog galimà atpažinimo sistemų vystymosi kryptis – hibridinių metodų, apjungiančių teigiamas skirtingų metodų savybes, kūrimas ir naudojimas.

- Sudarant atpažinimo sistemą tenka spręsti šiuos uždavinius: požymių sistemos parinkimas, žodžio ribų nustatymas (pavienių žodžių atpažinimo atveju) ir segmentavimas (ištisinės kalbos ir smulkesnių nei žodis vienetų atpažinimo atveju), klasifikacijos metodo parinkimas.

4

Atpažinimo sistemos realizacija

Šiame skyriuje pristatysime darbe sukurtus ir realizuotus sprendimus atpažinimo problemoms spręsti bei pavienių žodžių atpažinimo sistemą, kurioje tie sprendimai buvo realizuoti. Pagrindinis kuriamos sistemos vertinimo kriterijus – atpažinimo tikslumas, kurio didinimą pasirinkome svarbiausiu tikslu. Darbo metu nekeltas klasifikacijos arba tiesiog atpažinimo metodo, analizės tipo ir atstumo skaičiavimo klausimai. Visi pasiūlyti sprendimai iš esmės gali būti apjungti su bet kuriuo atpažinimo metodu ar signalo analizės būdu.

Kaip atpažinimo metodą pasirinkome dinaminį laiko skalės kraipymą. Pasirinkimą lėmė įsitikinimas, kad pavieniams žodžiams atpažinti dinaminis laiko skalės kraipymo metodas yra tinkamesnis dėl savo paprastumo, efektyvumo ir nedidelių reikalavimų apmokymo aibėms. Signalui analizuoti panaudojome tiesinės prognozės modelio ir tiesinės prognozės modelio kepstro analizės metodus.

Nagrinėdami atpažinimo sistemos realizaciją segmentais vadinsime segmentavimo metu gautas žodžio atkarpas, o garsais – segmentus, atitinkančius elementarius gramatikos vienetus – raides.

4.1. Galimos atpažinimo proceso tobulinimo kryptys

Kaip jau nagrinėjome praeitame skyriuje, pavienių žodžių atpažinimas naudojant dinaminį laiko skalės kraipymą ir dinaminį programavimą, pasižymi ir tam tikrais privalumais, ir trūkumais. Nenagrinėdami atpažinimo metodo ir naudoja-

mos požymių sistemos klausimų suformulavome tokias galimas pavienių žodžių atpažinimo proceso tobulinimo kryptis:

- **Smulkus nagrinėjamasis kalbos vienetas**. Esminis pavienių žodžių atpažinimo proceso trūkumas – nagrinėjamo kalbos vieneto trukmė. Parinkę smulkesnius atpažinimo vienetus galime sumažinti etalonų skaičių, kadangi smulkesni vienetai dažniau pasikartoja kalbos pavyzdžiuose. Be to, smulkesnis atpažinimo vienetas lemia trumpesnę vieno etalono analizę. Taigi smulkesnis sistemos nagrinėjamasis vienetas gali padėti sumažinti etalonų skaičių tam pačiam žodyno dydžiui bei paspartinti atpažinimo procesą.

- **Etalonų optimalumas**. Vienas iš atpažinimo proceso atsparumo atpažinimo sąlygoms (įvairių tipų triukšmams, naudojamai techninei įrangai, kalbėtojo būsenai bei pačiam kalbėtojui) didinimo būdų yra keleto etalonų kūrimas vienam kalbos pavyzdžiui. Tokio sprendimo trūkumas – M kartų padidinamas etalonų skaičių (čia M – kuriamų etalonų skaičius vienam kalbos pavyzdžiui). Todėl etalonai turėtų būti kuriami iš tam tikrais optimalumo kriterijais remiantis išrinktų pavyzdžių, taip siekiant padidėjusiu atpažinimo tikslumu kompensuoti išaugusį skaičiavimų kiekį.

- **Neperspektyvaus etalono atmetimas**. Įprastomis sąlygomis palyginimo procese visi etalonai nagrinėjami nuo pradžios iki galo. Turint omeny, kad dalis etalonų (kai kuriais atvejais didžioji dalis) yra neperspektyvūs rezultatų prasme, turėtų būti numatyta galimybė nagrinėti tarpinius palyginimo rezultatus. Procesą, kurio tarpiniai rezultatai būtų nepatenkinami, reikėtų nutraukti, taip atsisakant neperspektyvių etalonų analizės. Tai leistų sumažinti vidutinį vieno etalono analizės laiką, o tuo pačiu ir atpažinimo proceso trukmę.

- **Palyginimo proceso optimizavimas**. Proceso trukmę pavienių žodžių atpažinimo atveju lemia palyginimo proceso trukmė. Dinaminiam laiko skalės kraipymo metodui būdingas pasikartojantis atstumų tarp požymių vektorių skaičiavimas ir papildomų sąlygų vertinimas (pvz., lokalių paieškos krypties ir globalių trajektorijos apribojimų). Optimizavus palyginimo algoritmą, tokių pasikartojančių veiksmų skaičių galima būtų sumažinti.

Savo darbe mes sprendėme etalonų optimalumo ir neperspektyvių etalonų atmetimo klausimus. Panaudoję žodžių segmentavimą į garsus, pasiūlėme sumažinto kalbos vieneto sprendimą bei supaprastinome palyginimo procesą.

4.2. Kalbos atpažinimo ir segmentavimo sistema

Darbo metu sukurta pavienių žodžių atpažinimo ir segmentavimo sistema KAS – Kalbos Atpažinimas ir Segmentavimas. Sukurtojoje sistemoje įdiegti trys darbo

režimai: pavienių žodžių atpažinimo, žodžių segmentavimo ir žodžio segmentų (garsų) atpažinimo. Kiekvienam darbo režimui būdingi savi parametrai, valdymo elementų rinkiniai ir rezultatų pateikimas.
Sistemos darbas organizuotas sesijomis – kiekvienai sesijai sukuriama atskira direktorija, kurioje saugomi visi failai – žodžio ribų, požymių, žodyno. Darbo metu pasirinkus sesiją, į atmintį įkraunami atitinkamos direktorijos failų duomenys.

Sekančiuose poskyriuose atpažinimo sistemą pristatysime nagrinėdami kiekvieną režimą atskirai.

4.3. Žodžių atpažinimas

Pirmasis iš galimų sistemos KAS darbo režimų – pavienių žodžių atpažinimas. Atpažinimas nėra susietas su kalbos pavyzdžiu, todėl atpažinimui gali būti pateiktas tiek garsas, tiek garsų junginys, tiek žodžių junginys. Nagrinėdami sistemos realizaciją, sistemą traktuosime kaip pavienių žodžių atpažinimo.

4.3.1. Žodžių atpažinimo algoritmas

Pirmasis atpažinimo proceso etapas – kalbos signalo įvedimas. Įvestasis kalbos signalas apdorojamas, nustatomos žodžio ribos. Nustatytose žodžio ribose atliekama analizė, kurios rezultatas – požymių vektorių seka. Gautoji seka gali būti panaudota sistemai mokyti (sukurti ištarto žodžio etaloną) arba pateikta žodžiui atpažinti. Atpažinimo rezultatas – su nagrinėjamam žodžiui artimiausiu etalonu susieta fonetinė transkripcija. Sudaryto žodžių atpažinimo algoritmo schema pateikta 4.1 paveiksle.

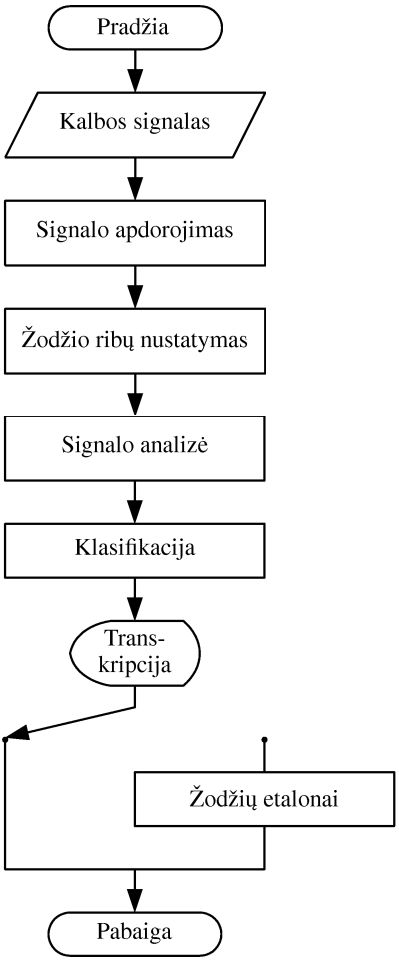
Toliau trumpai aptarsime kiekvieną žodžių atpažinimo algoritmo etapą ir priimtus sprendimus juose.

4.3.2. Kalbos signalo įvedimas

Sistemoje kalbos signalas įvedamas dviem būdais: įrašomas per mikrofoną arba nuskaitomas iš failo. Naudojama garso kokybė – 11025 Hz diskretizacijos dažnio, 16 bitų skiltiškumo, vieno kanalo garsas. Kaip šaltinis taip pat gali būti panaudotas impulsais moduliuoto garso failas (*Windows* operacinėje sistemoje – *wav* tipo).

4.3.3. Signalo apdorojimas

Kalbos signalui apdoroti sistemoje taikomos dvi procedūros: nuolatinės dedamosios atėmimas ir filtravimas I eilės RIR aukštų dažnių filtru.



4.1 pav. *Žodžių atpažinimo algoritmas*

Nuolatinės dedamosios atėmimas realizuotas kaip signalo reikšmių vidurkio atėmimas

$$\tilde{x}(n) = x(n) - \frac{1}{N}\sum_{n=1}^{N} x(n), \quad \text{kai } n = 1, 2, \ldots, N, \tag{4.1}$$

čia $\tilde{x}(n)$ – apdoroto signalo atskaita, $x(n)$ – apdorojama atskaita, N – signalo ilgis atskaitomis.

Nuolatinę dedamąją signale lemia naudojama garso įrašymo įranga ir jos nustatymai, todėl dedamosios pašalinimas sumažina įrangos įtaką signalo analizei.

Sekantis apdorojimo etapas – filtravimas aukštų dažnių filtras. Tam naudo-

jamas pirmos eilės ribotos impulsinės reakcijos filtras, kurio sistemos funkcija apibrėžiama (3.24) išraiška. Keičiamo filtro koeficiento reikšmė pagal nutylėjimą prilyginta 0,95. Filtravimas aukštų dažnių filtru „išlygina“ signalo spektrą, sumažindamas žemų dažnių įtaką signalo analizei.

4.3.4. Žodžio ribų nustatymas

Esminis pavienių žodžių atpažinimo proceso momentas, lemiantis atpažinimo tikslumą – žodžio ribų signale nustatymas. Sistemoje KAS realizuoti du žodžio ribų nustatymo metodai.

Pirmasis jų – energijos slenkstis (ES). Tai klasikinis ribų nustatymo metodas, žodžio ribomis įvardijantis momentus, kuriais signalo energijos reikšmės viršija iš anksto apibrėžtas slenksčio reikšmes. Žodžio pradžia aptinkama signalo energijos reikšmes nagrinėjant nuo signalo pradžios, pabaiga – nuo signalo pabaigos. Siekiant išvengti atsitiktinių energijos šuolių impulsų įtakos nagrinėjama keletas ne iš eilės einančių kadrų energijos reikšmės. Be to, signalo kadrų energijos reikšmės papildomai filtruojamos medianiniu filtru, taip siekiant sumažinti pavienių energijos impulsų šuolių įtaką skaičiavimams.

Kalbos signalą galima nagrinėti kaip atsitiktinį signalą su besikeičiančiomis savybėmis. Remdamiesi šia prielaida, žodžio riboms rasti pritaikėme atsitiktinių sekų savybių pasikeitimo momentų radimo metodą [A6], tardami, kad nagrinėjamame signale yra du pasikeitimo momentai – žodžio pradžia ir pabaiga. Signalui modeliuoti naudojamas tiesinės prognozės (autoregresijos) modelis.

4.3.4.1. Pasikeitimo momentų nustatymo uždavinys

Nagrinėkime atsitiktinę seką $x = \{x(1), x(2), \ldots, x(N)\}$, gaunamą tiesinės, diskretinės, laike kintančios sistemos išėjime. Sistema aprašoma tiesinės prognozės lygtimi

$$x(n) = -a_1(n) \cdot x(n-1) - \cdots - a_p \cdot x(n-p) + b(n) \cdot v(n).$$

Paprastumo dėlei tarkime, kad parametrai yra žinomi ir tenkina sąlygą

$$A(n) = \begin{cases} A_1, & \text{kai } n = 1, 2, \ldots, u_1; \\ A_2, & \text{kai } n = u_1 + 1, \ldots, u_2; , \\ A_3, & \text{kai } n = u_2 + 1, \ldots, N, \end{cases} \tag{4.2}$$

čia $A(n) = [a_1(n), a_2(n), \ldots, a_p(n), b(n)]$ – modelio parametrai, u_1 ir u_2 – šuoliško parametrų pasikeitimo momentai, tenkinantys sąlygą $p < u_1 < u_2 < N$. Žodžio ribų nustatymo uždavinys – rasti pasikeitimo momentų įverčius $\hat{u} = [\hat{u}_1, \hat{u}_2]$, kai turime konkrečią signalo x realizaciją.

4.3.4.2. Pasikeitimo momentų nustatymo uždavinio sprendimas

Pasikeitimo momentų įverčių radimas gali būti nagrinėjamas kaip klasifikavimo uždavinys, sprendžiamas maksimizuojant aposteriorinę klasifikavimo tikimybę. Pasikeitimo momentų rinkinio tikimybė esant konkrečiai signalo realizacijai apskaičiuojama pagal Bejeso formulę

$$P(u|x) = \frac{p(x|u) \cdot P(u)}{p(x)}, \tag{4.3}$$

čia $p(x|u)$ – sąlyginis signalo realizacijos x tikimybinis tankis fiksuotam pasikeitimo momentų rinkiniui, $P(u)$ – pasikeitimo momentų rinkinio tikimybė, $p(x)$ – signalo realizacijos x tikimybinis pasiskirstymas.

Tokiu atveju labiausiai galimas pasikeitimo momentų rinkinys bus

$$\hat{u} = \arg\max_{u} P(u|x), \tag{4.4}$$

t. y. pasikeitimo momentų rinkinio turime ieškoti maksimizuodami (4.3) išraišką. Įvertinus tai, kad konkrečiai realizacijai x išraiškos (4.3) vardiklis yra pastovus dydis, o tikimybė $P(u)$ yra nežinoma (ir laikoma pastovia), labiausiai galimas pasikeitimo momentų rinkinys bus

$$\hat{u} = \arg\max_{u} p(x|u). \tag{4.5}$$

Konkrečiai realizacijai x sąlyginis tikimybinis tankis $p(x|u)$ tampa pasikeitimo momentų rinkinio u tikėtinumo funkcija $l(u|x)$, o argumentai maksimizuojantys šią funkciją – maksimalaus tikėtinumo įverčiais

$$\hat{u} = \arg\max_{u} l(u|x). \tag{4.6}$$

Tikėtinumo funkcija $l(u|x)$ gali būti išreikšta kaip sąlyginio tankio funkcija $p(x|u)$ esant fiksuotai x realizacijai

$$\begin{aligned} l(u|x) = p\left(x(1), x(2), \ldots, x(p)\right) \cdot (2\pi)^{-(N-p)/2} \cdot b^{-(u_1-p)}(1) \times \\ \times\, b^{-(u_2-u_1)}(2) \times \cdot b^{-(N-u_2)}(3) \cdot \exp\Bigg(\frac{1}{2b^2(1)} \times \\ \times \sum_{n=p+1}^{u_1} \left(\sum_{j=0}^{p} a_j(1) x(n-j) \right)^2 - \frac{1}{2b^2(2)} \times \end{aligned}$$

$$\times \sum_{n=u_1+1}^{u_2} \left(\sum_{j=0}^{p} a_j(2)x(n-j) \right)^2 - \frac{1}{2b^2(3)} \times \\ \times \sum_{n=u_2+1}^{N} \left(\sum_{j=0}^{p} a_j(3)x(n-j) \right)^2 \Bigg). \tag{4.7}$$

Tiesioginis tikėtinumo funkcijos (4.7) maksimizavimas yra netikslingas dėl didelės skaičiavimų apimties, todėl tikėtinumo funkciją tenka pertvarkyti. Siekiant sandaugas pakeisti sumomis, tikėtinumo funkcija logaritmuojama

$$\hat{u} = \arg\max_u l(u|x) = \arg\max_u \log l(u|x), \tag{4.8}$$

čia

$$\begin{aligned} \log l(u|x) = & \log p\left(x(1), x(2), \ldots, x(p)\right) - \\ & - (N-p)/2 \log(2\pi) - (u_1 - p)\log b(1) - \\ & - (u_2 - u_1)\log b(2) - (N - u_2)\log b(3) - \\ & - \frac{1}{2b^2(1)} \sum_{n=p+1}^{u_1} \left(\sum_{j=0}^{p} a_j(1)x(n-j) \right)^2 - \\ & - \frac{1}{2b^2(2)} \sum_{n=u_1+1}^{u_2} \left(\sum_{j=0}^{p} a_j(2)x(n-j) \right)^2 - \\ & - \frac{1}{2b^2(3)} \sum_{n=u_2+1}^{N} \left(\sum_{j=0}^{p} a_j(3)x(n-j) \right)^2 . \end{aligned} \tag{4.9}$$

Išraiškoje (4.9) pašalinę dėmenis, nepriklausančius nuo pasikeitimo taškų, gausime naują funkciją $\theta(u|x)$, kurios maksimumo vieta sutampa su maksimizuojamos funkcijos $\log l(u|x)$ maksimumo vieta

$$\hat{u} = \arg\max_u \log l(u|x) = \arg\max_u \theta(u|x), \tag{4.10}$$

čia

$$\theta(u|x) = l_1(u_1|x) + l_2(u_2|x). \tag{4.11}$$

Kiekviena funkcija $l_i(u_i|x)$ – dalinė tikėtinumo funkcija – priklauso tik nuo

vieno pasikeitimo momento u_i ir išreiškiama

$$\begin{aligned} l_i(k|x) = &-(k-p)\log b(i) - (N-k)\log b(i+1) - \\ &- \frac{1}{2b^2(i)} \sum_{n=p+1}^{k} \left(\sum_{j=0}^{p} a_j(i) x(n-j) \right)^2 - \\ &- \frac{1}{2b^2(i+1)} \sum_{n=k+1}^{N} \left(\sum_{j=0}^{p} a_j(i+1) x(n-j) \right)^2, \\ &\text{kai} \quad i = 1, 2; \quad k = p+1, p+2, \ldots, N. \end{aligned} \tag{4.12}$$

Siekiant dar labiau sumažinti reikalingų skaičiavimo operacijų kiekį, dalinėms tikėtinumo funkcijoms skaičiuoti gautos rekurentinės išraiškos

$$\begin{aligned} l_i(k|x) = &\, l_i(k-1|x) - \log b(i) + \log b(i+1) - \\ &+ \frac{1}{2b^2(i)} \left(\sum_{j=0}^{p} a_j(i) x(n-j) \right)^2 + \\ &+ \frac{1}{2b^2(i+1)} \left(\sum_{j=0}^{p} a_j(i+1) x(n-j) \right)^2, \\ &\text{kai} \quad i = 1, 2; \quad k = 2, 3, \ldots, N. \end{aligned} \tag{4.13}$$

Kadangi pradinės sąlygos nepriklauso nuo pasikeitimo momentų, (4.13) išraiškai skaičiuoti galima naudoti nulines pradines sąlygas.

Funkcija $\theta(u|x)$ yra dviejų vieno kintamojo funkcijų suma, taigi jai maksimizuoti galima taikyti dinaminio programavimo metodą (DP). Tuo tikslu apibrėžiamos Belmano funkcijos

$$g_1(u_2|x) = \max_{\substack{u_1 \\ p<u_1<u_2}} l_1(u_1|x), \tag{4.14}$$

$$g_2(u_3|x) = \max_{\substack{u_2 \\ p+1<u_2<u_3}} \left[l_2(u_2|x) + g_1(u_2|x) \right], \tag{4.15}$$

$$\text{kai} \quad u_3 = p+3, p+4, \ldots, N.$$

Tuomet pasikeitimo momentai randami kaip minimalios Belmano funkcijų maksimumų argumentų reikšmės

$$\hat{u}_k = \min\left[\arg \max_{\substack{n \\ p+k<n<\hat{u}_{k+1}}} g_k(n|x)\right], \qquad (4.16)$$
$$\text{kai} \quad k = 2, 1; \quad \hat{u}_3 = N.$$

Nagrinėdami pasikeitimo momentų tikėtinumo funkcijos maksimizavimą darėme prielaidą, kad signalo tiesinės prognozės modelio parametrai yra žinomi. Nagrinėjant realius signalus modelio parametrai yra nežinomi. Tokiu atveju tikėtinumo funkcijai maksimizuoti galima naudoti apibendrintą matematinės vilties maksimizavimo algoritmą. Be to, tiesioginė signalo analizė gali būti pakeista signalo kadrų energijos analize. Būtent šie sprendimai priimti realizuojant žodžio ribų nustatymo metodą.

