VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS INSTITUTAS

Sigita LAURINČIUKAITĖ

LIETUVIŲ ŠNEKOS ATPAŽINIMO AKUSTINIS MODELIAVIMAS

Daktaro disertacija

Technologijos mokslai, Informatikos inžinerija (07T)

Vilnius, 2008

Disertacija rengta 2003–2008 metais Matematikos ir informatikos institute.

Darbo mokslinis vadovas

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

http://leidykla.vgtu.lt
VGTU leidyklos TECHNIKA 1488-M mokslo literatūros knyga

ISBN 978- 9955-28-271-6

Sigita Laurinčiukaitė

LIETUVIŲ ŠNEKOS ATPAŽINIMO AKUSTINIS MODELIAVIMAS
Daktaro disertacija
Technologijos mokslai, Informatikos inžinerija (07T)

Sigita Laurinčiukaitė

ACOUSTIC MODELLING OF LITHUANIAN SPEECH RECOGNITION
Doctoral Dissertation
Technological Sciences, Informatics Engineering (07T)

2008 05 15. 6 sp. l. Tiražas 20 egz.
Vilniaus Gedimino technikos universiteto leidykla „Technika“, Saulėtekio al. 11,
LT-10223 Vilnius, *http://leidykla.vgtu.lt*
Spausdino UAB „Baltijos kopija“, Kareivių g. 13B, 09109 Vilnius, *http://www.kopija.lt*

Reziumė

Darbas „Lietuvių šnekos atpažinimo akustinis modeliavimas“ yra skirtas lietuvių šnekos atpažinimo akustiniam modeliavimui. Darbe buvo tirtas žodžiais, skiemenimis, kontekstiniais skiemenimis, fonemomis ir kontekstinėmis fonemomis grįstas šnekos atpažinimas. Tyrimai atlikti izoliuotiems žodžiams ir ištisinei šnekai. Iki šiol lietuvių šnekos atpažinime populiariausi kalbos vienetai buvo fonema ir kontekstinė fonema, o kitų kalbos vienetų analizė nebuvo atliekama. Šiame darbe siekiama palyginti lingvistinio tipo kalbos vienetų gebėjimą modeliuoti šneką ir parodyti, kad kalbos vienetų analizė siūlo alternatyvius fonemai ir kontekstinei fonemai kalbos vienetus.

Darbe pasiūlyta metodika mišriam skiemenų ir fonemų akustiniam modeliavimui, naujas kalbos vienetas – pseudo-skiemuo; technologijos atskirų kalbos vienetų akustiniam modeliavimui (schemos, įrankiai, rekomendacijos). Eksperimentiniams tyrimams atlikti paruoštas izoliuotų žodžių garsynas ir sukurtos dvi ištisinės šnekos garsyno LRN versijos.

Ištyrus izoliuotų žodžių atpažinimą, akustinius modelius konstruojant žodžiams, nustatyta, kad modelių mokymo aibės dydis, akustinių modelių mokymo aibės turinys daro įtaką šnekos atpažinimo tikslumui. Pateikiamos rekomendacijos akustiniam modeliavimui žodžių pagrindu.

Ištyrus izoliuotų žodžių atpažinimą, akustinius modelius konstruojant žodžiams, skiemenims ir fonemoms, gauti rezultatai 98 ±1,8 % tikslumu siejami su skiemens tipo kalbos vienetais. Dėl skiemenų akustinio modeliavimo sudėtingumo jį rekomenduojama atlikti žodžiams.

Atlikus ištisinės šnekos atpažinimo tyrimus fonemų ir kontekstinių fonemų kalbos vienetų pagrindu išrinktos dvi fonemų aibės, kurios pasiekia didžiausią atpažinimo tikslumą (62 ±1,5 % ir 62 ±1,5 %). Modeliuojant ištisinės šnekos atpažinimą rekomenduojama naudoti fonemų aibę be (arba su) minkštumo žymių (-ėmis), su kirčio žymėmis ir išskaidžius dvigarsius į atskiras komponentes. Renkantis tarp paprastos fonemos ir kontekstinės fonemos dėl atpažinimo tikslumo siūloma rinktis kontekstinę fonemą, o jei svarbiau modeliavimo paprastumas – paprastą fonemą.

Tiriant skiemens tipo kalbos vienetus pagal pasiūlytą metodiką naujas kalbos vieneto tipas – pseudo-skiemuo padidina šnekos atpažinimo tikslumą (57 ±0,3 %) lyginant su fonemų akustiniais modeliais (52 ±0,3 %). Analizuojant metodikos etapus ištisinės šnekos atpažinimo tikslumą pavyko padidinti iki 67 ±1,4 %. Sukurti kontekstinių skiemenų akustiniai modeliai šnekos atpažinimo tikslumą padidina iki 72 ±1,4 %. Lyginant su kontekstinėmis fonemomis (76 ±1,3 %) kontekstinių skiemenų modeliavimo atpažinimo tikslumas mažesnis.

Abstract

This paper is devoted to an acoustic modelling of Lithuanian speech recognition. Word-, syllable-, contextual syllable-, phoneme- and contextual phoneme-based speech recognition was investigated. Investigations were performed for isolated words and continuous speech. The most popular sub-word units in Lithuanian speech recognition are phonemes and contextual phonemes, and research on other sub-word units is omitted. This paper aims to compare capacity of linguistic sub-word units to model speech and to demonstrate that investigation of sub-word units suggest using alternative sub-word units to phoneme and contextual phoneme.

The dissertation proposes a new methodology for acoustic modelling of syllables and phonemes, new sub-word unit – pseudo-syllable; technologies for acoustic modelling of separate sub-word units, including developed schemes, tools and recommendations. Speech corpus of isolated words was prepared and two versions of corpus of continuous speech LRN were developed for experimental research.

Investigation of recognition of isolated words and construction of acoustic models for words showed that a size of training set of acoustic models, a content of training set in regard to number of speakers have an influence on speech recognition accuracy. The recommendations for word-based acoustic modelling are given.

Investigation of recognition of isolated words and construction of acoustic models for words, syllables and phonemes showed that the best recognition results 98 ±1,8 % are achieved with sub-word unit of syllable. The complexity of syllable-based acoustic modelling prescribes sub-word unit type of word to use for acoustical modelling.

After investigation of phoneme-based and contextual phoneme-based recognition of continuous speech two sets of phonemes with the best speech recognition accuracy (62 ±1,5 % and 62 ±1,5 %)) were selected. Set of phonemes without (or with) softness of consonants, accent and splitting of diphthongs are recommended for acoustic modelling of phoneme- and contextual phoneme-based recognition of continuous speech. Contextual phoneme with regard to speech recognition accuracy or phoneme with regard to simplicity of acoustic modelling is recommended.

Investigation of recognition of continuous speech according to proposed methodology showed that new sub-word unit (pseudo-syllable) increase speech recognition accuracy (57 ±0,3 %) in comparison to phoneme models (52 ±0,3 %). Investigation of separate blocks in methodology allowed to increase speech recognition accuracy to 67 ±1,4 %. Contextual syllables-phonemes increase speech recognition accuracy to 72 ±1,4 %, but are inferior to contextual phonemes (76 ±1,3 %).

Žymėjimai

Sąvokos

Akustinis modelis – paslėptasis Markovo modelis, per mokymo procedūrą įsimenantis konkrečių šnekos signalų charakteristikas ir taip tampantis konkretaus šnekos signalo reprezentantu.
Atpažinimo sistemos ištekliai – akustiniai ir kalbos modeliai, kuriuos paruošus galima naudoti keliose ASA sistemose.
Automatinis šnekos atpažinimas (*automatic speech recognition*) – 1) žmogaus šnekos pavertimas tekstu naudojant kompiuterį; 2) atpažinimo sistemos darbo procesas, atliekant teksto atitikmens akustiniams signalams parinkimą.
Fenonas (*front end based phone*) – vienos būsenos diskretus paslėptasis Markovo modelis, skirtas trumpo akustinio intervalo arba konkretaus taško, išreiškiamo vektorių klasterizacijos eigoje gautų vektorių prototipu, kepstrinėje srityje reprezentacijai.
Fonema (*phoneme, phone*) – minimalus kalbos garsas, leidžiantis atskirti vieną ištartą žodį nuo kito.
Garsynas (*speech corpus*) – duomenų (garso įrašų ir jų anotacijų) rinkinys, skirtas šnekos atpažinimo modeliavimui.
Kalbos vienetas (*sub-word unit*) – pagrįstu būdu (priklausančiu nuo kalbos vieneto tipo) gauta vieno ar kelių rašytinių simbolių arba garsų kombinacija. Kalbos vienetas išreiškia tekstinį simbolį, o pagal jį sukurtas akustinis modelis – tekstinio simbolio reprezentaciją garsu.
Kontekstinė fonema (*contextual phone, triphone*) – fonema, kuriai (iš kairės ir/arba dešinės pusės) daro įtaką koartikuliacija.
Melų dažnių skalės kepstriniai požymiai (*Mel-frequency Cepstral Feature, MFCC*) – požymiai, gauti šnekos signalui pritaikius Furjė transformaciją, gautą spektrą praleidus pro Melo dažnių filtrų bloką, logaritmavus ir pritaikius kosinuso transformaciją.
Multonas (*multone*) – tiesinė fenonų kombinacija su multono koeficientais.
Paslėptasis Markovo modelis (*hidden Markov model, HMM*) – baigtinė būsenų, susietų perėjimo tikimybėmis, seka, naudojama signalų laikiniam ir spektriniam kitimui modeliuoti.
Pseudo-skiemuo – skiemuo, kuris naudojamas kito skiemens skaidymui į dalis.
Senonas (*state-dependent phone*) – vienos būsenos diskretus paslėptasis Markovo modelis, panašus į fenoną.
Skiemuo (*syllable*) – kalbos srauto atkarpa, kurios garsai sudaro minimalų artikuliacinį, akustinį ir funkcinį vienetą.
Šnekos tipai – šnekos klasifikacija pagal tarimo pobūdį. Išskiriami izoliuoti žodžiai (*isolated words*), rišlios frazės (*connected words*), ištisinė šneka (*continuous speech*), spontaniška šneka (*spontaneous speech*).

Žodynas (*vocabulary, lexicon*) – žodžių su transkripcijomis sąrašas, žodynas apibrėžia ASA sistemos atpažįstamus žodžius.
Transkripcija – nuosekli kalbos vienetų seka, sudaroma tik iš apibrėžtos kalbos vienetų aibės elementų.

Simboliai

Δc_i	pirmos eilės skirtumas tarp požymio vektoriaus reikšmių;
$\Delta\left(\Delta c_i\right)$	antros eilės skirtumas tarp požymio vektoriaus reikšmių;
$\mathbf{A}=\left\{a_{ij}\right\}$	perėjimo iš vienos būsenos į kitą tikimybių matrica PMM;
$\alpha_j\left(t\right)$	tiesioginė tikimybė;
$\mathbf{B}=\left\{b_j\left(\mathbf{o}_t\right)\right\}$	stebėjimų tikimybės tankio funkcija PMM būsenoje;
$\beta_j\left(t\right)$	atbulinė tikimybė;
c_i	požymių vektoriaus reikšmės;
c_{jl}	*l*-tojo mišinio svoris *j*-oje būsenoje;
e	šnekos segmento energijos reikšmė;
$L\left(\mathbf{O}\mid M\right)$	tikėtinumo išraiška PMM parametrų įvertinimo procedūroje;
$\lambda=\left(\mathbf{A},\mathbf{B},\pi\right)$	PMM apibūdinantis parametrų rinkinys;
M ir M^*	vienas AM ir AM seka (aibė);
M^{**}	visos galimos modelių sekos;
N	PMM būsenų skaičius;
$\mathbf{O}=\mathbf{o}_1,\mathbf{o}_2,...,\mathbf{o}_T$	požymių vektorių (stebėjimų) seka;
$P\left(\mathbf{O}\mid W\right)$	akustinio modeliavimo tikimybė;
$P\left(W\right)$	kalbos modeliavimo tikimybė;
π_i	pradinio buvimo PMM būsenoje tikimybė;
$Q=\left(q_1,q_2,...,q_N\right)$	PMM būsenų seka;
T	šnekos signalo ilgis;
W	vienas žodis;
$W^*=W_1,W_2,...,W_n$	žodžių seka (aibė);
W^{**}	visos galimos žodžių sekos;
$X=x_1,x_2,...,x_T$	šnekos signalo segmentų (kadrų) seka.

Santrumpos

AM	akustinis modelis;
AM_10_40	vieno kalbėtojo 50-ies izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 10 mokymo vienetų ir skirti testuoti 40 testinių vienetų;
AM_20_30	vieno kalbėtojo 50-ies izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 20 mokymo vienetų ir skirti testuoti 30 testinių vienetų;
AM_25_25	vieno kalbėtojo 50-ies izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 25 mokymo vienetus ir skirti testuoti 25 testinius vienetus;
AM_30_20	vieno kalbėtojo 50-ies izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 30 mokymo vienetų ir skirti testuoti 20 testinių vienetų;
AM_40_10	vieno kalbėtojo 50-ies izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 40 mokymo vienetų ir skirti testuoti 10 testinių vienetų;
AM_20_20	25 kalbėtojų 50-ies izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 20 x 25 mokymo vienetų ir skirti testuoti 20 testinių vienetų;
AM_1k	vieno kalbėtojo 99-ių izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 23 mokymo sesijas ir skirti testuoti to paties kalbėtojo 10 testavimo sesijų;
AM_1k+	vieno kalbėtojo 99-ių izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 23 mokymo sesijas. Kai kuriems akustiniams modeliams mokymo imtys padidintos specialiai. Akustiniai modeliai skirti testuoti to paties kalbėtojo 10 testavimo sesijų;
AM_4k	4 kalbėtojų 99-ių izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 65 mokymo sesijas ir skirti testuoti naujo kalbėtojo 10 testavimo sesijų;
AM_4k_ad	4 kalbėtojų 99-ių izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 65 mokymo sesijas, atliekant adaptaciją naujam kalbėtojui (tam naudojant 10 naujo kalbėtojo garso įrašų sesijų) ir skirti testuoti adaptuoto kalbėtojo 10 testavimo sesijų;
AM_5k	5 kalbėtojų 99-ių izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 88 mokymo sesijas ir skirti testuoti vieno iš tų kalbėtojų 10 testavimo sesijų;

AM_MKD	akustinių modelių aibė ištisinėje šnekoje esančių fonemų su priebalsių minkštumo ir su kirčių žymėmis modeliavimui (mišrieji dvigarsiai neskaidomi į atskiras komponentes);
AM_KD	akustinių modelių aibė ištisinėje šnekoje esančių fonemų be priebalsių minkštumo ir su kirčių žymėmis modeliavimui (mišrieji dvigarsiai neskaidomi į atskiras komponentes);
AM_D	akustinių modelių aibė ištisinėje šnekoje esančių fonemų be priebalsių minkštumo ir be kirčių žymių modeliavimui (mišrieji dvigarsiai neskaidomi į atskiras komponentes);
AM_MD	akustinių modelių aibė ištisinėje šnekoje esančių fonemų su priebalsių minkštumo ir be kirčių žymių modeliavimui;
AM_MK	akustinių modelių aibė ištisinėje šnekoje esančių fonemų su priebalsių minkštumo ir su kirčių žymėmis, bei mišriuosius dvigarsius išskaidžius į atskiras komponentes modeliavimui;
ASA	automatinis šnekos atpažinimas;
DTW	dinaminis laiko skalės kraipymas;
H_1	skiemenų ir fonemų aibė iš 293 elementų (227 skiemenys, 63 fonemos ir dvibalsiai, 3 pašalinių garsų žymenys), suformuota pagal elementų dažnį (>50) žodyne;
H_2	skiemenų ir fonemų aibė iš 289 elementų, išrenkant 223 skiemenis iš sąrašo, išrikiuoto pagal dažnį mokymo duomenyse. Skirtumas nuo H_1 – 45 skiemenys;
H_1_1P	skiemenų ir fonemų aibė iš 293 elementų , suformuota iš aibės H_1;
H_1_K	skiemenų ir fonemų aibė iš 235 elementų (227 skiemenys, 5 fonemų ir dvibalsių grupės, 3 pašalinių garsų žymenys). Skiemenys išlieka tapatūs bazinės aibės H_1 skiemenims, fonemos ir dvibalsiai suskirstomi į grupes;
H_1_M	skiemenų ir fonemų aibė iš 283 elementų (233 skiemenys, 47 fonemos ir dvibalsiai, 3 pašalinių garsų žymenys). Aibėje panaikintas priebalsių minkštumo žymuo;
H_1dPM	skiemenų ir fonemų aibė, kurioje skiemenys imami iš bazinės aibės H_1, o fonemų modeliai formuojami atskiru tyrimu;
H_1dTM	skiemenų ir fonemų aibė kurioje skiemenys imami iš bazinės aibės H_1, fonemų modeliai formuojami atskiru tyrimu, prijungiamos kontekstinės fonemos;
H_1P, H_2P	žodynai, suformuoti pagal skiemenų ir fonemų bazines aibes H_1 ir H_2, nebazinius skiemenis keičiant fonemomis;

H_1SP H_2SP	žodynai, suformuoti pagal skiemenų ir fonemų bazines aibes H_1 ir H_2, nebazinius skiemenis keičiant skiemenų ir fonemų seka;
HTK	įrankių paketas, skirtas akustiniam modeliavimui;
IS_PNK_F	modeliuojamos šnekos atpažinimo sistemos pavadinimas šifruojamas: ištisinė šneka, priklausoma nuo kalbėtojo, fonemos;
IS_PNK_S	modeliuojamos šnekos atpažinimo sistemos pavadinimas šifruojamas: ištisinė šneka, priklausoma nuo kalbėtojo, skiemenys;
IZ_NNK	modeliuojamos šnekos atpažinimo sistemos pavadinimas šifruojamas: izoliuoti žodžiai, nepriklausoma nuo kalbėtojo;
IZ_NNK_F	modeliuojamos šnekos atpažinimo sistemos pavadinimas šifruojamas: izoliuoti žodžiai, nepriklausoma nuo kalbėtojo, fonemos;
IZ_NNK_S	modeliuojamos šnekos atpažinimo sistemos pavadinimas šifruojamas: izoliuoti žodžiai, nepriklausoma nuo kalbėtojo, skiemenys;
IZ_NNK_Z	modeliuojamos šnekos atpažinimo sistemos pavadinimas šifruojamas: izoliuoti žodžiai, nepriklausoma nuo kalbėtojo, žodžiai;
IZ_NNK_ZSF	modeliuojamos šnekos atpažinimo sistemos pavadinimas šifruojamas: izoliuoti žodžiai, nepriklausoma nuo kalbėtojo, žodžiai-skiemenys-fonemos;
IZ_PNK	modeliuojamos šnekos atpažinimo sistemos pavadinimas šifruojamas: izoliuoti žodžiai, priklausoma nuo kalbėtojo;
IZ_RS	modeliuojamos šnekos atpažinimo sistemos pavadinimas šifruojamas: izoliuoti žodžiai, realios sąlygos;
LRN0, LRN0.1	ištisinės šnekos Lietuvos radijo naujienų garsynas ir jo versijos;
MFCC	Melų dažnių skalės kepstriniai koeficientai;
PA	skiemenų ir fonemų kalbos vienetų pagrindinė aibė, formuojama atrinkimo principu pagal siūlomą metodiką;
PMM	Paslėptieji Markovo modeliai;
SAMPA-LT	šnekos atpažinimui skirtas fonemų žymėjimo sistemos standartas;
SNR	signalo/triukšmo santykis;
T_AM_MKD	kontekstinių fonemų modelių aibė, suformuota pagal AM_MKD;
T_AM_KD	kontekstinių fonemų modelių aibė be priebalsių minkštumo žymių, bet su kirčiais, suformuota pagal AM_MKD;

T_AM_MK	kontekstinių fonemų modelių aibė, suformuota pagal AM_MK;
T_AM_K	kontekstinių fonemų modelių aibė su kirčiais, suformuota pagal AM_MK;
TIMIT	šnekos garsynas anglų kalbai;
WSJCAM0	šnekos garsynas anglų kalbai;
ZAT	žodžio atpažinimo teisingumas;
ZK	žodžio klaida;
ZT	žodžio tikslumas.

Turinys

1

Įvadas

1.1. Tiriamoji problema

Perėjimas nuo šnekos prie akustinių modelių kaip šnekos atpažinimo objektų grindžiamas tarpiniu elementu – kalbos vienetu. Vienas iš svarbiausių akustinio modeliavimo klausimų yra kalbos vienetų tipo ir konkretaus tipo aibės sudėtinių elementų parinkimas. Esant didelei kalbos vienetų įvairovei, tyrėjai, atlikdami tyrimus, pasirenka kurį nors vieną kalbos vienetų tipą ir jam kuria akustinius modelius. Lietuvių šnekos atpažinimo tyrimuose trūksta skirtingų kalbos vienetų tipų akustinio modeliavimo lygiagrečių tyrimų naudojant tą patį garsyną ir panašias modeliavimo schemas. Neaišku, ar atliekant detalesnį kalbos vienetų modeliavimą galima padidinti šnekos atpažinimo tikslumą. Todėl darbe nagrinėjamas įvairių kalbos vienetų akustinių modelių sudarymas, atliekami lyginamieji tyrimai.

1.2. Darbo aktualumas

Automatinis šnekos atpažinimas (ASA) yra sudėtinė balsinių technologijų, apimančių šnekos sintezę, asmens tapatybės vertinimą pagal jo balsą, dalis. Šių technologijų plėtra yra globali dėl žmogaus poreikio technologinius pasiekimus integruoti į kasdienį gyvenimą palengvinant buitį ir dėl unikalios balsinių technologijų savybės per šneką užtikrinti natūralų žmogaus santykį su mašina. Populiariausios ir labiausiai komercializuotos balsinių technologijų taikymo sritys yra diktavimas ir valdymas balsu, informacijos teikimas. Paprasčiausiu ASA sistemos pavyzdžiu galima laikyti diktavimo ar valdymo balsu produktus. Diktavimo balsu produktai paprastai įtraukia sistemos adaptavimo operaciją –

sistemą adaptavus naujam balsui, ji būna parengta darbui. Valdymo balsu produktai apima kompiuterio ir buitinių prietaisų valdymą. Sudėtingesnės sistemos, kuriose naudojamas visas balsinių technologijų paketas, naudojamos tiekti telekomunikacijos paslaugoms: informacijai apie orus, eismą, sporto varžybų rezultatus ir pan. Kita atpažinimo sistemų dalis patenka kaip papildomas elementas į bendro naudojimo produktus: automobilius, telefonus, žaidimus. Realiai veikiančius produktus naudoja tautos, kurių kalbos yra plačiai paplitusios ir kurioms komerciškai naudinga kurti minėtus produktus. Balsinių technologijų vystymas Lietuvai yra svarbus norint išlikti visaverte informacinės visuomenės nare.

Automatinis šnekos atpažinimas yra žmogaus šnekos pavertimas tekstu naudojant kompiuterį. Tikslesnį apibrėžimą pateikia Jurafsky (Jurafsky 2000): „Automatinis šnekos atpažinimas – kompiuterio programinės įrangos (sistemos) darbo procesas, atliekantis žodžių atitikmenų akustiniams signalams parinkimą“. ASA technologijų pagrindinis tikslas – sukurti mašinas, kurios galėtų *girdėti, suprasti, kalbėti* ir *veikti* pagal gautą informaciją. Komunikavimo objektas – šneka. Automatinis šnekos atpažinimas yra pirmoji grandis balsinių technologijų produktuose. Ji užtikrina pirminės originalios informacijos, kuri vėliau gali būti koreguojama, gavimą iš šnekos signalo. Pagrindinis reikalavimas šiai grandžiai – informacijos tikslumas ir patikimumas, pasiekiamas atskirų ASA sistemos dalių veikimo optimizavimu. Pagrindinėmis ASA sistemos dalimis laikoma:

- požymių išskyrimas iš šnekos signalo;
- garsinę informaciją reprezentuojančių modelių (pvz. akustinių modelių) formavimas;
- nežinomo ištarimo klasifikavimas vienam iš reprezentacinių modelių.

Šiame darbe nagrinėjamas reprezentacinių (akustinių) modelių formavimas. Kaip minėta, tai pirmoji atpažįstant šneką grandis, turinti pateikti žmogaus šneką atitinkantį tekstą. Vėliau kitomis technologijomis atpažintą tekstą galima tikslinti. Tuo būdu akustinių modelių formavimas, lemiantis šnekos atpažinimo tikslumą ir vėlesnių technologijų darbą, yra esminis atpažįstant šneką. Akustinių modelių aibės sudarymas atliekamas kiekvienai kalbai atskirai, nes yra susijęs su konkrečios kalbos garsų aibe ir specifika. Lietuvių šnekai yra atlikta įvairių akustinio modeliavimo tyrimų, apimančių daugiausiai fonemų, kontekstinių fonemų akustinį modeliavimą, sukonstruotos prototipinės šnekos atpažinimo sistemos, tačiau trūksta alternatyvių skiemenų, žodžių ir lyginamojo akustinio modeliavimo darbų.

1.3. Tyrimų objektas

Darbe modeliuojamos automatinio šnekos atpažinimo sistemos naudoja statistinį metodą, susijusį su paslėptaisiais Markovo modeliais (PMM). Mokymo

režime šiuose modeliuose užkoduojamos konkrečių šnekos signalų charakteristikos ir modelis tampa tą šnekos signalų aibę reprezentuojančiu modeliu, dar vadinamu akustiniu modeliu (AM). Kalbos vienetų tipo (fonemų, skiemenų, žodžių, kontekstinių fonemų, kontekstinių skiemenų) parinkimas AM kūrimui pagal mokymo duomenis, mokymas ir AM darbingumo tikrinimas yra šio darbo tyrimo objektai.