4.3.4.3. Žodžio ribų nustatymas iš trumpalaikės signalo energijos

Atlikę signalo $x(n)$ apdorojimą ir kadrų energijos skaičiavimus, turime energijos reikšmių seką $e = \{e(1), e(2), \ldots, e(K)\}$, kuri tarkime, yra atsitiktinė. Darome prielaidą, kad energijos reikšmės yra nepriklausomi normalieji atsitiktiniai dydžiai. Tuomet galime užrašyti

$$A(k) = \begin{cases} A_1 = N\left(\mu_1, \sigma_1^2\right), & \text{kai } k = 1, 2, \ldots, u_1; \\ A_2 = N\left(\mu_2, \sigma_2^2\right), & \text{kai } k = u_1 + 1, \ldots, u_2; \\ A_3 = N\left(\mu_3, \sigma_3^2\right), & \text{kai } k = u_2 + 1, \ldots, K, \end{cases} \qquad (4.17)$$

čia A_1 ir A_3 – aplinkos triukšmo (tylos atkarpų) energijos parametrai prieš ir po žodžio, A_2 – žodžio energijos parametrai, μ_i ir σ_i – kadrų energijos reikšmių atkarpose vidurkiai ir dispersijos, u_1 ir u_2 – parametrų pasikeitimo momentai, tenkinantys sąlygą $1 < u_1 < u_2 < K$.

Žodžio pradžią ir pabaigą atitinkančius parametrų pasikeitimo momentų įverčius $\hat{u} = [\hat{u}_1, \hat{u}_2]$ nustatysime maksimizuodami tikėtinumo funkcijos logaritmą (4.8). Šiuo atveju tikėtinumo funkcijos logaritmas išreiškiamas

$$\begin{aligned} \log l(u|x) = &-\frac{K}{2}\log(2\pi) - u_1 \log\sigma_1 - (u_2 - u_1)\log\sigma_2 - \\ &- (K - u_2)\log\sigma_3 - \frac{1}{2\sigma_1^2}\sum_{k=1}^{u_1}(x(k) - \mu_1)^2 - \\ &- \frac{1}{2\sigma_2^2}\sum_{k=u_1+1}^{u_2}(x(k) - \mu_2)^2 - \frac{1}{2\sigma_3^2}\sum_{k=u_2+1}^{K}(x(k) - \mu_3)^2. \end{aligned} \qquad (4.18)$$

Dalinės tikėtinumo funkcijos (4.12) tampa

$$\begin{aligned} l_i(k|x) = &-k\log\sigma_i - (K-k)\log\sigma_{i+1} - \\ &- \frac{1}{2\sigma_i^2}\sum_{m=1}^{k}\left(x(m)-\mu_i\right)^2 - \\ &- \frac{1}{2\sigma_{i+1}^2}\sum_{m=k+1}^{K}\left(x(m)-\mu_{i+i}\right)^2, \\ &\text{kai} \quad i=1,2; \quad k=1,2,\ldots,K; \end{aligned} \tag{4.19}$$

o jų rekurentinės išraiškos su nulinėmis pradinėmis sąlygomis

$$\begin{aligned} l_i(k|x) = &\, l_i(k-1|x) - \log\sigma_i + \log\sigma_{i+1} - \\ &- \frac{1}{2\sigma_i^2}\left(x(k)-\mu_i\right)^2 + \frac{1}{2\sigma_{i+1}^2}\left(x(k)-\mu_{i+1}\right)^2, \\ &\text{kai} \quad i=1,2; \quad k=2,3,\ldots,K. \end{aligned} \tag{4.20}$$

Analogiškai žinomų parametrų modelio atvejui, funkcijai maksimizuoti taikomas dinaminio programavimo metodas – pagal (4.12) išraišką skaičiuojamos Belmano funkcijos, o iš pastarųjų maksimumų naudojant (4.16) – pasikeitimo momentų, t. y. žodžio pradžios ir pabaigos, reikšmės.

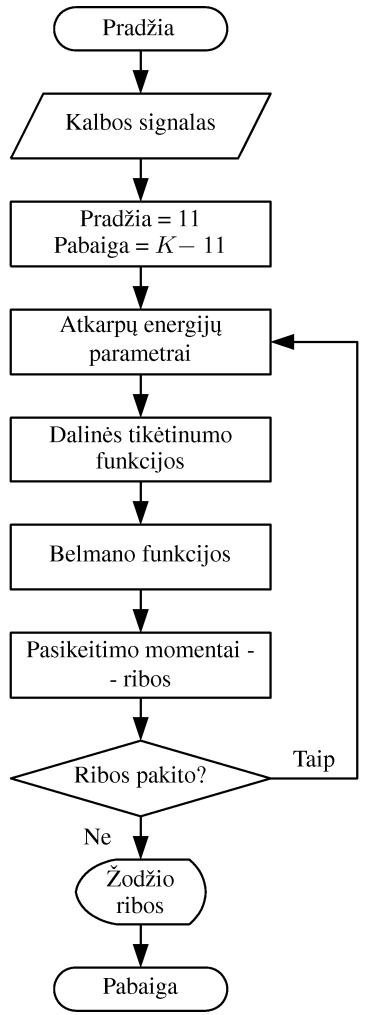
4.3.4.4. Žodžio ribų nustatymo algoritmas

4.2 paveiksle pateikiama žodžio ribų nustatymo algoritmo, sudaryto remiantis (4.19), (4.12) ir (4.16) išraiškomis, struktūra.

Naudojant apibendrintą matematinės vilties maksimizavimo algoritmą, reikalingi pradiniai ištarimo ribų įverčiai. Todėl aplinkos (tylos metu prieš žodį ir po) pradiniais parametrų įverčiais fiksuojami parametrai, apskaičiuoti iš fiksuoto ilgio atkarpų signalo pradžioje ir pabaigoje. Aplinkos parametrams vertinti fiksavome pirmuosius ir paskutiniuosius 11 signalo analizės kadrų. Žodžio energijos parametrų įverčiais laikėme parametrus, apskaičiuotus iš viso likusio signalo. Panaudojus šiuos parametrų įverčius, skaičiuojamos dalinės tikėtinumo funkcijos, Belmano funkcijos, gaunami pradiniai žodžio pradžios ir pabaigos taškų įverčiai. Remiantis gautais ribų įverčiais iš naujo vertinami tylos atkarpų ir žodžio energijų parametrai, nustatomos naujos žodžio ribų reikšmės ir tikrinama, ar jos pakito. Iteratyvus procesas tęsiamas tol, kol taškų įverčiai nustoja kisti.

4.3.4.5. Algoritmo efektyvumas

Tiesiogiai maksimizuojant tikėtinumo funkcijos logaritmą (4.7), tektų skaičiuoti tikėtinumo funkcijos reikšmę visiems galimiems žodžio ribų rinkiniams. Įvertinus tai, kad ribų variantų skaičius yra lygus ribų derinių skaičiui, o kiek-



4.2 pav. *Žodžio ribų nustatymo algoritmas*

vienos tikėtinumo funkcijos skaičiavimas reikalauja apytiksliai $2K$ operacijų (K – nagrinėjamo signalo kadrų skaičius), bendras operacijų skaičius, reikalingas tikėtinumo funkcijai maksimizuoti

$$\Lambda_T \approx \frac{K(K-1)}{2} \cdot 2K \approx K^3 - K^2 \approx K^3. \tag{4.21}$$

Funkcijai maksimizuoti pritaikius dinaminio programavimo principą, kiekvienam signalo kadrui tenka skaičiuoti dvi dalines tikėtinumo funkcijas, kiekvienai

atliekant apie 10 skaičiavimo operacijų, Belmano funkcijų skaičiavimas reikalauja dar apie $3K$ operacijų. Suminis operacijų skaičius taikant dinaminį programavimą

$$\Lambda_{DP} \approx 20K. \tag{4.22}$$

Įvertinus tai, kad žodžių įrašų trukmė (su fono intarpais žodžio pradžioje ir pabaigoje) paprastai būna apie 100–300 kadrų, dinaminio programavimo panaudojimo tikėtinumo funkcijai maksimizuoti pranašumas akivaizdus.

4.3.5. Kalbos signalo analizė

Sistemoje KAS realizuoti du signalo analizės metodai: tiesinės prognozės ir tiesinės prognozės kepstro. Signalo analizė vyksta kadrais, kiekvieną kadrą dauginant iš Hamming lango. Analizės kadro ilgis ir postūmis, nustatomi vartotojo, pagal nutylėjimą lygūs 250 (22,6 ms) ir 125 (11,3 ms) atskaitų atitinkamai.

TPM analizei naudojamas autokoreliacijos metodas, realizuotas Levinsono-Durbino algoritmu, pateiktu (3.18)–(3.23) išraiškose. Analizės eilė fiksuota ir lygi 10. TPMK analizė atliekama naudojant išraiškas (3.25)–(3.27). Analizės eilė yra keičiama, pagal nutylėjimą lygi 12. Be to, sudaryta galimybė atlikti kepstro analizę su vidurkio atėmimu ($\overline{\text{TPMK}}$)

$$c_i = c_i - \frac{1}{K}\sum_{k=1}^{K} c_k, \quad \text{kai } i = 1, 2, \ldots, K, \tag{4.23}$$

čia c_i – i-asis kepstro vektorius, K – signalo kadrų skaičius.

Siekdami išvengti nereikalingo kartojimosi, tolimesniuose skyriuose tiesinės prognozės modelio kepstro analizę vadinsime tiesiog kepstro analize, tą patį taikydami ir tiesinės prognozės kepstro analizei su vidurkio atėmimu.

4.3.6. Etalonai

Sistemoje realizuoti du žodžių etalonų kūrimo metodai: tiesioginio etalono kūrimo metodas ir klasterizavimu pagrįstas metodas – mokymas.

Tiesioginio etalono sukūrimo atveju įvestasis kalbos pavyzdys paskelbiamas etalonu be jokių papildomų procedūrų. Pagrindinis tiesioginio etalono sukūrimo pranašumas – paprastumas.

Mokyme etalonais paskelbiami klasterių centrai – pavyzdžiai, iki kurių gaunamas mažiausias vidutinis atstumas. Klasterių centrus atstovaujantys pavyzdžiai nusta-

tomi nagrinėjant visus galimus etalonų variantus

$$\{i_P^m\} = \arg \min_{i \in \{I_P^m\},\, j \neq i} \left[\frac{1}{P-m} \sum_i \min\{D_{ij}\} \right], \qquad (4.24)$$

kai $m = 1, 2, \ldots, M$.

Čia m – kuriamų etalonų skaičius, P – žodžių-kandidatų į etalonus skaičius, $\{i_P^m\}$ – m etalonų aibė, atrinkta iš P žodžių-kandidatų, $\{I_P^m\}$ – visų galimų m kandidatų iš P derinių aibė, $\{D_{ij}\}$ – atstumų tarp visų galimų kandidatų derinių ir likusių kandidatų aibė.

Pirmiausia nagrinėjami visi galimi vieno etalono variantai iš P ištarimų-kandidatų. Etalonu paskelbiamas pavyzdys, turintis mažiausią vidutinį atstumą su kitais kandidatais (atstumas skaičiuojamas naudojant dinaminį laiko skalės kraipymą). Po to nagrinėjami visos galimos poros, trejetai ir t. t. kiekvienu atveju į etalonus pasiūlant atitinkamą skaičių žodžių. Kuriamų etalonų skaičių pasirenka vartotojas.

Tiesioginiams apmokymo procedūros skaičiavimams reikalingų palyginimų skaičius išreiškiamas

$$\Lambda_T = (P-m)\frac{P!}{m!(P-m)!} = \frac{P!}{m!(P-m-1)!}. \qquad (4.25)$$

Panagrinėję išraišką galime matyti, kad prie tam tikrų mokymo aibės dydžio ir etalonų skaičiaus reikšmių, palyginimų skaičius galo išaugti iki šimtų ar net tūkstančių eilės. Sistemoje KAS maksimalus kuriamų etalonų skaičius M lygus 5, mokymo aibės dydis P – 10. Ribiniu atveju, sistemai apmokyti 5 etalonais prireiktų 1260 palyginimo operacijų. Įvertinus tai, kad skaičiavimuose iš esmės operuojama tų pačių atstumų skirtingomis kombinacijomis, palyginimų kiekį galima sumažinti iki

$$\Lambda = P(P-1). \qquad (4.26)$$

Mokymo privalumas – galimybė pasirinkti kuriamų etalonų skaičių ir mokymo (ištarimų-kandidatų) aibės atžvilgiu optimalios etalonų aibės suformavimas.

4.3.7. Palyginimas

Signalo analizės etapo rezultatas – požymių vektorių seka, atstovaujanti nežinomą pavyzdį, kurį reikia palyginti su etaloninius žodžius atstovaujančiomis sekomis siekiant gauti jų panašumo (kitaip atstumo) skaitmeninį įvertinimą. Sekoms palyginti naudojamas dinaminio laiko skalės kraipymo metodas, realizuotas naudojant dinaminio programavimo išraiškas (3.33)–(3.38). Požymių vektorių sekų sutapdinimo kreivė (3.7 b) pav.) formuojama naudojant Itakura lokalius krypties apribojimus (3.10 pav.). Žodžių pradžioms ir pabaigoms sutapdinti panaudoti galo taškų apribojimai, išreikšti (3.39)–(3.40).

Vektorių panašumui įvertini tiesinės prognozės modelio analizės atveju naudojamas tiesinės prognozės tikėtinumų santykis (3.28). Kepstro ir kepstro su vidurkio atėmimu analizių atveju naudojamas kvadratinis Euklido atstumas (3.29).

Standartiniame palyginimo procese pavyzdžių palyginimas atliekamas nuo pirmojo kadro iki paskutiniojo, palyginimo rezultatu fiksuojant paskutiniajame kadre gautą atstumą. Papildomai sudaryta galimybė atlikti palyginimą greituoju režimu. Greitajame režime tarpinis rezultatas – dalinis atstumas – palyginamas su slenksčio reikšme. Jei ji viršijama, pavyzdžių palyginimas nutraukiamas kaip neperspektyvus, palyginimo rezultatu paskelbiant begalinį atstumą

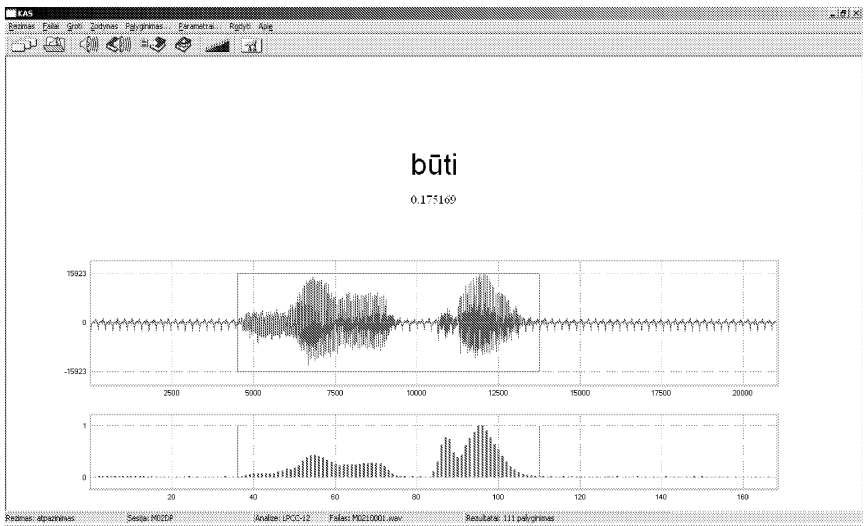
$$D_{RZ} = \begin{cases} \infty, & \text{kai } D_{RZ}(k) > D_{sl}; \\ D_{RZ}(K), & \text{kai } D_{RZ}(k) < D_{sl}. \end{cases} \tag{4.27}$$

Čia D_{RZ} – atstumas tarp etalono R ir nežinomojo pavyzdžio Z, $D_{RZ}(k)$ – dalinio atstumo įvertinimas k-ajame kadre, $D_{RZ}(K)$ – atstumo reikšmė paskutiniajame palyginimo kadre, D_{sl} – dalinio atstumo slenksčio reikšmė. k ir D_{sl} reikšmes pasirenka vartotojas.

Kaip parodė vėlesnis darbas atpažinimo sistema, greitajame palyginimo režime nutraukiama apie 80 % pavyzdžių palyginimų.

4.3.8. Žodžių atpažinimo rezultatai

Žodžių atpažinimo rezultatų iliustracija pateikta 4.3 paveiksle.



4.3 pav. *Atpažinimo rezultatų pateikimas*

Kaip rezultatas pateikiama artimiausio etalono transkripcija, skaitmeninė atstumo išraiška, įvesto signalo laiko ir energijos diagramos su žodžio ribomis. Jei žodžio riboms nustatyti naudotas energijos slenksčio metodas, energijos diagramoje papildomai atvaizduojamas slenksčio lygis. Papildomai pateikiama informacija apie sistemos darbo režimą, darbo sesiją, signalo analizės metodą, pasirinkto failo pavadinimą, atliktų palyginimų skaičių.

4.4. Žodžių segmentavimas

Antrasis sistemos darbo režimas – žodžių segmentavimas. Persijungus į šį režimą, pasikeičia valdymo elementų rinkinys ir parametrų dialogai. Segmentuoti galima žodžius tiek su tylos atkarpomis, tiek be jų.

4.4.1. Kalbos signalo įvedimas, apdorojimas ir analizė

Žodžių segmentavimo režime kalbos signalo įvedimas ir pradinis apdorojimas analogiškas atpažinimo režimui. Žodžiams segmentuoti sistemoje atliekama 10 eilės tiesinės prognozės modelio analizė. Kaip ir žodžių atpažinimo režime vartotojas gali nustatyti pageidaujamas pradinio apdorojimo koeficiento ir analizės kadro ilgio reikšmes.

4.4.2. Segmentavimas

Sistemoje KAS realizuoti du kalbos signalų segmentavimo metodai: tikėtinumo funkcijos maksimizavimo ir prognozės klaidos minimizavimo [A2]. Abu metodai remiasi ta pačia prielaida kaip ir žodžio ribų nustatymas metodas – kalbos signalas yra atsitiktinis signalas su besikeičiančiomis savybėmis [A1]. Tik šiuo atveju uždavinys sprendžiamas su sąlyga, kad nagrinėjamame signale yra ne du, o keletas pasikeitimo momentų.

Jeigu darysime prielaidą, kad kalbos signalas yra pseudostacionarus kalbos garsų atkarpose, galime teigti, kad pasikeitimo momentai atitiks ribas tarp garsų, o išskirtieji segmentai – garsus. Todėl segmentavimo metu išskirtąsias signalo atkarpas vadinsime tiek segmentais, tiek garsais.

4.4.2.1. Segmentavimo uždavinys

Segmentavimo atveju nagrinėjamoji atsitiktinė seka $x = \{x(1), x(2), \ldots, x(N)\}$ turi M pasikeitimo momentų ir aprašoma p-os eilės tiesinės prognozės lygtimi

$$x(n) = -a_1(n) \cdot x(n-1) - \cdots - a_p \cdot x(n-p) + b(n) \cdot v(n). \qquad (4.28)$$

Tuomet sekos modelio parametrų sąlyga (4.2) tampa

$$A(n) = \begin{cases} A_1, & \text{kai } n = 1, 2, \ldots, u_1; \\ A_2, & \text{kai } n = u_1 + 1, \ldots, u_2; \\ \ldots & \\ A_i, & \text{kai } n = u_{i-1} + 1, \ldots, u_i; \\ \ldots & \\ A_M, & \text{kai } n = u_{M-1} + 1, \ldots, u_M; \\ A_{M+1}, & \text{kai } n = u_M + 1, \ldots, N. \end{cases} \tag{4.29}$$

Čia M – parametrų pasikeitimo momentų skaičius, $u = [u_1, u_2, \ldots, u_M]$ – parametrų pasikeitimo momentų rinkinys, tenkinantis sąlygą $p < u_1 < u_2 < \cdots < N$.

Segmentavimo uždavinio tikslas – nustatyti pasikeitimo momentų įverčių rinkinį $\hat{u} = [\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_M]$. Segmentavimo metodą nulemia optimalumo kriterijus, naudojamas momentų įverčiams gauti.

4.4.2.2. Tikėtinumo funkcijos maksimizavimo metodas

Kaip ir atsitiktinio signalo atveju, šiame metode ieškomi pasikeitimo momentų įverčiai, maksimizuojantys tikėtinumo funkciją

$$\hat{u} = \arg\max_{u} l(u|x). \tag{4.30}$$

Kadangi šiuo atveju mes turime signalą su M pasikeitimo momentų, tikėtinumo funkcijos išraiška (4.7) tampa

$$\begin{aligned} l(u|x) = {} & p\left(x(1), x(2), \ldots, x(p)\right) \cdot (2\pi)^{-(N-p)/2} \cdot b^{-(u_1-p)}(1) \times \\ & \times \cdot b^{-(u_2-u_1)}(2) \times \cdots \times b^{-(N-u_M)}(M+1) \times \\ & \times \exp\left(\frac{1}{2b^2(1)} \sum_{n=p+1}^{u_1} \left(\sum_{j=0}^{p} a_j(1) x(n-j) \right)^2 - \right. \\ & - \frac{1}{2b^2(2)} \cdot \sum_{n=u_1+1}^{u_2} \left(\sum_{j=0}^{p} a_j(2) x(n-j) \right)^2 - \cdots - \\ & \left. - \frac{1}{2b^2(M+1)} \cdot \sum_{n=u_M+1}^{N} \left(\sum_{j=0}^{p} a_j(M+1) x(n-j) \right)^2 \right). \end{aligned} \tag{4.31}$$

Analogiškai nagrinėtajam atvejui, tikėtinumo funkcija yra pertvarkoma ją logaritmuojant ir pašalinant nuo pasikeitimo momentų nepriklausančius dedamuo-

sius. Gaunama nauja maksimizuojama funkcija

$$\theta(u|x) = l_1(u_1|x) + l_2(u_2|x) + \cdots + l_M(u_M|x), \tag{4.32}$$

čia $l_i(u_i|x)$ – dalinė tikėtinumo funkcija, apibrėžiama

$$\begin{aligned} l_i(k|x) = &-(k-p)\log b(i) - (N-k)\log b(i+1) - \\ &- \frac{1}{2b^2(i)} \sum_{n=p+1}^{k} \left(\sum_{j=0}^{p} a_j(i)x(n-j) \right)^2 - \\ &- \frac{1}{2b^2(i+1)} \sum_{n=k+1}^{N} \left(\sum_{j=0}^{p} a_j(i+1)x(n-j) \right)^2, \\ &\text{kai } i = 1, 2, \ldots, M; \quad k = p+1, p+2, \ldots, N. \end{aligned} \tag{4.33}$$

Palyginę su (4.12) matome, kad gautoji dalinės tikėtinumo funkcijos išraiška yra identiška dviejų pasikeitimo momentų atvejui, taigi $l_i(k|x)$ skaičiavimui galime naudoti rekurentinę tikėtinumo funkcijos išraišką (4.13) su nulinėmis pradinėmis sąlygomis. Funkcijai $\theta(u|x)$ maksimizuoti taikydami dinaminio programavimo principą, apibrėžiame Belmano funkcijas

$$\begin{aligned} g_i(u_{i+1}|x) = &\max_{\substack{u_i \\ p+i-1<u_i<u_{i+1}}} \left[l_i(u_i|x) + g_{i-1}(u_i|x) \right], \\ &\text{kai } i = 1, 2, \ldots, M;\ u_{i+1} = p+i+1, p+i+2, \ldots, N. \end{aligned} \tag{4.34}$$

Pastarosioms taip pat sudarytos rekurentinės skaičiavimo išraiškos

$$\begin{aligned} &g_i(u_{i+1}|x) = \max\left[g_i(u_{i+1}-1|x), (g_{i-1}(u_{i+1}-1|x) + l_i(u_{i+1}-1|x)) \right], \\ &\text{kai } i = 1, 2, \ldots, M;\ u_{i+1} = p+i+2, p+i+3, \ldots, N \end{aligned} \tag{4.35}$$

su pradinėmis sąlygomis

$$\begin{aligned} &g_i(p+i+1|x) = l_i(p+i|x) + g_{i-1}(p+1), \\ &\text{kai } i = 1, 2, \ldots, M. \end{aligned} \tag{4.36}$$

Tuomet maksimalaus tikėtinumo pasikeitimo momentų įverčiai, atitinkantys garsų ribas signale, randami

$$\begin{aligned} \hat{u}_i = &\min \left[\arg \max_{\substack{k \\ p+i \leqslant k \leqslant \hat{u}_{i+1}}} g_i(k|x) \right], \\ &\text{kai } i = M, M-1, \ldots, 1;\ \hat{u}_{M+1} = N. \end{aligned} \tag{4.37}$$

Segmentuojant realius kalbos signalus, garsų (pasikeitimo momentų) skaičius bei tiesinės prognozės modelio parametrai būna nežinomi. Vėlgi naudojamas matematinės vilties maksimizavimo algoritmas. Algoritmui reikalingus pasikeitimo momentų skaičių bei modelio parametrų reikšmes tenka nurodyti vartotojui.

4.4.2.3. Prognozės klaidos minimizavimo metodas

Jei modelyje (4.28) esantis stiprinimo koeficiento pasiskirstymas nežinomas, tuomet pasikeitimo momentų įverčiai nustatomi maksimizuojant neigiamą suminę prognozės klaidą (arba tiesiog minimizuojant prognozės klaidą). Šis metodas vadinamas prognozės klaidos minimizavimo metodu.
Pasikeitimo momentų minimalios prognozės klaidos įverčiai randami

$$\hat{u} = \arg\max_{u} E(x|u), \tag{4.38}$$

čia $E(x|u)$ – neigiama suminė prognozės klaida.

Neigiama suminė prognozės klaida išreiškiama

$$\begin{aligned} E(x|u) = &- \sum_{n=p+1}^{u_1} \left(\sum_{j=0}^{p} a_j(1)x(n-j) \right)^2 - \\ &- \sum_{n=u_1+1}^{u_2} \left(\sum_{j=0}^{p} a_j(2)x(n-j) \right)^2 - \cdots - \\ &- \sum_{n=u_M+1}^{N} \left(\sum_{j=0}^{p} a_{M+1}(1)x(n-j) \right)^2 . \end{aligned} \tag{4.39}$$

Pastaroji išraiška gali būti išskaidyta į vieno kintamojo funkcijų sumą

$$E(x|u) = e_1(u_1|x) + e_2(u_2|x) + \cdots + e_M(u_M|x) + D, \tag{4.40}$$

čia $e_i(k|x), i = 1, 2, \ldots, M$ – dalinė prognozės klaida, D – nuo pasikeitimų momentų nepriklausantis pastovus dydis.

Kaip ir tikėtinumo funkcijos maksimizavimo atveju, dalinėms funkcijoms skaičiuoti sudarytos rekurentinės išraiškos

$$e_i(k|x) = e_i(k-1|x) - \left(\sum_{j=0}^{p} a_j(i)x(k-j) \right)^2 +$$

$$+\left(\sum_{j=0}^{p} a_{j+1}(i)x(k-j)\right)^2, \quad \text{kai } i=1,2,\ldots,M;\ k=p+1,p+2,\ldots,N \tag{4.41}$$

su pradinėmis sąlygomis

$$e_i(p|x)=0, \quad \text{kai } i=1,2,\ldots,M. \tag{4.42}$$

Kadangi funkcija $e_i(k|x)$ yra keleto vieno kintamojo funkcijų suma, jai maksimizuoti taip pat galime naudoti dinaminį programavimą. Tuo tikslu vėl skaičiuojamos Belmano funkcijos, kurių rekurentinės išraiškos analogiškos tikėtinumo maksimizavimo atvejui

$$g_i(u_{i+1}|x)=\max\left[g_i(u_{i+1}-1|x), g_{i-1}(u_{i+1}-1|x)+e_i(u_{i+1}-1|x)\right], \quad \text{kai } i=1,2,\ldots,M;\ u_{i+1}=p+i+2,p+i+3,\ldots,N \tag{4.43}$$

su pradinėmis sąlygomis

$$g_i(p+i+1|x)=e_i(p+i|x)+g_{i-1}(p+1), \quad \text{kai } i=1,2,\ldots,M. \tag{4.44}$$

Minimalios prognozės klaidos pasikeitimo momentų įverčiai, atitinkantys garsų ribas signale, randami

$$\hat{u}_i=\min\left[\arg \max_{p+i\leqslant k\leqslant \hat{u}_{i+1}} g_i(k|x)\right], \quad \text{kai } i=M,M-1,\ldots,1;\ \hat{u}_{M+1}=N. \tag{4.45}$$

4.4.2.4. Metodų tarpusavio ryšys

Palyginę išraiškas (4.32) ir (4.13) su (4.40) ir (4.41), bei įvertinę tai, kad abiejuose metoduose analogiškai taikomas dinaminio programavimo principas (išraiškos (4.35), (4.37) ir (4.43), (4.45)), galime pamatyti, kad jeigu tikėtinumo funkcijos maksimizavimo metode stiprinimo koeficientą prilyginsime vienetui

$$b(i)=1, \quad \text{kai } i=1,2,\ldots,M+1, \tag{4.46}$$

dalinių tikėtinumo funkcijų išraiškos (4.13) tampa analogiškomis dalinių prognozės klaidų išraiškoms (4.41). Taigi tikėtinumo funkcijos maksimizavimo metode (TF) stiprinimo koeficientus prilyginę vienetui, gausime pasikeitimo momentų minimalios prognozės klaidos (MK) įverčius, t. y. abiem metodais gauti rezultatai sutaps.

4.4.2.5. Segmentavimo algoritmas

Segmentacijos uždavinys sudarytas ir abu segmentacijos metodai sukurti padarius prielaidą, kad signalo modelio parametrai žinomi, o pasikeitimo momentų skaičius yra žinomas. Segmentuojant realų kalbos signalą, nežinomas nei garsų skaičius žodyje, nei garsų modelių parametrai. Tuomet tikėtinumo funkcijai ir neigiamai suminei prognozės klaidai maksimizuoti vėl pritaikomas apibendrintasis matematinės vilties maksimizavimo metodas. Visų pirma priimama, kad garsų skaičius žodyje yra žinomas. Taip pat daroma prielaida, kad turime pradinę informaciją apie autoregresijos modelį, t. y. turime modelio parametrų pradines reikšmes. Tuomet naudodami pradines autoregresijos modelio parametrų vertes gauname pasikeitimo momentų įverčius. Pagal naująsias ribų reikšmes gauname patikslintus modelių parametrus, kuriuos panaudojame sekančioje pasikeitimo momentų nustatymo iteracijoje. Skaičiavimai tęsiami tol, kol signalo pasikeitimo momentų įverčiai nustoja kisti. Segmentų ribų įverčiai, gauti paskutinėje iteracijoje, laikomi ribomis tarp kalbos signalo garsų. Sudaryto segmentacijos algoritmo struktūra pateikta 4.4 paveiksle.

Pasikeitimo momentų skaičių ir pradines autoregresijos parametrų reikšmes tenka nurodyti vartotojui. Įvedus kalbos signalą, nurodomas numanomos garsų vietos signale, tuomet pasikeitimo momentų skaičius nustatomas vienu mažesniu už nurodytų garsų skaičių, o reikalingos tiesinės prognozės modelio parametrų reikšmės apskaičiuojamos iš analizės kadro trukmės atkarpų nurodytose signalo vietose.

4.4.2.6. Algoritmų efektyvumas

Tiesiogiai maksimizuojant tikėtinumo funkciją (4.31) ir neigiamą suminę prognozės klaidą (4.39), tektų nagrinėti visus galimus pasikeitimo momentų derinius (abiem metodais gaunasi vienodas galimų variantų skaičius). Įvertinus tai, kad paprastai signalo atskaitų skaičius yra žymiai didesnis už pasikeitimo momentų skaičių, o maksimizuojamos funkcijos vienai reikšmei apskaičiuoti reikia apie $2pN$ operacijų (tikėtinumo funkcijos reikšmės apskaičiavimas pareikalautų apie $3M$ operacijų daugiau nei suminės prognozės klaidos), operacijų skaičius reikalingas tiesiogiai maksimizuoti funkciją

$$\Lambda_T = \frac{N!}{M!(N-M)!} \cdot 2pN \approx \frac{N^M}{M!} \cdot 2pN \approx 2p\frac{N^{M+1}}{M!}, \qquad (4.47)$$

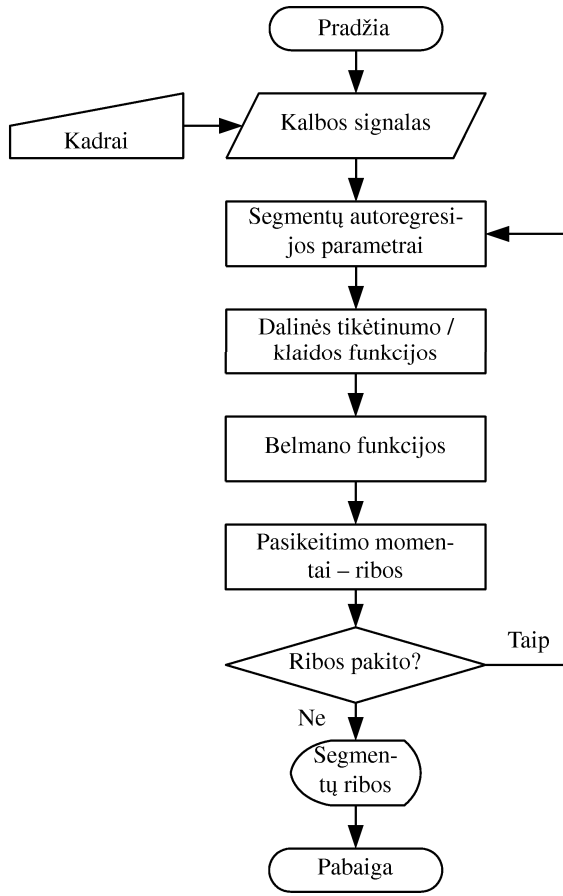
čia N – signalo atskaitų skaičius, M – pasikeitimo momentų skaičius, p – modelio eilė.

Funkcijai maksimizuoti naudojant dinaminį programavimą, tiek tikėtinumo funkcijos, tiek suminės prognozės klaidos atveju dalinėms funkcijoms apskaičiuoti reikia apie $4pMN$ operacijų, Belmano funkcijoms – apytiksliai N operacijų, taigi

bendras reikalingų operacijų skaičius

$$\Lambda_{DP} \approx 4pMN^2. \tag{4.48}$$

Dinaminio programavimo metodo pranašumas prieš tiesioginius skaičiavimus akivaizdus, turint omeny, kad atskaitų skaičius gali siekti kelias dešimtis tūkstančių.

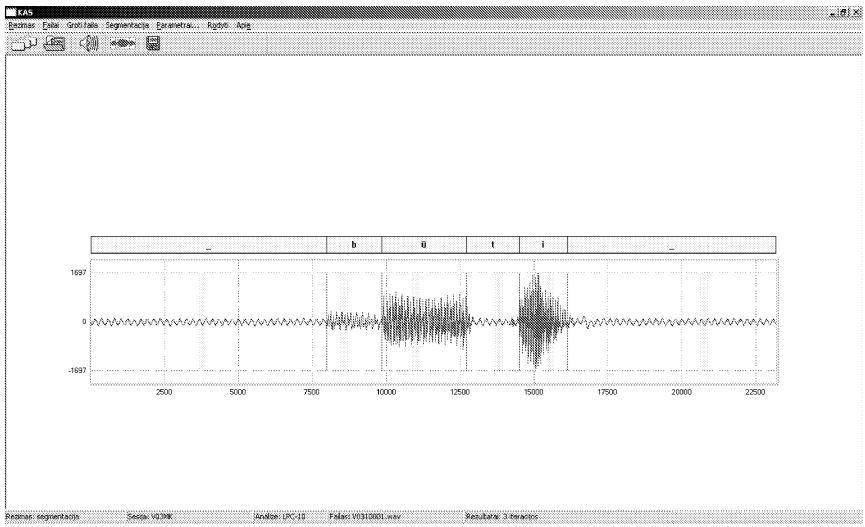


4.4 pav. *Sudarytasis segmentavimo algoritmas*

4.4.3. Segmentavimo rezultatai

4.5 paveiksle pateiktas žodžio segmentavimo rezultatų pateikimo pavyzdys. Kaip rezultatai pateikiama pilka spalva išskirti vartotojo pasirinkti kadrai pradi-

nėms parametrų reikšmėms vertinti ir garsų ribos signale. Papildomai pateikiamas darbo sesijos pavadinimas, pasirinkto failo pavadinimas ir skaičiavimo iteracijų skaičius. Vartotojas turi galimybę išskirtiems segmentams suteikti fonetinę transkripciją ir visus segmentavimo rezultatus išsaugoti failuose.



4.5 pav. *Žodžio segmentavimo rezultatų pateikimas*

4.5. Žodžio segmentų atpažinimas

Išskirtieji žodžio segmentai gali būti panaudoti kaip akustiniai vienetai atpažinimo sistemoje. Jeigu kiekvienam jų sukurtume etaloną, o segmentavimo į garsus procesą įdiegtume kaip pradinį palyginimo proceso etapą, galėtume realizuoti žodžio garsų atpažinimo procesą. Remiantis šia idėja realizuotas trečiasis sistemos KAS darbo režimas – žodžio segmentų atpažinimas.

4.5.1. Prielaidos žodžio garsams atpažinti

Žodžių atpažinimas segmentais padėtų išvengti bene rimčiausio pavienių žodžių atpažinimo, naudojant dinaminį laiko skalės kraipymą, trūkumo – atpažinimo proceso trukmės priklausomybės nuo žodyno dydžio.

Idealiu atveju segmentavimo rezultatas būtų garsų, kurių skaičius kalboje yra baigtinis, ribos. Taigi elementarios žodžio segmentų atpažinimo sistemos etalonų kiekis būtų lygus vartojamos kalbos garsų skaičiui ar jo kartotiniui (tuo atveju, kai kiekvienam garsui kuriama keletas etalonų). Praktiškai stabilus ir tikslus garsų ri-

bų nustatymas vargu ar įmanomas, todėl etalonų skaičius būtų didesnis už kalbos garsų skaičių. Tačiau net ir šiuo atveju galima tikėtis, kad tam tikromis aplinkybėmis (didelis apmokomų žodžių kiekis, pakankamai tikslus ir stabilus signalo segmentavimas į garsus) segmentų etalonų skaičius bus mažesnis nei apmokymui panaudotų žodžių skaičius. Be to, realizavus nepriklausomą nuo ištarto žodžio tikslų ir stabilų segmentavimą ir efektyvią požymių vektorių klasifikaciją, galima tikėtis, kad tokia sistema teisingai sugebės atpažinti žodį (ar bent jo dalį), neesantį sistemos žodyne. Galimybė nors ir teorinė, tačiau nepaneigtina.

Antra, tikėtina išskirto garso trukmė – nuo keliolikos iki keliasdešimt milisekundžių, o tai yra bent jau keletą kartų mažiau nei vieno žodžio trukmė. Šios savybės įtaka atpažinimo procesui būtų dvejopa:

- Sumažėtų vieno etalono analizės trukmė. Tai reikštų spartesnę viso etalonų rinkinio analizę.
- Sumažėtų atminties kiekis, reikalingas etalonams saugoti.

Kita vertus, varijuojanti segmentų trukmė iškelia analizės kadro trukmės klausimą. Naudojant fiksuotos trukmės analizės kadrą, segmentų analizė kraštutiniu atveju gali duoti ne ką mažiau duomenų nei ištisinių žodžių analizė. Taip bus prarandamas esminis segmentų atpažinimo privalumas – galimas etalonų skaičiaus fiksavimas didėjant sistemos žodynui. Savo darbe garsų ribų ieškojome kaip segmentų tiesinės prognozės modelio parametrų pasikeitimo momentų, darydami prielaidą, kad signalo atkarpa tarp pasikeitimo momentų yra pseudostacionari. Tos pačios prielaidos galima laikytis ir realizuojant segmentų atpažinimą – nagrinėti segmento trukmės analizės kadrą. Tokiu atveju kiekvienas segmentas sistemoje būtų atvaizduojamas vienu požymių vektoriumi. Tokio sprendimo nauda yra dvejopa:

- Požymių vektorių klasifikacija apsiribotų paprasčiausio atstumu tarp vektorių skaičiavimu – išnyksta vektorių sekų sutapdinimo laike poreikis.
- Vėlgi stipriai sumažėtų etalonams saugoti reikalingas atminties kiekis.

Be abejo, segmento atvaizdavimas vienu požymių vektoriumi neabejotinai sumažintų klasifikacijos tikslumą. Todėl kitas įmanomas analizės kadro ilgio atvejis – fiksuotos trukmės kadras. Tokiu atveju kiekvienas segmentas būtų atvaizduojamas nuo vieno iki keleto (ar net keliolikos) vektorių seka ir vektorių klasifikacijos klausimas taptų sudėtingesnis nei pirmuoju atveju.