1.4. Darbo tikslas

Darbo tikslas yra pagal skirtingiems kalbos vienetams sukurtus akustinius modelius atlikti lyginamuosius šnekos atpažinimo tyrimus, kurie leistų pasiūlyti įvairių kalbos vienetų akustinio modeliavimo technologijas ir įvertinti kalbos vienetų akustinių modelių efektyvumą ir panaudojimo galimybes.

1.5. Darbo uždaviniai

Remiantis darbo tikslu suformuluoti šie uždaviniai:

1. Sudaryti akustinio modeliavimo schemas atsižvelgiant į kalbos vieneto (žodžių, fonemų, skiemenų, kontekstinių fonemų ir kontekstinių skiemenų) ir šnekos tipą (izoliuoti žodžiai, ištisinė šneka). Šias schemas naudoti akustinio modeliavimo tyrimams.
2. Paruošti garsynus, reikalingus eksperimentiniams tyrimams, ir trūkstamas technologijas ir įrankius sukurtųjų schemų blokų realizavimui.
3. Eksperimentais ištirti akustinių modelių kūrimą lingvistinio kriterijaus būdu gaunamiems kalbos vienetams bei sukurtųjų akustinių modelių efektyvumą ir pritaikomumą įvairių šnekos tipų atpažinimui.
4. Atlikus tyrimus pateikti įvairių kalbos vienetų akustinio modeliavimo technologijas akustinių modelių kūrimui automatinio šnekos atpažinimo sistemoms.

1.6. Tyrimų metodai

Šiame darbe naudoti skaitmeninio signalų apdorojimo, paslėptųjų Markovo modelių teorijos, matematinės statistikos, lietuvių kalbos gramatikos ir fonetikos metodai ir sąvokos. Disertacijos rezultatai gauti naudojant įrankių paketą HTK (HTK 2003), darbą palengvinančius autorės sukurtus automatinius įrankius ir darbo autorės sukurtus ar paruoštus izoliuotų žodžių ir ištisinės kalbos LRN0 (Laurinčiukaitė *et al.* 2006) garsynus.

1.7. Mokslinis darbo naujumas

Gauti moksliniai rezultatai yra šie:

- Pagal lyginamųjų šnekos atpažinimo tyrimų rezultatus, gautus naudojant žodžių, skiemenų, fonemų, kontekstinių skiemenų ir kontekstinių fonemų akustinius modelius, pateikiamos kalbos vienetų akustinio modeliavimo technologijos.
- Sukurta mišrios skiemenų ir fonemų akustinių modelių aibės kūrimo metodika, pagal ją atliekant eksperimentinius testavimus.
- Pasiūlytas naujas kalbos vienetas – pseudo-skiemuo, gerinantis šnekos atpažinimo tikslumą, lyginant su lingvistiškai apibrėžiamais vienetais.
- Sukurti akustiniai modeliai (ištekliai), naudojami konkrečioje automatinio šnekos atpažinimo sistemoje ir galintys pagerinti šnekos atpažinimo rezultatus.

1.8. Darbo rezultatų praktinė reikšmė

Balso technologijos, kurių dalis yra ASA sistemos, naudojamos siekiant švietimo (mokymo procesų tobulinimas), specialių priemonių neįgaliesiems kūrimo, buities ir darbo palengvinimo (mašinų valdymas balsu, telekomunikacijos) tikslų. Šnekos atpažinimo tyrimai lengvina šių technologijų tobulinimą.

Šio darbo rezultatai buvo taikomi vykdant „Lietuvių kalbos informacinėje visuomenėje 2000–2006 m.“ programą.

Atliktų tyrimų rezultatai kaip rekomendacijos gali būti taikomi kuriant konkrečias lietuvių kalbos automatinio šnekos atpažinimo sistemas kalbos vienetų parinkimo ir akustinių modelių sudarymo klausimais.

Sukurti kalbos vienetų akustiniai modeliai gali būti taikomi, kaip automatinio šnekos atpažinimo sistemos ištekliai.

1.9. Ginamieji disertacijos teiginiai

1. Skiemenų ir fonemų akustinių modelių aibės sudarymo metodika, leidžianti sistemingai ištirti įvairių kalbos vienetų akustinių modelių aibes.
2. Autorės sukurtos technologijos kalbos vienetų akustiniam modeliavimui, šnekos atpažinimo schemų blokų realizavimui, kalbos vienetų ir žodynų apdorojimo automatizavimui, leidžiančios praktiškai realizuoti mokymą ir šnekos atpažinimą.

3. Sukurti žodžių, fonemų, skiemenų, kontekstinių fonemų ir kontekstinių skiemenų akustiniai modeliai, tinkantys naudojimui įvairiose šnekos atpažinimo sistemose.
4. Ištisinės lietuvių šnekos garsyno LRN versijos LRN0 ir LRN0.1, leidžiančios atlikti visapusiškus šnekos atpažinimo tyrimus.

1.10. Darbo rezultatų aprobavimas

Darbo rezultatai buvo paskelbti šešiose mokslinėse konferencijose Lietuvoje bei užsienyje. Disertacijos tematika yra išspausdinti 6 moksliniai straipsniai: du žurnaluose, įtrauktuose į Mokslinės informacijos instituto duomenų bazės pagrindinį sąrašą: (Thomson ISI Master Journal List) (Laurinčiukaitė, Lipeika 2007), (Thomson Scientific (ISI)) (Laurinčiukaitė, Lipeika 2006); vienas – tarptautinėse duomenų bazėse (INSPEC) referuotame žurnale (Laurinčiukaitė *et al.* 2006); trys – tarptautinių konferencijų leidiniuose (Šilingas *et al.* 2006, Šilingas *et al.* 2004, Laurinčiukaitė 2004).

Tarpiniai disertacijos rezultatai pristatyti šiuose konferencijų pranešimuose:

- 2003 m. Kaunas „Informacinės technologijos 2003“;
- 2004 m. Šiauliai „Žmogaus ir kompiuterio sąveika“;
- 2004 m. Vilnius „Lietuvos matematikų draugijos XLIV konferencija“;
- 2004 m. Ryga (Latvija) „Baltic Human and Language Technologies 2004“;
- 2006 m., 2007 m. Vilnius „Electronics and electrical engineering“; tarptautinės konferencijos *Elektronika 2006* sekcijoje *Sistemų inžinerija, kompiuterių technologija* gautas IEEE diplomas ir antroji premija geriausio jaunojo mokslininko pranešimo konkurse;
- seminaruose Matematikos ir informatikos institute (MII), Vilniaus pedagoginiame universitete (VPU).

1.11. Disertacijos struktūra

Disertaciją sudaro įvadas, 5 skyriai, literatūros sąrašas (102 nuorodos) ir 3 priedai. Pagrindinė darbo dalis 108 puslapiai.

Įvade pateikiamas tyrimo objektas, temos aktualumas, darbo tikslas ir uždaviniai, tyrimo metodai, mokslinis naujumas, darbo rezultatų praktinė reikšmė, darbo aprobavimas konferencijose, darbo struktūra ir turinys.

Antrame skyriuje pateikiama kalbos vienetų tipų, pagal kuriuos konstruojami akustiniai modeliai, apžvalga, apimanti šiuo metu pasaulyje naudojamus pagal automatinį ir lingvistinį kriterijus gautus kalbos vienetus. Išsamiau kaip vieni iš

darbo objektų aprašomi pagal lingvistinį kriterijų gautieji kalbos vienetai, nurodant jų reprezentatyvumą, lankstumą ir sudėtingumą. Apžvelgiami Lietuvoje atlikti šnekos atpažinimo tyrimai ir darbai, akcentuojant akustinio modeliavimo sritį ir formuluojant būsimą šio darbo problemą.

Trečiame skyriuje suformuluojamas darbo tikslas, pateikiamos naudojamos žinios: aprašomi statistiniai šnekos atpažinimo metodai, paslėptasis Markovo modelis. Toliau pateikiama bendra šnekos atpažinimo sistemos schema, šios schemos detalizavimas skirtingiems šnekos tipams ir kalbos vienetams, naujai pasiūlytos fonemų ir skiemenų akustinių modelių aibės formavimo metodikos aprašymas.

Ketvirtame skyriuje pateikiami tyrimams naudojamų garsynų aprašai, jų kūrimo specifika.

Penktame skyriuje pateikiami atlikti tyrimai, rezultatai ir išvados.

Išvadose pateikti apibendrinti disertacijos rezultatai.

Trijuose prieduose pateikiami LRN garsyno fonetinės sistemos vienetų sąrašas, dviejų tyrimuose nagrinėtų skiemenų ir fonemų aibių H_1, H_2 sąrašai, paslėptojo Markovo modelio mokymo ir atpažinimo procedūrų aprašai.

1.12. Padėka

Šio darbo autorė nuoširdžiai dėkoja moksliniam vadovui doc. dr. A. L. Lipeikai už konsultacijas, vadovavimą ir pagalbą rengiant disertaciją.

Ypatinga padėka reiškiama kolegoms: M. Filipovič, M. Skripkauskui, D. Šilingui, T. Lygutui – už nuolatinę pagalbą, konsultacijas, pastabas, kantrybę ir palaikymą rengiant darbą; lietuvių kalbos specialistei L. Skulskytei už darbe esančių klaidų ištaisymą ir vertingas pastabas.

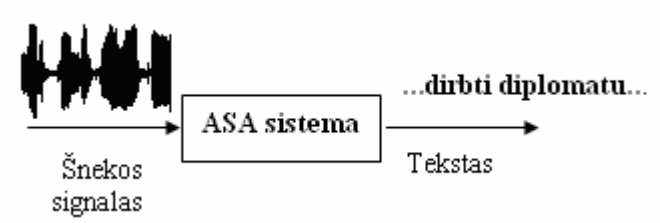
Taip pat dėkoju Matematikos ir informatikos instituto Atpažinimo procesų skyriaus vadovui prof. habil. dr. L. Telksniui ir kolegoms už palaikymą ir pagalbą.

2

Akustinio modeliavimo problematika

Pagrindinis automatinio šnekos atpažinimo (ASA) sistemų kūrimo ir tobulinimo tikslas yra sukurti mašinas, kurios galėtų *girdėti, suprasti, kalbėti* ir *veikti* pagal balsu gautą informaciją. Įprastinės darbo su kompiuteriu priemonės, kaip klaviatūra ir pelė, būtų papildytos dar vienu komunikavimo būdu – šneka. Tokių mašinų sukūrimas yra daugiapakopis procesas, kuriame pirmoji pakopa yra šnekos atpažinimas, t. y. kompiuteriu žmogaus šnekos pavertimas tekstu. Tikslesnį apibrėžimą pateikia Jurafsky (2000), kur ASA apibrėžiamas, kaip atpažinimo sistemos darbo procesas atliekant žodžių atitikmenų akustiniams signalams parinkimą (*system for mapping acoustic signals to a string of words*). Šnekos atpažinimas galimas tik įvedus tris prielaidas-principus (Levinson, Roe 1990). Pirmasis – informacija, esanti šnekos signale tiksliai atvaizduojama trumpalaikio spektro amplitude. Ši prielaida leidžia pasikliauti požymiais, gautais iš spektro, ir panaudotais sprendimo priėmimo operacijoje. Antras principas – šnekos signalo turinys, o taip pat ir prasmė gali būti išreikšti rašytine forma simbolių iš baigtinės abėcėlės eilute. Ši prielaida leidžia tikėtis, kad ištarimo turinys ir prasmė, atlikus akustinio signalo pakeitimą fonetinių simbolių seka, nenukentės. Trečias principas – šnekoje garsų suvokimas, gramatinės ir semantinės struktūros yra persipynusios. Ši prielaida šalia akustinės informacijos leidžia naudoti papildomą informaciją apie kalbos struktūras, kai vien akustinė informacija būna dviprasmiškuma.

Abstraktus šnekos atpažinimo procesas kompiuteriu vaizduojamas 2.1 paveiksle. Šnekos atpažinimo procesas prasideda šnekos signalo įvedimu į ASA sistemą ir baigiasi tekstinio atitikmens akustiniam pateikimu.



2.1 pav. *Šnekos atpažinimo procesas*

Iš tikrųjų šnekos atpažinimo procesas nėra paprastas. ASA sistemoje atliekamos gauto akustinio signalo apdorojimo, požymių išskyrimo, atpažinimo objektų alternatyvų paieškos ir sprendimo priėmimo operacijos. Šioje schemoje nėra pavaizduotas paruošiamasis etapas, kurio metu formuojama minėta atpažintinų objektų (reprezentacinių modelių arba akustinių modelių) aibė. Tuo požiūriu automatinio šnekos atpažinimo procesą galima laikyti tolygiu klasifikavimui. Abiejų procesų pradžioje analizuojant duomenų imtį suformuojami pradinių duomenų reprezentacijos modeliai, kurie naudojami klasifikavimo ar atpažinimo sprendimui priimti. Reprezentacinis modelis šnekos atpažinime vadinamas akustiniu modeliu (AM). Struktūros ir turinio požiūriu AM priklauso nuo ASA **metodų ir iš signalo išskiriamų požymių**. Šiame darbe AM pagrindas yra paslėptasis Markovo modelis, kurio pagrindiniai elementai yra būsenų tolydiniai skirstiniai, aprašomi vidurkių ir dispersijų vektoriais, ir perėjimų tarp būsenų matrica. Požymių vektoriai – Melų dažnių skalės kepstriniai požymiai. AM kaip visavertės reprezentacinių modelių aibės turinys priklauso nuo **konkrečios kalbos struktūros, garsinės bazės** ir jos elementų specifinių savybių, darančių įtaką šnekos atpažinimo rezultatams. Akustinio modeliavimo tikslu tampa šių dviejų minėtų tyrimo aspektų suderinimas siekiant suformuoti AM aibę, darančią teigiamą įtaką šnekos atpažinimo tikslumui. Jei sukūrus AM jie būtų stabilūs ir tinkami visiems šnekos atpažinimo atvejams, atpažinimo problema būtų triviali. Atsiranda dar vienas, trečias faktorius, veikiantis AM. Tai didžiulis **šnekos signalo kintamumas dėl kalbėtojų įvairovės, kalbėjimo greičių, kontekstų skirtumų, akustinių terpės sąlygų**. Šie trys pagrindiniai faktoriai leidžia akustinį modeliavimą atlikti įvairiais aspektais. Šiame darbe atpažinimo metodas ir požymių tipas yra fiksuojami. Pagrindiniu darbo objektu tampa akustinių modelių priklausomybės nuo lietuvių kalbos struktūros, jos elementų specifinių savybių tyrimas, siekiant surasti pageidautinas priklausomybes šnekos atpažinimo tikslumo požiūriu. Tam tikra dalimi tyrimas susijęs ir su signalo kintamumo dėl kalbėtojų įvairovės faktoriumi. Tai įgalina padaryti tyrimo metu naudoti garsynai.

Akustinių modelių priklausomybė nuo kalbos struktūros, jos elementų specifinių savybių pasireiškia akustinių modelių aibės sudarymo specifiškumu.

Perėjimas nuo šnekos prie akustinių modelių kaip šnekos atpažinimo objektų grindžiamas tarpiniu elementu – kalbos vienetu, išskiriamu iš rašytinės kalbos. Kalbos vienetui sukuriamas jį reprezentuojantis akustinis modelis. Tuo būdu kalbos vienetas išreiškia tekstinį simbolį, o pagal jį sukurtas akustinis modelis – tekstinio simbolio reprezentaciją garsu.

Čia išryškėja pagrindinis klausimas – kaip parinkti kalbos vienetus, sudaryti jų aibę, tuo pačiu metu didinant šnekos atpažinimo tikslumą? Pradėti reikia nuo tikslesnio *kalbos vieneto sąvokos* apibrėžimo, išskiriant galimus kalbos vienetų tipus, paskui nustatyti, kuriuos kalbos vienetų tipus verta nagrinėti toliau, ir juos aprašyti.

2.1. Akustinio modeliavimo vienetai

Šiuolaikinėse šnekos atpažinimo sistemose akustiniai modeliai yra kuriami kalbos vienetams mažesniems nei žodis. Kalbos vienetais gali būti tiek lingvistiškai apibrėžiami vienetai, kaip fonema, skiemuo, tiek įvairios fonemų kombinacijos, gaunamos taikant automatinius darbo su akustiniais duomenimis metodus. Visiems šiems kalbos vienetams, esantiems sudėtinėmis žodžio dalimis, apibūdinti anglų kalboje vartojamas terminas *sub-word*. Lietuviškas atitikmuo šiame darbe yra *kalbos vienetai*. Taigi, kalbos vienetu bus vadinama pagrįstu būdu (priklausančiu nuo kalbos vieneto tipo) gauta vieno ar kelių rašytinių simbolių arba garsų kombinacija.[1]

ASA sistemoje operuojama minėtais kalbos vienetais, kurių skaičius yra baigtinis ir kažkuriuo būdu apibrėžiamas prieš ASA sistemos modeliavimą. Žodžio atpažinimas yra pagrįstas atitinkamos sekos kalbos vienetams sukurtų AM atpažinimu. Žodžiai ir juos atitinkančios kalbos vienetų sekos laikomos ASA sistemos žodyne. ASA sistemos žodyno žodžio transkripcija – kalbos vienetų nuoseklus junginys. Ši kombinacija gali būti sudaryta tiek iš skiemenų, tiek iš fonemų ar fenonų. Žodžio transkripcija taip pat gaunama pagrįstu būdu, priklausančiu nuo kalbos vieneto tipo.

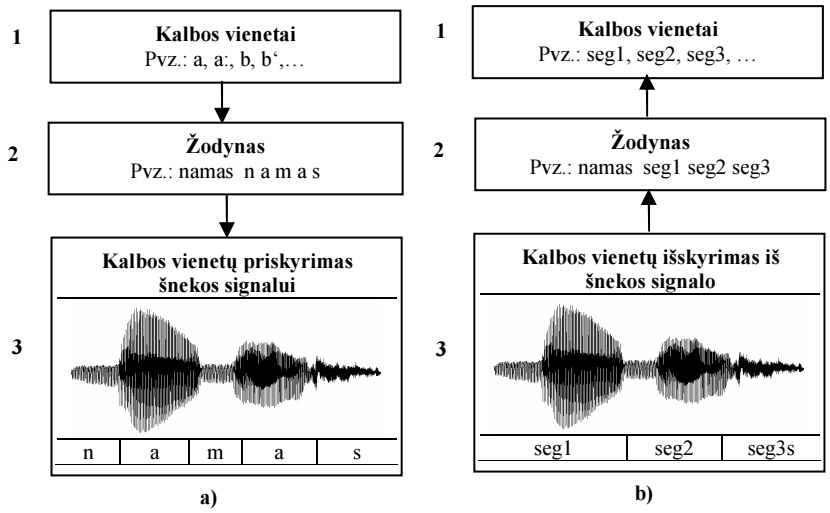
Vienas iš svarbiausių akustinio modeliavimo klausimų yra minėtų kalbos vienetų tipo ir konkretaus tipo aibės sudėtinių elementų parinkimas. **Galimi du kalbos vienetų tipo parinkimo būdai** (Černocký 2002): kalbos vienetų išankstinis (apriorinis) apibrėžimas **remiantis lingvistiniu kriterijumi** ir **automatinis atrinkimas** naudojant mokymo duomenų klasterizacijos metodus. Būtina apžvelgti šiuos būdus, jais gaunamus kalbos vienetų tipus; išvardinti parinkimo būdų trūkumus bei privalumus.

[1] Šiame darbe laikomasi lingvistinės kalbos vieneto kilmės požiūrio.

2.1.1. Automatinio ir lingvistinio kalbos vienetų tipų naudojimo argumentai

Kalbos vienetų aibės parinkimas remiantis lingvistiniu (dar vadinamu klasikiniu) kriterijumi reiškia, kad kalbos vienetai ar jų gavimas yra apibrėžiamas iš anksto, aprioriai. Šis darbas yra atliekamas konkrečios kalbos specialistų, pateikiančių galutinius kalbos vienetų sąrašus ar taisykles jiems gauti. Šnekos atpažinimo tyrėjai dažnai tyrimus atlieka naudodami garsynus, kuriuose jau naudojamas vienas ar kitas kalbos vienetų sąrašas, ir nesusimąstydami naudoja paruoštus kalbos vienetų sąrašus. Iš tikrųjų nėra aišku, ar kalbos specialistų sudaryti kalbos vienetų sąrašai tinka šnekos atpažinimui. Šiuo atveju šnekos tyrėjas turi mažesnę galimybę kontroliuoti vieną iš šnekos atpažinimo faktorių – kalbos vienetų aibės sudarymą.

Automatinis kalbos vienetų parinkimo būdas remiasi algoritmu, kuriuo kalbos vienetai „ištraukiami“ iš šnekos signalo. Čia sudaroma galimybė taikyti kuo įvairesnius algoritmus. Skirtingas lingvistinių ir automatinių kalbos vienetų tipų susiejimas su šnekos signalu yra vaizduojamas 2.2 paveiksle.



2.2 pav. *a) Lingvistinių ir b) automatinių kalbos vienetų tipų susiejimo su šnekos signalu būdai*

Siejant lingvistiniu būdu gautus kalbos vienetus su šnekos signalu didžiausia problema yra tiksli šnekos signalo segmentacija pagal pateiktą kalbos vienetų seką, vaizduojamą 2.2 a) paveikslo trečiame bloke. Automatiniu būdu gautų kalbos vienetų problema atsiranda tada, kai jiems bandoma surasti atitikmenis lingvistinių kalbos vienetų aibėje, t. y. 2.2 b) paveikslo pirmajame bloke.

Šnekos tyrėjo pasirinkimas naudoti vieno ar kito tipo kalbos vienetus grindžiamas pasitikėjimu. Pagrindinė kritika, išsakoma lingvistinio tipo kalbos vienetus naudojantiems tyrėjams susijusi su populiariausiu kalbos vienetu – fonema. Abejojama ar fonemos yra pagrindiniai ir tinkamiausi vienetai aprašyti akustinį signalą. Visuotinai manoma (McLennan *et al.* 2003, Černocký 1998), kad tarp leksinės ir akustinės šnekos reprezentacijos dar yra tarpinis pereinamasis lygis, bet nėra įrodyta, kad šiam lygiui priklauso fonemos. Daugiau ar mažiau foneminės ir kitos reprezentacijos yra susiję su lingvistine, tekstine reprezentacija. Ši sąsaja formavosi šimtmečius ir negalima manyti, kad akustinės reprezentacijos atveju galima pasielgti analogiškai. Į šią kritiką galima atsakyti tuo pačiu, t. y. net ir taikant automatiniu būdu gautus kalbos vienetus, galutinis žingsnis turi būti atpažinto teksto transformacija į žmogui suprantamą tekstą, kuris remiasi lingvistiniais kalbos vienetais. Tiesioginis lingvistinių kalbos vienetų taikymas sumažina darbų apimtį. Šiuo metu vis labiau linkstama prie hipotezės, kad žmogaus supratimas (*perception*) priklauso nuo netiesiškų sąsajų tarp įvairių lygių informacijos, kaip: skiemenys, žodžiai, sakinio struktūra, pokalbio tema (Greenberg 1996). Taip šnekos atpažinime pradeda dominuoti žodžių ir skiemenų kalbos vienetai.

Tokioms balso technologijoms, kaip kodavimas ir kalbėtojo verifikavimas, apriorinis kalbos vienetų parinkimo būdas nėra naudingas (Černocký 1998). Net suvokiant šnekos foneminės reprezentacijos naudingumą šnekos atpažinime, kitas užduotis sprendžiant ši nauda nėra akivaizdi. Kodavime simbolinis aprašas tėra tarpinis žingsnis tarp įėjimo ir sintezuoto signalų, todėl patogiau simbolinį aprašą sieti su signalu nei su tekstu. Kalbėtojo verifikavime signalas skaidomas į klases pagal diskriminacines savybes, todėl klasių konstravimą patogiau grįsti verifikavimo efektyvumu. Taip ryškėja kalbos vienetų taikymo galimybės, priklausančios nuo užduoties.

Pagrindinis automatiškai parenkamų kalbos vienetų būdo trūkumas yra tas, kad taip suformuotos aibės nėra nekintančios, o pritaikytos konkrečiai duomenų aibei. Kiekvieną kartą konstruojant naują ASA sistemą reikia iš naujo ieškoti kalbos vienetų aibės. Kita problema atsiranda siekiant susieti garsinės kilmės kalbos vienetus su baigtine tekstinių simbolių aibe. Kadangi kalbos vienetų yra labai daug, reikalingi papildomi algoritmai. Kad automatiškai parenkamų kalbos vienetų uždavinys yra sudėtingas, rodo tai, kad lietuvių šnekai kol kas dar nėra sukurta jais grįsta šnekos atpažinimo sistema.

Galima būtų daryti išvadą, kad lingvistinio tipo kalbos vienetų taikymas šnekos atpažinime reikalauja mažiau darbo ir laiko sąnaudų, nes galutiniai sąrašai ir taisyklės buvo pradėti sudarinėti anksčiau, nei atsirado kompiuteris, ir buvo gauti ilgomis kalbos tyrėjų pastangomis. Tuo tarpu automatinio tipo kalbos vienetai palyginti su lingvistiniais tiriami trumpesnį laiką ir dar nėra išanalizuoti konkrečioms kalboms.

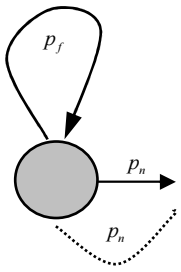
Čia pateikta kritika nesuteikia pirmenybės kažkuriam vienam kalbos vienetų tipui.

2.1.2. Automatinis kalbos vienetų parinkimo būdas

Automatinio kalbos vienetų parinkimo būdai remiasi duomenimis (*data driven*) ir stengiasi naudoti kuo mažiau apriorinių žinių. Šie metodai yra nepriklausomi nuo kalbos, t. y. tik akustiniai duomenys lemia kalbos vienetų aibės turinį. Tai leidžia sukurtus algoritmus naudoti įvairioms kalboms.

Automatiniu būdu gautų kalbos vienetų tipų pavyzdžiais gali būti fenonai (Bahl *et al.* 1993a), senonai (Hwang, Huang, 1992), multonai (Bahl *et al.* 1993, 1996). Nors šie kalbos vienetai gauti tarsi-automatiniu būdu, pradiniame etape jie remiasi lingvistine informacija (kaip foneminės žodžių transkripcijos), kuri naudojama PMM modelių, tarnaujančių naujų kalbos vienetų konstravimui, įverčių formavimui. Nauji kalbos vienetai, o kartu ir modeliai yra PMM sudėtinės dalys – būsenos, kurioms taikomos įvairios transformacijos.

Fenonas (*front end based phone*) apibrėžiamas kaip vienos būsenos diskretus PMM, skirtas trumpo akustinio intervalo arba konkretaus taško, išreiškiamo vektorių klasterizacijos eigoje gautų vektorių prototipu, kepstrinėje srityje reprezentacijai (Bahl *et al.* 1993a). Jokia lingvistinė informacija nėra naudojama. Fenono modelio topologija vaizduojama 2.3 paveiksle. Fenono parametrus sudaro viena būsena ir trys perėjimo tikimybės: p_s – pasilikimo būsenoje, p_f – perėjimo į kitą būseną ir p_n – peršokimo per fenoną (nulinis perėjimas). Pirmosios dvi tikimybės išreiškiamos vektoriais *s* ir *f*.



2.3 pav. *Fenono struktūra*

Žodžio modelis – nuosekliai sujungtų fenonų seka. Pagrindinė žodžio transkripcija fenonais (*fenonic baseform*) sudaroma darant prielaidą, kad fenonų

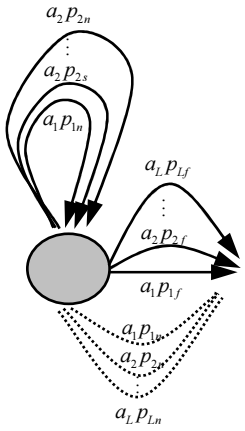
modeliai jau yra sukurti, ir ieškant to žodžio mokymo duomenų imtyje *geriausios* fenonų sekos:

$$\hat{u} = \arg\max_{\{u\}} p\left(W^1, W^2, ..., W^n \mid u\right) = \arg\max_{\{u\}} \prod_{i=1}^{n} p\left(W^i \mid u\right), \tag{2.1}$$

čia $W^1, W^2, ..., W^n$ – n to paties žodžio tarimo variantų, $\{u\}$ – visos įmanomos fenonų sekos, $\hat{u}$ – atrinktoji fenonų žodžio transkripcija. Tariant žodžių sekos $W^1, W^2, ..., W^n$ tarpusavio nepriklausomumą, jungtinė tikimybė gali būti pakeista sandauga, kaip pavaizduota (2.1) formulėje.

Analogiškas vienetas fenonui yra senonas (*state-dependent phone*) – tai taip pat vienos būsenos PMM, nors nebūtinai diskretus. Šis kalbos vienetas atsirado siekiant gauti tikslesnius kontekstinių fonemų modelius, klasterizuojant atskiras jų būsenas ir išrenkant klasterį reprezentuojantį modelį. Išsamus senonų aprašymas ir palyginimas su fenonais jau seniai išanalizuotas (Hwng 1993).

Dar vienas iš fenonų išvestinis vienetas yra multonas (*multone*). Šis vienetas įvestas siekiant leisti nežymiai kisti žodžio tarimui. Multoną galima įsivaizduoti kaip tiesinę fenonų kombinaciją su multono koeficientais a_i, $i = 1, 2, ..., L$. Multono struktūra vaizduojama 2.4 paveiksle.



2.4 pav. *Multono struktūra*

Multono parametrai įvertinami keturiais žingsniais: fenono inicializavimas, multono modelių kūrimas, multono koeficientų įvertinimas nekeičiant fenono parametrų, kai kurių multono modelių perėjimo tikimybių atsisakymas.

Kiti metodai iš akustikos gautiems kalbos vienetams neturi specifinių pavadinimų. Galima paminėti tokius tyrimus, kaip: segmentinis modeliavimas (Fukada *et al.* 1996, Ostendorf *et a*l. 1996), polinominių

trajektorijų modeliavimas (Bacchiani *et al.* 1996). Šie metodai atlieka akustinių duomenų segmentavimą, remdamiesi didžiausiu tikėtinumu. Segmento tikėtinumas išreiškiamas vieno Gausinio modelio visiems segmentams skaičiavimu. Tuo būdu segmento vidurkis nėra konstanta, bet slenka tam tikra trajektorija. Ši trajektorija yra išreiškiama kiekvieno segmento polinominių koeficientų ir laiko matricomis. Atlikus segmentavimą, taikoma klasterizavimo ir iteracinio įvertinimo operacijos. Daugiau apie segmentinius modelius galima rasti (Černocký 1998, Ostendorf *et al.* 1996).

Kalbos vienetai gaunami kaip rezultatas tiriant automatinį šnekos segmentavimą (Hu *et al.* 1996). Šis mechanizmas veikia tokiu principu: atliekamas automatinis šnekos segmentavimas ir gauti rezultatai palyginami su eksperto nustatytais segmentais. Vietose, kur segmentavimo algoritmas dažniausiai klysta, atliekamas segmento ribos panaikinimas sujungiant gretimus garsus ir suformuojant ištisinį neskaidomą kalbos vienetą. Tokiais kalbos vienetais paprastai tampa tokių garsų, kaip: balsis-balsis, sklandusis priebalsis-balsis junginiai.

Deligne ir Bimbot (1997) šaltinis pateikia dar vieną kalbos vienetų formavimo būdą, susijusį su tam tikru būdu gauto pereinamojo atitikmens nuo akustinių duomenų prie teksto kvantavimu. Pereinamąjį atitikmenį sudaro laikinės dekompozicijos (*Temporal Decomposition*, TD) modeliu (Atal 1983) suformuotos stebėjimo vektorių sekos. TD modeliuoja spektro kitimą, šnekos segmentą nusakydamas kaip vektorių iš tam tikros aibės tiesinę kombinaciją. Sukvantavus šiuos vektorius ergodiniu PMM ir vektorius pakeitus kvantavimo klasę nusakančiu simboliu, gaunama pastarųjų eilutė, nusakanti akustinį signalą. Bet kuri n, $n \le n_{\max}$ ilgio šių simbolių kombinacija, pasikartojanti daugiau kartų, nei yra užfiksuotas slenkstis, skelbiama kalbos vienetu ir jam sukuriamas AM.

Kalbos vienetais galima laikyti ir fonetinius požymius, vėliau į juos integruojant aukštesnio lygio prozodinę sritį – skiemenį, kaip daro Kirchoff (Kirchoff 1996). Ji fonetinius požymius suskirsto į 6 aibes pagal artikuliacines savybes: artikuliacijos vietą, eilę, būdą ir t. t. Kiekvieną iš šių aibių sudaro elementai, dar vadinami požymiais-reikšmėmis, pvz.: artikuliacijos vietą – lūpiniai, dantiniai, liežuvio priešakiniai, viduriniai ir užpakaliniai požymiai-reikšmės. ASA sistema sudaryta iš trijų modulių: požymių atpažinimo, sinchronizavimo ir šnekos atpažinimo. Požymių atpažinimo etape akustinis signalas analizuojamas visų 6 požymių grupių atžvilgiu lygiagrečiai gaunant 6 požymių-reikšmių sekas. Sinchronizavimo etape požymių-reikšmių sekos yra papildomos informacija apie skiemenų ribas ir taip vienodiems skiemenims sinchronizuojamos požymių sekos. Šio etapo pabaigoje gaunami skiemenų šablonai. Atpažinimo metu vertinant, kiek testinis pavyzdys artimas etalonui, naudojamos visos 6 požymių-reikšmių sekos.

2.1.3. Lingvistiniu kriterijumi paremtas kalbos vienetų parinkimo būdas

Fonetika – mokslas apie garsinę kalbos sandarą, kurią galima nagrinėti keliais aspektais: artikuliaciniu, akustiniu, fonologiniu ir ortoepiniu. Kalbą nagrinėjant artikuliaciniu aspektu, tyrimo objektu tampa kalbos padargų veikla; akustiniu – kalbos padargų sukelti virpesiai; fonologiniu – fonetiniai kalbos elementai; ortoepiniu – tarties normos (Pakerys 2003). Šnekos atpažinimui aktualesni yra trys pirmieji aspektai. Šiame darbe šnekos atpažinimo tikslais šneka nagrinėta akustiniu ir fonologiniu aspektais.

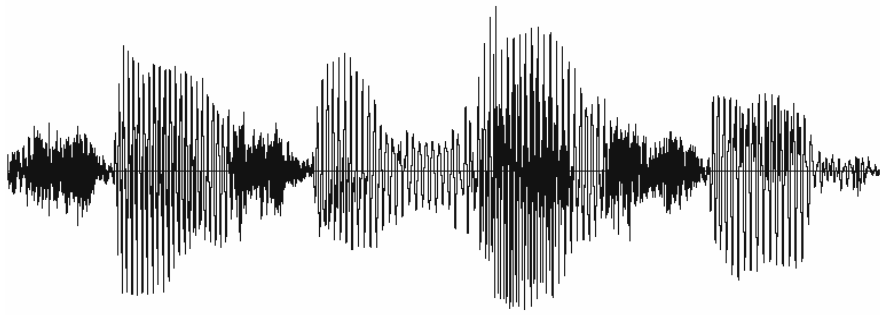
Fonologija nagrinėja fonetinius kalbos elementus, kurie gali būti segmentiniai (fonema, skiemuo) ir prozodiniai (kirtis, priegaidė, intonacija). Akustinė fonetika dažniausiai spektriniais metodais tiria patį garsą. Tad fonologija atsako į klausimą: kiek ir kokių fonetinių vienetų yra kalboje, o akustinė fonetika pateikia šių vienetų apibūdinimą spektriniais metodais. Šie du kalbos analizės būdai buvo panaudoti: 1) kalbos vienetų parinkimo etape – formuojamų kalbos vienetų aibių elementai buvo parenkami atlikus kalbos analizę fonologiniu aspektu, 2) kalbos vienetus atitinkančių akustinių modelių mokymo metu – įgyvendintas akustinės fonetikos tikslas. Kadangi mokymo metu duotų kalbos vienetų spektriniai atitkmenys gaunami automatiniu būdu, toliau aprašoma fonologinė kalbos analizė.

Jau buvo minėta, kad fonetinių vienetų skaičius pagal pasirinktą segmentinį elementą (šiame darbe nagrinėjami tik segmentiniai elementai) lemia akustinių modelių skaičių. Papildomi fonetiniai vienetai ir modeliai gaunami fonetinius vienetus jungiant su prozodiniais kirčio požymiais. Kadangi visos kalbos turi savas fonetinių vienetų sistemas, skirtingų kalbų ASA sistemų naudojamos kalbos vienetų ir tuo pačiu akustinių modelių aibės skiriasi tiek skaičiumi, tiek atskirų jos elementų (kad ir tų pačių skirtingoms kalboms) spektrine išraiška. Pagrindiniai kalbos vienetų tipai, kurie bus nagrinėti šiame darbe yra: fonemų, kontekstinių fonemų, skiemenų, kontekstinių skiemenų ir žodžių. Kaip šiais kalbos vienetų tipais galima išskaidyti šnekos signalą, vaizduoja 2.5 paveikslas.

Pabrėžiant kalbos vienetų parinkimo svarbą, toliau pateikiamas fonetinių vienetų pagal segmentinį elementą apibūdinimas. Šie elementai ir buvo analizuojami darbe. Huang (2001) suformulavo tam tikrus reikalavimus kalbos vienetams. Kuo tiksliau kalbos vienetai atitinka tuos reikalavimus, tuo tikslesni gaunami pagal juos kuriami akustiniai modeliai. Tie reikalavimai yra:

- segmentinis elementas turi būti tikslus, reprezentuojantis akustines savybes, pasitaikančias įvairiose terpėse;
- turi būti pakankamai pavyzdžių, reikalingų suformuoti tikslią jo reprezentaciją – akustinį modelį;
- segmentinis elementas turi būti lankstus ir pritaikomas formuojant nežinomus žodžius.

Toliau segmentiniai kalbos vienetai trumpai apibūdinami ir analizuojami pagal pateiktas charakteristikas.



a)	**š'**	**e:**	**š'**	**a**	**z'**	**d'**	**e**	**š'**	**i**	**m**	**t**	
b)	**še**		**šias**			**de**		**šimt**				
c)	**šešiasdešimt**											

2.5 pav. *Šnekos signalo „šešiasdešimt“ įrašas ir jo skaidiniai a) fonemų, b) skiemenų, c) žodžių tipo segmentiniais vienetais. Kontekstinių fonemų segmentai sutampa su fonemų segmentais, dėl to kontekstinių fonemų tipo skaidinys nevaizduojamas*

Fonema (*phoneme*) – minimalus kalbos garsas, leidžiantis atskirti vieną ištartą žodį nuo kito. Kalbos garsinis vienetas (*phone*) – fonemos akustinė išraiška[2]. Pvz. fonema /k/ žodžiuose *kalnas* ir *niekšas* yra skirtingos akustinės išraiškos – skirtingi jų kalbos garsiniai vienetai (kodėl ir kaip keičiasi fonemų akustinės išraiškos aptariama toliau).

Lietuvių kalboje fonemos skirstomos į 3 grupes: balsius, priebalsius ir dvigarsius. Analogiškai dalinama ir kitose kalbose. Pabrėžtina, kad fonemų sistemos atspindi kalboje egzistuojančius tik balsius ir priebalsius, bet ne dvigarsius. Kalbininkai (Pakerys 2003, Vaitkevičiūtė 2001, Girdenis 2003) savo darbuose pateikia fonemų aibes, kurios iš principo sutampa, o svarstymų atsiranda dėl klasifikacijos ir vieno ar kito garso priklausymo sistemai. Pagal esamas lietuvių kalbos gramatines taisykles (Ambrazas *et el*. 2005) pagrindine fonemų sistema laikoma sistema, suformuota Girdenio, kurioje yra 56 fonemos[3]. Įvairių fonemų sistemų analizė, kreipiant dėmesį, kaip kalbininkai traktuoja vieną ar kitą fonemą, padeda išsiaiškinti, ar fonemų akustinės savybės yra pakankamai skirtingos, kad joms būtų galima kurti atskirus akustinius modelius.

Šnekos atpažinimo tyrimams buvo pasiūlytas VDU kompiuterinės lingvistikos centre sukurtas SAMPA-LT fonemų sistemos

[2] Anglų kalboje šios sąvokos yra aiškiai skiriamos, tuo tarpu lietuvių kalboje dvi skirtingas sąvokas įvardija tas pats žodis *fonema*. Todėl vengdami dviprasmybių dėl naujos sąvokos, toliau naudojamas vienas terminas – fonema.

[3] Ši ir kitos toliau minimos fonemų sistemų lentelės pateikiamos prieduose.

standartas (Raškinis 2003b). Nuo standartinės Girdenio sistemos ši sistema skiriasi viena papildoma fonemų pora, kurią kalbininkai laiko konkrečios fonemos /n/ alofonu (ši sąvoka aiškinama toliau).

Lietuvių kalboje fonemų skaičius yra ribotas ir akustinių modelių taikomumas yra aukštas, tačiau dėl mažos fonemos trukmės akustinės garso savybės gali būti atspindimos ne visiškai tiksliai.

Kontekstinė fonema. Šnekamojoje kalboje grynos fonemos pasitaiko retai, jų junginiai sudaro didesnius segmentus – skiemenis ir žodžius. Sudarant šiuos junginius tarp dviejų fonemų vyksta perėjimas, nusakomas terminu – koartikuliacija. Tai procesas, kurio metu kaimyniniai garsai daro įtaką vienas kito tarimui. Fonemų aibė atspindi idealų-abstraktų garsų rinkinį – etaloną, nusakantį pagrindinius skirtumus tarp garsų. Kalbant apie fonemą kontekste vartojamas terminas kontekstinė fonema (*context dependent phone, triphone*). Šnekos atpažinime dėl koartikuliacijos didesnis vaidmuo tenka kontekstinei fonemai, o ne jos etalonui. Fonetikai vartoja kitą terminą – alofonas, norėdami pabrėžti kontekstinės fonemos ryšį su konkrečios fonemos etalonu. Alofonas apibrėžiamas kaip vienos fonemos galimi tarimo variantai, turintys panašių akustinių ir artikuliacinių savybių (Girdenis 2003). Iš alofonų aibės foneminei abstrakcijai reprezentuoti išrenkamas tas alofonas, kurio tarimo dažnis yra didžiausias. Grubiai skaičiuojant, kiekviena lietuvių kalbos fonema gali turėti $57 \times 58 \times 57$ alofonų (trijų fonemų kombinacija, vidurinei įgyjant bet kokią reikšmę iš 58 fonemų aibės, o dviem iš jų galint įgyti bet kokią reikšmę iš 57 fonemų aibės). Gautasis skaičius nusako akustinių modelių skaičių. Akivaizdu, kad tokio dydžio akustinių modelių aibei reikia didelės mokymo imties.

Pastebėta, kad kai kurios fonemos daro panašią įtaką savo kaimynėms, pvz.: /b/ ir /p/ vienodai modifikuoja po jų einančių balsių tarimą. Suradus panašų kontekstinį poveikį darančias kontekstines fonemas, jų akustinių modelių tam tikriems parametrams priskiriama ta pati spektrinė reprezentacija – parametrai naudojami ir viename, ir kitame modeliuose – siejami (*to tie*). Taip galima išspręsti akustinių modelių skaičiaus ir mokymo duomenų trūkumo problemas. Taigi, pagrindinis tikslas yra susieti akustinių modelių parametrus. Tam reikia surasti siejimo objektus – kontekstines fonemas su panašiu akustiniu kontekstu, t. y. jas sugrupuoti, klasterizuoti. Tiksliau, siekiant didesnio reprezentatyvumo (Huang *et al.* 2001), yra klasterizuojamos ne pačios kontekstinės fonemos, o jų būsenos. Yra galimi du klasterizavimo būdai: grindžiant duomenimis ir sprendimų medžio naudojimas. Šie klasterizavimo būdai skiriasi stiliumi (klasterių dalinimu ar jungimu) ir metrika (Euklidinis atstumas ar logaritminė tikimybė). Duomenimis grindžiamas klasterizavimas iš pradžių daro prielaidą, kad visų kontekstinių fonemų akustinių modelių būsenos yra atskiri klasteriai. Akustinių modelių būsenos yra jungiamos į klasterius pagal tai, kam tyrėjas atiduoda pirmenybę (pvz.: visų kontekstinių fonemų „a“ centrinių

būsenų sujungimas į vieną klasterį). Klasterių jungimas vyksta skaičiuojant Euklidinį atstumą tarp klasterių porų ir išrinkus mažiausią Euklidinį atstumą juo siejama pora yra sujungiama. Jungimas į klasterius vyksta tol, kol klasterių skaičius nukrenta iki tam tikro slenksčio arba didžiausio klasterio dydis pasiekia nustatytą ribą. Euklidinio atstumo skaičiavimo formulės skiriasi ir priklauso nuo PMM.

Sprendimų medis yra binarus medis, klasterizavimą vykdantis atsakymų į binarius klausimus būdu. Iš pradžių visos akustinių modelių būsenos yra sudedamos į vieną klasterį. Nuosekliai užduodant tyrėjo paruoštus klausimus vyksta klasterių dalinimo procesas. Klasteriai dalijami remiantis tikimybiniu matu, tikimybę skaičiuojant pagal Gausinių skirstinių vidurkio, dispersijos ir buvimo būsenose skaičių. Pritaikius kiekvieną klausimą iš klausimyno klasterio dalijimui, skaičiuojama mokymo duomenų tikimybė. Išrenkamas tas klausimas ir dalijamas tas klasteris, su kuriais buvo pasiektas tikimybės padidėjimas. Klausimai turi aprėpti visą galimą konteksto įtakos fonemai erdvę. Parametrai susiejami toms kontekstinių fonemų būsenoms, kurios atsiduria viename galutiniame klasteryje. Sprendimų medis yra svarbus tuo, kad leidžia klasterizuoti, o paskui suformuoti bendrą vaizdą toms kontekstinėms fonemoms, kurių nebuvo mokymo duomenyse; taip pat geresnį vaizdą įgauna tie kontekstinių fonemų modeliai, kurie turėjo mažas mokymo imtis. Akustinių modelių parametrų siejimo būdu mažinamos skaičiavimo įrenginio atminties sąnaudos ir apeinama mokymo imties nepakankamumo problema.

Kad kontekstinės fonemos atitika akustinį reprezentatyvumą, mokymo imties kiekį ir taikymo lankstumą yra identiška fonemoms. Skirtumas yra tas, kad siekiant spręsti akustinio reprezentatyvumo problemą nukenčia skyriaus pradžioje pateikti kalbos vienetų gausumo ir kalbos vienetų lankstumo sudarant nežinomą žodį punktai.

Skiemuo. Skiemuo suprantamas, kaip pereinamasis segmentinis vienetas tarp fonemos ir žodžio. Tai „kalbos srauto atkarpa, kurios garsai sudaro minimalų artikuliacinį, akustinį ir funkcinį vienetą“ (Pakerys 2003). Lietuvių kalboje skiemens pagrindu būna balsiai, dvibalsiai ar mišrieji dvigarsiai. Šalia šliejasi priebalsiai. Šnekos tyrėjus skiemuo domina dėl tos priežasties, kad tai yra pirmoji plotmė, kurioje pasireiškia koartikuliacija, t. y. garsai skiemens viduje veikia vienas kitą labiau, nei tie patys garsai atskirti skiemens riba. Kaip tik čia iškyla problema, būdinga ne vien lietuvių kalbai, – skiemens ribų problema. Fonetikai skiemenų riboms nustatyti vadovaujasi keliomis teorijomis: ekspiracijos teorija, kalbos padargų raumenų įtampos teorija ir sonoringumo teorija (Pakerys 2003). Šnekos atpažinimui labiau tinka pastaroji, kadangi ji susijusi su akustinių šnekos savybių tyrimu. Sonoringiausi yra balsiai, o kadangi jie yra skiemens pagrindas, žodžiai į skiemenis skirstomi pagal tai, kiek sonoringumo taškų jame surandama. Ne visada lengva suskirstyti žodį skiemenimis, kai greta atsiduria panašaus sonoringumo garsai. Fonetikai šią problemą sprendžia funkcine skiemens teorija

– skiemuo prasideda maksimaliu priebalsių junginiu, galimu skiemens pradžioje (Pakerys 2003).

Lietuvių kalbos skiemenavimui yra sukurta tik viena programa (Fotonija 1996), kuri žodžių skaidymą skiemenimis atlieka lietuvių kalbos taisyklių pagrindu. Skiemenuojama gali būti ir remiantis akustiniu signalu. Šiuo metu dar nėra sukurtas įrankis automatiniam lietuvių kalbos žodžių skaidymui skiemenimis pasinaudojant signalo akustinėmis savybėmis. Šiame darbe žodžių skaidymui skiemenimis naudotas taisyklių rinkinys – aiški instrukcija kaip žodį skiemenuoti.

Lietuvių kalboje esančių skiemenų skaičiaus skaitinė išraiška nėra žinoma. Kaip ir kontekstinė fonema, jis ribojamas atsižvelgiant į mokymo duomenis.

Skiemens, kaip ir toliau aprašomo žodžio, kaip kalbos vieneto naudojimas nėra paplitęs dėl sudėtingų akustinių modelių aibės elementų parinkimo būdų ir ne tokio lankstaus taikomumo kaip fonemos, nors kokybės atžvilgiu skiemenų akustiniai modeliai yra daug tikslesni dėl ilgesnės jų trukmės.

Skiemuo irgi ne visada atitinka skyriaus pradžioje išvardintus reikalavimus, nes išryškėja mokymo imties ir akustinių modelių taikymo lankstumo problemos, nors akustinio reprezentatyvumo klausimas čia nekyla. Analogiška situacija yra ir su žodžio atveju.

Žodis. Žodžio sąvoka intuityviai suprantama daugelio indoeuropiečių prokalbės atstovų. Apytiksliai jis gali būti apibrėžtas kaip „leksinis vienetas, kuriam konvencijos būdu priskirta reikšmė ir kuris gali būti naudojamas įvairiose sintaksinėse kombinacijose, priklausomai nuo to, kokiai kalbos daliai jis priklauso“ (Girdenis 2003). Indoeuropiečių fonetinėse kalbose rašymo sistema remiasi abėcėle, kur vienas simbolis atitinka vieną garsą, o pagrindiniai rašymo sistemos elementai – žodžiai, sudaromi iš abėcėlės simbolių. Tekste tarp žodžių yra įprasta palikti tarpus, o kalbant žodžiai suliejami – gaunama ištisinė šneka. Kitos kalbų grupės remiasi kitomis rašymo sistemomis, kaip: logografine, ideografine. Logografinėje rašymo sistemoje naudojamos žodžio morfemos, dažnai skiemenys, todėl sąvoka „žodis“ kalbant apie šios rašymo sistemos grupės kalbas netinka. Ši sistema paplitusi Azijos šalyse (Kinijoje, Japonijoje, Korėjoje, Vietname), kuriose leksiniu vienetu tampa ne tik žodis, bet kartu ir skiemuo.