Realizuodami žodžių atpažinimą garsais, mes pasirinkome pirmąjį analizės kadro trukmės variantą – lygų išskirto segmento trukmei, tuo pačiu klausimą apie analizės kadro parinkimą palikdami ateities darbams.

Kitas žodžių atpažinimo segmentais momentas – klaidos. Galimi keli klaidų tipai:

- Sukeitimo klaida gautųsi, jei garsas būtų atpažintas kaip kitas garsas. Tokios klaidos šaltinis – požymių klasifikatorius.

- Ištrynimo (pašalinimo) klaida būtų gaunama susiliejus gretimiems garsams ir taip prarandant vieną ar daugiau garsų. Šiuo atveju klaidos šaltiniu įvardintume žodžio segmentavimo etapą.

- Įterpimo klaida įvyktų atpažintame žodyje atsiradus papildomiems garsams. Tai vėlgi galimo klaidingo segmentavimo pasekmė.

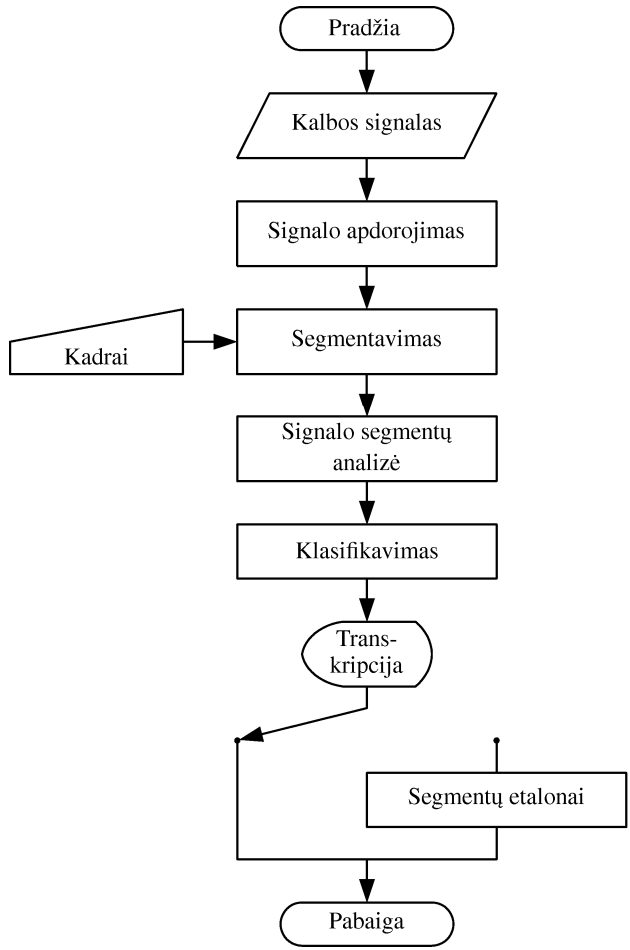
Sukeitimo klaidos atveju tektų naudoti lingvistinį atpažinimo rezultato apdorojimą – ieškoti klaidų ir jas taisyti. Ištrynimo klaidų kiekį sumažinti galima išskirtoms garsų poroms ar trejetams priskirti atitinkamą dvigarsių ar trigarsių fonetinę transkripciją, t. y. gautume atpažinimo vieneto sustambėjimą. Įterpimo klaidos atveju vėlgi padėtų lingvistinis rezultatų apdorojimas. Taigi žodžių atpažinimo garsais atveju, jei būtų daromos vieno garso atpažinimo klaidos, reikalingas lingvistinis apdorojimas būtų pakankamai žemo lygio – iš esmės tektų ieškoti ir taisyti gramatines klaidas. Jei būtų klystama dviejų ir daugiau garsų atveju, lingvistinis apdorojimas taptų daug sunkesnis ir net komplikuotas. Pavienių žodžių atpažinimo atveju gaunamos klaidos būtų sukeitimo tipo ir joms taisyti tektų nagrinėti žodžio kontekstą – lingvistinis apdorojimas būtų aukštesnio lygio.

Kadangi mūsų atveju garsų skaičių žodyje nurodo vartotojas, įterpimo klaida yra neįmanoma. Taigi žodžio garsų atpažinime turėtume gauti dviejų tipų klaidas – sukeitimo ir ištrynimo.

Žodžių atpažinimo segmentais sistemai bus būdingas ir trūkumas – segmentavimo įtaka atpažinimo rezultatams. Nestabilūs segmentavimo rezultatai gali nulemti prastą atpažinimo tikslumą, o atpažinimo klaida savo prigimtimi tampa sudėtingesnė.

4.5.2. Žodžio segmentų atpažinimo algoritmas

Žodžio segmentų atpažinimui realizuoti panaudotos sukurtosios segmentavimo (prognozės klaidos minimizavimo metodas) ir palyginimo procedūros. Įvestasis kalbos signalas apdorojamas – atimama nuolatinė dedamoji ir atliekamas filtravimas I eilės RIR aukštų dažnių filtru. Vartotojui nurodžius tikėtinas garsų vietas, signalas segmentuojamas. Išskirtuose segmentuose atliekama analizė, kurios rezultatas – žodžio segmentus atstovaujančios požymių vektorių sekos. Galima atlikti vieną iš dviejų analizių: tiesinės prognozės modelio ir kepstro. Kepstro analizės su vidurkio atėmimu netaikėme, manydami kad keleto vektorių (o būtent tiek vektorių šiuo atveju atvaizduos vieną žodį) vidurkinimas gali lemti didelius nuokrypius vektorių reikšmėse. Išskirtieji požymiai gali būti pateikti atpažinti arba panaudoti etalonams kurti. Pateikus vektorių sekas atpažinti, kaip atpažinimo rezultatas pateikiamas artimiausių etaloninių garsų fonetinės transkripcijos. Sudarytoji segmentų atpažinimo algoritmo struktūra pateikta 4.6 paveiksle.



4.6 pav. *Žodžio segmentų atpažinimo algoritmas*

4.5.3. Žodžio segmentų atpažinimo rezultatai

4.7 paveiksle pateikta žodžio segmentų atpažinimo režimu dirbančios sistemos iliustracija.

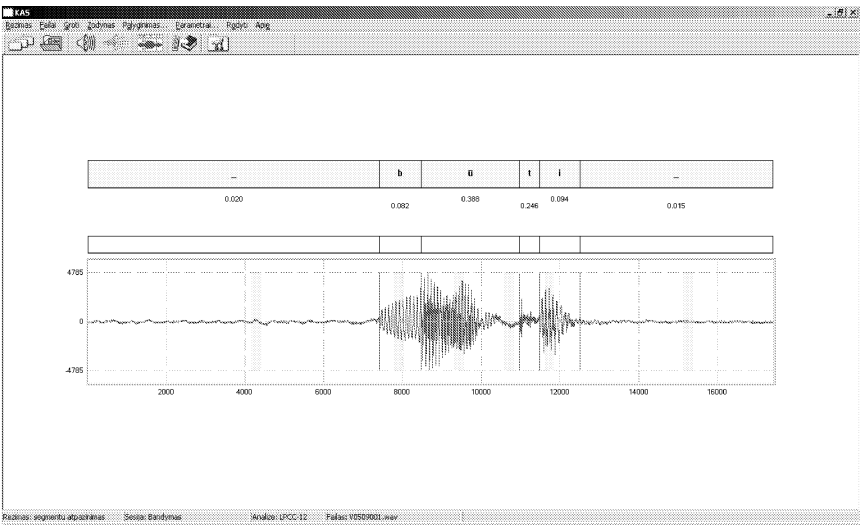
Kaip atpažinimo rezultatas pateikiama žodžio laiko diagrama su segmentų ribomis, atpažintų segmentų fonetinės transkripcijos ir atstumai iki etalonų, laukeliai segmentams transkribuoti. Papildomai pateikiamas darbo sesijos pavadinimas, naudojamas analizės tipas ir eilė, pasirinkto garso failo pavadinimas, nurodomi segmentacijos metu atliktų iteracijų ir palyginimo operacijų skaičiai.

4.6. Kitos atpažinimo ir segmentavimo realizacijos

Ruošiant disertaciją žodžių ribų nustatymo ir žodžių segmentavimo metodai buvo realizuoti ir kitose sistemose. Buvo sukurtos pavienių žodžių atpažinimo sistema *Atpazinimas* ir žodžių segmentavimo sistema *Segmentacija*. Pavienių žodžių atpažinimo sistema *Atpazinimas* skirta atpažinimo procesui vizualizuoti ir analizuoti. Joje be atpažinimo rezultatų pateikiama papildoma informacija – pilni palyginimo rezultatai, dinaminio laiko skalės kraipymo kreivė, dalinių atstumų normuotos reikšmės. Buvo sukurti keli praktiniai sistemos taikymai. Vienas jų buvo skirtas dialogams su kompiuteriu realizuoti. Su kiekvienu etalonu susiejamas garso įrašas, kuris tarnauja kaip atsakymas į atpažintą žodį. Kitas pritaikymas skirtas interneto naršyklei valdyti ir programoms iškviesti balsu [A3]. Šioje kiekvienas etalonas susietas su konkrečia operacinės sistemos komanda, kuri įvykdoma atpažinus etaloną.

Žodžių segmentavimo sistema *Segmentacija* taip pat buvo skirta procesui vizualizuoti. Joje be segmentavimo rezultatų papildomai pateikiami dalinių tikėtinumų ir Belmano funkcijų grafikai, nurodomas ryšys tarp signalo ir Belmano funkcijų pasikeitimo momentų.

Šios atpažinimo ir segmentavimo sistemos buvo įtrauktos į 2000–2006 m. programos „Lietuvių kalba informacinėje visuomenėje“ automatinio lietuvių kalbos atpažinimo tiriamuosius darbus. *Atpazinimas* versija dialogams kurti pristatyta mokslo, inovacijų ir aukštųjų technologijų parodoje „Mokslas 2004“, o versija, skirta naršyklei valdyti – informacinių technologijų parodoje „Infobalt 2007“.



4.7 pav. *Žodžio segmentų (garsų) atpažinimo rezultatų pateikimas*

4.7. Ketvirtojo skyriaus apibendrinimas

- Sukurta sistema KAS, dirbanti trimis režimais – žodžių atpažinimo, žodžių segmentavimo ir žodžio segmentų atpažinimo.
- Žodžio riboms nustatyti sukurtas metodas, ribas aptinkantis kaip signalo savybių pasikeitimo momentus. Metodas realizuotas naudojant dinaminį programavimą.
- Etalonams kurti sistemoje pritaikytas klasterizavimo principas, minimizuojantis vidutinį atstumą iki klasterių centrų (atstumus skaičiuojant dinaminio laiko skalės kraipymo metodu).
- Žodžiams segmentuoti pritaikytas kalbos signalo tiesinės prognozės modelio parametrų pasikeitimo momentų aptikimo principas, realizuotas naudojant dinaminį programavimą.
- Padarius prielaidą, kad išskirtieji žodžių segmentai yra mažai kintančios atkarpos, realizuotas žodžio segmentų atpažinimas. Priimta prielaida lemia tai, kad segmentų palyginimas apsiriboja atstumo tarp segmentų vektorių skaičiavimu.

5

Atpažinimo sistemos tyrimas

Šiame skyriuje pateikiami eksperimentinių tyrimų rezultatai. Atliktų eksperimentų tikslas – nustatyti pasiūlytų sprendimų darbingumą ir jų įtaką atpažinimo sistemos darbingumui. Eksperimentais tirtas žodžio ribų nustatymo metodo darbingumas, žodžio ribų nustatymo metodo ir sistemos mokymo įtaka atpažinimo tikslumui, segmentavimas naudojant pasiūlytuosius metodus bei žodžio segmentų atpažinimas.

5.1. Eksperimentų sąlygos

Visi minėtieji eksperimentai atlikti asmeniniu kompiuteriu naudojant sukurtąją sistemą. Siekiant užtikrinti vienodas eksperimentų sąlygas, visi žodžiai į sistemą buvo įvedami iš garso failų. Kadangi tyrimų tikslas buvo nustatyti sukurtų metodų ir skaičiavimų principų įtaką sistemos darbingumui, klausimas dėl darbo parametrų (tokių kaip analizės kadro ilgis ir poslinkis, analizių eilės, pradinio apdorojimo koeficientas, atpažinimo slenksčiai ir t. t.) optimalumo nekeltas, t. y. parametrų reikšmės nustatytos remiantis tik autoriaus praktine patirtimi. Eksperimentai, susiję su atpažinimu, atlikti naudojant visus tris įdiegtus signalo analizės metodus.

Tyrimų rezultatai pateikiami lentelėse bei grafine forma. Lentelėse pateikiami kalbėtojų individualūs ir vidutiniai, grafikuose - tik vidutiniai rezultatai. Pasikliautinieji intervalai skaičiuoti su pasikliovimo lygmeniu 0,95. Lentelėse intervalų apatinis ir viršutiniai rėžiai pateikiami skliausteliuose šalia nagrinėjamo dydžio.

5.2. Eksperimentų duomenys

Eksperimentuose panaudota pavienių žodžių įrašų bazė, sukaupta Matematikos ir informatikos instituto Atpažinimo procesų skyriuje.
Visi įrašai atlikti įvairiose akustinėse aplinkose, naudojant nevienodą techninę įrangą – asmeninius kompiuterius su skirtingo lygio garso plokštėmis, mikrofonu. Žodyną sudaro 111 įvairios trukmės skirtingų formų žodžių (priedas A). Žodžius įkalbėjo 10 žmonių: 5 vyrai ir 5 moterys. 4 vyrai ir 4 moterys žodžius ištarė po 10 kartų, taip gaunant 10 tarimo sesijų po 111 žodžių kiekvienam kalbėtojui, likę du po vieną kartą – iš viso 9102 žodžių. Kiekvienam kalbėtojui suteiktas kodas, nurodantis kalbančiojo lytį (M – moteris, V – vyras) ir kalbančiojo numerį. Kalbančiųjų charakteristikos ir įrašų santykis signalas-triukšmas (ST) pateikta 5.1 lentelėje.

5.1 lentelė. *Kalbėtojų ir įrašų charakteristikos*

Kalbėtojas	Gimtoji kalba	Amžius, m.	ST, dB
M1	lietuvių	22	12,9
M2	lietuvių	23	20,5
M3	lietuvių	23	22,3
M4	lietuvių	26	19,0
M5	lietuvių	22	27,2
V1	lietuvių	23	17,1
V2	lietuvių	23	12,4
V3	lietuvių	23	16,4
V4	lietuvių	28	23,8
V5	lietuvių	22	24,4

Kalbėtojų M5 ir V5 įrašai naudoti tik nuo kalbėtojo nepriklausomo atpažinimo eksperimentuose kaip testinės aibės.

5.3. Žodžio ribų nustatymo tyrimas

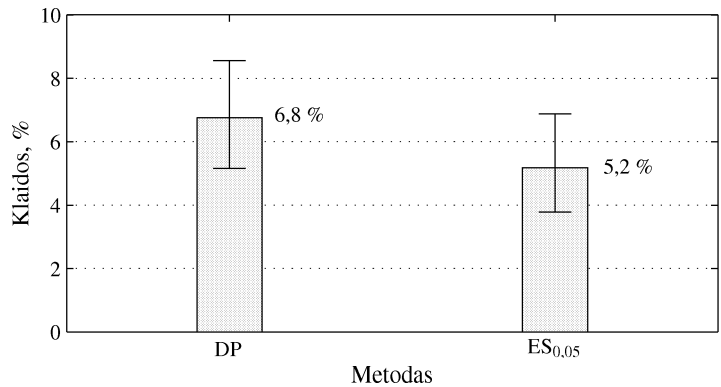
Tiriant žodžio ribų nustatymą buvo fiksuojamas ribų nustatymo klaidos. Klaidos fiksuotos lyginant rezultatus su rankiniu būdu nustatytomis žodžio ribomis. Rezultatams palyginti visi eksperimentai buvo atlikti su abiem žodžio ribų nustatymo metodais – energijos slenksčiu (ES) ir automatiniu nustatymu (pastarąjį, siekdami išvengti pilno metodo pavadinimo vartojimo ir pabrėžti realizacijos principą, toliau vadinsime dinaminio programavimo metodu – DP).

5.3.1. Žodžio ribų nustatymo tikslumo tyrimas

Šiame eksperimente tirtas žodžio ribų nustatymo tikslumas, fiksuojant nustatymo klaidą. Klaida buvo laikomas atvejis, jei išskirtoji riba nuo rankiniu būdu gautosios buvo nutolusi daugiau nei per 20 % žodžio trukmės. Siekiant sudaryti vienodas eksperimento sąlygas skirtingiems kalbėtojams, energijos slenksčio metode buvo naudojama vienoda slenksčio reikšmė visiems kalbėtojams – 0,05. Eksperimentams panaudoti kalbėtojų M1–M4 ir V1–V4 10-osios tarimo sesijos įrašai. Eksperimento rezultatai pateikti 5.2 lentelėje ir 5.1 paveiksle.

5.2 lentelė. *Individualūs ir vidutiniai žodžių ribų nustatymo tikslumo rezultatai*

Kalbėtojas	Klaidos, %	
	DP	ES
M1	16,2 (9,9 ÷ 24,5)	4,5 (1,5 ÷ 10,2)
M2	2,7 (0,5 ÷ 7,7)	0,9 (0 ÷ 4,9)
M3	10,8 (5,7 ÷ 18,2)	3,6 (0,9 ÷ 8,9)
M4	9,9 (5,0 ÷ 17,1)	6,3 (2,6 ÷ 12,6)
V1	9,0 (4,4 ÷ 15,9)	4,5 (1,5 ÷ 10,2)
V2	0,9 (0 ÷ 4,9)	7,2 (3,2 ÷ 13,7)
V3	1,8 (0,2 ÷ 6,4)	2,7 (0,5 ÷ 7,7)
V4	2,7 (0,5 ÷ 7,7)	11,7 (6,4 ÷ 19,2)
Vidut.	6,8 (5,2 ÷ 8,6)	5,2 (3,8 ÷ 6,9)



5.1 pav. *Žodžio ribų nustatymo rezultatai*

Vidutiniai rezultatai rodo, kad energijos slenksčio naudojimas žodžio riboms nustatyti duoda šiek tiek geresnius rezultatus, tačiau gautasis skirtumas nėra didelis

– apie 1,5 %. Bendru atveju, 5–7 % klaidų lygį laikytume vidutiniais. Individualūs kalbėtojų rezultatai dėsningumu nepasižymi – vieno kalbėtojo atveju tiksliau veikė ribų nustatymas naudojant dinaminį programavimą, kito atveju – energijos slenkstį. Kadangi dinaminio programavimo metode vertinami kalbos signalo (ir fono triukšmo) energijos parametrai, darome prielaidą, kad rezultatams įtakos galėjo turėti tiek kalbančiojo energetinės savybės, tiek signalo kokybė, t. y. santykis signalas-triukšmas, fono triukšmo tipas.

5.3.2. Triukšmo įtakos ribų nustatymui tyrimas

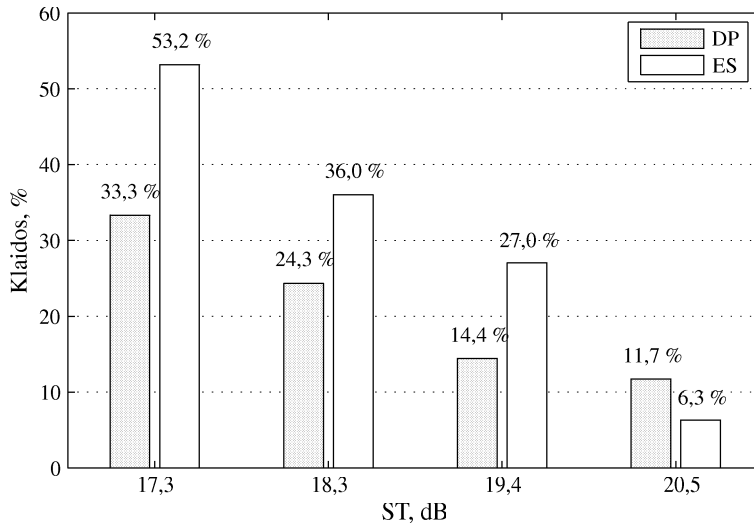
Šiame eksperimente tirta dinaminė žodžių ribų nustatymo metodų savybė – tikslumo kitimas kintant nagrinėjamo signalo santykiui signalas-triukšmas. Eksperimentams naudoti M3 kalbėtojo įrašai – jų santykis signalas-triukšmas buvo didžiausias, o ribų nustatymo tikslumo rezultatai artimiausi vidutiniams. Keičiant signalo kokybę, santykio signalas-triukšmas reikšmė buvo mažinama apytiksliai 1 dB žingsniu. Kad būtų užtikrintas energijos slenksčio darbingumas (pasikeitus triukšmo lygiui signale, metodas galėjo nebeveikti), energijos slenksčio reikšmė buvo parenkama kiekvienai signalas-triukšmas reikšmei atskirai. Siekiant užfiksuoti metodų jautrumą signalo kokybės kitimui, sugriežtintas ribų nustatymo klaidos kriterijus. Klaida laikyti atvejai, kai ribos nustatomos su paklaida, viršijančia 10 % žodžio trukmės arba prarandama daugiau nei pusę pirmo ar paskutinio garso. Eksperimento rezultatai pateikiami 5.3 lentelėje ir 5.2 paveiksle.