Žodžiams šiuolaikinėse atpažinimo sistemose akustiniai modeliai nėra kuriami, išskyrus vienskiemenius trumpus ir dažnus žodžius, bet žodžio sąvoka yra svarbi. Nors šnekos atpažinimo sistemos atpažįsta smulkesnius žodžio elementus, kaip: skiemuo, kontekstinė fonema ar fonema – atpažinimo įrankiui pateikiamas minėtų elementų junginys, reprezentuojantis žodį. Taip apribojama atpažinimo erdvė, o atpažinimo rezultatai pateikiami suprantama forma – žodžių atpažinimo tikslumu. Žodžio elementų akustinių modelių vertinimo prasme toks būdas yra neinformatyvus, nes neleidžia įvertinti atskirų akustinių modelių tikslumo reprezentuojamiesiems vienetams.

Apibendrintos nagrinėtų kalbos vienetų savybės pateiktos 2.1 lentelėje.

2.1 lentelė. *Kalbos vienetų tipų savybių apibūdinimas*[4]

Kalbos vienetų tipas	Kalbos vienetų tipų savybės		
	Tikslumas	Gausumas	Lankstumas
Fonema	--	++	++
Kontekstinė fonema	-	-	-
Skiemuo	+	-	-
Žodis	++	--	--

Populiariausi lingvistiniai kalbos vienetai yra paprastos ir kontekstinės fonemos, skiemenys. Tiriant šiuos kalbos vienetus gaunami tai vienų, tai kitų kalbos vienetų privalumą patvirtinantys argumentai. Fonema ir kontekstinė fonema paprastai naudojamos ASA sistemose (Veeravalli 2005), todėl bet koks ASA sistemos tobulinimas kalbos vienetų atžvilgiu yra lyginamas su pagrindine – fonemomis ar kontekstinėmis fonemomis grįsta ASA sistema. Daugelis darbų skirta skiemens kalbos vieneto naudojimo įtvirtinimui arba hibridiniam skiemenų-fonemų naudojimui. Tiriant telefoninę šneką iš skiemenų, kontekstinių fonemų ir įvairių jų kombinacijų buvo išskirtas vienas derinys *skiemenys* + *kontekstinės fonemos* + *vienskiemeniai žodžiai*, kuris pagerina kontekstinėmis fonemomis pasiektą atpažinimo rezultatą iki 50,9 % (Ganapathiraju *et al.* 2001). Skaitytinės šnekos atpažinime, grįstame skiemenimis ir fonemomis, skiemenų naudojimas užtikrina didesnį atpažinimo tikslumą – 61,1 %, o fonemų naudojimas – 35,4 % (Rhys *et al.* 1997). Svarbus tyrimas buvo atliktas nurodant skiemens kalbos vieneto tipo panaudojimo kryptį – vardų atpažinimą. Sethy (2002) palygino vardų atpažinimo rezultatus, atpažinimą grindžiant fonemomis ir skiemenimis. Atpažinimas vien tik skiemenimis siekė 80 % atpažinimo tikslumą, fonemomis – 63 %, skiemenimis-fonemomis – 75 %.

Kitų darbų tikslas – modifikuoti esamus kalbos vienetus: šalia kalbos vienetų naudoti papildomą nefonetinę informaciją (Reichl *et al.* 1999), akcentuoti vienas kalbos vieneto dalis labiau nei kitas (Fosler-Lussier *et al.* 1999, Greenberg 1998, Greenberg, Chang 2000), tirti kalbos vieneto AM topologijas, parametrų struktūrą (Lockwood, Blanchet 1993, Smyth 1997, Hain, Woodland 1999, Hain, Woodland 2000).

Kalbos vienetai, naudojami (Reichl, Chou 1999) tyrimuose, yra įprastos kontekstinės fonemos, tačiau jų būsenų klasterizavimo procedūra modifikuojama įvedant papildomą klasterizavimo kriterijų. Standartinis kontekstinių fonemų ar jų būsenų klasterizavimo kriterijus yra kontekstas, t. y. tiriama, kokie garsai supa nagrinėjamą fonemą. Autoriai teigia, kad šalia šio kriterijaus galima naudoti papildomą, nefonetinį kriterijų, kaip kad jie naudoja lytį ar fonemos poziciją žodyje.

[4] Kalbos vienetų vertinimas (++,+,-,--) yra grįstas jo lyginimu su likusiais kalbos vienetais. Naudojami žymenys: ++ stipriai pasižymi savybe, + gerai, – vidutiniškai, -- silpnai.

Skiemens kalbos vienetas paprastai dalinamas į tris segmentus: pradžią (*onset*), branduolį (*nucleus*) ir pabaigą (*coda*). Buvo atliekami ištisinės ir skaitytinės šnekos tyrimai (Fosler-Lussier *et al.* 1999, Greenberg 1998), kuriais buvo bandoma atsakyti į klausimą, kiek stabili yra kiekviena skiemens dalis. Pagal šiuos tyrimus, nepriklausomai nuo šnekos stiliaus skiemens pradžia yra laikoma stabiliausiu dariniu, kuriame fonemų seka išlieka ta pati. Skiemens branduolio balsis gali būti keičiamas į kurį nors kitą balsį, o pats nestabiliausias darinys – skiemens pabaiga. Skiemens pabaigoje fonemų seka ne visada išlieka, priklauso nuo kalbėjimo stiliaus. Ši skiemens dalis dažniausiai yra praleidžiama, netariama ir jos nebuvimas nedaro įtakos skiemens branduoliui. Greenberg ir Chang (2000) parodė, kad yra glaudus ryšys tarp to, kaip sėkmingai atpažįstamas žodis ir kaip teisingai identifikuojama skiemens pradžia. Visi šie tyrimai akustinio modeliavimo metu linkę daugiau dėmesio skirti skiemens pradžios stiprinimui.

Šnekos atpažinime fonemos modeliuojamos naudojant fiksuotą PMM topologiją. Ši topologija pasižymi 3–5 PMM būsenomis, kurios išreiškiamos stebėjimų skirstiniais, ir nenuliniais tapačiais perėjimais tarp būsenų. Nėra atsižvelgiama į fonemos kontekstą, pačios fonemos savybes. Tyrėjai (Lockwood, Blanchet 1993, Smyth 1997, Hain, Woodland 1999, Hain, Woodland 2000) nagrinėjo įvairias PMM topologijų gavimo schemas panaudodami akustinius duomenis. Štai Lockwood ir Blanchet tokius PMM parametrus, kaip: būsenų, Gausinių mišinių skaičius, būsenų perėjimo matrica – nustato korekcinėje mokymo procedūroje pagal atpažinimo klaidą. Pavyzdžiui, jei įvyksta vienos fonemos pakeitimas kita – pridedamas mišinys, jei įvyksta fonemos įterpimas – pridedama nauja būsena, jei fonema praleidžiama – pridedamas naujas perėjimas (Lockwood, Blanchet 1993).

2.2. ASA sistemų charakteristikų įtakos šnekos atpažinimui aptarimas

Standartinės ASA sistemų charakteristikos yra šios: šnekos tipas, žodyno dydis, priklausomumas/nepriklausomumas nuo kalbėtojo, ryšio kanalai ir aplinka. ASA sistemas įmanoma lyginti, jei šie parametrai sutampa.

Žodyno dydis. Žodynas apibrėžia ASA sistemos atpažįstamus žodžius. Teigiama (Deller *et al.* 1993), kad atpažinimo sudėtingumas didėja logaritmiškai, augant žodžių skaičiui žodyne. ASA sistemas galima klasifikuoti pagal žodyno dydį – mažos, vidutinės ir didelės apimties žodynai. Literatūroje žodyno dydis skaičiais apibrėžiamas labai įvairiai ir sąvokos siejimas su skaitine išraiška nėra nusistovėjęs. Viename literatūros šaltinyje mažos apimties žodynas yra iki 20, o didelės virš 20 000 žodžių (Cole *et al.* 1998). Kitame šaltinyje teigiama, kad mažos apimties žodyne turėtų būti 1–99 žodžiai (Deller *et al.* 1993), vidutinės –

100–999, o didelės – virš 1000 žodžių. Daug paprasčiau ASA sistemą priskirti vienai ar kitai kategorijai pagal sistemos taikymo sritį. Mažos apimties žodyno sistemos naudojamos tokių uždavinių, kaip: kreditinės kortelės ar telefono numerio skaitmenų atpažinimo užduotims; vidutinio – eksperimentinėms sistemoms laboratorijose, modeliuojant ištisinės kalbos atpažinimą; didelės apimties žodyno sistemos naudojamos kuriant komercinius produktus, realizuojant teksto diktavimo ar korespondencijos uždavinius.

Priklausomumas/nepriklausomumas nuo kalbėtojo. Viena iš ASA sistemų svarbių charakteristikų – atpažįstamų kalbėtojų skaičius. Sistema, sukurta vieno kalbėtojo poreikiams, vadinama nuo kalbėtojo priklausoma sistema; jei ja gali naudotis daugiau kalbėtojų – tai nuo kalbėtojo nepriklausanti sistema. Pirmosios sistemos didesnis atpažinimo tikslumas paaiškinamas mažesnėmis variacijomis šnekos signaluose. Universalumo prasme antroji sistema yra geresnė, nes nereikalauja pakartotinio akustinių modelių kūrimo naujam kalbėtojui. Norint padidinti atpažinimo tikslumą atliekama kalbėtojo adaptacija esamai sistemai, t. y. sistemos akustiniai modeliai modifikuojami pagal kalbėtojo balso charakteristikas. Ši operacija gali vykti darbo su sistema sesijos metu (*on-line*) arba kalbėtojas prieš pradėdamas naudotis sistema gali būti paprašytas įrašyti garsinę informaciją pagal pateiktą tekstą (*off-line*).

Ryšio kanalai ir aplinka. Vienas iš pagrindinių reikalavimų ASA sistemoms – atsparumas (robastiškumas) ryšio kanalo iškraipymams ir aplinkos triukšmams. Tiek ryšio kanalai, tiek aplinka šalia šnekos signalo įkomponuoja papildomos informacijos, kuri yra ne tik nenaudinga, bet ir iškraipo patį šnekos signalą. Tokią informaciją įprasta vadinti triukšmu. Triukšmą galima suklasifikuoti į: biuro aplinkos, kanalų, transporto ir pramonės (Junqua, Haton 1996). Galimas ir daug smulkesnis klasifikavimas pagal jo pasiskirstymą laiko ir dažnių srityse (periodiniai, impulsiniai, plačiajuosčiai ir pan.). Triukšmui signale išmatuoti naudojamas signalo/triukšmo santykio matas SNR (*signal-to-noise ratio*) ir jo modifikacijos.

Teigiama (Junqua, Haton 1996), kad ASA sistemų darbingumas nenukenčia SNR mažėjant iki 25 dB, bet triukšmo lygiui didėjant atpažinimas blogėja. Su triukšmu kovojama įvairiais būdais: jis šalinamas, naudojami atsparūs požymiai, akustiniai modeliai adaptuojami triukšmui.

Šnekos tipai. *Izoliuoti žodžiai (isolated words*). Izoliuotų žodžių atpažintuvui reikalaujama pateikti ištarimą (tai gali būti atskiras žodis arba frazė), iš abiejų pusių ribojamą kitų signalo charakteristikų (tyla, žemo lygio triukšmas). Šie atpažintuvai veikia klausymo/neklausymo režimais, realizuojamais signalo galo taškų nustatymo algoritmais (*an endpoint detection algorithm*). Taip identifikuojami minėtų signalų charakteristikų pasikeitimo momentai. Iš kalbėtojo reikalaujama tarp ištarimų daryti trumpas pauzes kiekvieną žodį tariant izoliuotai nuo kitų, kaip atsirado ir pavadinimas. Šis kalbėjimo būdas patogus, jei reikalinga atpažinti atskiras komandas, bet netinkamas kitoms atpažinimo

sistemų taikymo sritims. *Rišlios frazės* (*connected words*) yra pereinamasis variantas nuo izoliuotų žodžių prie ištisinės šnekos. *Ištisinė šneka* (*continuous speech*) paprastai būna skaitytinė. Ištisinės šnekos atpažintuvams keliamas reikalavimas – atpažinti šneką, kurioje tarp žodžių ar frazių pauzės gali būti arba dažniausiai jų nėra. Kalbėtojas gali kalbėti beveik natūraliai. Panaudojimo sritis – diktavimas kompiuteriui. *Spontaniška šneka* (*spontaneous speech*) kasdienė šneka. Spontaninį šnekėjimo būdą galima apibūdinti jo savybėmis. Tai nėra sklandi šneka, dažnai joje yra ilgų pauzių, žodžių fragmentų (pradėtų ir nebaigtų žodžių), neteisingų ištarimų, mikčiojimų, pasikartojimų, sintaksiškai neteisingų sakinių. Tai paaiškinama tuo, kad kalbėdamas žmogus kartu mąsto. Dažnai tokia šneka lydima papildomų triukšmo šaltinių, kaip: juokas, čepsėjimas ir pan. Spontaniška šneka pasižymi ir dideliu skaičiumi žodžių, nesančių sistemos žodyne. Dėl to ją atpažinti tiek akustiškai, tiek gramatiškai yra sudėtinga.

Izoliuotų žodžių tipas šnekos atpažinime. Izoliuotų žodžių atpažinimas palyginti su ištisinės ar spontaniškos šnekos atpažinimu laikomas lengvesne užduotimi. Nuo jų prasideda ir šnekos atpažinimo raida: nuo paprastesnių dalykų einama prie sudėtingesnių – nuo izoliuotų žodžių atpažinimo prie ištisinės šnekos atpažinimo. Ankstyvosios izoliuotų žodžių atpažinimo sistemos rėmėsi pavyzdžiais grįstu šnekos atpažinimo metodu (3.1.2 skyrelis). Standartinis tokios atpažinimo sistemos modelis susidėjo iš požymių išskyrimo, pavyzdžių palyginimo ir sprendimo priėmimo blokų, kaip pavaizduota 3.1 paveiksle.

Šis šnekos atpažinimo modelis buvo praktiškas dėl kelių priežasčių: 1) invariantiškumo žodynams, vartotojams, pavyzdžių palyginimo algoritmams, sprendimo taisyklėms, 2) lengvo įgyvendinimo ir gero praktinio veikimo (Rabiner, Levinson 1981). Tokių sistemų etaloniniais ir testiniais pavyzdžiais vadinama požymių pavidalo šnekos signalo spektrinė informacija. Pavyzdžiams palyginti ji taikoma tiesiogiai skaičiuojant atstumą dinaminio laiko skalės kraipymo būdu (*Dynamic Time Warping*, **DTW**). Laiko ašių iškraipymai atliekami taikant dinaminį programavimą. Pagal tai, ar atpažinimo sistema priklausoma/nepriklausoma nuo kalbėtojo, skiriasi etalonų formavimo procesas – mokymas ir sprendimo taisyklė. Jei atpažinimo sistema priklausoma nuo kalbėtojo, užtenka vieno ar dviejų kalbėtojo ištarimų spektrinės informacijos. Jei atpažinimo sistema nepriklauso nuo kalbėtojo, atliekamas vektorinis kvantavimas, t. y. daugelio kalbėtojų ištarimų spektrinės informacijos pavyzdžiai suskirstomi į mažesnį pavyzdžių skaičių. Šio tipo sistema yra veikiama tokių veiksnių, kaip: žodyno sudėtingumas, veikiantis atpažinimo tikslumą; perdavimo kanalai, darantys įtaką atpažinimo tikslumui ir šnekos signalo galo taškų nustatymui; pačios sistemos sudėtingumas, turintis įtakos sistemos atsako laikui. Tokio tipo atpažinimo sistemų, besiskiriančių žodyno dydžiu (641–5000 žodžių) ir jo struktūra (žodis neskaidomas ar skaidomas į mažesnius vienetus), priklausomumu/nepriklausomumu nuo kalbėtojo, įvairiais kitais parametrais ir modifikacijomis (kalbos modelio naudojimu) buvo sukonstruota

daug (pavyzdžiui Sugamura *et al.* 1983, Jelinek *et al.* 1985). Šių izoliuotų žodžių atpažinimo sistemų tikslumas svyravo tarp 94 % ir 96,3 %.

Kitas izoliuotų žodžių atpažinimo etapas susijęs su netiesioginiu spektrinės informacijos panaudojimo atpažinimui būdu. Didelės apimties garsynų atsiradimas paskatino naudoti statistinius metodus ir naują spektrinės išraiškos būdą (3.2 poskyris). Izoliuotų žodžių atpažinimui pradedami naudoti PMM (Murviet, Weintraub 1988), neuroniniai tinklai (Sakoe *et al*. 1989). Atpažinimo tikslumas yra įvairus 40 %–99,3 %.

Lietuvoje izoliuotų žodžių atpažinimas taip pat yra dviejų krypčių: tiesiogiai naudojant spektrinę informaciją šnekos pavyzdžiams palyginti (Lipeika *et al.* 2002) ir naudojant statistinius modelius spektrinės informacijos apibendrinimui (Filipovič, Lipeika 2004, Raškinis, Raškinienė 2003b).

Ištisinės šnekos atpažinimas. Ištisinės šnekos atpažinimo sistemos palyginti jas su izoliuotų žodžių atpažinimo sistemomis yra sudėtingesnės. Izoliuotų žodžių atpažinimo sistemose lengviau išskirti žodžio pradžią ir pabaigą, nes paprastai juos žymi pauzės. Ištisinės šnekos atpažinimo sistemose šios ribos yra neaiškios, nežinomos ir jų nustatymas yra viena iš papildomų užduočių atpažinimo sistemai. Antras skirtumas yra žodyno dydis, nes ištisinės šnekos atpažinimo sistemos neįsivaizduojamos be didelių žodynų (*Large Vocabulary Continuous Speech Recognition*, **LVCSR**), todėl šnekos atpažinimo tikslumui didinti naudojami kalbos modeliai.

Holmes (2001) pateikia 1996–2000 metų šiuolaikinių ištisinės ir spontaniškos šnekos atpažinimo sistemų tikslumą, kuris svyruoja tarp 69,6 % ir 98 %. Panašus ištisinės šnekos atpažinimo tikslumas 74,5–95,6 % buvo dar 1992–1995 metais (Gauvain, Lamel 1996). Tai rodo, kad ištisinės šnekos atpažinimas netobulėja, išlieka dar vis toje pačioje stadijoje.

Ištisinės šnekos atpažinimui vien žodžių kalbos vienetai nėra naudojami dėl mažų tokių modelių mokymo imčių. Atskirai sukurti žodžių AM gali būti integruoti į galutinę AM aibę.

2.3. Akustinio modeliavimo vieta lietuvių šnekos kalbinių technologijų tyrimuose

Lietuvių šnekos tyrimai vyksta įvairiomis kryptimis, todėl kartais sunku susidaryti bendrą vaizdą, koks kalbinių technologijų išsivystymo lygis jau pasiektas. Atskiri darbai nurodo, kiek toli pažengta konkrečioje srityje.

Galima identifikuoti tokias būdingiausias lietuvių šnekos tyrimo sritis: skaitmeninių signalų apdorojimas, kalbančiojo identifikavimas ir verifikavimas, akustinis modeliavimas, kalbos modelio tyrimai, šnekos sintezavimo tyrimai, šnekos tyrimai lingvistiniu aspektu. Visi šie tyrimai skirti šnekos atpažinimo ir

sintezavimo blokams sukurti, kuriuos sujungus būtų galima kalbėti apie šnekos atpažinimo-sintezės sistemą.

Lietuvių šnekos tyrimai prasidėjo tik atsiradus minimalioms lietuvių šnekos bazėms. Iki to laiko šnekos atpažinimas buvo taikomas daugiausiai rusų kalbai. Pirmieji tokie tyrimai buvo skirti analizuoti požymių vektorių sudėties (dinaminių ir statinių komponenčių atžvilgiu) ir dydžio, mokymo imties dydžio įtaką atpažinimui, fonemų diskriminatyvumą (Rudžionis 1987). Nors šie tyrimai nėra tiesiogiai susiję su lietuvių šneka, pasiekti rezultatai buvo svarbūs dėl įgytos patirties, vėliau tyrėjai atliko daug tyrimų lietuvių šnekai. Ir dabar tiriant šnekos atpažinimo sistemos charakteristikų pokyčius, kur reikalingos didelės duomenų imtys, tyrėjai naudoja kitoms kalboms sukurtus garsynus TIMIT, WSJCAM0 (Šilingas, Telksnys 2004).

Skaitmeninio signalo apdorojimo srityje dirbama nuo seno. Atsitiktinių procesų charakteristikų pasikeitimų tyrimai (Telksnys 1987, Lipeika, Lipeikienė 1992) buvo pritaikyti šnekos atpažinimo procese ieškant šnekos pradžios ir galo taškų (Lipeika, Lipeikienė 2003, Lipeika, Tamulevičius 2004).

Išsamūs tyrimai buvo atliekami identifikuojant kalbantįjį. Teoriniais ir eksperimentiniais tyrimais buvo tikrinamas matų atstumui tarp pseudo-stacionarių šnekos segmentų skaičiavimo būdų darbingumas (Lipeika, Lipeikienė 1993), kalbančiojo identifikavimo metodai, kaip: grįstas vidutinio atstumo tarp klasterių skaičiavimu, grįstas vektoriniu kvantavimu ir grįstas balso trakto ir žadinimo signalo tiesinės prognozės parametrais (Lipeika, Lipeikienė 1995, Lipeika, Lipeikienė 1996). Šių tyrimų rezultatai sėkmingai panaudoti sprendžiant kalbančiojo identifikavimo ir verifikavimo problemą (Lipeika, Lipeikienė 1999).

Jei iš pradžių apsiribota skaitmeninių signalų apdorojimo tyrimais, tai šiuo metu daug dėmesio skiriama šnekos tyrimams lingvistiniu aspektu. Rezultatai gali gerėti tik nagrinėjant pačią kalbą: jos struktūrą, ypatumus ir kuriant papildomą kalbos analizės bloką. Šių tyrimų tiesioginė paskirtis yra automatinis vertimas, bet gauti rezultatai gali būti naudojami ir šnekai atpažinti, siekiant suvokti atpažįstamo sakinio prasmę ir užtikrinant gyvą bendravimą dialogo forma su kompiuteriu. Pagrindiniais etapais galima įvardinti morfologinę, sintaksinę ir semantinę analizę. Morfologinės analizės įrankis jau yra sukurtas (Zinkevičius 2000), bet kol kas nėra pritaikytas šnekos atpažinimui. Sintaksinė lietuvių kalbos sakinio struktūra yra analizuota tik teoriškai (Šveikauskienė 2005), bet praktiškai nėra taikyta.

Pradiniai darbai vyksta bandant sujungti šnekos atpažinimo sistemą su vaizdo atpažinimo sistema ir taip gerinant šnekos atpažinimą (Kaukėnas *et al.* 2006).

Šnekos sintezės tyrimus galima suskaidyti į kelis etapus: teksto skaidymas skiemenimis[5], kirčiavimas, tekstinių simbolių pavertimas fonetiniais vienetais[6] ir garso generavimas (Kasparaitis 1999). Kiekvienas iš šių etapų buvo nuosekliai tiriamas, tolesnį etapą grindžiant prieš tai buvusio etapo žiniomis. Buvo sukurtos formalios taisyklės automatiniam žodžių kirčiavimui ir tekstinių simbolių vertimui fonetiniais vienetais, sintezatorius (Kasparaitis 2000, Kasparaitis 2001), analizuojami fonetinių vienetų tipai, tinkami lietuvių kalbos sintezei (Kasparaitis 2005). Atskirai buvo atliekami žodžių transkribavimo į fonemų tipo kalbos vienetų aibę tyrimai (Skripkauskas, Telksnys 2006).

Daug teorinių ir eksperimentinių tyrimų buvo atlikta akustinio modeliavimo srityje. Visų jų tikslas buvo sukurti šnekos atpažinimo sistemas. Tirti tokie atpažinimo metodai, kaip: dirbtiniai neuroniniai tinklai (Filipovič 2005), dinaminis laiko skalės kraipymas (Lipeika *et al.* 2002), paslėpti Markovo modeliai (Raškinis, Raškinienė 2003a, 2003b), sukurtas naujas projekcija grįstas šnekos atpažinimo metodas (Noreika, Rudžionis 1991). Daug tyrimų atlikta ieškant kiekvienos modeliuojamos atpažinimo sistemos geriausių parametrų, kuriuos galima padalinti į tris grupes (Raškinis, Raškinienė 2003a):

- požymių tipą apibrėžiantys parametrai (požymių tipai, dažnio diapazonas, požymių koeficientų skaičius ir kt.);
- PMM mokymo procesą veikiantys parametrai (fonemų rinkiniai, PMM topologijos, Gausinių mišinių skaičius būsenoje, kontekstinių fonemų klasterizavimo būdai);
- atpažinimo procesą veikiantys parametrai (kalbos modelis).