5.3 lentelė. *Ribų nustatymo tikslumo priklausomybė nuo signalo kokybės*

ST, dB	Slenkstis	Klaidos, %	
		DP	ES
17,3	0,06	33,3	53,2
18,3	0,05	24,3	36,0
19,4	0,03	14,4	27,0
20,5	0,02	11,7	6,3

Kintant santykiui signalas-triukšmas tenka keisti energijos slenksčio žodžio riboms nustatyti reikšmę. Tačiau net ir keičiant reikšmes nepavyksta užtikrinti stabilaus ribų nustatymo tikslumo. To priežastis išskirtume dvi – subjektyviąją ir objektyviąją. Subjektyvioji nestabilumo priežastis – nėra garantijos, jog parinktoji slenksčio reikšmė yra optimali turimam santykiui signalas-triukšmas. Objektyvioji priežastis – kintant santykiui signalas-triukšmas, kinta energetinės kalbos pavyzdžio savybės, kurios nulemia žodžio ribų aptikimą. Kaip pavyzdžius galima būtų paminėti žodžio pradžioje ar pabaigoje esančių nevokalizuotų „s“ ir „š“, sprogstamųjų „p“ ir „t“ paskendimą triukšme, atskirų garsų energiją viršijančią slenksčio reikšmę. To pasekoje slenksčio reikšmė puikiai tinkanti vieniems žo-

džiams, visiškai netinka kitiems – atsiranda poreikis slenkstį adaptuoti ne tik prie signalo ir triukšmo santykio, bet ir žodžio pobūdžio (vokalizuotu prasideda žodis ar ne, sprogstamuoju ar nesprogstamuoju ir pan.). Dinaminio programavimo atveju subjektyvių priežasčių nestabilumui nėra. Todėl pagrindiniu klaidų šaltiniu galime laikyti kintančią signalo kokybę.



5.2 pav. *Žodžio ribų nustatymo tikslumo priklausomybė nuo signalo kokybės*

Kaip ir pirmojo eksperimento metu dinaminis programavimas esant didžiausiai santykio signalas-triukšmas reikšmei tikslumu nusileido energijos slenksčio metodui, tačiau visais kitais atvejais jį lenkė. Tai rodo, jog dinaminis programavimas yra atsparesnis triukšmo lygiui signale. Dinaminio programavimo metodo klaidų lygis gali kilti esant didelėms santykio signalas-triukšmas reikšmėms. Tą būtų galima paaiškinti matematinės vilties maksimizavimo metodo taikymu žodžio ribų tikėtinumo funkcijai maksimizuoti. Pradiniai parametrų įverčiai gaunami iš fiksuotų, laisvai parinktų atkarpų (fono prieš ir po žodžio ir likusios dalies). Esant žemam fono triukšmo lygiui, pradiniai įverčiai gali gautis nekorektiški algoritmo požiūriu. Galimos tokios situacijos sprendimas – minimalių galimų įverčių reikšmių uždavimas.

Apibendrinant žodžio ribų nustatymo tyrimo rezultatus galime teigti, jog dinaminio programavimo naudojimas žodžio riboms nustatyti negarantuoja geresnių rezultatų, lyginant su energijos slenksčiu, tačiau užtikrina ribų nustatymo proceso automatizavimą (netenka keisti proceso parametrų) ir didesnį atsparumą signalo kokybės kitimui.

5.4. Žodžių atpažinimo tyrimas

Šiuose eksperimentuose tirta žodžių ribų nustatymo metodų, etalonų kūrimo metodų įtaka atpažinimo sistemos darbingumui. Darbingumas buvo vertinamas dviem kriterijais – teisingu atpažinimu ir neteisingu atmetimu. Teisingo atpažinimo atvejį turime, kai gautasis pavyzdžių atstumas neviršija nustatyto slenksčio reikšmės, o etalono fonetinė transkripcija sutampa su ištartuoju tekstu, neteisingo atmetimo atvejį – kai transkripcija atitinka tekstą, tačiau atstumas viršija slenksčio reikšmę, t. y. atmetamas teisingas žodis. Pastarąjį kriterijų pasirinkome, nes manome, jog padidinti teisingo atpažinimo lygį pirmiausia galima mažinant neteisingo atmetimo lygį.

5.4.1. Žodžio ribų nustatymo metodo įtakos atpažinimui tyrimas

Šiame eksperimente buvo tiriamas žodžių atpažinimo tikslumas ir neteisingo atmetimo lygis žodžio riboms nustatyti naudojant dinaminį programavimą (DP) ir energijos slenkstį (ES). Etalonams kurti naudotas tiesioginis etalono sukūrimo būdas, įvedant pirmosios tarimo sesijos įrašus, testine aibe naudoti dešimtosios sesijos įrašai. Atpažinimo rezultatu fiksuoti teisingo atpažinimo ir neteisingo atmetimo atvejai. Tiesinės prognozės modelio analizės (TPM) atveju naudotas atpažinimo slenkstis 0,7, tiesinės prognozės kepstro (TPMK) ir tiesinės prognozės kepstro su vidurkio atėmimu ($\overline{\text{TPMK}}$) analizių atveju – 0,5. Žodžio riboms nustatyti naudota energijos slenksčio reikšmė – 0,05. Teisingo atpažinimo rezultatai pateikti 5.4 lentelėje ir 5.3 paveiksle, o neteisingo atmetimo – 5.5 lentelėje ir 5.4 paveiksle. Lentelėse ties energijos slenksčio ir analizės metodų pavadinimais pateiktos naudotos slenksčio reikšmės.

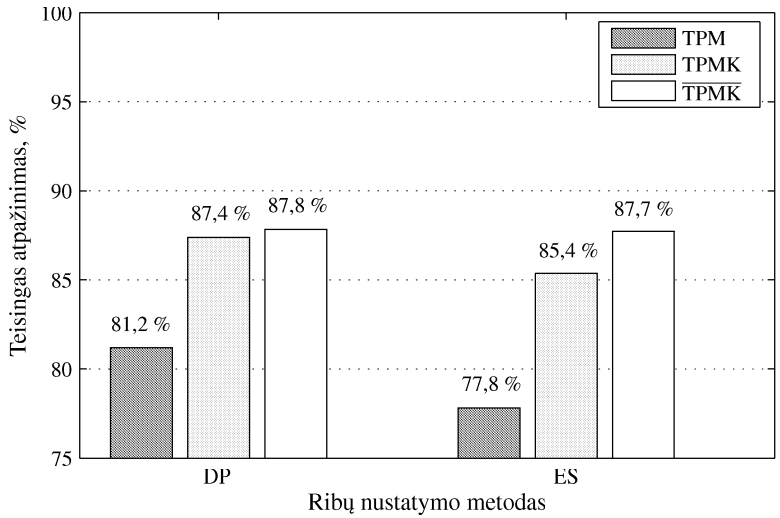
Nepaisant to, kad DP tikslumu nusileido energijos slenksčio metodui, žodžių atpažinimo su dinaminiu programavimu žodžio riboms nustatyti tikslumas gautas iki 3,5 % didesnis. Taigi dinaminio programavimo netikslumai nustatant žodžio ribas nėra pakankamai dideli, kad nulemtų atpažinimo rezultatus. Kita vertus, nagrinėdami individualius kalbėtojų atpažinimo rezultatus, galime matyti, jog žemiausias atpažinimo, žodžio riboms nustatyti naudojant DP, lygis gautas kalbėtojui M1, kuriam ir žodžio ribų nustatymo rezultatai buvo prasčiausi (5.2 lentelė). Šis momentas patvirtina aukščiau išsakytą teiginį, kad teisingas žodžių ribų nustatymas turi labai didelę įtaką atpažinimo rezultatams.

Žodžio ribų nustatymas naudojant dinaminį programavimą leido pasiekti geresnius rezultatus ir neteisingo atmetimo lygio požiūriu – vidutinis atmestų teisingų žodžių lygis sumažėjo 0,7–2 %. Nagrinėdami individualius kalbėtojų rezultatus matome, kad naudojant DP, net 6 kalbėtojų atveju kepstro analizė su vidurkio atėmimu leido išvengti teisingų žodžių atmetimo (naudojant energijos slenkstį – 4 kalbėtojų atveju), kepstro analizės atveju tai pasiekta 4 kalbėtojų atveju naudojant

DP (naudojant energijos slenkstį neteisingo atmetimo išvengti nepavyko).

5.4 lentelė. *Teisingas atpažinimas naudojant skirtingus žodžio ribų nustatymo metodus*

Kalbėtojas	Teisingas atpažinimas, %					
	DP			$ES_{0,05}$		
	$TPM_{0,7}$	$TPMK_{0,5}$	$\overline{TPMK}_{0,5}$	$TPM_{0,7}$	$TPMK_{0,5}$	$\overline{TPMK}_{0,5}$
M1	78,4	81,1	79,3	80,2	87,4	89,2
M2	77,5	90,9	93,7	64,9	84,7	90,9
M3	81,9	85,6	85,6	85,6	90,1	90,1
M4	86,5	90,9	90,9	83,8	90,1	90,9
V1	75,7	85,6	83,8	79,3	82,9	81,9
V2	78,4	82,9	87,4	74,8	74,8	84,7
V3	78,4	84,7	86,5	73,9	87,4	87,4
V4	92,8	97,3	95,5	80,2	85,6	86,5
Vidut.	81,2	87,4	87,8	77,8	85,4	87,7



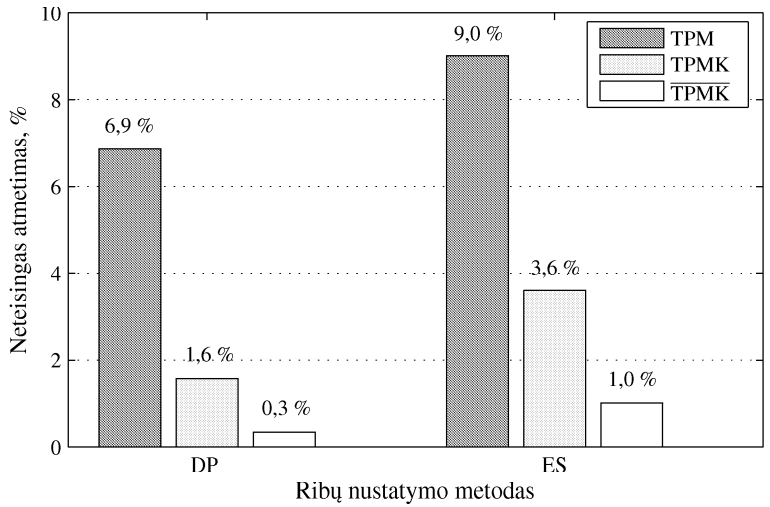
5.3 pav. *Teisingas atpažinimas naudojant skirtingus žodžių ribų nustatymo metodus*

Tiek teisingo atpažinimo, tiek neteisingo atmetimo atveju, dinaminio programavimo panaudojimas žodžio riboms nustatyti labiausiai paveikė tiesinės prognozės analizės rezultatus, tuo tarpu kai kuriais kepstro ir kepstro su vidurkio atėmimu analizių atveju žodžio ribų nustatymo metodo pakeitimas beveik neįtakojo rezul-

tatų. Viena vertus, tai gali reikšti, kad požymių sistemos atsparumas triukšmams ir kalbančiajam leidžia sumažinti žodžio ribų nustatymo netikslumų įtaką atpažinimo rezultatams. Kita vertus, tai gali liudyti apie pasirinktų atpažinimo slenksčio reikšmių neoptimalumą tiesinės prognozės analizės atveju.

5.5 lentelė. *Neteisingas atmetimas atpažinime naudojant skirtingus žodžio ribų nustatymo metodus*

Kalbėtojas	**Neteisingas atmetimas, %**					
	DP			**ES**$_{0,05}$		
	TPM$_{0,7}$	**TPMK**$_{0,5}$	$\overline{\textbf{TPMK}}_{0,5}$	**TPM**$_{0,7}$	**TPMK**$_{0,5}$	$\overline{\textbf{TPMK}}_{0,5}$
M1	1,8	0,9	0	5,4	1,8	0
M2	16,2	4,5	0,9	27,0	9,0	1,8
M3	6,3	1,8	1,8	6,3	2,7	2,7
M4	5,4	0,9	0	8,1	2,7	0
V1	4,5	0	0	1,8	0,9	1,8
V2	7,2	0	0	3,6	3,6	1,8
V3	10,8	4,5	0	14,4	6,3	0
V4	2,7	0	0	5,4	1,8	0
Vidut.	6,9	1,6	0,3	9,0	3,6	1,0



5.4 pav. *Neteisingas atmetimas atpažinime naudojant skirtingus žodžių ribų nustatymo metodus*

5.4.2. Etalono kūrimo tyrimas

Šiame eksperimente tirta etalonų kūrimo metodų įtaka atpažinimo tikslumui. Eksperimento tikslas – palyginti etalonų kūrimo metodus ir parinkti optimalų etalonų kiekį vienam žodžiui. Etalonams kurti naudoti abu įdiegtieji metodai – tiesioginis metodas ir mokymas. Vienam žodžiui buvo kuriama nuo 1 iki 5 etalonų (abiem metodais) ir tiriamas atpažinimo tikslumas. Šįkart neteisingo atmetimo lygis nefiksuotas, kadangi tikslas yra nustatyti optimalų etalonų skaičių, t. y. minimalų etalonų skaičių, leidžiantį pasiekti maksimalų atpažinimo tikslumą. Eksperimentui pasirinktas kalbėtojas M1, kadangi jo atveju gavome mažiausią atpažinimo tikslumą, žodžio ribas nustatinėjant dinaminio programavimo metodu (5.4 lentelė). Etalonams kurti naudotos pirmosios įrašų sesijos (tiesioginio kūrimo metodui reikėjo 5, mokymui – 9), atpažinimui testuoti – 10-osios įrašai. Eksperimentai atlikti visiems analizės tipams, užduodant tas pačias atpažinimo slenksčio reikšmes kaip ir ankstesniame eksperimente (jos pateiktos lentelėje). Žodžio riboms nustatyti naudotas dinaminis programavimas. Eksperimento rezultatai pateikti 5.6 lentelėje ir 5.5 paveiksle.

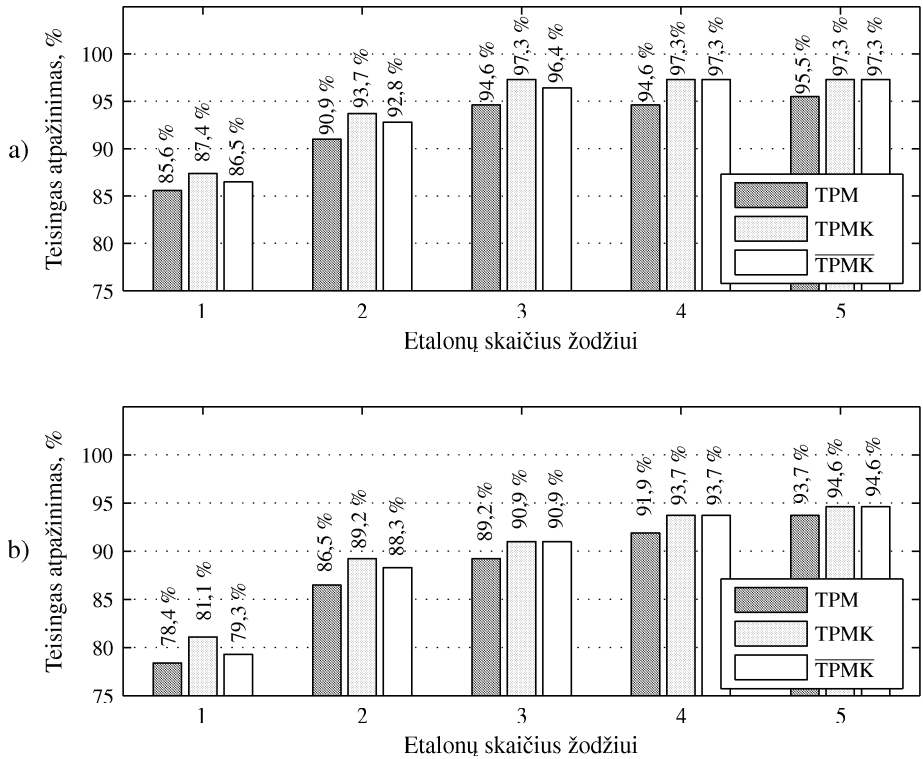
5.6 lentelė. *Teisingo atpažinimo priklausomybė nuo etalonų skaičiaus vienam žodžiui*

Etalonų skaičius žodžiui	**Teisingas atpažinimas, %**					
	Mokymas			**Tiesioginis metodas**		
	TPM$_{0,7}$	**TPMK**$_{0,5}$	$\overline{\textbf{TPMK}}_{0,5}$	**TPM**$_{0,7}$	**TPMK**$_{0,5}$	$\overline{\textbf{TPMK}}_{0,5}$
1	85,6	87,4	86,5	78,4	81,1	79,3
2	90,9	93,7	92,8	86,5	89,2	88,3
3	94,6	97,3	96,4	89,2	90,9	90,9
4	94,6	97,3	97,3	91,9	93,7	93,7
5	95,5	97,3	97,3	93,7	94,6	94,6

Matome, jog etalonams kurti naudojant tiesioginio kūrimo metodą, atpažinimo proceso tikslumo priklausomybė nuo etalonų skaičiaus yra beveik tiesinio pobūdžio. Jei bandytume išrinkti optimalų etalonų skaičių, juo galėtume pasirinkti bet kurį etalonų kiekį – visais atvejais etalonų skaičiaus padidėjimas pakelia atpažinimo tikslumo lygį. Naudojant mokymą, teisingo atpažinimo lygiui augant etalonų skaičiui būdingas „įsisotinimas“ – kuriant daugiau nei 3 etalonus, teisingo atpažinimo lygis beveik nekinta. Taigi šio eksperimento atveju optimalus skaičius yra 3 etalonai vienam žodžiui.

Lyginant etalonų kūrimo metodų rezultatus tarpusavyje matosi, kad tiesioginiu metodu sudarant net 5 etalonus vienam žodžiui, atpažinimo tikslumas, gaunamas naudojant mokymą 3 etalonais, nepasiekiamas. Be abejonės, jei tiesioginiu metodu būtų kuriama 6 ir daugiau etalonų (šiuo konkrečiu atveju tai būtų 9 etalonai – tiek kiek įrašų sunaudota sistemai apmokyti vienu žodžiu), atpažinimo su mokymu

tikslumas būtų pasiektas, tačiau tai reikštų keletą kartų didesnį žodyną, o dinaminiame laiko skalės kraipymo metode tai tolygu keletą kartų ilgesniam atpažinimo procesui.



5.5 pav. *Teisingo atpažinimo priklausomybė nuo etalonų skaičiaus vienam žodžiui, etalonus kuriant: a) mokymo būdu ir b) tiesioginio kūrimo metodu*

Siekdami įsitikinti sistemos apmokymo trimis etalonais vienam žodžiui efektyvumu, sekančiame eksperimente atliksime sistemos, naudojančios dinaminį programavimą žodžio riboms nustatyti ir mokymą etalonams kurti, atpažinimo tyrimą.

5.4.3. Atpažinimo, naudojant DP riboms nustatyti ir mokymą, tyrimas

Šiame eksperimente atpažinimo tikslumą fiksuojame 2 kriterijais – teisingo atpažinimo ir neteisingo atmetimo lygiais. Siekdami įvertinti automatinio ribų

nustatymo ir mokymo procedūros įtaką atpažinimui, eksperimentą atliksime nuo kalbėtojo priklausomo ir nepriklausomo atpažinimo sąlygomis, o rezultatus palyginsime su paprasčiausios atpažinimo sistemos atveju, kai žodžio riboms nustatyti naudojamas DP metodas, o etalonams kurti – tiesioginis metodas (kuriamas vienas etalonas vienam žodžiui). Eksperimentas atliekamas su visų kalbėtojų įrašais, sistemai mokinti skiriant pirmųjų devynių įrašo sesijų duomenis. Žodžio riboms nustatyti naudojamas dinaminis programavimas, kuriamų etalonų skaičius vienam žodžiui – 3. Signalui taikomi visi trys analizės būdai, naudojant atitinkamas atpažinimo slenksčių reikšmes.

Pirmiausia atliktas nuo kalbančiojo priklausomo atpažinimo tyrimas, kurio teisingo atpažinimo rezultatai pateikiami 5.7 lentelėje ir 5.6 paveiksle, neteisingo atmetimo – 5.8 lentelėje ir 5.7 paveiksle.

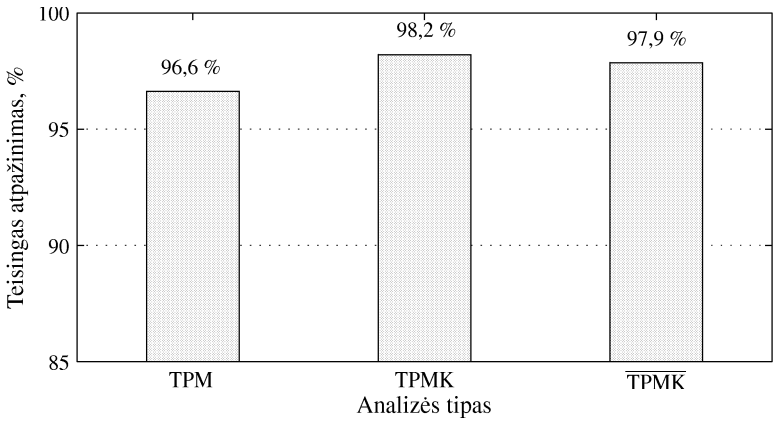
5.7 lentelė. *Teisingas atpažinimas naudojant DP ir mokymą*

Kalbėtojas	**Teisingas atpažinimas, %**		
	TPM$_{0,7}$	**TPMK**$_{0,5}$	$\overline{\textbf{TPMK}}_{0,5}$
M1	93,7	97,3	96,4
M2	96,4	97,3	96,4
M3	98,2	100	99,1
M4	96,4	98,2	97,3
V1	97,3	98,2	99,1
V2	98,2	99,1	99,1
V3	94,6	95,5	95,5
V4	98,2	100	100
Vidut.	96,6	98,2	97,9

Kaip matome, teisingas atpažinimas, naudojant dinaminį programavimą riboms nustatyti ir mokymą etalonams kurti, šoktelėjo iki 96–98 % lygio. Palyginę šiuos rezultatus su paprasčiausiu atpažinimo sistemos atveju (5.4 lentelė), matome, kad teisingo atpažinimo tikslumas padidėjo nuo 10 % kepstro analizės atveju iki 15 % tiesinės prognozės atveju. Be to, tiesinės prognozės modelio analizė savo rezultatais beveik prilygo kepstro ir kepstro su vidurkio atėmimu analizių rezultatams. Reikėtų turėti omeny, kad pasirinktų atpažinimo slenksčio reikšmių optimalumas nenagrinėtas, todėl nereikėtų atmesti galimybės, kad atpažinimo rezultatai kistų parenkant kitas atpažinimo slenksčio reikšmes.

Neteisingo atmetimo lygis sumažėjo nuo 1 % kepstro analizės atveju iki 10 % tiesinės prognozės atveju. Nors tiesinės prognozės analizės atveju gaunamas neteisingo atmetimo lygis yra didesnis už kepstro analizės atvejais gaunamus, tačiau yra pakankamai mažas. Kepstro analizių atveju gautas neteisingo atmetimo lygis yra paneigtinai mažas. Remdamiesi šiais rezultatais drįstame teigti, jog nuo kalbėtojo priklausomo atpažinimo atveju dažnai akcentuojamas tiesinės prognozės modelio

silpnumas kalbos atpažinime gali būti pašalintas naudojant papildomas procedūras (mūsų atveju tai buvo mokymo procedūra).



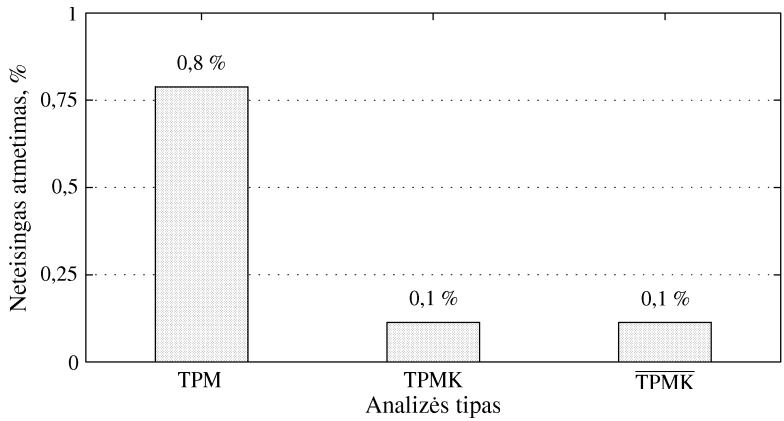
5.6 pav. *Teisingas atpažinimas naudojant DP ir mokymą*

5.8 lentelė. *Neteisingas atmetimas naudojant DP ir mokymą*

Kalbėtojas	**Neteisingas atmetimas, %**		
	TPM$_{0,7}$	**TPMK**$_{0,5}$	$\overline{\textbf{TPMK}}_{0,5}$
M1	0,9	0	0
M2	0,9	0	0
M3	1,8	0	0
M4	0	0	0
V1	0	0	0
V2	0,9	0	0
V3	0,9	0,9	0,9
V4	0,9	0	0
Vidut.	0,8	0,1	0,1

Antroje eksperimento dalyje atliktas nepriklausomo nuo kalbėtojo atpažinimo tyrimas. Šiuo atveju testinėmis aibėmis naudotos M5 ir V5 įrašai (mokyme jie nenaudoti). Tiesinės prognozės modelio analizės atveju parinktas atpažinimo slenkstis – 1,9, kepstro analizės atveju – 1, kepstro su vidurkio atėmimu atveju – 0,7. Teisingo atpažinimo ir neteisingo atmetimo rezultatai (lyginamosios ir tiriamosios sistemų) pateikiami 5.9 ir 5.10 lentelėse, bei 5.8 paveiksle.

Kaip matome, sistemos mokymas leido padidinti vidutinį teisingo atpažinimo lygį 10–11 % ir pasiekti maksimalų 68,5 % reikšmę. Tačiau panagrinėję individua-



5.7 pav. *Neteisingas atmetimas naudojant DP ir mokymą*

5.9 lentelė. *Teisingas atpažinimas nepriklausomame nuo kalbėtojo atpažinime*

Kalbėtojas	**Teisingas atpažinimas, %**					
	1 etalonas			**Mokymas 3 etalonais**		
	$\mathbf{TPM}_{1,9}$	$\mathbf{TPMK}_{1}$	$\overline{\mathbf{TPMK}}_{0,7}$	$\mathbf{TPM}_{1,9}$	$\mathbf{TPMK}_{1}$	$\overline{\mathbf{TPMK}}_{0,7}$
M1	17,1	30,6	48,6	16,2	34,2	65,8
M2	13,5	23,4	54,9	15,3	20,7	63,1
M3	18,0	27,0	60,4	49,5	57,7	79,3
M4	43,2	62,2	62,2	53,2	72,1	68,5
V1	68,5	63,1	72,1	85,6	83,8	86,5
V2	0	5,4	55,9	3,6	9,0	55,9
V3	21,6	20,7	49,5	33,3	27,9	64,9
V4	49,5	53,2	54,9	62,2	60,4	63,9
Vidut.	28,9	35,7	57,3	39,9	45,7	68,5

lius kalbėtojų rezultatus matome, jog jie labai priklauso nuo kalbėtojo ir teisingo atpažinimo pokytis siekė 70 %. Be to, kalbėtojų M1 ir M2 atvejais rezultatai, gauti naudojant mokymą, buvo prastesni nei naudojant tiesioginį vieno etalono sukūrimą. Taigi galime daryti išvadą, jog apmokymo aibės požiūriu optimalių etalonų parinkimas negarantuoja didesnio teisingo atpažinimo lygio nuo kalbėtojo nepriklausomame atpažinime.

Neteisingo atmetimo rezultatai savo pobūdžiu labai panašūs į teisingo atpažinimo rezultatus – tiesinės prognozės ir kepstro atveju gauti dideli rezultatų svyravimai tarp kalbėtojų (iki 50 %), o kalbėtojų M1, M2, V2 ir V3 atvejais mokymo

panaudojimas lėmė aukštesnį neteisingo atmetimo lygį. Taigi nepriklausomo nuo kalbėtojo atpažinimo atveju mokymo panaudojimas negarantuoja ne tik didesnio teisingo atpažinimo, bet ir mažesnio neteisingo atmetimo lygio. Tiksliausi ir stabiliausi nepriklausomo nuo kalbėtojo atpažinimo rezultatai gauti naudojant kepstro su vidurkio atėmimu analizę (tiek teisingo atpažinimo, tiek neteisingo atmetimo požiūriu).

5.10 lentelė. *Neteisingas atmetimas nepriklausomame nuo kalbėtojo atpažinime*

Kalbėtojas	Neteisingas atmetimas, %					
	1 etalonas			Mokymas 3 etalonais		
	$\mathbf{TPM}_{1,9}$	$\mathbf{TPMK}_1$	$\overline{\mathbf{TPMK}}_{0,7}$	$\mathbf{TPM}_{1,9}$	$\mathbf{TPMK}_1$	$\overline{\mathbf{TPMK}}_{0,7}$
M1	41,4	30,6	26,1	47,7	34,2	18,0
M2	51,4	44,1	23,4	53,2	51,4	18,0
M3	45,0	45,9	18,0	33,3	25,2	8,1
M4	28,8	16,2	18,0	21,6	8,1	16,2
V1	0	0	0	0	0	0,9
V2	27,0	17,1	0,9	36,0	20,7	0,9
V3	20,7	18,9	2,7	24,3	17,1	0,9
V4	1,8	0,9	0	0,9	0,9	0
Vidut.	27,0	21,7	11,1	27,1	19,7	7,9

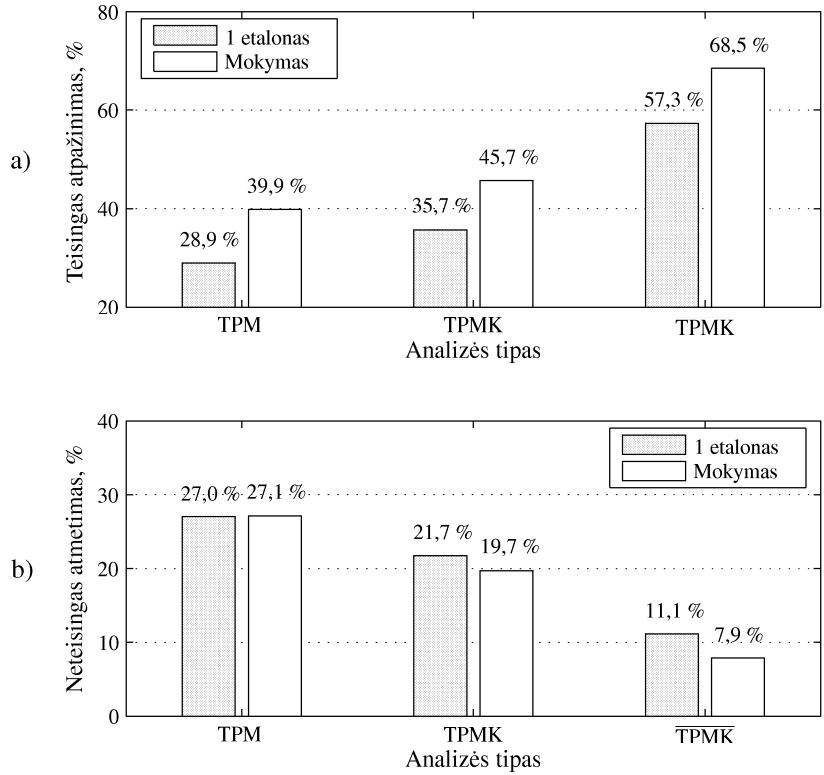
Taigi pasiūlymas optimizuoti atpažinimo procesą optimizuojant etalonų kūrimą pilnai pasiteisino tik priklausomo nuo kalbėtojo atpažinimo atveju – panaudojus klasterizavimu pagrįstą sistemos apmokymą, rezultatai, gaunami naudojant skirtingus analizės metodus, praktiškai susilygino, o klaidų lygis tesiekė keletą procentų. Nepriklausomo nuo kalbėtojo atpažinimo procesas buvo labai nestabilus – tiesinės prognozės ir kepstro analizių atveju atpažinimo tikslumo ir neteisingo atmetimo lygio svyravimo nesumažino netgi mokymas trimis etalonais vienam žodžiui.

Baigdami norėtume akcentuoti tai, kad šie rezultatai gauti nekeičiant ir nemodifikuojant signalo analizės ar požymių klasifikacijos metodo.

5.5. Žodžių segmentavimo tyrimas

Šiuose eksperimentuose tirti sukurtieji žodžių segmentavimo metodai – tikėtinumo funkcijos maksimizavimo ir prognozės klaidos minimizavimo metodai. Pagrindinis segmentavimo rezultato vertinimo kriterijus buvo išskirtų garsų skaičius. Jeigu išskirtų garsų skaičius prilygdavo vartotojo nurodytam (tikrajam garsų skaičiui žodyje), segmentavimas buvo laikomas sėkminga, nepaisant garsų ribų netikslumo. Kalbos signalui filtruoti aukštų dažnių filtru naudotas koeficientas 0,95,

analizei naudotas 250 atskaitų (22,6 ms) kadras.



5.8 pav. *Nepriklausomo nuo kalbėtojo atpažinimo rezultatai: a) teisingas atpažinimas; b) neteisingas žodžių atmetimas*

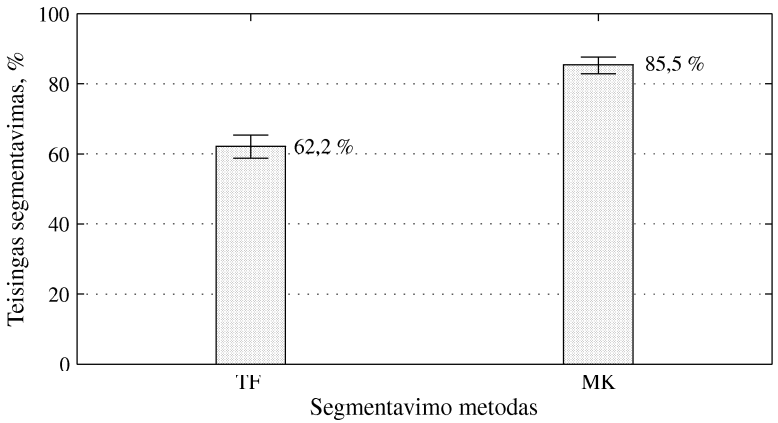
5.5.1. Segmentavimo tikslumo tyrimas

Segmentavimo tikslumo eksperimente fiksuotas tikėtinumo funkcijos maksimizavimo (TF) ir prognozės klaidos minimizavimo (MK) metodų tikslumas – segmentavimo klaidų lygis, klaidų tipai ir išskirtų garsų ribų paklaidos. Klaidas skirstėme į vieno, dviejų, trijų, keturių ir daugiau nei keturių garsų klaidas, taip įvardindami klaidingo segmentavimo atvejus, kai būdavo gaunama vienu garsu mažiau nei yra žodyje, dviem, trimis, keturiomis ir daugiau nei keturiomis. Taip įvykdavo susiliejus gretimams garsams – nepavykus rasti staigaus pasikeitimo tarp gretimų garsų parametrų. Išskirtųjų ribų reikšmes lyginome su rankiniu būdu išskirtomis

garsų ribomis ir vertinome jų neatitikimą. Eksperimentas atliktas su kalbėtojų M1–M4 ir V1–V4 dešimtosios tarimo sesijos įrašais. Segmentavimo algoritmų tikslumo tyrimo rezultatai pateikti 5.11–5.12 lentelėse ir 5.9–5.10 paveiksluose.

5.11 lentelė. *Individualūs ir vidutiniai segmentavimo metodų rezultatai*

Kalbėtojas	**Teisingas segmentavimas, %**	
	TF metodas	**MK metodas**
M1	64,9 (54,9 ÷ 73,8)	83,8 (75,5 ÷ 90,1)
M2	83,8 (75,6 ÷ 90,1)	94,6 (88,6 ÷ 98,0)
M3	87,4 (79,8 ÷ 92,9)	86,5 (78,6 ÷ 92,2)
M4	82,9 (74,5 ÷ 89,4)	84,7 (76,6 ÷ 90,8)
V1	51,4 (40,9 ÷ 61,3)	83,8 (75,5 ÷ 90,1)
V2	30,6 (22,2 ÷ 40,3)	89,2 (81,8 ÷ 94,3)
V3	45,9 (36,2 ÷ 56,3)	81,1 (72,5 ÷ 87,9)
V4	50,5 (39,9 ÷ 60,4)	80,2 (71,4 ÷ 87,2)
Vidut.	62,2 (58,8 ÷ 65,4)	85,5 (82,9 ÷ 87,7)



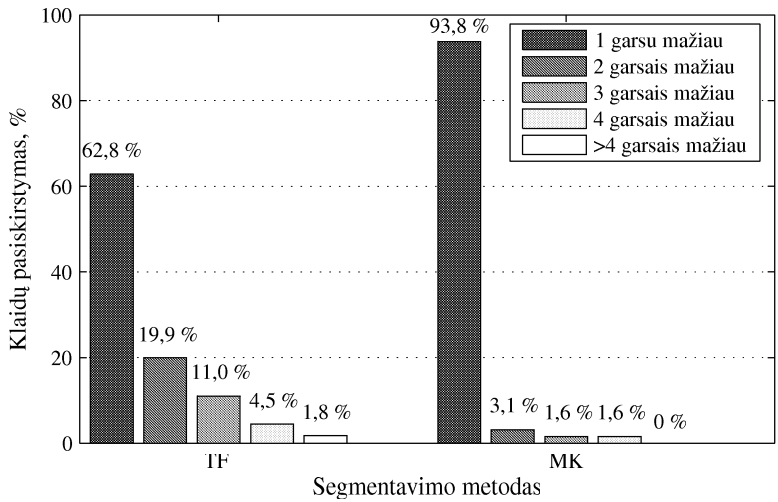
5.9 pav. *Vidutiniai segmentavimo metodų rezultatai*

Nagrinėdami metodų teisingo segmentavimo rezultatus matome, jog klaidos prognozės klaidos minimizavimo metodas veikė pakankamai stabiliai (lygis svyravo apie 80–90 %) ir lenkė tikėtinumo funkcijos maksimizavimo metodą, kurio teisingo segmentavimo lygis kito nuo 30,6 % V2 kalbėtojo atveju iki 87 % M3 atveju. Sprendimo konvergavimo požiūriu nė vienas metodas nebuvo pranašesnis – TF metodo atveju garsų riboms nustatyti reikėjo vidutiniškai 3–6 iteracijų, MK

atveju – vidutiniškai 3–5. Garsų skaičius žodyje (žodžio ilgis) iteracijų skaičiui įtakos nepastebėta.

5.12 lentelė. *Segmentavimo klaidų prarandant garsus pasiskirstymas*

Kalbėtojas	Klaidos prarandant garsus, %									
	TF metodas					MK metodas				
	1	2	3	4	>4	1	2	3	4	>4
M1	84,6	15,4	0	0	0	100	0	0	0	0
M2	83,3	16,7	0	0	0	100	0	0	0	0
M3	64,3	7,1	21,4	0	7,1	80,0	20,0	0	0	0
M4	84,2	10,5	5,3	0	0	100	0	0	0	0
V1	81,5	11,1	7,4	0	0	100	0	0	0	0
V2	40,3	36,4	10,4	10,4	0	100	0	0	0	0
V3	65,0	10,0	20,0	5,0	0	100	0	0	0	0
V4	43,6	27,3	16,4	7,3	5,5	81,8	0	9,1	9,1	0
Viso	62,8	19,9	11,0	4,5	1,8	93,8	3,1	1,6	1,6	0



5.10 pav. *Segmentavimo klaidų prarandant garsus pasiskirstymas*

Beje, pasitaikė atvejų, kuomet segmentavimas patekdavo į uždarą ciklą – remiantis iteracijos metu gautų atkarpų parametrais, naujai gauti garsų ribų įverčiai sutapdavo su ankstesnės iteracijos rezultatais. Tokių ciklišku skaičiavimų rezultatas būdavo pakaitomis besikartojantis dviejų paskutiniųjų prieš patenkant į ciklą

iteracijų rezultatas. TF metode tokie atvejai sudarė apie 3 %, MK metode – apie 6,1 %. Tačiau šie ciklai problemų nesudaro, kadangi juo lengva aptikti lyginant n-iosios ir $(n-1)$-iosios iteracijų (čia n – nagrinėjamosios iteracijos numeris) rezultatus.