Atliekant šių parametrų optimizavimą fonemų atpažinimu grindžiamoje atskirai tariamų žodžių atpažinimo sistemoje, naudojant 1 valandos trukmės garsyną, ZK pavyko sumažinti nuo 48–19 % iki 3–9 % (Raškinis, Raškinienė 2003a). Analogiška sistemos parametrų paieška atliekama tiriant rišlių frazių atpažinimą ir pasiekiant 6–21 % ZK (Raškinis, Raškinienė 2004), naudojant ne fonemų, o perėjimo tarp fonemų modelius ir pasiekiant 18–51 % ZK (Štrimaitis 2004). Nors buvo naudotas tas pats 1 valandos trukmės garsynas, kaip ir Raškinių tyrimuose, bet gautas žemesnis atpažinimo procentas sietinas su kitų kalbos vienetų tipo naudojimu. Šis tyrimas kartu nurodo ir kitą kryptį, kuria gali būti atliekami šnekos atpažinimo tyrimai, t. y. modeliuojamų kalbos vienetų paieška. Lietuvių šnekos atpažinimą įprasta grįsti fonemomis, todėl atliekami įvairių fonemų rinkinių tyrimai (Raškinis, Raškinienė 2003a, Šilingas *et al.* 2004b), stengiamasi atrasti būdų, kaip geriau diskriminuoti fonemas (Driaunys *et. al.* 2005). Yra tyrimų, kurie nagrinėja fonemomis grįsto

[5] Kasparaičio sukurtas skaidymo skiemenimis algoritmas (Kasparaitis 2004) buvo naudojamas ir šiame darbe.

[6] Straipsnyje tekstinių simbolių vertimas fonetiniais vienetais vadinamas transkribavimu. Šiame darbe transkribavimo samprata yra kita – žodžio, parašyto tekstiniais simboliais, pavertimas pasirinktų kalbos vienetų seka.

atpažinimo alternatyvas – kvazi-fonemas, formuojamas pagal akustinius duomenis (Skripkauskas 2006). Šiam darbui artimą tematiką nagrinėjo Šilingas, tyręs fonemas, kontekstines fonemas, skiemenis ir jų galimus derinius (Šilingas 2005). Vis dėl to minėtame darbe pagrindiniu kalbos vienetu išlieka fonema, kontekstinė fonema, o skiemenų kalbos vienetai pridedami, jei viršijamas nurodytas slenkstis. Minėto darbo rezultatai parodė, kad skiemenų prijungimas prie fonemų aibės nežymiai padidina atpažinimo tikslumą, bet nepralenkia kontekstinėmis fonemomis pasiektų rezultatų.

Alternatyvų DTW ir PMM naudojančiam šnekos atpažinimo algoritmams sukūrė ir realizavo Noreika ir Rudžionis. Šis algoritmas remiasi šnekos signalo segmentavimu į statines-dinamines fonemos tipo atkarpas pagal segmentavimo funkcijos minimumus ir maksimumus. Algoritmas buvo modifikuojamas ir optimizuojamas segmentų skaičiaus, požymių tipų atžvilgiu (Rudžionis *et al*. 1995). Jis toliau taikytas tiriant fonemų diskriminatyvumą ir siekiant jį padidinti. Buvo tirti lingvistiniai kalbos vienetai, t. y. priebalsių ir balsių kombinacijos naudojant įvairius diskriminantinius metodus (Rudžionis *et al.* 1999). Gauti rezultatai nurodo būdą, kaip geriau diskriminuoti fonemas; identifikuoja ir paaiškina sudėtingus konkrečių junginių atpažinimo atvejus.

Lietuvių kalbos modelio, skirto integravimui į šnekos atpažinimo sistemą, kūrimas vyksta lėtai. Daugelį darbų šioje srityje yra atlikę vos pora tyrėjų. Kalbos modelio kūrimas yra ypač sudėtinga užduotis dėl to, kad kalbą labai veikia šnekos specifika ir struktūra. Lietuvių kalba yra sudėtingai kaitoma ir turi kintančią žodžių tvarką, todėl metodai, taikomi kitų kalbų kalbos modelių kūrimui, netinka. Buvo nagrinėti keturi kalbos modelių kūrimo būdai: įprastas-standartinis n-gramų, besiremiantis žodžių suskirstymu į klases (Vaičiūnas *et al.* 2004), besiremiantis žodžio skaidymu į prasmines ir morfologines dalis (Vaičiūnas, Raškinis 2003), besiremiantis kelių kalbos modelių kūrimu pagal teksto žanrus ir jų jungimu į jungtinį kalbos modelį (Vaičiūnas, Raškinis 2005b). Visa tai apibendrinama atlikus kalbos modelio integraciją į šnekos atpažinimo sistemą (Vaičiūnas, Raškinis 2005a). Ši integracija leido padidinti šnekos atpažinimo tikslumą 15 %.

Bendras lietuvių kalbos tyrimų vaizdas pateiktas norint pademonstruoti, kad akustiniu modeliavimu užsiima ir tie tyrėjai, kurių tiesioginiai tyrimų tikslai yra visai kiti – sintezės, atpažinimo metodų, kalbos modelių tyrimai. Atlikdami akustinį modeliavimą jie naudoja fiksuotus dydžius (kalbos vieneto tipą, jų rinkinį, požymius, AM parametrus, mokymo schemą ir pan.). Tyrėjai, kurių tiesioginis tikslas yra akustinis modeliavimas, minėtus dydžius laiko kintančiais ir modeliuodami ASA sistemą stengiasi juos optimizuoti atpažinimo tikslumo atžvilgiu.

Išskiriant kalbos vienetų akustinio modeliavimo tyrimus iš bendro konteksto, svarbiausi tyrimai pateikiami 2.2 lentelėje.

2.2 lentelė. *ASA sistemų, skirtų kalbos vienetų tyrimams, apžvalga*

ASA sistemos tyrimo tikslas	ZT	Šnekos tipas	Kalbos vienetas
Kitoms kalboms			
Kalbos vienetų tipų modeliavimo tyrimas (Rhys *et al.* 1997)	35–61 %	ištisinė šneka	fonema, skiemuo ir fonema
Hibridinių kalbos vienetų modeliavimo tyrimas (Ganapathiraju *et al.* 2001)	50,9 %	telefoninė šneka	kontekstinė fonema, skiemuo
Kalbos vienetų tipų modeliavimo tyrimas (Sethy 2002)	63–80 %	asmenvardžiai	fonema, skiemuo ir fonema, skiemuo
Lietuvių kalbai			
Fonemų diskriminantiniai tyrimai, naudojant įvairius diskriminantinius metodus (Rudžionis *et al.* 1999)	77–84 %	garsų junginiai	fonema
ASA sistemos parametrų optimizavimas (Raškinis, Raškinienė 2003a)	91–97 %	izoliuoti žodžiai	kontekstinė fonema
ASA sistemos parametrų optimizavimas (Raškinis, Raškinienė 2004)	79–94 %	rišlios frazės	kontekstinė fonema
Kalbos vieneto modeliavimo tyrimas (Štrimaitis 2004)	49–82 %	rišlios frazės	perėjimo tarp fonemų modelis
ASA sistemos parametrų optimizavimas, fonemų rinkinių modeliavimo tyrimas (Filipovič 2005)	80–91 %	izoliuoti žodžiai	fonema
Fonemų, kontekstinių fonemų rinkinių modeliavimo tyrimas (Šilingas et al. 2004)	57–77 %	ištisinė šneka	fonema, kontekstinė fonema
Fonemų, kontekstinių fonemų rinkinių modeliavimo tyrimas (Šilingas et al. 2006)	81–84 %	ištisinė šneka	kontekstinė fonema
ASA sistemos parametrų optimizavimas, hibridinių kalbos vienetų modeliavimo tyrimas (Šilingas et al. 2006)	74–83 %	ištisinė šneka	fonema, skiemuo, kontekstinė fonema, kontekstinis skiemuo

Iš pirmo žvilgsnio matyti didelė šnekos atpažinimo tikslumo (ZT) įvairovė, pavyzdžiui, to paties kalbos vieneto – skiemens – akustinis modeliavimas skirtinguose tyrimuose (Rhys et al. 1997) ir (Sethy 2002) pateikia skirtingus šnekos atpažinimo tikslumo rezultatus 61 % ir 80 %. Šis pavyzdys rodo, kad skirtingų ASA sistemų atpažinimo tikslumo rezultatų taip paprastai palyginti negalima. Lyginant reikia atsižvelgti į šnekos tipą, akustinių modelių aibės dydį, mokymo imčių dydį, ASA sistemos sudėtingumą, ASA sistemos parametrų optimizacijos etapo buvimą ir daugelį kitų parametrų. Žvelgiant į kalbos vienetų akustinio modeliavimo tyrimus lietuvių kalbai, matyti, kad nors ir dominuoja fonemos ir kontekstinės fonemos kalbos vienetai, ASA sistemų rezultatų lyginimas yra sunkiai įmanomas. Taigi, buvo minėta, kad lietuvių šnekos

atpažinimo procese dominuoja fonemos ir kontekstinės fonemos, o automatiniu būdu gaunamų kalbos vienetų ar hibridinių kalbos vienetų modeliavimas nėra atliekamas, kas daroma kitoms kalboms atpažinti. Tiesa, pirmasis darbas, lyginant kelis kalbos vienetų tipus ir naudojant hibridinę kalbos vienetų aibę jau pasirodė (Šilingas et al. 2006). Šis darbas buvo atliktas bendradarbiaujant su šios disertacijos autore. Reikia paminėti, kad dr. Šilingas ir disertacijos autorė kartu tyrė tuos pačius kalbos vienetų tipus (fonema, kontekstinė fonema, skiemuo), tačiau dr. Šilingą labiau domino ASA sistemos sudėtingumo (parametrų optimizavimas), atpažinimo tikslumo aukščiausios ribos (įjungiant kalbos modelį, rankinį atpažinto teksto taisymą) klausimai. Disertacijos autorės dėmesio objektas buvo kalbos vienetų tipų atskiras modeliavimas, atliekant lyginimą pagal atpažinimo tikslumą ir pritaikomumo sudėtingumą.

2.4. Antrojo skyriaus rezultatai ir išvados

1. Perėjimas nuo šnekos prie akustinių modelių kaip šnekos atpažinimo objektų grindžiamas tarpiniu elementu – kalbos vienetu. Kalbos vienetas išreiškia tekstinį simbolį, o pagal jį sukurtas akustinis modelis – tekstinio simbolio reprezentaciją garsu. Vienas iš svarbiausių akustinio modeliavimo klausimų yra kalbos vienetų tipo ir konkretaus tipo aibės sudėtinių elementų parinkimas.
2. Remiantis lingvistiniu kriterijumi išskiriami šie kalbos vienetų tipai: paprastos ir kontekstinės fonemos, žodžiai, skiemenys. Populiariausios yra paprastos ir kontekstinės fonemos, prijungiant papildomą kirčio požymį. Kiti kalbos vienetų tipai iki šiol buvo mažai nagrinėjami dėl didelio jų skaičiaus ir nesusiformavusios jų akustinio modeliavimo tradicijos.
3. Lietuvių šnekai atlikti akustinio modeliavimo tyrimai apima požymių tipų, modelių parametrų parinkimo tyrimus; taip pat atlikti fonemų, kontekstinių fonemų ir fragmentiniai skiemenų akustinio modeliavimo tyrimai. Pasaulyje atliekamų analogiškų tyrimų dėmesio centre atsiduria kelių kalbos vienetų tipų akustinių modelių lyginimas, įvairių jų kombinacijų sudarymas, naujų pritaikymo sričių paieška. Tyrinėjant lietuvių šneką pasigendama išsamesnio kalbos vienetų tipų akustinio modeliavimo tyrimų.
4. Atlikus literatūros analizę kalbos vienetų parinkimo ir jų akustinio modeliavimo aspektu, iškilo šie analizuojami tolesniuose disertacijos skyriuose probleminiai klausimai:
   - Lietuvių šnekos atpažinimo tyrimuose dominuoja fonemos ir kontekstinės fonemos kalbos vienetų tipai ir jų akustinis modeliavimas neanalizuojant alternatyvių (skiemens, kontekstinio skiemens, žodžio) kalbos vienetų tipų.

- Lietuvių šnekos atpažinimo tyrimuose trūksta skirtingų kalbos vienetų tipų akustinio modeliavimo lygiagrečių tyrimų naudojant tą patį garsyną ir panašias modeliavimo schemas.
- Neaišku, ar atliekant detalesnį kalbos vienetų modeliavimą galima padidinti šnekos atpažinimo tikslumą.

3

Šnekos atpažinimo sistemų struktūrų aptarimas

Ankstesniame skyriuje buvo aptartas šio darbo problemos kontekstas ir pati problema. Šiame skyriuje pirmiausia pristatomas darbe naudotas statistinis šnekos atpažinimo metodas, akustinių modelių pagrindas – paslėptieji Markovo modeliai. Toliau pateikiama bendra modeliuojamos šnekos atpažinimo sistemos struktūra ir atlikti jos pritaikymai konkretiems kalbos tipams ir kalbos vienetams ASA sistemos modeliavimo schemų pavidalu.

3.1. Šnekos atpažinimo metodų evoliucija

Šnekos atpažinimo metodus galima suskirstyti į tris pagrindines grupes: akustinių-fonetinių (*acoustic-phonetic*), grįstų pavyzdžiais (*pattern recognition*) ir dirbtinio intelekto (*artificial intelligence*) (Rabiner, Juang 1993). Pradiniame šnekos atpažinimo tyrimų etape buvo paplitę akustiniai-fonetiniai metodai. Vystantis skaičiavimo technikai – pavyzdžiais grįsti metodai.

3.1.1. Akustiniai-fonetiniai metodai

Šnekos atpažinimo istorijoje vyksta nuolatinė išskirtinių, invariantiškų triukšmui, iškraipymams, kalbėtojams šnekos požymių paieška (Fant 1973). Akustiniai-fonetiniai metodai apibūdina šnekos atpažinimo etapą, kuriame universalieji požymiai ieškoti tokiose šnekos garsų savybėse, kaip: garso vokalizacijos lygis, pagrindinis tonas, energija, formantės, nosinis garso tarimo

būdas. Kiekvienas garsas pasižymi tam tikru išvardintų savybių rinkiniu. Manoma, kad akustinį signalą suskaidžius į segmentus, pasižyminčius tam tikru savybių deriniu, ir atlikus fonetinių vienetų tiems segmentams priskyrimą, galima identifikuoti šnekos signalą.

Šie metodai remiasi akustinės fonetikos teorija, teigiančia, kad šnekamojoje kalboje egzistuoja baigtinis skaičius fonetinių vienetų, apibūdinamų savybėmis, aptinkamomis šnekos signale. Pagrindiniai šių metodų etapai:

- Akustinio signalo segmentavimas ir fonetinių vienetų priskyrimas segmentams. Prieš šį etapą iš signalo išskirti požymiai pakeičiami identifikuotų savybių rinkiniu. Pagal savybių derinius signalas segmentuojamas, kartu uždedant konkretaus fonetinio vieneto žymę.
- Žodžių sekos iš fonetinių vienetų sekos gavimas. Paskui pagal fonetinių žymių seką bandoma surasti teisingą žodį arba žodžių seką, kurie atitinka tam tikras sąlygas (žodžiai priklauso apibrėžtam žodynui, žodžių seka tenkina sintaksės reikalavimus ir turi semantinę prasmę).

Šie metodai reikalauja gero fonetinių vienetų akustinių savybių išmanymo. Savybės, pagal kurias atliekama segmentacija ir fonetinių vienetų žymių priskyrimas, dažniausiai parenkamos remiantis intuicija. Didelių problemų kyla atliekant fonetinių vienetų sekų dekodavimą į žodžius. Visa tai kelia abejonių dėl atpažinimo rezultatų nepatikimumo.

3.1.2. Pavyzdžiais grįsti atpažinimo metodai

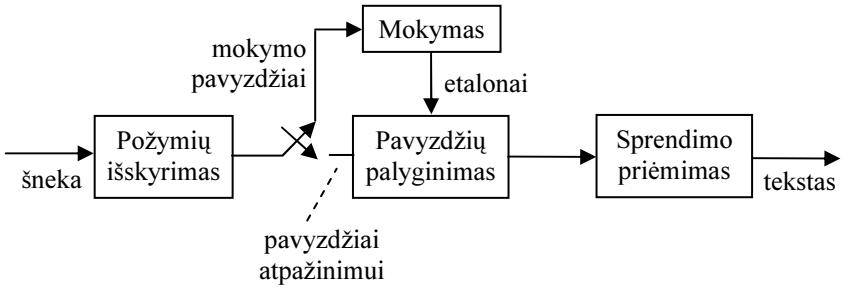
Pavyzdžiais grįsti atpažinimo metodai[7] yra labiausiai paplitę. Galima daryti prielaidą, kad tai yra susiję su mažesne paklaida parenkant svarbiausias akustinio signalo savybes. Kitaip nei akustiniuose-fonetiniuose metoduose, čia sprendimas priimamas automatiškai analizuojant didelius duomenų kiekius ir identifikuojant stabilias ir pasikartojančias savybes. Išskiriami du etapai: mokymo ir atpažinimo. Mokymo procedūros metu vyksta akustinių modelių kūrimo procesas, mokymas. Atpažinimo etape nežinomas ištarimas lyginamas su visais modeliais ar jų kombinacijomis ir atrenkamas geriausias modelis tam tikro atstumo prasme. 3.1 paveiksle pateikiama apibendrinta pavyzdžiais grįsto atpažinimo schema.

Į atpažinimą galima išskirti tris požiūrius, besiskiriančius akustinio modelio formavimo būdu mokymo etape ir algoritmu pavyzdžių palyginimo etape (Rabiner, Juang 1993).

Atpažinimas atliekant nežinomo ištarimo palyginimą su etalonais. Čia etalonai arba akustiniai modeliai yra pakankamai primityvūs, nes jiems atstovauja iš etalonais laikomų šnekos signalų išskirtų *n*-mačių požymių vektorių sekos. Pavyzdžių palyginimo etape šnekos pavyzdys, kurį reikia atpažinti, yra palyginamas su visais akustiniais modeliais atstumo tarp jų skaičiavimo prasme.

[7] Šiame darbe sąvoka *pavyzdys* metodo apibrėžime atitinka *akustinio modelio* sąvoką.

Mažiausias atstumas rodo, kad nežinomas ištarimas yra artimiausias konkrečiam modeliui ir pastarasis gali jį reprezentuoti.



3.1 pav. *Atpažinimo, grįsto pavyzdžiais, schema*

Atpažinimas naudojant paslėptuosius Markovo modelius. Šio tipo atpažinimui naudojami statistiniai šnekos atpažinimo metodai. Akustiniai modeliai yra parametriniai, todėl mokymo metu naudojamos didelės duomenų imtys. Atpažinimo etape vyksta analogiškos pirmajam atpažinimo būdui lyginimo ir sprendimo operacijos, tačiau tikimybių plotmėje. Šio darbo pagrindas irgi yra paslėptieji Markovo modeliai, todėl išsamesnis šio metodo aprašas pateikiamas 3.2 poskyryje.

Atpažinimas naudojant dirbtinius neuroninius tinklus (*artificial neural networks*). Šiuo atveju akustinių modelių nėra, jų funkciją atlieka specialiai apmokytas dirbtinių neuronų tinklas. Tam taip pat reikalingos didelės duomenų imtys.

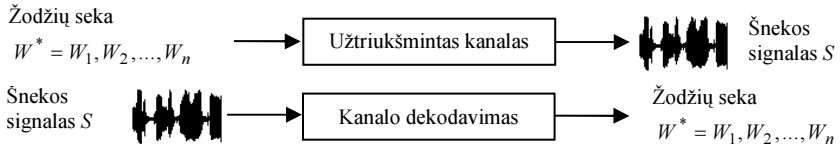
Šie šnekos atpažinimo metodai, tiek atskirai, tiek jungiant kelis požiūrius yra paplitę labiausiai.

3.2. Statistiniai šnekos atpažinimo metodai

Pagrindinės statistinių šnekos atpažinimo metodų savybės yra duomenimis grįstas (*data driven*) šnekos modeliavimas ir apriorinių žinių apie šneką naudojimas.

Statistiniai šnekos atpažinimo metodai formuluojami tikimybių teorijos kategorijomis ir remiasi tikimybine informacijos kilme. Šių šnekos atpažinimo metodų pagrindas yra Bajeso sprendimų teorija ir į šnekos atpažinimą žiūrima kaip į klasifikavimo procesą. Čia sprendimas priimamas remiantis tiek aposteriorinėmis žiniomis, gautomis iš stebėjimo duomenų, tiek apriorinėmis žiniomis apie galimus žodžius (klases). Modeliuojant ar kuriant Bajeso sprendimų teorija grįstą ASA sistemą, reikia įvertinti šiuos abu parametrus – apriorines žodžių tikimybes ir sąlygines klasių tikimybes. Bahl (1983) pateikia

šnekos generavimo ir atpažinimo modelius, vadinamus šaltinio-kanalo (*source-channel*) modeliais (3.2 paveikslas), kurie suformuoja konkretų Bajeso sprendimų teorijos pritaikymo šnekos atpažinimui vaizdą.



3.2 pav. *Šnekos generavimo ir atpažinimo modeliai (pagal Bahl et al. 1983)*

Pagal šiuos modelius generuojant šneką, šaltinis generuoja žodžių seką W^* ir jos atitikmenį – garso signalą *S*. Žodžių seka W^* gali būti iš vieno žodžio. Pašalinių išorės faktorių poveikis, pasireiškiantis garso signalo iškraipymu, vaizduojamas užtriukšmintu kanalu. Šnekos atpažinimas formuluojamas kaip maksimalios aposteriorinės (*maximum a posteriori)* tikimybės radimas – dekodavimas. Šnekos signalas *S* išreiškiamas požymių vektorių seka $\mathbf{O} = \mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_T$, čia *T* priklauso nuo šnekos signalo ilgio. Dekodavimo problema yra žodžių sekos W^* išrinkimo iš visų galimų žodžių sekų W^{**} didžiausios tikimybės prasme problema:

$$W^*_{\max} = \arg\max_{W^* \in W^{**}} P\left(W^* \mid \mathbf{O}\right). \tag{3.1}$$

Performulavimas Bajeso taisyklės pagrindu leidžia tikimybę (3.1) išskaidyti į kelias komponentes:

$$\arg\max_{W^* \in W^{**}} P\left(W^* \mid \mathbf{O}\right) = \arg\max_{W^*} \frac{P\left(\mathbf{O} \mid W^*\right) P\left(W^*\right)}{P\left(\mathbf{O}\right)} \approx \tag{3.2}$$

$$\approx \arg\max_{W^*} P\left(\mathbf{O} \mid W^*\right) P\left(W^*\right). \tag{3.3}$$

Apriorinė akustikos tikimybė $P\left(\mathbf{O}\right)$, esanti išraiškos (3.2) vardiklyje atmetama, nes išraiškos maksimizavimui įtakos neturi ir išlieka konstanta visoms galimoms žodžių sekoms W^{**}. Pagrindine šnekos atpažinimo išraiška tampa išraiška (3.3). $P\left(\mathbf{O} \mid W^*\right)$ žymi aposteriorinę akustinių stebėjimų sekos tikimybę ir vadinama akustiniu procesoriumi. $P\left(W^*\right)$ žymi apriorinę žodžių sekos tikimybę ir vadinama lingvistiniu procesoriumi. Modeliuojant reikia įvertinti abu tuos tikimybinius skirstinius, kurių pirmojo modeliavimas vadinamas akustiniu, o

antrojo – kalbos modeliavimu. Pagrindinė sąlyga, kad ASA sistema veiktų, yra akustinio modeliavimo etapas. Jo metu pagal dideles duomenų imtis formuojamas $P(\mathbf{O}|W^*)$ skirstinys. Šiame darbe to skirstinio šaltinis yra paslėptasis Markovo modelis. Pagal jo parametrus kiekvienam žodžiui skaičiuojama konkreti tikimybė.

3.3. Paslėptasis Markovo modelis

Paslėptasis Markovo modelis modeliuoja atsitiktinius procesus. Atsitiktinis procesas juda būsenų seka, kiekvienoje būsenoje generuodamas kokį nors įvykį. Skiriami:

- Stebimas Markovo modelis. Jis pasižymi tuo, kad pagal įvykių seką galima surasti būsenų seką. Čia viena būsena atitinka vieną įvykį.
- Paslėptasis Markovo modelis. Pagal įvykių seką negalima atstatyti būsenų sekos. Čia įvykis yra tikimybinė būsenos funkcija, nes viena būsena atitinka kelis įvykius, pasikartojančius kitose būsenose.

Daugeliui procesų modeliuoti dėl jų sudėtingumo yra taikomas paslėptasis Markovo modelis. Jis tinka ir šnekos atpažinimo modeliavimui, nes šnekos signalas yra laike kintantis stochastinis (atsitiktinis) procesas.