Prognozės klaidos minimizavimo metodo pranašumas matosi ir nagrinėjant segmentavimo klaidų prarandant garsus pasiskirstymą. Tikėtinumo funkcijos maksimizavimo metode vieno garso praradimo klaidos sudaro apie du trečdalius visų klaidų, neretas ir dviejų garsų praradimo atvejis. Prognozės klaidos minimizavimo metode dauguma segmentavimo klaidų padaryta prarandant vieną garsą, kitos klaidos tesudarė keletą procentų.

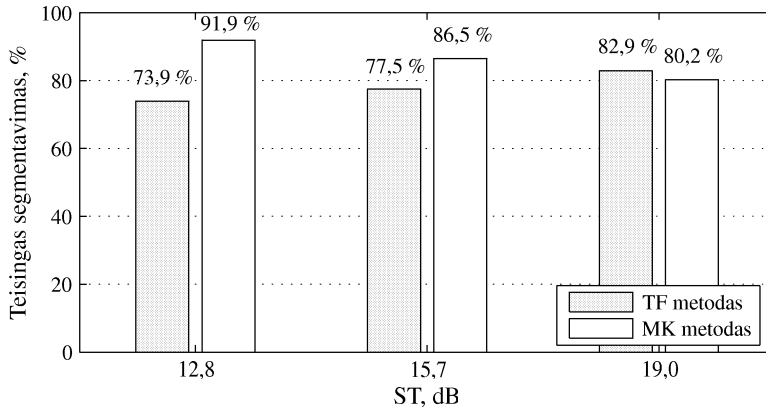
Garsų ribų nustatymo tikslumo požiūriu MK metodas taip pat buvo pranašesnis, jo ribų nustatymo paklaidos vidurkis buvo apie 43,3 ms. TF metodo atveju nustatomų ribų paklaidos vidurkis siekė 50,5 ms – beveik 17 % didesnis už MK metodo. Pastebėta ir charakteringų ribų nustatymo paklaidų atskiriems kalbėtojams. Pvz., kalbėtojo M2 atveju naudojant TF metodą, dažna ir nemaža pirmojo garso pradžios paklaida buvo gaunama žodžiams, prasidedantiems sprogstamuoju garsu. V1 kalbėtojo atveju dažnai buvo prarandams garsas „s“ žodžio pabaigoje. To priežastį įvardintume signalo kokybę – santykį signalas-triukšmas, kuriam esant nepakankamam, kai kurie garsai „paskęsdavo“ triukšme – savo parametrais supanašėdavo su fono triukšmu.

5.5.2. Triukšmo įtakos segmentavimui tyrimas

Šiame eksperimente tyrėme signalo kokybės įtaką segmentavimo rezultatams – kaip kinta segmentavimo rezultatas didėjant triukšmo lygiui signale. Testine aibe pasirinkti kalbėtojo M4 įrašai dėl gerų segmentavimo (abiem metodais) rezultatų ankstesniame eksperimente ir tarties natūralumo. Manome, jog kalbėjimo būdas gali turėti įtakos segmentavimo rezultatams, todėl siekdami realios kalbos segmentavimo įvertinimo, tarimo natūralumą taikėme kaip reikalavimą testinei aibei. Eksperimento metu fiksuotas teisingo segmentavimo faktas, reiškiantis, kad segmentavimo metu gautas garsų skaičius yra lygus žodžio garsų skaičiui. Eksperimento rezultatai pateikiami 5.13 lentelėje ir 5.11 paveiksle.

5.13 lentelė. *Segmentavimo rezultatų priklausomybė nuo signalo kokybės*

ST, dB	**Teisingas segmentavimas, %**	
	TF metodas	**MK metodas**
12,8	73,9	91,9
15,7	77,5	86,5
19,0	82,9	80,2



5.11 pav. *Segmentavimo rezultatų priklausomybė nuo signalo kokybės (kalbėtojo M4 atveju)*

Rezultatai rodo, kad tikėtinumo funkcijos maksimizavimo metodo rezultatai augant triukšmo lygiui signale prastėja. Tuo tarpu prognozės klaidos minimizavimo metodo rezultatai gerėja. Kadangi metodai tarpusavyje skiriasi tik tiesinės prognozės modelio stiprinimo koeficiento kvadrato naudojimu TF metode, galime daryti prielaidą, kad būtent jis yra tokių rezultatų priežastis.

Apibendrindami teigiame, kad segmentavimas naudojant prognozės klaidos minimizavimo metodą pasirodė beesantis tikslesnis – gauta žymiai mažiau klaidų, o ir dauguma klaidų pasireiškė vieno garso praradimu. Tuo tarpu TF metodas klydo dažniau, o vidutinis prarandamų žodžio garsų skaičius buvo didesnis. Be to, TF metodo rezultatai buvo pakankami nestabilūs ir žymiai kito tarp kalbėtojų. MK metodo rezultatai buvo daug stabilesni ir rezultatų skirtumas tarp kalbėtojų nebuvo didelis. Tokias segmentavimo metodų savybes galima paaiškinti tiesinės prognozės modelio stiprinimo koeficientu kvadratu TF metodo išraiškose. Koeficiento reikšmė paprastai būna didelė (tūkstančių eilės), todėl jis stipriai įtakoja metodo skaičiavimus ir netgi nedideli koeficiento reikšmių pokyčiai (dėl skirtingų kalbėtojų, besiskiriančios signalo kokybės) gali stipriai paveikti skaičiavimo rezultatus.

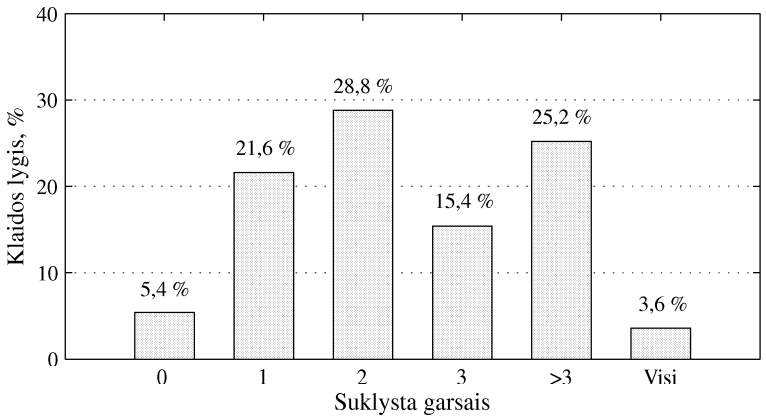
5.6. Žodžio segmentų atpažinimo tyrimas

Šiame eksperimente tirtas žodžio segmentų (garsų) atpažinimo tikslumas (ištarimams segmentuoti pasirinktas prognozės klaidos minimizavimo metodas). Testine aibe dėl gerų segmentavimo rezultatų naudoti kalbėtojo M2 įrašai. Eksperimento metu žodžiai susegmentuoti į garsus ir sužymėti (signalo segmentams pri-

skirta fonetinė garso transkripcija). Garsams žymėti panaudojome lietuvių gramatikos simbolius. Tylos atkarpoms (prieš ir po žodžio) žymėti panaudotas pabraukimo simbolis „_“. Toks žymėjimas leido išvengti ilgo žymėjimo proceso (kuris naudojant standartines transkribavimo sistemas būtų sudėtingesnis), o atpažinimo rezultatus gáuti paprasta tekstine forma. Garsų etalonai sukurti naudojant tiesioginį kūrimo metodą. Etalonai kurti visiems išskirtiems garsams, nepaisant segmentavimo tikslumo – susiliejusiems garsams buvo priskiriama atitinkama dvigarsio transkripcija, netikslių ribų atveju buvo priskiriama transkripcija garso, kuris turėtų būti pagal žodžio garsų seką. Atpažinimo fazėje testinės sesijos žodžių įrašai segmentuoti į garsus, pastaruosius vėliau atpažįstant. Eksperimento metu fiksuoti sukeitimo ir ištrynimo klaidų lygiai, teisingai atpažintų žodžio garsų kiekis (tylos atkarpos prie garsų nepriskirtos). Eksperimento rezultatai pateikti 5.14 lentelėje ir 5.12 paveiksle.

5.14 lentelė. *Žodžio garsų atpažinimo rezultatai*

Analizė	**Klaidų lygis neatpažįstant garsų skaičiaus žodyje, %**					
	0	**1**	**2**	**3**	**>3**	**Visi**
$TPMK_{0,5}$	5,4	21,6	28,8	15,4	25,2	3,6



5.12 pav. *Žodžio garsų atpažinimo rezultatai*

Lentelės stulpeliuose esantys skaičiai nurodo neatpažintų garsų skaičių žodžiuose (nesigilinant, buvo tai sukeitimo klaida ar ištrynimo). Taigi 0 garsų atvejis atitinka pilno žodžio atpažinimą, visų garsų atvejis – visišką žodžio neatpažinimą, kai neatpažintas nė vienas garsas.

Matome, jog dažniausiai pasitaikė dviejų garsų neatpažinimo klaidos, jų lygis sudarė beveik trečdalį visų klaidų. Teisingo atpažinimo ir visiško neatpažinimo atvejai sudarė panašią dalį. Vidutiniškai viename žodyje buvo padaroma 2,5 sukeitimo ir beveik 0,1 ištrynimo klaidų, atpažįstant 57,8 % žodžio.

Prisimindami savo teiginį apie lingvistinio apdorojimo poreikį, gautiesiems atpažinimo rezultatams, o tiksliau vieno garso neatpažinimo klaidos atvejams, pritaikėme patį elementariausią lingvistinį apdorojimą – automatizuotą gramatikos tikrinimą. Tikrinimo rezultatu buvo priimamas pirmasis pasiūlytas ištaisymo variantas. Eksperimento rezultatas – teisingo atpažinimo lygis šoktelėjo iki 15,3 %, neatpažinto vieno garso klaidų lygis smuko iki 11,7 % – gautas maždaug 10 % absoliutus pokytis (palyginus su rezultatais 5.14 lentelėje).

Išskirtume dvi žodžio segmentų atpažinimo tikslumo didinimo kryptis, kuriomis tektų dirbti, siekiant realizuoti efektyvų segmentų atpažinimą. Viena jų – segmentų atpažinimo tobulinimas, tobulinant segmentavimo procesą, parenkant požymių sistemas, segmentų tarpusavio panašumo įvertinimo metodus ir pan. Kita vertus, atpažinimo tikslumas gali būti padidintas taikant lingvistinį apdorojimą. Jei darytumėme prielaidą, kad mes galime realizuoti lingvistinį atpažinimo rezultatų apdorojimą, sugebantį pilnai pašalinti vieno garso klaidas ar net sumažinti dviejų garsų klaidų lygį, atpažinimo tikslumą (mūsų eksperimento atveju) galima būtų padidinti keliolika ar net keliasdešimt procentų.

5.7. Penktojo skyriaus apibendrinimas

- Ištirtas sukurtojo metodo žodžio ribų nustatymo tikslumas ir jo priklausomybė nuo signalo kokybės.
- Ištirta etalonų kūrimo, naudojant klasterizavimą, įtaka atpažinimo tikslumui.
- Eksperimentiškai patvirtintas pavienių žodžių atpažinimo tikslumo augimas įdiegus sukurtąjį žodžio ribų nustatymo metodą ir klasterizacija pagrįsto etalonų kūrimą.
- Ištirtos sukurtųjų segmentavimo tikėtinumo maksimizavimo ir prognozės klaidos minimizavimo metodų savybės: segmentavimo tikslumas, klaidos tipų pasiskirstymas, segmentavimo rezultatų priklausomybė nuo signalo kokybės.
- Eksperimentiškai iliustruota žodžių atpažinimo segmentais galimybė. Suformuluoti pasiūlymai atpažinimui garsais vystyti.

6

Rezultatai ir išvados

Atlikę pavienių žodžių atpažinimo, naudojant dinaminio laiko skalės kraipymo metodą, tyrimą, suformulavę metodo trūkumus ir pasiūlę sprendimus jiems pašalinti, gautus darbo rezultatus apibendriname:

1. Sukurtas automatinis žodžio ribų nustatymo metodas, pasižymintis stabilumu bei atsparumu signalo kokybės kitimui. Eksperimentų metu gautas 6,8 % klaidų lygis buvo maždaug ketvirtadaliu didesnis nei energijos slenksčio metodo. Tačiau atpažinimo klaidų lygis naudojant pasiūlytąjį metodą buvo iki 3,5 % mažesnis nei energijos slenksčio atveju. Todėl teigiame, kad svarbiausia žodžio ribų nustatymo savybė – stabilumas. O stabilus ir triukšmams atsparus žodžių ribų nustatymas leidžia sumažinti atpažinimo klaidų lygį.

2. Etalonams kurti pasiūlytas klasterizavimo principas, minimizuojantis vidutinį atstumą iki klasterių centrų (atstumas skaičiuojamas naudojant dinaminį laiko skalės kraipymą). Sistemos, apmokytos 3 etalonais kiekvienam žodyno žodžiui (naudojant klasterizacijos principą), atpažinimo tikslumas buvo vidutiniškai 2,5 % didesnis už sistemos su 5 kiekvieno žodžio etalonais (etalonus kuriant tiesiogiai, be jokios atrankos procedūros). Taigi papildomų atrankos procedūrų naudojimas mokyme 2–3 % procentais padidina sistemos atpažinimo tikslumą su mažesniu etalonų skaičiumi.

3. Įdiegus pasiūlytuosius žodžio ribų nustatymo ir etalonų kūrimo metodus nuo kalbėtojo priklausomo pavienių žodžių atpažinimo tikslumo absoliutus padidėjimas buvo 10–19 %, nepriklausomo nuo kalbėtojo atpažinimo –

10–11 %. Taigi kalbos atpažinimo sistemos tikslumas gali būti padidintas ne modifikuojant atpažinimą metodą, o optimizuojant atskirus atpažinimo etapus.

4. Pasiūlyta žodžio garsų ribų nustatymo metodika, kuria remiantis sukurti du metodai žodžiams segmentuoti – tikėtinumo funkcijos maksimizavimo ir prognozės klaidos minimizavimo. Eksperimentų metu tikėtinumo maksimizavimo metodo segmentavimo tikslumas siekė 62,2 %, prognozės klaidos minimizavimo metodo – 85,5 %. Taigi garsų ribos signale gali būti ieškomos kaip kalbos signalo tiesinės prognozės modelio parametrų pasikeitimo momentai.

5. Suformuluota žodžių atpažinimo garsais idėja. Šiuo atveju atpažinimo procesas vykdomas dviem etapais – žodis segmentuojamas į garsus, pastaruosius bandant atpažinti. Eksperimentų metu visi žodžio garsai teisingai atpažinti 15,3 % žodžių, suklysta vienu garsu – 11,7 % žodžių. Pagrindinis tokio atpažinimo organizavimo privalumas – elementarus dviejų pavyzdžių palyginimas ir smulkus atpažinimo vienetas, leidžiantis sumažinti dideliam žodynui reikalingų etalonų kiekį.

Disertacijos darbo rezultatai parodė, kad tolimesni pavienių žodžių atpažinimo tyrimai turėtų būti nukreipti žodžių atpažinimo garsais idėjai tobulinti – parinkti optimalų atpažinimo vienetą, signalo analizės metodą, automatizuoti segmentavimo procesą, optimizuoti panašumo įvertinimo procedūrą lygiagrečiai taikant lingvistinį apdorojimą.

Literatūros sąrašas

[1] ADAMS, J. W. A new optimal window. *IEEE Transactions on Signal Processing*, 1991, t. 39, nr. 8, p. 1753–1769.

[2] AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 1974, t. 19, nr. 6, p. 716–723.

[3] ATAL, B. S.; HANAUER, S. L. Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acoustical Society of America*, 1971, t. 50, nr. 2, p. 637–655.

[4] BAHL, L. R.; BROWN, P. F.; SOUZA, P. F. D.; MERCER, R. L. Maximum mutual information estimation of hidden Markov model parameters for speech recognition. Iš *IEEE Inernational Conference on Acoustics, Speech, and Signal Processing ICASSP'86*, 1986, t. 11, p. 49–52.

[5] BAHL, L. R.; BROWN, P. F.; SOUZA, P. V. D.; MERCER, R. L.; PICHENY, M. A. A method for the construction of acoustic Markov models for words. *IEEE Transactions on Speech and Audio Processing*, 1993, t. 1, nr. 4, p. 443–452.

[6] BAKER, J. K. The Dragon system - an overview. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1975, t. 23, nr. 1, p. 24–29.

[7] BALBONAS, D.; DAUNYS, G. Fonemų klasifikavimas panaudojant garso ir vaizdo informaciją. *Elektronika ir elektrotechnika*, 2005, t. 61, nr. 5, p. 74–77.

[8] BURG, J. *Maximum entropy spectral analysis.* PhD thesis, Stanford university, 1975.

[9] BURR, D. J. Experiments on neural net recognition of spoken and written text. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1988, t. 36, nr. 7, p. 1162–1168.

[10] COLLA, A.; SCAGLIOLA, C.; SCIARRA, D. A connected speech recognition system using a diphone-based language model. Iš *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP'85*, 1985, t. 10, p. 1229–1232.

[11] COOLEY, J. W.; LEWIS, P. A. W.; WELCH, P. D. Historical notes on the fast Fourier transform. *IEEE Transactions on Audio and Electroacoustics*, 1967, t. 15, nr. 2, p. 76–79.

[12] COOLEY, J. W.; TUKEY, J. W. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 1965, t. 19, nr. 90, p. 297–301.

[13] CRAVERO, M.; PIERACCINI, M.; RAINERI, F. Definition and evaluation of phonetic units for speech recognition by hidden Markov models. Iš *IEEE Inernational Conference on Acoustics, Speech, and Signal Processing ICASSP'86*, 1986, t. 11, p. 2235–2238.

[14] DAVID, E. E.; SELFRIDGE, O. G. Eyes and ears for computers. *Proc. of IRE*, 1962, t. 50, nr. 5, p. 1093–1101.

[15] DAVID, S.; RAMAMURTHI, B. Two-sided filters for frame-based prediction. *IEEE Transactions on Signal Processing*, 1991, t. 39, nr. 4, p. 789–794.

[16] DAVIS, K. H.; BIDDULPH, R.; BALASHEK, S. Automatic recognition of spoken digits. *The Journal of the Acoustical Society of America*, 1952, t. 24, nr. 6, p. 637–642.

[17] DAVIS, S. B.; MERMELSTEIN, P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1980, t. 28, nr. 4, p. 357–366.

[18] DELLER, J. R.; HANSEN, J. H. L.; PROAKIS, J. G. Discrete-time processing of speech signals. IEEE Press, Piscataway, 2000. ISBN 0-7803-5386-2.

[19] DELSARTE, P.; GENIN, Y. V. The split Levinson algorithm. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1986, t. 34, nr. 3, p. 470–478.

[20] DENES, P. The design and operation of the mechanical speech recognizer at University college London. *Journal of the British Institution of Radio Engineers*, 1959, t. 19, nr. 4, p. 219–234.

[21] DENES, P.; MATTHEWS, M. V. Spoken digit recognition using time-frequency pattern matching. *The Journal of the Acoustical Society of America*, 1960, t. 32, nr. 11, p. 1450–1455.

[22] DOMATAS, A.; RUDŽIONIS, A. Raspoznovanije opornich fonetičeskich elementov. *Statističeskije problemi upravlenija*, 1986, t. 72, p. 38–45.

[23] EPHRAIM, Y.; DEMBO, A.; RABINER, L. R. A minimum discrimination information approach for hidden Markov modeling. Iš *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP'87*, 1987, t. 12, p. 25–28.

[24] FAUNDEZ-ZANUY, M.; MCLAUGHLIN, S.; ESPOSITO, A.; HUSSAIN, A.; SCHOENTGEN, J.; KUBIN, G.; KLEIJN, W. B.; MARAGOS, P. Nonlinear speech processing: Overview and applications. *Control and Intelligent Systems*, 2002, t. 30, nr. 1, p. 1–10.

[25] FILIPOVIČ, M. Atskirai pasakytų žodžių atpažinimo, naudojant neuroninius tinklus, tyrimas. Iš *Informacinės technologijos 2003*, 2003, p. 10–20.

[26] FILIPOVIČ, M. *Lithuanian isolated word recognition using hybrid artificial neural networks and hidden Markov models approach.* PhD thesis, Vytautas magnus university and Institute of mathematics and informatics, 2005.

[27] FUJIMURA, O. Syllable as a unit of speech recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1975, t. 23, nr. 1, p. 82–87.

[28] FURUI, S. Speaker-independent isolated word recognition using dynamic features of speech spectrum. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1986, t. 34, nr. 1, p. 52–59.

[29] GIACHIN, E. Spoken language dialogue. Iš R., C.; J., M.; H., U.; A., Z.; V., Z.; G., V.; A., Z.; redaktoriai, *Survey of the state of the art in human language technology*, p. 241–244. Cambridge University press, London, 1996.