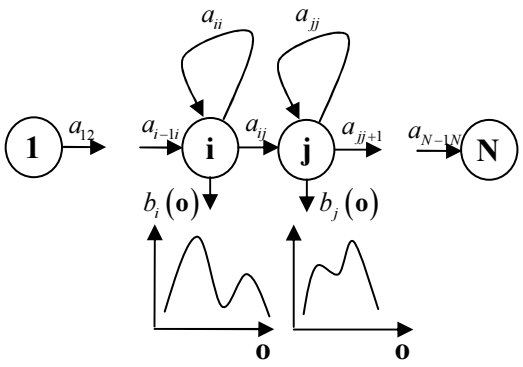
Paslėptasis Markovo modelis (Rabiner 1989) apibrėžiamas kaip baigtinė būsenų, susietų perėjimo tikimybėmis, seka, naudojama signalų laikiniam ir spektriniam kitimui modeliuoti. Paslėptas Markovo modelis aprašomas nusakant penkis dydžius:

- Būsenų $Q=(q_1,q_2,...,q_N)$ skaičių N.
- Skirtingų stebėjimų būsenoje skaičių D – diskrečiu atveju arba stebėjimų $\mathbf{O}=\mathbf{o}_1,\mathbf{o}_2,...,\mathbf{o}_T$ pasiskirstymo tankį – tolydžiu atveju.
- Perėjimo iš vienos būsenos į kitą tikimybių matricą $\mathbf{A}=\{a_{ij}\}$, $i,j=1,...,N$, $\forall i,j \;\; a_{ij}\geq 0$ ir $\forall i \sum_{j=1}^{N} a_{ij}=1$.
- Stebėjimų tikimybinius skirstinius $\mathbf{B}=\{b_j(k)\}$, $k=1,...,D$, $j=1,...,N$ – diskrečiu atveju arba stebėjimų tikimybės tankio funkcijas $\mathbf{B}=\{b_j(\mathbf{o}_t)\}$, $j=1,...,N$, $t=1,...,T$ – tolydžiu atveju.
- Pradinio (*initial*) buvimo būsenoje tikimybes $\pi_i=P(q_1=i)$, $i=1,...,N$. Šnekos atpažinimo modeliavime pradine būsena laikoma pirmoji paslėptojo Markovo modelio būsena $\pi_1=1$.

Šie penki dydžiai nusako konkretų paslėptąjį Markovo modelį, nors tam užtenka nusakyti rinkinį $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$. Kadangi π yra konstanta visiems PMM, naudojamiems šnekai atpažinti, rinkinys paprastėja iki $\lambda = (\mathbf{A}, \mathbf{B})$.

Šnekos atpažinimui naudojama iš kairės į dešinę (*left-to-right*) nukreipta PMM topologija. Čia galimi perėjimai tarp būsenų yra nukreipti į dešinę, į priekį (*forward transition*), galimas ir likimas toje pačioje būsenoje (*self transition*). Struktūrinis PMM ir jį apibrėžiantys parametrai pateikti 3.3 paveiksle.

Teigiant, kad šnekos signalą ar jo dalis gali reprezentuoti PMM, šnekos signalas pradedamas traktuoti kaip būsenų seka. Pačios būsenos nėra stebimos, o paslėptos, bet apie jų seką galima spręsti iš kalbos signalo. Darant prielaidą, kad garsas gali būti skaidomas į tris dalis – pradžią, nekintančią dalį ir pabaigą, garsą įprasta aprašyti trimis būsenomis. Modelio pradžioje ir pabaigoje pridedama po vieną papildomą neemituojančią (*non-emitting*) būseną, kurios sujungia kelis modelius.



3.3 pav. *Struktūrinis paslėptojo Markovo modelio vaizdas*

PMM tampa akustiniu modeliu tik po mokymo, kurio metu pagal stebėjimų duomenis **O** įvertinami modelio parametrai $\lambda = (\mathbf{A}, \mathbf{B})$, t. y. perėjimo iš būsenos į būseną tikimybių matrica ir stebėjimų tikimybės tankio funkcijos kiekvienai modelio būsenai. Ji yra išreiškiama stebėjimų arba požymių vektorių tikimybės tankio funkcija, nusakančia konkretaus požymio vektoriaus tikimybę. Tikimybinis skirstinys gali būti įvairių formų.

Tolydus skirstinys (*continuous mixture density*). Šnekos signalas yra tolydinės kilmės, todėl šnekos atpažinimui naudingiau naudoti šio tipo skirstinius. Šiame darbe buvo naudojami tolydūs tikimybiniai skirstiniai, kuriuos reprezentuoja Gauso mišinio tikimybės tankio funkcijos. Tada požymių

vektoriaus $\mathbf{o}_t, t = 1,...,T$ priklausymo tikimybiniams skirstiniams $b_j(\mathbf{o}_t)$ tikimybės apskaičiuojamos pagal formulę:

$$b_j(\mathbf{o}_t) = \sum_{l=1}^{L} c_{jl} N(\mathbf{o}_t; \boldsymbol{\mu}_{jl}, \Sigma_{jl}), j = 1,...,N, \quad (3.4)$$

čia L – mišinių skaičius, c_{jl} – l-tojo mišinio svoris j-oje būsenoje, $\forall j \sum_{l=1}^{L} c_l = 1$ ir $N(\cdot; \boldsymbol{\mu}, \Sigma)$ – Gauso mišinio tikimybės tankio funkcija su vidurkių vektoriumi $\boldsymbol{\mu}$ ir kovariacijos matrica Σ, t. y.

$$N(\mathbf{o}_t; \boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2} (\mathbf{o} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{o} - \boldsymbol{\mu}) \right), \quad (3.5)$$

n – požymio vektoriaus $\mathbf{o}$ dydis.

Diskretus skirstinys (*discrete density*). Čia reikalinga kvantavimo operacija, kurios metu stebėjimų vektoriai pervedami iš tolydžios į diskrečią erdvę. Dėl kvantavimo atsiranda paklaidų.

Pusiau-tolydaus (*semi-continuous density*) ar susietų mišinių skirstinys (*tied-mixture density*). Kitaip nei auksčiau vardinamieji skirstinių tipai, kurie yra skirtingi modeliams ir atsieti vienas nuo kito, šio tipo skirstiniai gali būti bendri (susieti) visoje modelių erdvėje.

Vietoj tikimybinių mišinių gali būti naudojami neuroniniai tinklai (hibridinėse sistemose), autoregresiniai skirstiniai, diskrečių tikimybių lentelės ir t. t.

3.2 poskyryje nagrinėtas statistinis šnekos atpažinimo metodas. Jame gauta išraiška (3.3) yra pagrindinė šnekos atpažinimo procese. Šioje išraiškoje naudoti abstraktūs žodžių sekos ir žodžių sekų aibės žymenys W^* ir W^{**}. ASA sistemoje žodis ir žodžių aibė pakeičiami jas reprezentuojančia statistinių-tikimybinių žodžių modelių seka M^* (modelio pagrindas – paslėptasis Markovo modelis) ir modelių sekų aibe M^{**}. Tuomet išraišką (3.3) galima transformuoti:

$$\arg\max_{M^*} P(\mathbf{O} \mid M^*) P(M^*). \quad (3.6)$$

Tolesniame aprašyme kalbos modelio reikšmės įtaka šioje išraiškoje nenagrinėjama, todėl laikoma konstanta, o paslėptųjų Markovo modelių sekos $M^* \in M^{**}$ atpažinimas keičiamas nuosekliu kiekvieno modelio $M \in M^*$ atpažinimu. Taigi, akustinis modelis atpažinimo proceso metu turi pateikti stebėjimo duomenų – požymių vektorių sekos tikimybę. Tikimybės

maksimizavimo procesas yra požymių vektorių sekos atitikimo kiekvienam modeliui $M \in M^*$ tikimybės $P(\mathbf{O}|M)$ skaičiavimas ir didžiausios reikšmės išrinkimo procesas. Atskiram modeliui $M \in M^*$ ši tikimybė yra (Rabiner, Juang 1993):

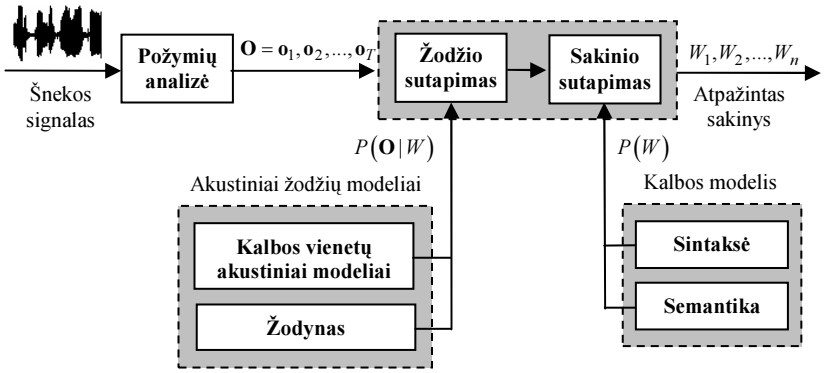
$$\begin{aligned} P(\mathbf{O}|M) &= \sum_{q_1,q_2,\ldots,q_N} P(\mathbf{O}|\boldsymbol{q},M)P(\boldsymbol{q}|M) = \\ &= \sum_{q_1,q_2,\ldots,q_N} \pi_{q_1} \cdot b_{q_1}(\mathbf{o}_1) \cdot a_{q_1q_2} \cdot b_{q_2}(\mathbf{o}_2) \cdot \ldots \cdot a_{q_{N-1}q_N} \cdot b_{q_N}(\mathbf{o}_T). \end{aligned} \tag{3.7}$$

Šios tikimybės skaičiavimas yra sudėtingas, todėl apeinamas taikant tiesioginį-atbulinį (*forward-backward*) algoritmą (Rabiner, Juang 1993).

Čia buvo apibrėžtas PMM, tuo tarpu mokymo ir atpažinimo procesų aprašymas pateikiamas priede C.

3.4. ASA sistemos struktūra

Statistiniais šnekos atpažinimo metodais besiremiančių ir smulkesniais nei žodis kalbos vienetais grįstų ASA sistemų struktūros yra bendros visoms kalboms. Šiame darbe modeliuojamos ASA sistemos struktūra remiasi bendra ASA sistemos struktūra, pateikiama 3.4 paveiksle. Ši struktūra vaizduoja ištisinės šnekos atpažintuvą. Atskirai sakomų žodžių atpažintuvas gaunamas atmetus sakinio sutapimo bloką. Pagrindiniai ASA sistemos elementai yra: požymių analizė, akustinių ir kalbos modelių paruošimas – mokymas, žodžio ir sakinio atpažinimas.



3.4 pav. *ASA sistemos struktūra (Rabiner, Juang 1993)*

Pateiktoje ASA sistemos struktūroje reikia išskirti išteklius – akustinius ir kalbos modelius. Jeigu sistemos struktūra yra bendra visoms kalboms, tai

akustinių ir kalbos modelių turinys yra priklausomi nuo konkrečios kalbos ir jos specifikos. Kiekvienos kalbos ištekliai yra ruošiami atskirai žmonių, turinčių žinių apie tą kalbą. Akustiniai ir kalbos modeliai irgi paruošiami atskirai ir vėliau gali būti naudojami keliose ASA sistemose.

3.4.1. Šnekos analizė ir požymių išskyrimas

Požymių analizės modulio funkcija – šnekos signalą parametrizuoti į požymių vektorių seką išlaikant atpažinimui svarbią informaciją apie ištariamus garsus. Išskirti požymiai turi pasižymėti tam tikromis savybėmis: 1) geru panašių kalbos garsų diskriminatyvumu, 2) patogumu tolesniam naudojimui, 3) statistinėmis savybėmis, invariantiškomis kalbėtojų ir terpių įvairovei, 4) panašumu į žmogaus klausos aparato naudojamus požymius. Skirtingų požymių tipų naudojimas implikuoja faktą, kad požymių tipo, pasižyminčio visomis išvardintomis savybėmis, nėra. Populiariausi parametrinių vektorių tipai yra gaunami iš tiesinės prognozės (*Linear Predictive Coding*), iš kepstrinės analizės – Melų skalės kepstrinių požymių (*Mel-frequency Cepstral Feature*, **MFCC**) ir iš percepcinio tiesinio prognozavimo (*Perceptual Linear Prediction*).

Šnekos signalo dažnių srities analizės (spektrinė analizė) tikslas – iš signalo išskirti požymius. Ji gali būti atliekama naudojant vieną iš šešių spektrinės analizės algoritmų (Picone 1993), kurie kildinami iš Furjė transformacijos (*Fourier Transform*) ir tiesinės prognozės (*Linear Prediction*) spektro analizės metodų. Šie metodai skiriasi šnekos signalo traktavimo būdu. Iš tiesinės prognozės metodo kildinami spektro analizės būdai vadinami parametriniais, nes daro prielaidą, kad šnekos signalas nurodo kokį nors modelį: autoregresijos (*autoregression*), slenkančio vidurkio (*moving average*). Algoritmai, naudojantys Furjė transformaciją, šnekos signalą analizuoja neparametriniu būdu – be išankstinių prielaidų. Labiausiai paplitę yra juostinių filtrų ir kepstriniai spektro analizės būdai. Picone dominuojančiais spektro analizės būdais ir požymiais įvardija kepstrinius spektro analizės būdus ir iš greitosios Furjė transformacijos gautus kepstrinius koeficientus (Picone 1993). Anksčiau dominavę tiesinės prognozės spektro analizės būdai, palyginti su kepstriniais nesugeba iš šnekos signalo išgauti svarbios akustinės informacijos (Davis, Mermelstein 1980).

Visi minėti metodai grindžiami nestacionaraus šnekos signalo išskaidymu į pseudo-stacionarius fiksuoto ilgio 25–30 ms persidengiančius šnekos signalo segmentus (kadrus) $X = x_1, x_2, ..., x_T$ ir požymių vektorių išskyrimu kiekvienam segmentui $\mathbf{O} = \mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_T$. Kiekvienas vektorius $\mathbf{o}_i$, $i = 1, 2, ..., T$ susideda iš pasirinkto skaičiaus požymių vektoriaus reikšmių c_i, $i = 1, 2, ..., n$. Gauti požymių vektoriai vadinami statiniais ir parodo signalo segmento charakteristikas. Tarp segmentų vykstantį charakteristikų kitimą apibūdina

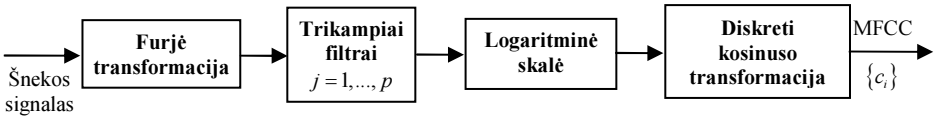
dinaminiai požymiai, vadinami pirmos ir antros eilės išvestinėmis ((3.8) ir (3.9) formulės) arba delta (Δ) požymiais:

$$\Delta c_i = c_{i+1} - c_i,\ i = 1,2,...,n-1, \tag{3.8}$$

$$\Delta\left(\Delta c_i\right) = \Delta c_{i+1} - \Delta c_i,\ i = 1,2,...,n-1. \tag{3.9}$$

Kepstrinių požymių naudojimas kartu su dinaminiais požymiais tapo vadinamuoju *de facto* standartu (Picone 1993), todėl toliau aprašomi dažniausiai taikomi, o taip pat naudoti atliekant tyrimus ir šiame darbe Melų dažnių skalės kepstriniai požymiai.

Melų dažnių skalės kepstriniai požymiai. Pastaruoju metu spektrinėje analizėje naudojamos Melų ir Barkų dažnių skalės. Jos imituoja žmogaus klausos aparato darbe naudojamą dažnių skalę spektrinės informacijos apdorojimui. Iš įprastos dažnių skalės, kurioje garsai matuojami hercais (Hz) prie šios pereinama taikant specialią formulę (pateikiama vėliau). Melų skalėje dažniai pasiskirsto nevienodai, todėl atsiranda du terminai: tiesinė ir netiesinė dažnių skalės. Netiesinių dažnių skalių pagrindu suformavus juostinius filtrus ir apskaičiavus kepstrinius požymius, gaunami Melų ar Barkų skalės kepstriniai požymiai. Naudojant tiesinę dažnių skalę kepstriniai požymiai gali būti gaunami iš tiesinės prognozės koeficientų, naudojant rekurentines išraiškas.



3.5 pav. *Melų dažnių skalės kepstrinių koeficientų skaičiavimo schema*

Šiame darbe buvo naudoti Melų dažnių skalės kepstriniai požymiai, gauti naudojant juostinius filtrus. MFCC apskaičiuojami perėjus 3.5 paveiksle pavaizduotus schemos blokus. Pasirenkami dydžiai yra filtrų skaičius bloke p ir požymių vektoriaus dydis – MFCC skaičius K.

Pirmiausia šnekos signalo segmentui pritaikoma Furjė transformacija (1 blokas) ir gaunamas signalo segmento spektras. Spektras praleidžiamas pro Melo dažnių filtrų bloką, išreiškiamą (2 blokas):

$$Mel\left(f\right) = 2595 \quad \log_{10}\left(1 + \frac{f}{700}\right), \tag{3.10}$$

čia f – dažnis tiesinėje dažnių skalėje. Vėliau spektras logaritmuojamas (3 blokas) ir jam taikoma kosinuso transformacija (4 blokas):

$$c_i = \sum_{j=1}^{p} m_j \cos\left(\frac{\pi i}{p}\left(j-0,5\right)\right),\ \ i=1,...,K\,. \tag{3.11}$$

čia m_j – po spektro logaritmavimo gautos reikšmės ir K – MFCC skaičius. Tokiu būdu gaunamas pasirinkto skaičiaus MFCC. Toliau detalizuojama atskiro požymio vektoriaus struktūra.

MFCC tipo vieno požymio vektoriaus struktūra. Šiame darbe remiantis empiriniais tyrimais (Picone 1993, Laurinčiukaitė 2003) pasirinktas MFCC požymių vektorių tipas.

Minėta, kad kiekvienam 25 ms trukmės šnekos signalo X segmentui $x_1, x_2, ..., x_T$ išskiriamas atitinkamas požymių vektorius $\mathbf{o}_i$, $i=1,2,...,T$. Kiekvienas toks vektorius gali susidėti iš tam tikro skaičiaus pagrindinių ir pasirenkamų komponentų: MFCC požymių, energijos, Δ požymių. Požymių vektoriaus komponentės nurodytos 3.1 lentelėje.

3.1 lentelė. *Darbe naudotų požymių vektorių komponenčių aprašas*

Požymių vektoriaus komponenčių pavadinimai	**Žymenys**
12 MFCC koeficientų	$c_1, c_2, ..., c_{12}$
1 segmento energijos reikšmė	$e, e = \ln\sum_{i=1}^{N} x_i^2$, N – segmento ilgis
12 Δ požymių	$\Delta c_1, \Delta c_2, ..., \Delta c_{12}$
1 segmento energijos reikšmės pokytis	Δe
12 $\Delta(\Delta)$ požymių	$\Delta(\Delta c_1), \Delta(\Delta c_2), ..., \Delta(\Delta c_{12})$
1 segmento energijos reikšmės pokytis	$\Delta(\Delta e)$

Atlikus nuoseklų šių komponenčių apjungimą, gaunamas 39 komponenčių vektorius:

$$\mathbf{o}_i = \left(c_{1_i}, c_{2_i}, ..., c_{12_i}, e_i, \Delta c_{1_i}, \Delta c_{2_i}, ..., \Delta c_{12_i}, \Delta e_i, \Delta\left(\Delta c_{1_i}\right), \Delta\left(\Delta c_{2_i}\right), ..., \Delta\left(\Delta c_{12_i}\right), \Delta\left(\Delta e_i\right)\right),$$

čia $i=1,2,...,T$.

3.4.2. Akustiniai žodžių modeliai

Čia aprašomi 3.4 paveiksle pavaizduoto bloko *Akustiniai žodžių modeliai* elementai: kalbos vienetų akustiniai modeliai ir žodynas.

Kalbos vienetų, skirtų modeliavimui, parinkimas. Jau 2.1 poskyryje buvo apibrėžta kalbos vieneto sąvoka: kalbos vienetais šnekos atpažinimo požiūriu vadinami elementai, iš kurių konstruojamas reikšmę turintis žodis (konkretus žodis taip pat bus vadinamas kalbos vienetu). Skiriamos:

- besiremiančios lingvistiniu kriterijumi aibės (pvz.: fonemos, skiemenys);
- kompiuterio generuotos aibės.

Kalbos vienetų aibės gali būti baigtinės suskaičiuotos (fonemos) ir baigtinės nesuskaičiuotos (žodžiai, skiemenys, kitais kriterijais gauti kalbos signalo segmentai). Iš pastarųjų aibių atrenkami tie elementai, kuriuos galima rasti ASA sistemos kūrimui naudojamuose duomenyse – garsyne.

Parenkant kalbos vienetų tipą ir aibes atsižvelgiama į ASA sistemos taikymo sritį – reikia atpažinti ištisinę šneką ar atskirus žodžius. Tai nulemia ASA sistemos sudėtingumą. Šiuolaikinėse ASA sistemose yra paplitęs kalbos vienetų, atitinkančių fonetines klases, naudojimas. Apriorinė prielaida apie lingvistinių kalbos vienetų akustines charakteristikas nėra taikoma. Akustinės charakteristikos sužinomos mokymo metu ir užkoduojamos lingvistinius kalbos vienetus atitinkančiuose akustiniuose modeliuose. Kadangi lingvistinių kalbos vienetų aibė perdengia visą konkrečios šnekos garsų aibę ir kiekvienas kalbos žodis gali būti ištartas jungiant tų garsų kombinacijas, tai akustiniai atitinkamų kalbos vienetų modeliai leidžia modeliuoti kiekvieną kalbos žodį ASA tikslais.

Pagrindiniais kiekvienos kalbos lingvistiniais vienetais yra laikomos fonemos. Pagal jas kuriami nuo konteksto nepriklausančių fonemų akustiniai modeliai, nes tariama, kad lingvistinis kontekstas šiuose modeliuose neatsispindi. Nuo konteksto priklausančios fonemos, ilgesnės trukmės lingvistiniai kalbos vienetai – skiemenys, žodžiai – alternatyvūs pasirinkimai fonemoms.

Kalbos vienetų akustinis modeliavimas. Kalbos vienetų akustinio modeliavimo sąvoka nusako lingvistinius kalbos vienetus atitinkančių akustinių modelių kūrimo ir mokymo (modelių parametrų įvertinimo) procesus pagal mokymo duomenis – tinkamai paruoštą garsyną, kuriame minėti kalbos vienetai pasirodo pakankamą skaičių kartų. Vienas iš pagrindinių klausimų mokymo procese – mokymo imties klausimas. Mokymo imtis visada yra baigtinė, vadinasi, vieni kalbos vienetai pasikartos rečiau už kitus, todėl mokymo metu gauti akustiniai modeliai gali būti nepakankamai reprezentatyvūs.

Mokymo etapo pradžioje nustatoma akustinio modelio parametrų struktūra. Šiame darbe tai yra PMM (3.3 poskyryje), kurio parametrai įgyja reikšmes mokymo metu pagal pateiktą mokymo medžiagą. Formuojant PMM struktūras pagrindinė informacija:

- stebimų vektorių tipas ir dydis;
- būsenų skaičius;
- mišinio komponenčių skaičius kiekvienai emituojančiai būsenai;
- perėjimo matricos reikšmių struktūra.

Visuose tyrimuose akustinių modelių struktūra išliko tokia pati. Darbe buvo tiriami tolydiniai tikimybiniai skirstiniai, reprezentuojami Gauso mišinio tikimybės tankio funkcijomis (3.3 poskyris).

Žodynas. Žodynas – tai ASA sistemos naudojamų žodžių sąrašas su žodžių transkripcijomis (nuosekliomis kalbos vienetų sekomis). Žodynas sudaromas tik iš apibrėžtos kalbos vienetų aibės elementų. Jame yra ir akustinių modelių deriniai, naudojami žodžiui atpažinti. Žodynas naudojamas tiek mokymo, tiek atpažinimo etapuose. ASA sistemų žodynai formuojami pagal jau esančius skaitmeninio formato kalbos transkripcijų žodynus.

Sakoma, kad žodyne esančios transkripcijos yra nepriklausomos nuo akustinių duomenų, kai jie nedaro įtakos transkripcijų sudarymui. Taip atsiribojama nuo situacijos, kai analizuojant leksinį kintamumą reikia atsižvelgti ne tik į kalbos vienetų aibes, bet ir į atskiro žodžio tarimo ypatumus. Šiame darbe naudoti nuo akustinių duomenų nepriklausantys žodynai.

Vienas iš būdų, kuriais žodyną siekiama padaryti lankstesnį, – žodyno papildymas alternatyviomis žodžių tarimo transkripcijomis, kurių skaičius parodo žodžio tarimo įvairovę.

Atpažįstamų žodžių tinklas. Tinklas apibrėžia žodžių aibę, kurią atpažinimo sistema yra pajėgi atpažinti. Žodyno ir tinklo žodžių aibių dydžiai nesutampa, nes žodyne yra viso garsyno žodžiai. Žodynas naudojamas ir mokymo, ir atpažinimo etapuose. Atpažįstamų žodžių tinkle esančių žodžių skaidiniai kalbos vienetais yra surandami žodyne.

3.4.3. Kalbos modelis

Toliau apžvelgiamas 3.4 paveiksle pavaizduotas blokas *Kalbos modelis*. Šio bloko pagrindinė funkcija – atrinkti optimalų žodžių rinkinį – sakinį, atitinkantį akustiką. ASA vyksta ir be šio bloko elementų (dar įvardinamų, kaip gramatika): sintaksės ir semantikos, tačiau jų integravimas didina šnekos atpažinimo tikslumą. ASA sistemos gramatika išreiškiama įvairiomis formomis: nuo konteksto nepriklausančia gramatika (*context-free grammar*), N-gramų žodžių tikimybėmis (*N-gram word probabilities*), žodžių poromis.

Šiame darbe buvo tirti tik atskirai sakomų žodžių ir ištisinės šnekos atpažinimas, bet ne kalbos modeliavimas. Pirmuoju atveju kalbos modelio naudojimas netikslingas, nes atpažįstami ne sakiniai, bet žodžiai. Atpažįstant ištisinę šneką koncentruotasi į akustinių modelių analizės bloką. Kalbos modelis naudojamas tik tam, kad būtų sukuriama standartinė nuo konteksto nepriklausanti gramatika, kurioje išvardinami ASA sistemos atpažįstami žodžiai.

3.4.4. Mokymo ir atpažinimo etapai

Mokymo etape atliekamas akustinių modelių mokymo-įverčių formavimo procesas. Jo pradžioje jau turi būti galutinai suformuoti akustinių modelių aibės kiekvieno modelio parametrų struktūra su pirminėmis reikšmėmis, žodynas ir papildomi pagalbiniai sąrašai. Mokymo algoritmo esmė pateikiama priede C.

Modelio mokymas yra modelio parametrų, maksimizuojančių stebėjimų sekos tikimybę, paieška. Tai yra iteratyvi maksimalaus tikėtinumo vertinimo procedūra, kurioje naudojamos ėjimo į priekį ir ėjimo atgal procedūros. Mokymo metu kiekviena mokymo sekos komponentė (požymių vektoriai) priskiriama būsenoms. Tiriant vektorių savybes suformuojama kiekvienos būsenos stebėjimų aibė, aprašoma skirstiniu. Iš tikrųjų apmokymas yra kelių pakopų:

- modelio įverčių (perėjimo tikimybių, Gausinių stebėjimų skirstinių) inicializavimo. Jis atliekamas Viterbi algoritmu, kuris iteratyviai ieško požymių vektorių priklausymo vienai ar kitai būsenai didžiausios tikimybės – ieškoma labiausiai tikėtina būsenų seka, atitinkanti mokomąjį pavyzdį taip suformuojant anksčiau minėtų būsenų Gausinių stebėjimų skirstinius;
- tikrojo mokymo – jau esamų įverčių tikslinimo. Jis atliekamas Baum-Welch algoritmu.

Atpažinimo metu daroma prielaida, kad nežinomą ištarimą atitinkanti tiriamų požymių vektorių seka yra gaunama-generuojama PMM sekos. Skaičiuojamos to ištarimo atitikimo akustinių modelių deriniams tikimybės ir labiausiai tikėtina modelių kombinacija identifikuoja ištarimą. Atpažinimo procesas grindžiamas Viterbi algoritmu. Šis algoritmas gali būti vizualizuotas kaip geriausio sprendimo paieška matricoje, kurioje vertikaliai išdėstomos PMM būsenos, o horizontaliai – požymių vektoriai. Šis algoritmas aprašomas priede C.

3.5. Tyrimuose modeliuotų ASA sistemų schemos

Ankstesniame 3.4 poskyryje buvo aptarta bendroji ASA sistemos struktūra ir jos atskiros dalys. Ši struktūra nurodo tik pagrindinius akustinių modelių paruošimo ir atpažinimo procesų principus. Atliekant konkrečius tyrimus, šią struktūrą reikia detalizuoti. Taigi, šiame poskyryje bendrosios ASA sistemos struktūros pagrindu siūlomos kelios konkrečios schemos: izoliuotų žodžių žodžiais, ištisinės šnekos fonemomis ar skiemenimis grįstos ASA sistemų modeliavimo schemos. Šios schemos yra taikomos tyrimams, aprašytiems 5 skyriuje.

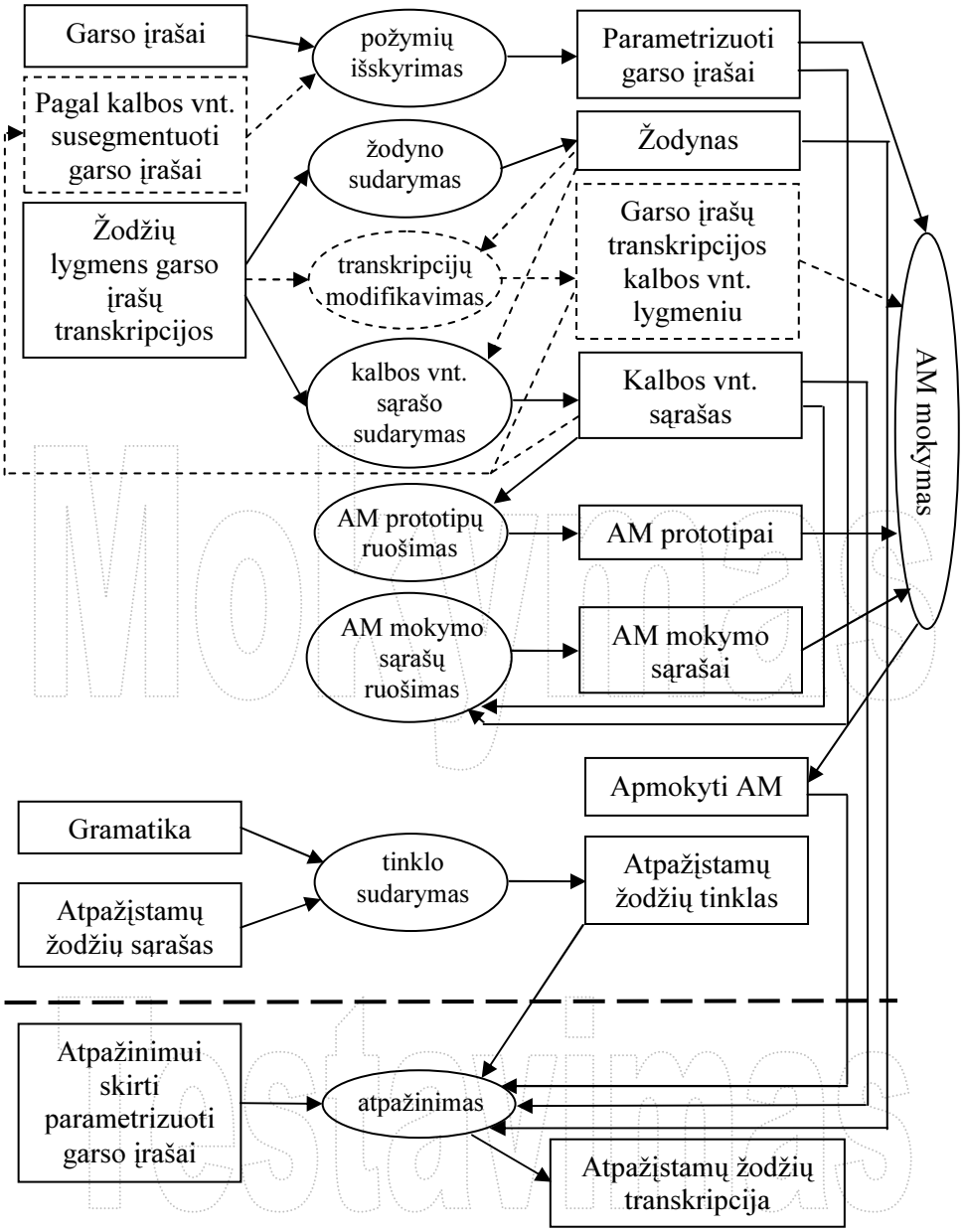
Kadangi tyrimams buvo naudojamas HTK įrankių paketas, tai ir šios schemos yra pritaikytos tam paketui. Pateikiamos schemos yra sudarytos remiantis bendrąja ASA modeliavimo schema, įrankių paketo HTK apraše teikiamomis rekomendacijomis ir akustinio modeliavimo šiuo paketu patirtimi. Šios schemos negali būti laikomos vienintelėmis galimomis akustinio modeliavimo schemomis.

3.5.1. Izoliuotų žodžių atpažinimo sistemos modeliavimas

Izoliuotų žodžių atpažinimui pritaikant paslėptuosius Markovo modelius ir šnekos atpažinimą grindžiant žodžio kalbos vienetu, buvo sudaryta ASA modeliavimo schema, pateikta 3.6 paveiksle (paveiksle stačiakampiai vaizduoja procesams pateikiamus ar po procesų gaunamus objektus, ovalai vaizduoja procesus). Pagal šią schemą pradedant modeliuoti izoliuotų žodžių atpažinimo sistemą, kiekvienas garso įrašas parametrizuojamas *Požymių išskyrimo* bloke, o iš garso įrašų transkripcijų žodžių lygmenyje *Žodyno* ir *Kalbos vienetų sąrašo sudarymo* blokuose išskiriami žodynas ir kalbos vienetų sąrašas. Vėliau pagal kalbos vienetų sąrašą sukuriami juos atitinkančių AM prototipai ir paruošiamos mokymo imtys *AM prototipų* ir *AM mokymo sąrašų ruošimo* blokuose. Pastaraisiais etapais gauti rezultatai kartu su parametrizuotais garso įrašais yra pradiniai duomenys AM mokymui *AM mokymo* bloke. Šio etapo svarbiausias rezultatas – gaunami apmokyti AM. Mokymo etapas pasibaigia atpažįstamų žodžių tinklo sudarymu. Testavimo etape vieninteliame bloke *Atpažinimas* vyksta atpažinimui skirtų parametrizuotų garso įrašų vertimas į žodžių lygmens transkripciją, šiam procesui naudojant mokymo metu gautus atpažįstamų žodžių tinklą, apmokytus AM, kalbos vienetų sąrašą ir žodyną.

Procesų sukeitimas vietomis pateiktoje izoliuotų žodžių atpažinimo sistemos modeliavimo schemoje nėra galimas, nes tolesni procesai galimi tik esant tam tikriems įeigos duomenims. Visus procesus, išskyrus *Žodyno sudarymą, keitimą, AM prototipų ruošimą* ir *Garso įrašų segmentavimą,* realizuoja HTK įrankiai.

Izoliuotų žodžių atpažinimą grindžiant fonemų ar skiemenų kalbos vienetais, galima taikyti dvi ASA sistemos modeliavimo schemas, besiskiriančias požiūriu į mokymo procesą ir reikiamų duomenų mokymui prieinamumu. Kai kalbos vienetų segmentų ribos garso įraše yra nustatytos, taikoma pirmoji schema (viena akustinių modelių inicializavimo ir viena akustinių modelių mokymo iteracijos); kitu atveju – antroji (dėl modeliuojamo segmento ribų neapibrėžtumo vykdomas pasirinktas mokymo iteracijų skaičius, nors negalima garantuoti, kad paskutinės iteracijos metu modelio apmokymui naudotas faktinis modeliuojamo vieneto segmentas). Pirmoji schema vaizduojama tame pačiame 3.6 paveiksle, įvedus papildomas procedūras ir blokus. Šioje schemoje, pirmiausia, reikia žodyne esantiems žodžiams paruošti transkripcijas pasirinktų kalbos vienetų lygmeniu. Naudojant pakeistą žodyną, *Transkripcijų modifikavimo* ir *Kalbos vienetų sąrašo sudarymo blokuose* gaunama garso įrašų transkripcijos kalbos vienetų lygmeniu ir kalbos vienetų sąrašas. Šie objektai naudojami garso įrašų segmentavimui. Antroji schema aprašyta kitame skyrelyje, pristatant ištisinės šnekos fonemomis ar skiemenimis grįstą atpažinimo sistemos modeliavimą.



3.6 pav. *Izoliuotų žodžių atpažinimo sistemos, grįstos žodžių kalbos vienetais, mokymo ir testavimo etapų detalizavimas. Brūkšninės rodyklės vaizduoja papildomų procesų ir objektų atsiradimą mokymo procedūroje modeliuojant skiemenimis ar fonemomis grįstą izoliuotų žodžių atpažinimo sistemą*

Kontekstinių fonemų ir skiemenų naudojimas izoliuotų žodžių atpažinime būtų įmanomas, jei AM mokymo imtys būtų didelės. Tuo atveju reikėtų taikyti kitame skyrelyje pateiktą kontekstinėmis fonemomis ar kontekstiniais skiemenimis grindžiamą ištisinės šnekos atpažinimo sistemos modeliavimo schemą. Šio darbo atveju izoliuotų žodžių garsynas buvo nedidelis, modeliuoti minėtų kalbos vienetų nebuvo įmanoma, todėl atskira kontekstinių fonemų ir skiemenų modeliavimo schema izoliuotų žodžių atveju nepateikiama.

3.5.2. Ištisinės šnekos atpažinimo modeliavimas

Ištisinės šnekos atpažinimui naudojant paslėptuosius Markovo modelius ir atpažinimą grindžiant fonemų ir skiemenų kalbos vienetais, buvo sudaryta ASA modeliavimo schema, vaizduojama 3.7 paveiksle. Lyginant šią schemą su izoliuotų žodžių žodžiais grįstam atpažinimui sudaryta ASA modeliavimo schema (3.6 paveikslas) išskirtinas: 1) sudėtingesnis AM mokymo procesas, atliekant Gausinių mišinių būsenose didinimo 3 iteracijas, kiekvieną kartą atliekant 3 mokymo iteracijas, 2) galimas kalbos vienetų sąrašo tikslinimas, darantis įtaką žodyno ir transkripcijų modifikavimo etapams. Visi kiti esminiai blokai išlieka tapatūs abiems schemoms. Štai trumpas šios schemos aprašymas.

Pirmiausia, kiekvienas garso įrašas parametrizuojamas *Požymių išskyrimo* bloke. *Transkripcijų modifikavimo* bloke, pasinaudojant žodynu, žodžių lygmens garso įrašų transkripcijos virsta pasirinktų kalbos vienetų lygmens transkripcijomis. Reikia pastebėti, kad šiuo atveju daroma prielaida, kad žodynas jau yra paruoštas, t.y. kiekvienas žodyno žodis turi transkripciją pasirinktų kalbos vienetų seka. Žodynas paprastai pateikiamas su garsynu; taip pat jis gali būti kokiu nors būdu modifikuotas. *Kalbos vienetų sąrašo sudarymo* bloke išskiriamas kalbos vienetų sąrašas. Duotame paveiksle objektas *Kalbos vienetų sąrašas* išskirtas kita spalva, nes sąrašą galima modifikuoti, taip keičiant žodyną, garso įrašų transkripcijas ir visą tolesnę sistemos modeliavimo eigą. Ištisinės šnekos skiemenimis grįstame atpažinime *Kalbos vnt. sąrašo* bloke vykdomos tam tikros sąrašo modifikacijos pagal kalbos vienetų atrinkimo metodiką, pristatomą kitame skyrelyje.

Paskui sukuriamas vienas AM prototipas, jis inicializuojamas ir klonuojamas visiems kalbos vienetų sąraše esantiems kalbos vienetams. Mišriu skiemenų ir fonemų kalbos vienetų naudojimo atveju AM gali skirtis būsenų skaičiumi. Tolesniame apmokyme aktualiais tampa du procesai: *AM mišinių skaičiaus didinimas* ir *AM mokymas*. Šie procesai vienas po kito gali būti kartojami neribotą skaičių kartų. Tyrimų metu buvo atliekama po 3 mokymo iteracijas didinant mišinių skaičių iki 4. Mokymo etapas pasibaigia atpažįstamų žodžių tinklo sudarymu. Testavimo etape *Atpažinimas* vyksta atpažinimui skirtų parametrizuotų garso įrašų vertimas transkripcija žodžių lygmeniu, šiam procesui

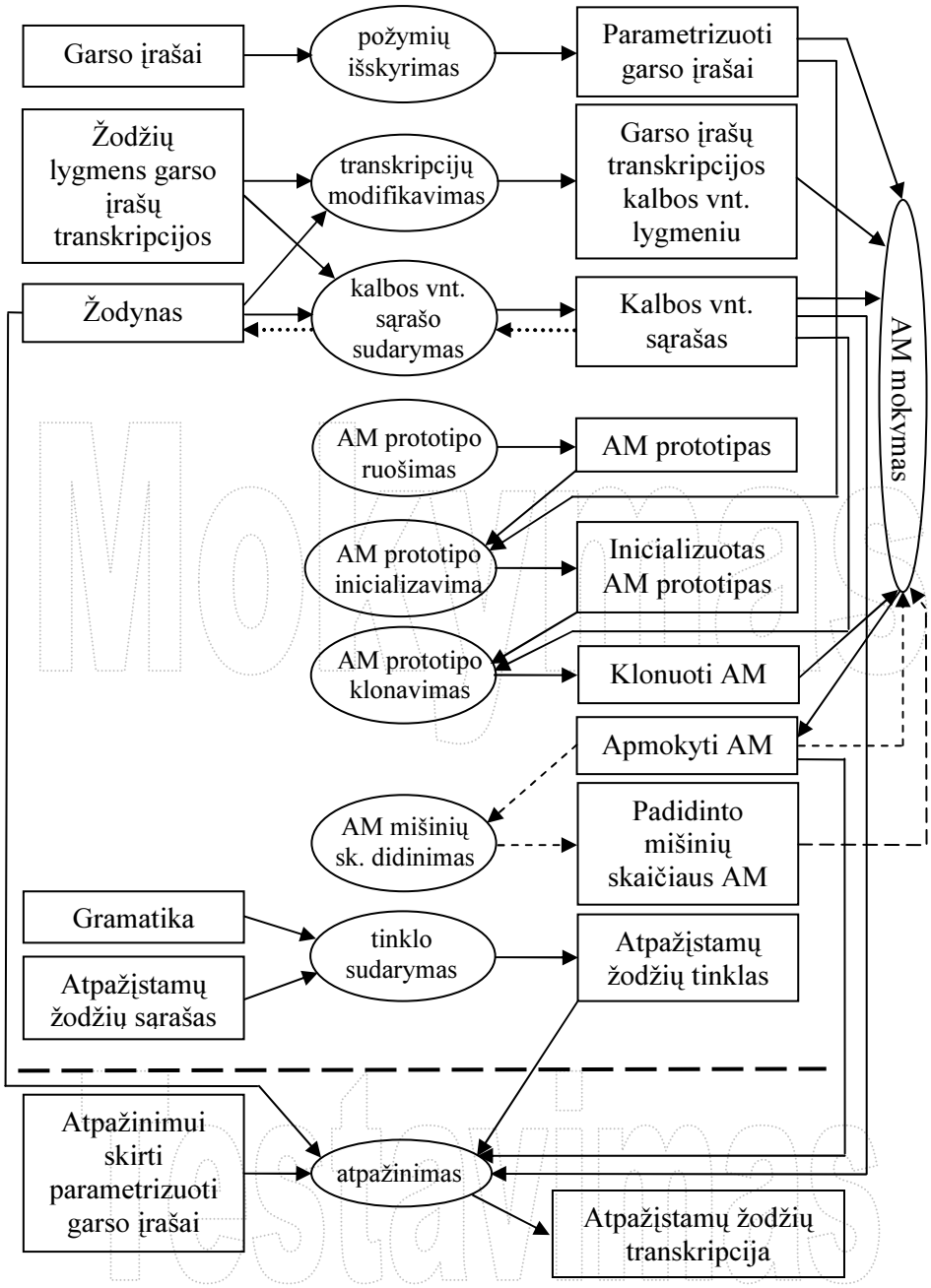
naudojant mokymo metu gautą atpažįstamų žodžių tinklą, apmokytus AM, kalbos vienetų sąrašą ir žodyną.

Pateiktoje ištisinės šnekos atpažinimo sistemos modeliavimo schemoje, procesų sukeitimas vietomis taip pat negalimas, kaip ir 3.6 paveiksle pavaizduotos schemos atveju. Tokie procesai, kaip *Žodyno paruošimas*, *Kalbos vienetų sąrašo modifikavimas*, *AM prototipo ruošimas*, *AM prototipo klonavimas* juos realizuojančių įrankių neturėjo, todėl dalis jų buvo paruošta.

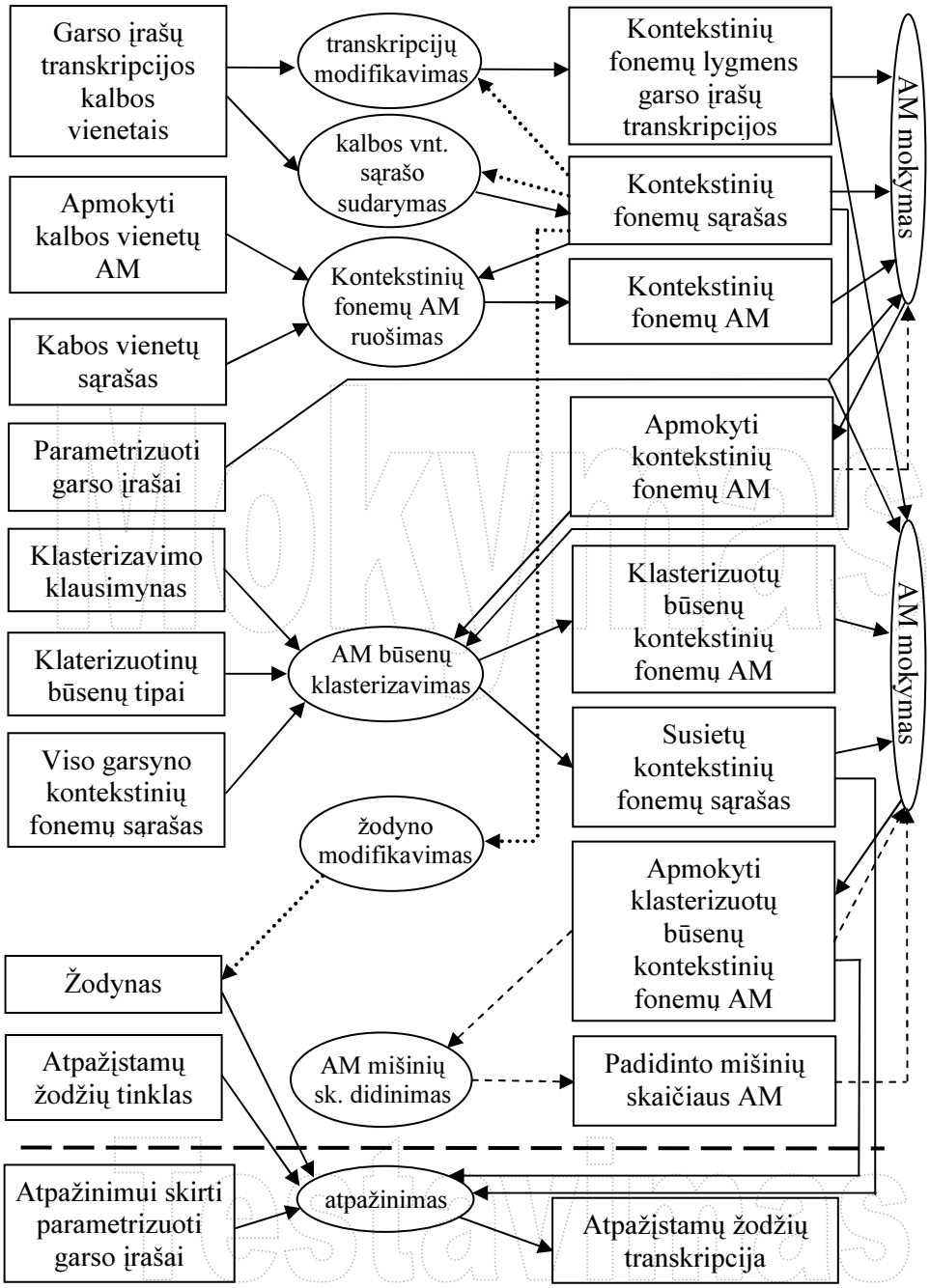
Ištisinės šnekos atpažinimą grindžiant kontekstinėmis fonemomis ir kontekstiniais skiemenimis, sudaryta ASA modeliavimo schema, pavaizduota 3.8 paveiksle. Šią schemą reikia taikyti tik po 3.7 paveiksle pavaizduotos ASA modeliavimo schemos *Mokymo etapo*, turint vieno Gausinio mišinio apmokytus AM. Kontekstinių AM mokymo etapą galima dalinti į dvi dalis, pirmoje suformuojant ir mokant kontekstinius AM iš mokymo duomenų, o antroje dalyje atliekant AM būsenų klasterizavimą ir taip atsižvelgiant į mokymo duomenyse nerastus kontekstinius AM.

Akustinį modeliavimą atliekant pagal šią schemą, garso įrašų transkripcijas fonemų ar skiemenų kalbos vienetų pagrindu reikia keisti į transkripcijas kontekstinių skiemenų ar fonemų pagrindu *Transkripcijų modifikavimo* bloke. Tuo pačiu metu *Kalbos vienetų sąrašo sudarymo* bloke sudaromas kontekstinių fonemų ar skiemenų sąrašas. Jei šis sąrašas papildomai keičiamas, vėl reikia atlikti transkripcijų ir žodyno modifikaciją. Tolesniame *Kontekstinių fonemų AM ruošimo* bloke pagal anksčiau apmokytus kalbos vienetų AM, kalbos vienetų sąrašą ir kontekstinių fonemų ar skiemenų sąrašą paruošiami kontekstinių kalbos vienetų AM. Gautieji AM yra mokomi *AM mokymo* bloke. Nors po šio etapo kiekviena mokymo duomenyse esanti kontekstinė fonema ar skiemuo turi savo AM modelį, tačiau atpažįstamuose duomenyse gali atsirasti nežinoma kontekstinė fonema ar skiemuo.

Dėl šių priežasčių yra atliekamas *AM būsenų klasterizavimas.* Po šio etapo daugelio AM parametrai yra susiejami, tai turėtų padidinti AM tikslumą, nors nežinomos kontekstinės fonemos ar skiemens problema nėra išspręsta (tyrimo metu būsenų klasterizavimas yra atliekamas ir mokymo, ir testavimo duomenims). Po klasterizavimo yra *AM mokymo* ir *AM mišinių skaičiaus didinimo* operacijos, atliekamos norimą skaičių kartų. Tyrimo metu buvo atliekama po 3 mokymo iteracijas didinant mišinių skaičių iki 4. Po mokymo proceso vykstantis *Atpažinimo* procesas atlieka ištisinės šnekos atpažinimą kontekstinių fonemų ar skiemenų pagrindu.