[30] GOLD, B. Word-recognition computer program. Technical Report 452, MIT Research Laboratory of Electronics, 1966.

[31] GOLD, B.; LIPPMANN, R. P. A neural network for isolated-word recognition. Iš *IEEE International Conference on Acoustics, Speech, and Signal Proceesing ICASSP'88*, 1988, t. 1, p. 44–47.

[32] GOLD, B.; MORGAN, N. Speech and audio signal processing: Processing and reception of speech and music. John Wiley & sons, inc., New York, 2000. ISBN 0-471-35154-7.

[33] GRAY, A. H.; MARKEL, J. D. Distance measures for speech processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1976, t. 24, nr. 5, p. 380–391.

[34] HA, Y. H.; PEARCE, J. A. A new window and comparison to standard windows. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1989, t. 37, nr. 2, p. 298–301.

[35] HARRIS, F. J. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 1978, t. 66, nr. 1, p. 51–83.

[36] HAYES, M. H. Statistical digital signal processing and modeling. John Wiley & sons, inc., Hoboken, 1996. ISBN 0-471-59431-8.

[37] HEIDEMAN, M. T.; JOHNSON, D. H.; BURRUS, C. S. Gauss ant the history of the fast Fourier transform. *IEEE ASSP Magazine*, 1984, t. 1, nr. 4, p. 14–21.

[38] HERMANSKY, H.; HANSON, B. A.; WAKITA, H. Perceptually based linear predictive analysis of speech. Iš *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP'85*, 1985, t. 10, p. 509–512.

[39] HERMANSKY, H.; MORGAN, N. RASTA processing of speech. *IEEE Transactions on Speech and Audio Processing*, 1994, t. 2, nr. 4, p. 578–589.

[40] HERMANSKY, H.; TSUGA, K.; MAKINO, S.; WAKITA, H. Perceptually based processing in automatic speech recognition. Iš *IEEE Inernational Conference on Acoustics, Speech, and Signal Processing ICASSP'86*, 1986, t. 11, p. 1971–1974.

[41] HERSCHER, M. B.; COX, R. B. Source data entry using voice input. Iš *IEEE international conference on acoustics, speech, and signal processing, ICASSP'76*, 1976, t. 1, p. 190–193.

[42] HORI, C.; FURUI, S. Automatic speech summarization based on word significance and linguistic likelihood. Iš *International Conference on Acoustics, Speech, and Signal Processing ICASSP'00*, 2000, t. 3, p. 1579–1582.

[43] HUGHES, G. W. The recognition of speech by machine. Technical Report 395, MIT Research Laboratory of Electronics, 1961.

[44] HWANG, M. Y.; HUANG, X. Subphonetic modeling with Markov states – senone. Iš *International Conference on Acoustics, Speech, and Signal Processing ICASSP-92*, 1992, t. 1, p. 33–36.

[45] ICHIKAWA, A.; NAKANO, Y.; NAKATA, K. Evaluation of various parameter sets in spoken digits recognition. *IEEE Transactions on Audio and Electroacoustics*, 1973, t. 21, nr. 3, p. 202–209.

[46] ITAKURA, F. Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1975, t. 23, nr. 1, p. 67–72.

[47] JELINEK, F. Continuous speech recognition by statistical methods. *Proceedings of the IEEE*, 1976, t. 64, nr. 4, p. 532–556.

[48] JELINEK, F. A real-time, isolated-word, speech recognition system for dictation transcription. Iš *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP'85*, 1985, t. 10, p. 858–861.

[49] JUANG, B.-H.; RABINER, L. R.; WILPON, J. G. On the use of bandpass liftering in speech recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1987, t. 35, nr. 7, p. 947–954.

[50] JURGUTIS, M.; UNDZĖNAS, V. Raspoznovanije izolirovanich slov na osnove peresečenija nulia nizkich i visokich poriadkov. Iš *Avtomatičeskoje raspoznovanije sluchovich obrazov: Materiali seminara ARSO-15*, 1989, p. 129–130.

[51] KANDRATAVIČIUS, I. *Lietuvių kalbos žodžių atpažinimo realaus laiko kompiuterijoje tyrimas.* PhD thesis, Kauno Technologijos universitetas, 2001.

[52] KAUKĖNAS, J.; NAVICKAS, G.; TELKSNYS, L. Audiovizualinė vartotojo ir programinės įrangos sąsaja. Iš *Informacinės technologijos 2006*, 2006, p. 69–75.

[53] KHAN, S. U.; SHARMA, G.; RAO, P. R. K. Speech recognition using neural networks. Iš *International Conference on Industrial Technology 2000*, 2000, t. 1, p. 432–437.

[54] KIVARAS, R. J. Primenenije skritich cepei Markova dlia raspoznovanija reči. Iš *Techničeskaja kibernetika*, 1987 32, p.

[55] KIVARAS, R. J.; UNDZĖNAS, V. J. Issledovanije metodov raspoznovanija slov nezavisimo ot diktora. Iš *Avtomatičeskoje raspoznovanije sluchovich obrazov: Materiali seminara ARSO-15*, 1989, p. 105–106.

[56] KLATT, D. H. Review of the ARPA speech understanding project. *The Journal of the Acoustical Society of America*, 1977, t. 62, nr. 6, p. 1345–1366.

[57] LAINE, U. K.; KARJALAINEN, M.; ALTOSAAR, T. Warped linear prediction (WLP) in speech and audio processing. Iš *International Conference on Acoustics, Speech, and Signal Processing ICASSP-94*, 1994, t. 3, p. 349–352.

[58] LAURINČIUKAITĖ, S. Atskirai pasakytų lietuvių kalbos žodžių atpažinimas, remiantis paslėptaisiais Markovo modeliais. Iš *Informacinės technologijos 2003*, 2003, p. 21–24.

[59] LEE, K.-F.; HON, H.-W. Large-vocabulary speaker-independent continuous speech recognition using HMM. Iš *International Conference on Acoustics, Speech, and Signal Processing ICASSP-88*, 1988, t. 1, p. 123–126.

[60] LESSER, V. R.; FENNELL, R. D.; ERMAN, L. D.; REDDY, D. R. Organization of the Hearsay II speech understanding system. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1975, t. 23, nr. 1, p. 11–24.

[61] LEVIN, E. Word recognition using hidden control neuron architecture. Iš *International Conference on Acoustics, Speech, and Signal Processing ICASSP-90*, 1990, t. 1, p. 433–436.

[62] LIPEIKA, A. Formantiniai požymiai atpažįstant kalbą. *Informacijos mokslai*, 2005, t. 34, p. 215–219.

[63] LIPEIKA, A.; LIPEIKIENĖ, J.; TELKSNYS, L. Development of isolated word speech recognition system. *Informatica*, 2002, t. 13, nr. 1, p. 37–46.

[64] LIPPMANN, R. P. An introduction to computing with neural nets. *IEEE Acoustics, Speech and Signal Processing Magazine*, 1987, t. 4, nr. 2, p. 417–425.

[65] MAKHOUL, J. Linear prediction: a tutorial review. *Proceedings of the IEEE*, 1975, t. 63, nr. 4, p. 561–580.

[66] MAKHOUL, J. Spectral linear prediction: Properties and applications. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1975, t. 23, nr. 3, p. 283–296.

[67] MAKHOUL, J. New lattice methods for linear prediction. Iš *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'76*, 1976, t. 1, p. 462–465.

[68] MAKINO, S.; KAWABATA, T.; KIDO, K. Recognition of consonant based on the perceptron model. Iš *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP'83*, 1983, t. 8, p. 738–741.

[69] MARKEL, J. D.; GRAY, A. H. Linear prediction of speech. Springer-Verlag, Berlin, 1976. ISBN 3-540-07563-1.

[70] MYERS, C. S.; RABINER, L. R. A level building dynamic time warping algorithm for connected word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1981, t. 29, nr. 2, p. 284–297.

[71] NOLL, A. M. Short-time spectrum and „cepstrum" techniques for vocal-pitch detection. *The Journal of the Acoustical Society of America*, 1964, t. 36, nr. 2, p. 296–302.

[72] NUTTALL, A. Some windows with very good sidelobe behavior. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1981, t. 29, nr. 1, p. 84–91.

[73] OLSON, H. F.; BELAR, H. Phonetic typewriter. *IRE Transactions on Audio*, 1957, t. 5, nr. 4, p. 90–95.

[74] OPPENHEIM, A. V.; SCHAFER, R. W.; STOCKHAM, T. G. Nonlinear filtering of multiplied and convolved signals. *Proceedings of the IEEE*, 1968, t. 56, nr. 8, p. 1264–1291.

[75] PAKERYS, A. Lietuvių bendrinės kalbos fonetika. Enciklopedija, Vilnius, 2003. ISBN 9986-433-32-0.

[76] PALIWAL, K. K. On the performance of the quefrency-weighted cepstral coefficients in vowel recognition. *Speech Communication*, 1982, t. 1, nr. 2, p. 151–154.

[77] PALIWAL, K. K. A study of line spectrum pair frequencies for speech recognition. Iš *International Conference on Acoustics, Speech, and Signal Processing ICASSP-88*, 1988, t. 1, p. 485–488.

[78] PALIWAL, K. K.; RAO, P. V. S. Evaluation of various linear prediction parametric representations in vowel recognition. *Signal Processing*, 1982, t. 4, nr. 4, p. 323–327.

[79] PARZEN, E. Some recent advances in time series modeling. *IEEE Transactions on Automatic Control*, 1974, t. 19, nr. 6, p. 723–730.

[80] PATAŠIUS, J. Ispolzovanije raspoznovanija izolirovanich slov v avtomatizirovanich informacionich sistem. *Statističeskije problemi upravlenija*, 1980, t. 42, p. 47–61.

[81] PATAŠIUS, J. V. Praktičeskaja realizacija sistemi dlia analiza i raspoznovanija reči. Iš *Avtomatičeskoje raspoznovanije sluchovich obrazov: Materiali seminara ARSO-13*, 1984, t. 1 116, p.

[82] PATEL, S. A lower-complexity Viterbi algorithm. Iš *International Conference on Acoustics, Speech, and Signal Processing ICASSP-95*, 1995, t. 1, p. 592–595.

[83] PAUL, D. B. A speaker-stress resistant HMM isolated word recognizer. Iš *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP'87*, 1987, t. 12, p. 713–716.

[84] PAUL, D. B. The Lincoln robust continuous speech recognizer. Iš *International Conference on Aucoustics, Speech, and Signal Processing ICASSP-89*, 1989, t. 1, p. 449–452.

[85] PIERCE, J. R. Whither speech recognition? *The Journal of the Acoustical Society of America*, 1969, t. 46, nr. 4, p. 1049–1051.

[86] RABINER, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989, t. 77, nr. 2, p. 257–286.

[87] RABINER, L. R.; ATAL, B. S.; SAMBUR, M. R. LPC prediction error – analysis of its variation with the position of the analysis frame. *IEEEASSP*, 1977, t. 25, nr. 5, p. 434–442.

[88] RABINER, L. R.; JUANG, B.-H. Fundamentals of speech recognition. Prentice-Hall, New Jersey, 1993. ISBN 0-13-285826-6.

[89] RABINER, L. R.; ROSENBERG, A. E.; LEVINSON, S. E. Considerations in dynamic time warping algorithms for discrete word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1978, t. 26, nr. 5, p. 575–582.

[90] RABINER, L. R.; SCHAFER, R. W. Digital processing of speech signals. Prentice-Hall, New Jersey, 1978. ISBN 0-13-213603-1.

[91] RAŠKINIS, A.; RAŠKINIENĖ, G.; KAZLAUSKIENĖ, A. VDU bendrinės lietuvių šnekos universalus anotuotas garsynas. Iš *Informacinės technologijos 2003*, 2003, p. 28–43.

[92] REDDY, D. R. Computer recognition of connected speech. *The Journal of the Acoustical Society of America*, 1967, t. 42, nr. 2, p. 329–347.

[93] REDDY, D. R.; ERMAN, L. D. A model and a system for machine recognition of speech. *IEEE Transactions on Audio and Electroacoustics*, 1973, t. 21, nr. 3, p. 229–238.

[94] RILEY, M. D.; LJOLJE, A. Automatic generation of detailed pronunciation lexicons. Iš H., L. C.; K., S. F.; K., P. K.; redaktoriai, *Automatic speech and speaker recognition: Advanced topics*, p. 1–17. Kluwer Academic publishers, Boston, 1996.

[95] RISSANEN, J. Modeling by shortest data description. *Automatica*, 1978, t. 14, nr. 5, p. 465–471.

[96] RUDŽIONIS, A.; RUDŽIONIS, V. E. Izoliuotų žodžių atpažinimas vidurkinant fonetiškai segmentuotus kalbinių signalų parametrus. Iš *Informacinės technologijos - 96*, 1996, p. 168–174.

[97] RUDŽIONIS, A.; RUDŽIONIS, V. E. Lithuanian speech database LTDIGITS. Iš *LREC 2002: Third International Conference on Language Resources and Evaluation*, 2002, p. 877–882.

[98] RUDŽIONIS, V. E. *Speech recognition by phonetic units.* PhD thesis, Kaunas University of technology, 1998.

[99] SAKOE, H. Two-level DP-matching – A dynamic programming-based pattern matching algorithm for connected word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1979, t. 27, nr. 6, p. 588–595.

[100] SAKOE, H.; CHIBA, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1978, t. 26, nr. 1, p. 43–49.

[101] SAWAI, H.; WAIBEL, A.; MIYATAKE, M.; SHIKANO, K. Spotting japanese CV-syllables and phonemes using the time-delay neural networks. Iš *International Conference on Acoustics, Speech, and Signal Processing ICASSP-89*, 1989, t. 1, p. 25–28.

[102] SCHAFER, R. W.; RABINER, L. R. Design of digital filter banks for speech analysis. *The Bell System Technical Journal*, 1971, t. 50, nr. 10, p. 3097–3115.

[103] SCHWARTZ, R.; CHOW, Y.; KIMBALL, .; ROUCOS, S.; KRASNER, M.; MAKHOUL, J. Context-dependent modeling for acoustic-phonetic recognition of continuous speech. Iš *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP'85*, 1985, t. 10, p. 1205–1208.

[104] SCHWARTZ, R.; CHOW, Y.; ROUCOS, S.; KRASNER, M.; MAKHOUL, J. Improved hidden Markov modeling of phonemes for continuous speech recognition. Iš *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP'84*, 1984, t. 9, p. 21–24.

[105] SHOLTZ, P. N.; BAKIS, R. Spoken digit recognition using vowel-consonant segmentation. *The Journal of the Acoustical Society of America*, 1962, t. 34, nr. 1, p. 1–5.

[106] SMITH, J. E. K.; KLEM, L. Vowel recognition using a multiple discriminant function. *The Journal of the Acoustical Society of America*, 1961, t. 33, nr. 3 358, p.

[107] STECKNER, M. C.; DROST, D. J. Fast cepstrum analysis using Hartley transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1989, t. 37, nr. 8, p. 1300–1302.

[108] STEIGLITZ, K.; DICKINSON, B. Computation of the complex cepstrum by factorization of the z-transform. Iš *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'77*, 1977, t. 2, p. 723–726.

[109] STERN, R. M. Robust speech recognition. Iš R., C.; J., M.; H., U.; A., Z.; V., Z.; G., V.; A., Z.; redaktoriai, *Survey of the state of the art in human language technology*, p. 17–23. Cambridge University press, London, 1996.

[110] TIERNEY, J. A study of LPC analysis of speech in additive noise. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1980, t. 28, nr. 4, p. 389–397.

[111] TOHKURA, Y. A weighted cepstral distance measure for speech recognition. Iš *IEEE Inernational Conference on Acoustics, Speech, and Signal Processing ICASSP'86*, 1986, t. 11, p. 761–764.

[112] VINCIUK, T. K. Raspoznovanije slov ustnoij reči metodami dinamičeskovo programirovanija. *Kibernetika*, 1968, t. 1, p. 81–88.

[113] VITERBI, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 1967, t. 13, nr. 2, p. 260–269.

[114] WACHTER, M. D.; MATTON, M.; DEMUYNCK, K.; WAMBACQ, P.; COOLS, R.; COMPERNOLLE, D. V. Template-based continuous speech recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 2007, t. 4, nr. 15, p. 1377–1390.

[115] WAIBEL, A.; HANAZAWA, T.; HINTON, G.; SHIKANO, K.; LANG, K. J. Phoneme recognition using time-delay neural networks. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1989, t. 37, nr. 3, p. 328–339.

[116] WEINTRAUB, M.; MURVEIT, H.; COHEN, M.; PRICE, P.; BERNSTEIN, J.; BALDWIN, G.; BELL, D. Linguistic constraints in hidden Markov model based speech recognition. Iš *International Conference on Acoustics, Speech, and Signal Processing ICASSP-89*, 1989, t. 2, p. 699–702.

[117] WHITE, G. M.; NEELY, R. B. Speech recognition experiments with linear predication, bandpass filtering, and dynamic programming. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1976, t. 24, nr. 2, p. 183–188.

[118] WOLF, J. J.; WOODS, A. The HWIM speech understanding system. Iš *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP'77*, 1977, t. 2, p. 784–787.

[119] WONG, K. F.; LEUNG, S. H.; NG, H. C. An eigendecomposition based two sided linear prediction model for robust speech recognition. Iš *International Symposium on Speech, Image Processing and Neural Networks ISSIPNN '94*, 1994, t. 1, p. 249–252.

[120] WOODLAND, P.; EVERMANN, G. HTK history., 2002 [Žiūrėta: 2006.10.03].

[121] ZWICKER, E.; FLOTTORP, G.; STEVENS, S. S. Critical band width in loudness summation. *The Journal of the Acoustical Society of America*, 1957, t. 29, nr. 5, p. 548–557.

Autoriaus publikacijų disertacijos tema sąrašas

[1A] LIPEIKA, A.; TAMULEVIČIUS, G. Segmentation of nonstationary signals. Iš *Biomedical Engineering*, 2004, p. 37–40. ISBN 9955-09-739-6.

[2A] LIPEIKA, A.; TAMULEVIČIUS, G. Segmentation of words into phones. *Electronics and Electrical Engineering*, 2006, t. 1, nr. 65, p. 11–15. ISSN 1392-1215.

[3A] TAMULEVIČIUS, G. Interneto naršyklės valdymas balsu. Iš *Informacinės technologijos 2007*, 2007, p. 67–70.

[4A] TAMULEVIČIUS, G.; LIPEIKA, A. Žodžių atpažinimo sistemos kūrimas. Iš *Lietuvos matematikos rinkinys*, 2003, t. 43, p. 292–296. ISSN 0132-2818.

[5A] TAMULEVIČIUS, G.; LIPEIKA, A. Dynamic time warping based speech recognition system. Iš *Human Language Technologies. The Baltic Perspective*, 2004, p. 156–161.

[6A] TAMULEVIČIUS, G.; LIPEIKA, A. Žodžio pradžios ir galo nustatymas atpažįstant atskirai sakomus žodžius. *Elektronika ir elektrotechnika*, 2005, t. 2, nr. 58, p. 61–64. ISSN 1392–1215.

Priedas

Žodynas

1. būti
2. kuris
3. galėti
4. visas
5. kaip
6. Lietuva
7. kitas
8. turėti
9. savas
10. darbas
11. žmogus
12. metai
13. labai
14. vienas
15. nebūti
16. reikėti
17. žinoti
18. didelis
19. tačiau
20. teisė
21. laikas
22. diena
23. dabar
24. pagal
25. valstybė
26. jeigu
27. respublika
28. nustatyti
29. dalis
30. įstatymas
31. straipsnis
32. įmonė
33. žodis
34. norėti
35. kalba
36. šalis
37. sudaryti
38. asmuo
39. naujas
40. sistema
41. sakyti
42. todėl
43. kartas
44. gauti
45. aukštas
46. žemė
47. metas
48. vieta
49. niekas
50. įvairus
51. lietuviai
52. svarbus

53. vaikas
54. gerai
55. prieš
56. tarp
57. dažnai
58. skirti
59. veikla
60. eiti
61. atlikti
62. pasakyti
63. gyventi
64. priimti
65. valstybinis
66. mokslas
67. akis
68. geras
69. atvejis
70. dirbti
71. antras
72. mažas
73. miestas
74. ranka
75. bendras
76. įstaiga
77. mokykla
78. teismas
79. kalbėti
80. forma
81. bankas
82. tada
83. kultūra
84. sąlyga
85. viskas
86. tyrimas
87. vanduo
88. matyti
89. grupė
90. priemonė
91. vyriausybė
92. būdas
93. naudoti
94. medžiaga
95. nors
96. procesas
97. pasaulis
98. ūkis
99. kiek
100. rašyti
101. nulis
102. du
103. trys
104. keturi
105. penki
106. šeši
107. septyni
108. aštuoni
109. devyni
110. pradžia
111. pabaiga