3.7 pav. *Ištisinės šnekos atpažinimo sistemos, grįstos skiemenimis ar fonemomis, mokymo ir testavimo etapų detalizavimas. Brūkšninės rodyklės vaizduoja su AM mokymu susijusių procedūrų galimą kartojimą. Taškinės linijos vaizduoja galimą kalbos vienetų sąrašo modifikavimą ir jo pasekmes*



3.8 pav. *Ištisinės šnekos atpažinimo sistemos, grįstos kontekstiniais skiemenimis ar fonemomis, mokymo ir testavimo etapų detalizavimas. Brūkšninės rodyklės*

vaizduoja su AM mokymu susijusių procedūrų galimą kartojimą. Taškinės linijos vaizduoja kalbos vienetų sąrašo (skiemenų atveju) modifikavimą ir jo pasekmes

Pateiktoje ištisinės šnekos atpažinimo sistemos modeliavimo schemoje, procesų sukeitimas vietomis, kaip ir ankstesnėse schemose, negalimas. Šioje schemoje daroma prielaida, kad būsenų klasterizavimo klausimynas yra žinomas. Iš tikrųjų šį klausimyną reikia parengti kiekvienai kalbai atskirai. Tyrimo metu klasterizavimo klausimynas buvo sukurtas remiantis analogiškais kitų kalbų klausimynais ir atsižvelgiant į lietuvių kalbos specifiką. Kontekstinių skiemenų modeliavimo atveju kontekstinių skiemenų ir fonemų sąrašas buvo modifikuojamas, kas paveikė *Transkripcijų modifikavimą* ir *Žodyno modifikavimą* sukuriant specialius šiuos procesus realizuojančius įrankius.

Pristačius tris detalizuotas ASA sistemų modeliavimo schemas pagal šnekos ir kalbos vienetų tipus, toliau pereinama prie 3.7 paveiksle pateiktos schemos *Kalbos vienetų sąrašo* bloko analizės.

3.5.3. Skiemenų ir fonemų kalbos vienetų aibės formavimo metodika

Metodiką realizuojanti struktūra[8] (Laurinčiukaitė, Lipeika 2007), skirta skiemenų ir fonemų kalbos vienetų aibės formavimui, vaizduojama 3.9 paveiksle. Ši struktūra nusako veiksmus, kurių metu iš suskiemenuoto žodyno suformuojama šnekos atpažinimui skirta baigtinė skiemenų ir fonemų kalbos vienetų aibė ir ją atitinkantis žodynas, vėliau naudojami AM kūrime ir mokyme. Detalus tyrimo aprašymas pateikiamas vėliau; dabar – trumpas kiekvieno metodikos etapo esmės apibūdinimas.

Pirmasis blokas struktūroje yra žodyno skiemenavimas, kuriam atlikti buvo parengta programinė įranga[9].

Antrajame bloke atliekama automatinė suskiemenuotų žodyno dalių korekcija, t. y. atsižvelgiant į gretimus skiemenis koreguojamas priebalsių minkštumas ir skardėjimas, duslėjimas.

Trečiajame bloke iš žodyno išgaunamas jame esančių skiemenų ir fonemų sąrašas. Šiame sąraše gali būti ne visos lietuvių kalboje esančios fonemos, o tik tos, kurios skiemenavimo procese yra laikomos skiemenimis. Galutinai fonemų sąrašas yra papildomas septintame bloke, kuriame suformuojama skiemenų ir fonemų kalbos vienetų aibė. Pagal kalbos vienetus kuriami akustiniai modeliai turi būti reprezentatyvūs, o tai reiškia, kad pavyzdžiui, trijų kalbos vieneto egzempliorių mokymo duomenyse nepakaks reprezentatyviems akustiniams modeliams gauti. Tuo būdu reikia tikrinti sąraše esančių skiemenų ir fonemų

[8] Toliau kalbant apie struktūrą jos blokai bus identifikuojami nurodant bloko numerį, pvz. kokybinio kriterijaus blokas – 6.1 blokas.

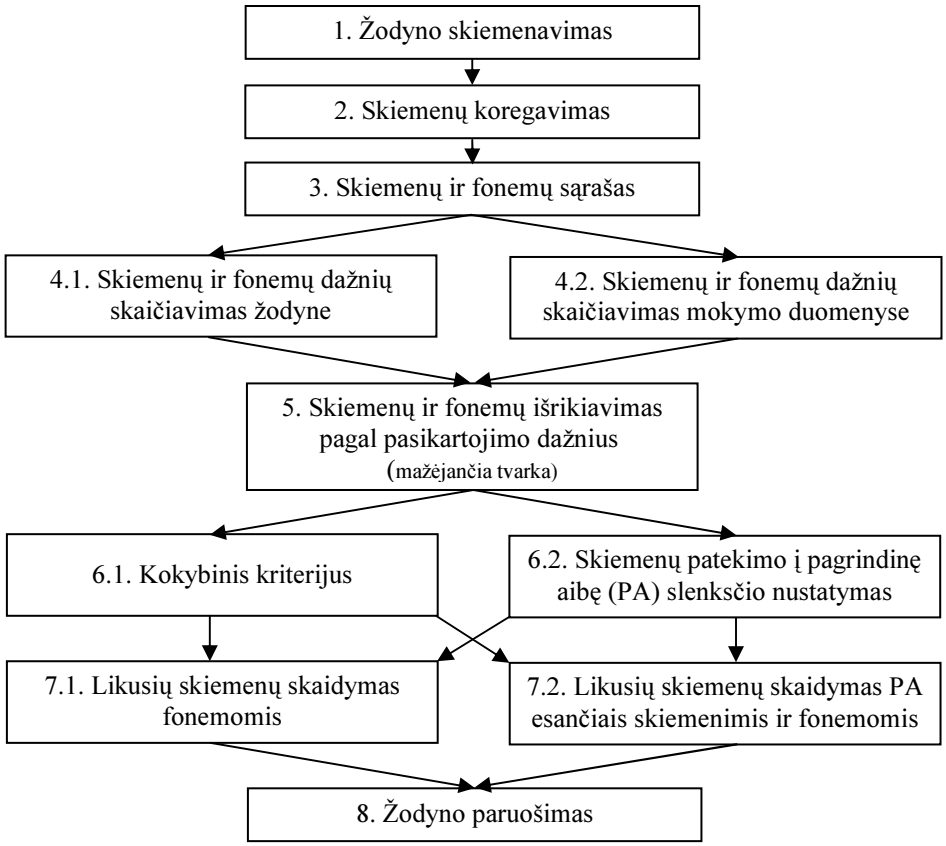
[9] Skiemenavimas atliekamas pagal Kasparaičio pateiktą algoritmą (Kasparaitis 2004).

mokymo imtis. Čia didesnis dėmesys skiriamas skiemenims, nes nepriklausomai nuo fonemų mokymo imčių, akustiniai modeliai joms bus kuriami.

Ketvirtasis blokas skyla į du, kadangi mokymo imtis galima skaičiuoti žodyne arba mokymo duomenyse esantiems kalbos vienetams.

Penktajame bloke sudaromas skiemenų ir fonemų sąrašas su mokymo imčių dydžiais jų mažėjimo tvarka.

Šeštajame bloke reikia pasirinkti kriterijų, pagal kurį skiemenys atrenkami į vadinamąją pagrindinę (PA), bazinę skiemenų ir fonemų aibę. Galimi du kriterijai: 1) kokybinis – atsižvelgti į skiemenų struktūrą, skiemenis sudarančių fonemų specifiką ir 2) kiekybinis – fiksuoti mokymo imties dydį, kurį viršijus kalbos vienetas tampa atrinktu į pagrindinę aibę.



3.9 pav. *Skiemenų ir fonemų kalbos vienetų aibės sudarymo schema*

Septintajame bloke reikia nuspręsti, kaip pasielgti su likusiais skiemenimis, kurie nepateko į pagrindinę aibę. Čia galimi du būdai: 1) likusius skiemenis

skaidyti fonemomis ir pagrindinėje aibėje nesančias fonemas prijungti prie pagrindinės aibės ir 2) likusius skiemenis skaidyti pagrindinėje aibėje esančiais skiemenimis ir fonemomis, naujai atsiradusias fonemas prijungiant prie pagrindinės aibės. Antrasis būdas yra naujas tuo, kad nukrypstama nuo griežto lingvistinių kalbos vienetų apibrėžimo ir pagrindinėje aibėje esantys skiemenys nėra skiemenys griežtąja šio žodžio prasme[10]. Septintajame bloke atliekamas galutinis žodyno sutvarkymas pagal pagrindinėje aibėje esančius kalbos vienetus tuo pačiu naikinant kitus skiemenis jų išskaidymo būdu.

Atlikus visus struktūros nurodytus žingsnius gaunama (pagrindinė) skiemenų ir fonemų kalbos vienetų aibė, kuriai yra kuriami akustiniai modeliai, ir pagal šią aibę sutvarkytas žodynas. Daugelio pateiktos struktūros blokų veiksmams atlikti yra sukurti programiniai įrankiai.

3.6. Šnekos atpažinimo įvertinimo matai

Darbe naudojamas kalbos vienetų akustinio modeliavimo efektyvumas matuojamas šnekos atpažinimo tikslumu. Šnekos atpažinimą įprasta vertinti keliais matais, kurie išreiškiami procentais. Vienas iš jų yra *žodžio atpažinimo teisingumas (*trumpinys – **ZAT**, *angl. – Word Correct Rate*):

$$\mathbf{ZAT} = \frac{\text{teisingai atpažintų žodžių skaičius}}{\text{bendras žodžių skaičius}} \times 100\,\%. \qquad (3.12)$$

Šis matas yra geras vertinant izoliuotų žodžių atpažinimo procesą, bet ištisinės šnekos atpažinime labiau paplitęs *žodžio klaidos (*trumpinys – **ZK**, *angl. – Word Error Rate*) ar *žodžio tikslumo* (trumpinys – **ZT**, *angl. – Word Accuracy*) matai:

$$\mathbf{ZK} = \frac{\text{pakeitimai} + \text{trynimai} + \text{įterpimai}}{\text{bendras žodžių skaičius}} \times 100\ \%, \qquad (3.13)$$

$$\mathbf{ZT} = 100\,\% - \mathbf{ZK}, \qquad (3.14)$$

čia *pakeitimai* – šnekos atpažinimo metu etaloninių žodžių, atpažintų, kaip kiti žodžiai, skaičius; *trynimai* – šnekos atpažinimo metu neatpažintų ir praleistų etaloninių žodžių skaičius; *įterpimai* – šnekos atpažinimo metu tarp etaloninių žodžių įterptų papildomų žodžių skaičius. ZK ir ZT matai leidžia atsižvelgti į

[10] Antruoju būdu gautoje pagrindinėje aibėje esančius skiemenis reikėtų vadinti pseudo-skiemenimis (pseudo-skiemuo – skiemuo, kuris naudojamas kito skiemens skaidymui į dalis), tačiau toliau vartojama skiemens sąvoka.

šnekos atpažinimo metu vykstantį neleistiną žodžių pakeitimą kitu žodžiu, ištrynimą ar papildomų žodžių įterpimą. Pavyzdžiui kuo daugiau žodžių įterpiama, tuo labiau padidėja tikimybė, kad tarp jų bus teisingas žodis – didėja **ZAT**, o ZT žemėja. Realią situaciją šiuo atveju atspindi ZT Tuo būdu ZT ar ZK leidžia objektyviau įvertinti šnekos atpažinimą. Akivaizdu, kad ZT bus dažniausiai žemesnis už **ZAT**.

Šiame darbe vartojama *atpažinimo tikslumo* sąvoka izoliuotų žodžių atpažinimo atveju reikš **ZAT**, o ištisinės šnekos atveju – **ZT**.

Pasikliautinieji intervalai. Siekiant įvertinti šnekos atpažinimo rezultatų patikimumą, rezultatams buvo skaičiuojami pasikliautinieji intervalai su 95 % pasikliovimo tikimybe. Reikia surasti ε, kad $\check{Z}T\text{-}\varepsilon \le \check{Z}T \le \check{Z}T+\varepsilon$.

Darbe buvo taikytos dvi šnekos atpažinimo rezultatų gavimo metodikos: 1) remiantis kryžminio tikrinimo principu iš *n* testinių aibių atpažinimo rezultatų išvedant vidurkį ir 2) atpažįstant vieną testinę aibę. Taikant skirtingas šnekos atpažinimo rezultatų gavimo metodikas, buvo naudojami du pasikliautinųjų intervalų skaičiavimo būdai.

Atpažįstant *n* skirtingų aibių, naudota normaliojo skirstinio parametrų, kai nežinoma dispersija, pasikliautinųjų intervalų radimo formulė. Pagal šią formulę matematinės vilties (*n* atpažinimo rezultatų ZT vidurkio) įverčio tikslumas yra kvantilio $t_{\alpha/2;n-1}$ ir matematinės vilties nepaslinktojo įverčio *s* santykio su kvadratine šaknimi iš imties dydžio *n* sandauga, t. y.:

$$\varepsilon = t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}, \tag{3.15}$$

čia

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(l_i - l_{\text{vid}}\right)^2}{n-1}}, \tag{3.16}$$

čia l_i – *i*-jo matavimo reikšmė, l_{vid} – visų matavimo reikšmių vidurkis. Kvantiliai $t_{\alpha/2;n-1}$, kai α=0,05 randami iš Stjudento skirstinio lentelės.

Atpažįstant vieną testinę aibę, naudota normaliojo skirstinio aproksimavimo binominiu skirstiniu pasikliautinųjų intervalų radimo formulė. Pagal šią formulę ZT įverčio tikslumas yra argumento, kuriam standartinio normaliojo skirstinio $\mathrm{N}(0,1)$ reikšmė lygi duotai pasikliovimo tikimybei, ir standartinio nuokrypio *s* santykio su kvadratine šaknimi iš imties dydžio *n* sandauga, t. y.:

$$\varepsilon = z_{1\text{-}\alpha}\frac{s}{\sqrt{n}}, \tag{3.17}$$

čia

$$s = \sqrt{\check{Z}T\left(1-\check{Z}T\right)}\,. \tag{3.18}$$

3.7. Trečiojo skyriaus rezultatai ir išvados

1. Statistiniais šnekos atpažinimo metodais besiremianti bendra automatinės šnekos atpažinimo sistemos struktūra išskiria akustinių modelių mokymo etapą ir atpažinimo etapą. Akustiniai modeliai kuriami fonetiniams vienetams (fonemoms, skiemenims, žodžiams, kontekstinėms fonemoms, kontekstiniams skiemenims). Akustinių modelių pagrindas yra iš kairės į dešinę nukreiptos modelio topologijos paslėptasis Markovo modelis. Šių modelių turinys priklauso nuo konkrečios kalbos ir jos specifikos, todėl kiekvienai kalbai modeliai yra ruošiami atskirai. Paruošti akustiniai modeliai vėliau gali būti naudojami keliose automatinio šnekos atpažinimo sistemose.
2. Statistiniais šnekos atpažinimo metodais besiremiančių automatinių šnekos atpažinimo sistemų struktūros yra bendros visoms kalboms, tačiau šią struktūrą reikia tikslinti konkretaus empirinio tyrimo metu. Pagal šnekos ir kalbos vienetų tipus buvo pateiktos 3 ASA sistemų modeliavimo schemos:
   - Žodžiais grįsta izoliuotų žodžių atpažinimo sistemos modeliavimo schema, atskiru atveju tinkanti fonemomis ar skiemenimis grįstam izoliuotų žodžių atpažinimui.
   - Fonemomis ar skiemenimis grįsta ištisinės šnekos atpažinimo sistemos modeliavimo schema, atskiru atveju tinkanti fonemomis ar skiemenimis grįstam izoliuotų žodžių atpažinimui.
   - Kontekstinėmis fonemomis ar kontekstiniais skiemenimis grįstos ištisinės šnekos atpažinimo sistemos modeliavimo schema.
3. Pateiktose schemose tam tikroms procedūroms atlikti sukurti įrankiai (AM klonavimo, žodynų ir transkripcijų modifikavimo kontekstinių skiemenų atveju), specialiai paruošti tam tikri objektai (žodynai, AM prototipai, klasterizavimo klausimynas AM būsenų klasterizavimui).
4. Skiemenimis ir fonemomis grįstos ištisinės šnekos atpažinimo sistemos modeliavimo atveju pasiūlyta metodika skiemenų ir fonemų atrinkimui, žodyno sudarymui. Sukurta didelė dalis šią metodiką realizuojančių įrankių.

4

Tyrimuose naudotų lietuvių šnekos garsynų kūrimas

Galima įvardinti kelias priežastis, kodėl garsynai yra tapę svarbūs kuriant ASA sistemas ir kodėl svarbu juos rinkti. Pirmiausia, realiai veikiančių ASA sistemų prototipai yra laboratorijose sukurtos ASA sistemos. ASA sistemos vertinamos pagal tam tikrus kriterijus, kaip šnekos atpažinimo tikslumas, sistemos robastiškumas aplinkos sąlygoms ir kalbėtojams su skirtingais kalbėjimo stiliais, sistemos atpažįstamų žodžių tinklas, sistemos panaudojimo galimybės ir pan. Kaip minėta 2.2 poskyryje, ASA sistemas nusako tam tikros charakteristikos, pagal kurias sistema yra vertinama. Šios charakteristikos tiriamos laboratorijose. Laboratorijose atliekami ASA sistemų modeliavimai būtų neįmanomi, jei nebūtų garsynų – didelės apimties šnekos signalų imčių. Garsynų įvairovė (pagal šnekos tipą, įrašymo sąlygas, kalbėtojų įvairovę) leidžia atlikti išsamesnius tyrimus ir to pagrindu kurti universalesnes ASA sistemas.

Antra garsynų rinkimo priežastis yra daugiau susijusi su statistinius atpažinimo metodus taikančiais tyrėjais. Šiais metodais grįstoms ASA sistemoms būdingas mokymo etapas. Mokymo metu reprezentatyviems akustinių modelių parametrams suformuoti reikalingos didelės mokymo imtys ir tuo pačiu dideli garsynai.

Toliau bus apžvelgta, kokie garsynai yra prieinami šnekos atpažinimo tyrimams Lietuvoje, charakterizuoti šiame darbe naudoti garsynai. Garsynas, prie kurio kūrimo buvo prisidėta, buvo publikuotas (Laurinčiukaitė *et al.* 2006).

4.1. Lietuvių šnekos garsynai

Lietuvoje žmonių grupės, užsiimančios vien lietuvių šnekos garsynų kūrimu, nėra. Tuo tenka rūpintis patiems šnekos tyrėjams. Šiuo metu garsynus renka ir ruošia Matematikos ir Informatikos institutas (MII), Vytauto Didžiojo (VDU), Kauno technologijos (KTU) ir Vilniaus (VU) universitetai. Esantys garsynai skiriasi anotacijos lygiu (vieni anotuoti fonemų lygiu, kiti žodžių ar sakinių), apimtimi (0,5–21 val.), žodyne esančių žodžių skaičiumi (50–32 000) ir kalbėtojų skaičiumi (1–350). Detalesnę informaciją galima rasti (Šilingas *et al.* 2004a).

Garsyną kaip produktą apsprendžia jį ruošianti žmonių grupė. Svarbu, kad į grupę patektų kuo įvairesnės specializacijos žmonių – lingvistų, programuotojų, vartotojų. Tada galima tikėtis, kad garsynas atspindės vartotojų poreikius, atitiks galiojančias kalbos normas ir bus aprūpintas programine įranga, leidžiančia lanksčiai dirbti su garsyne esančiais duomenimis. Išskirtiniu požiūriu į lietuvių šnekos garsinės sistemos ypatumus pasižymi VDU kuriami garsynai Jie anotuojami keliais lygiais, pagrįstai parenkant fonetinių vienetų sistemą (Raškinis *et al.* 2003a). Matematikos ir Informatikos institutas labiau specializuojasi kurti šnekos atpažinimo technologijas ir jas tirti, todėl čia ruošiami garsynai tenkina tik pagrindinius programinių įrankių reikalavimus.

Iki šiol didžiausias dėmesys teko atskirai sakomų lietuvių kalbos žodžių tyrimams, todėl esamos bazės atspindi šiuos poreikius – buvo renkami atskirų žodžių ištarimų įrašai. Šiuo metu jau pradedamos rinkti frazės, pereinama prie ištisinės kalbos rinkimo.

Toliau aprašyti garsynai yra sukurti Matematikos ir Informatikos institute. Tai ištisinės lietuvių kalbos Lietuvos radijo naujienų LRN garsynas ir izoliuotų žodžių garsynas.

4.2. LRN – ištisinės lietuvių kalbos Lietuvos radijo naujienų garsynas

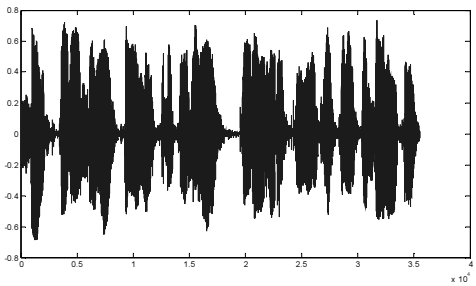
Šis garsynas didele dalimi buvo kuriamas, nuolat tobulinamas ir papildomas šio darbo autorės (Laurinčiukaitė *et al.* 2006). Darbo autorė buvo atsakinga už garsyno kūrimo proceso sudarymą; garsyno struktūros, pagrindinių garsyno charakteristikų ir fonemų sistemos parinkimą; darbo užduočių paruošimą ir paskirstymą darbo grupės nariams (tarp kurių buvo ir pati); atliktų darbų tikrinimą ir garsyno aprašymą.

Darbo rašymo metu garsynas kito kelis kartus išleidžiant LRN0 ir LRN0.1 versijas. Tyrimuose buvo naudota LRN0 versija. Kelių versijų naudojimas neleistų tyrimų palyginti. Garsynai LRN0 ir LRN0.1 skiriasi duomenų kiekiu ir skirtinga garsyno struktūra, bet jie susiję tuo, kad LRN0.1 yra

LRN0 tęsinys. Taigi, toliau pateikiamas garsyno LRN0 kūrimo aprašymas tinka ir garsynui LRN0.1. Ten, kur tarp šių garsynų atsiranda skirtumų, jie pateikiami.

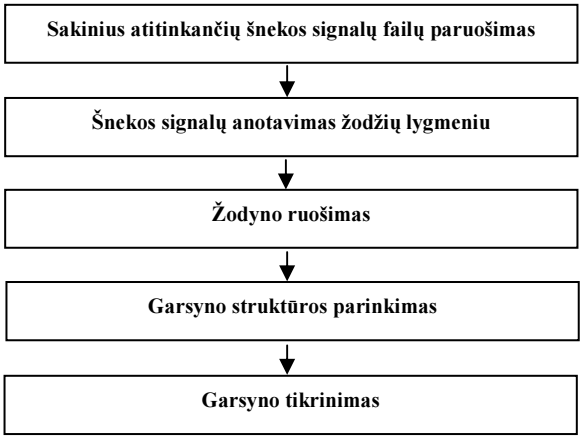
4.2.1. Garsyno kūrimo procesas

Garsynas LRN0 apima 10 valandų šnekos įrašų. Šnekos signalai įrašyti tiesiogiai iš Lietuvos radijo 1 programos (LR1) transliuojamų Lietuvos radijo žinių laidų pagal susitarimą tarp Lietuvos radijo ir televizijos (LRT) bei MII 2003–2004 metais. Šnekos signalų charakteristikos: vieno signalo trukmė – 12 min, diskretizavimo dažnis – 11 kHz, kanalai – mono, kvantavimo skyra – 16 bitų. Įrašyti šnekos signalai yra aukštos kokybės, tarimas aiškus ir teisingas. Į pradinę LRN0 garsyno versiją buvo atrinktas 141 įrašas, kurie rankiniu būdu buvo suskaidyti į sakinius. Iš Lietuvos radijo buvo gauti ir šnekos signalus atitinkantys tekstai, panaudoti sakinius atitinkančių signalų anotacijoms žodžių lygiu. Tekste esantys sutrumpinimai ir skaitmenys buvo perrašomi pilnais žodžiais, sintaksės žymės šalinamos.

RZ10625_18_5.wav – garso įrašas

RZ10625_18_5.txt – anotacija žodžių lygiu
šiuos pareigūnus seimo pritarimu skiria ir atleidžia prezidentas

4.1 pav. *Garso įrašo ir jo anotacinio failo poros pavyzdys*

Paruošus sakinius atitinkančių šnekos signalų ir juos anotuojančių failų poras (4.1 paveiksle pateiktas garso įrašo ir anotacinio failo poros pavyzdys), gautas garsyne esančių žodžių sąrašas. Sukurtas žodžių tarimo fonemų lygiu žodynas. Iš pradžių žodynas buvo kuriamas rankiniu būdu pagal lietuvių kalbos tarimo ir kirčiavimo taisykles. Vėliau sukurtas automatinis transkribavimo įrankis, besiremiantis lietuvių kalbos taisyklėmis ir skaitmenizuotais žodynais. Žodyną sudaro virš 18 000 žodžių. Garsyno versijoje LRN0.1 žodyno apimtis padidėjo. Vėliau tyrimuose panaudotas automatizuotas žodyno transformavimas pagal kitus kalbos vienetų rinkinius.



4.2 pav. *LRN0 garsyno konstravimo etapai*

Paruošus žodyną buvo parenkama garsyno struktūra ir atliekamas garsyno tikrinimas atliekant eksperimentinį tyrimą. Visi minėti garsyno konstravimo etapai pateikti 4.2 paveiksle.

4.2.2. Garsyno charakteristikos

Garso įrašų turinys. Garso įrašų turinys apima svarbiausius vietos ir užsienio politinius, ekonominius, kultūros bei sporto įvykius. Aiškios ribos tarp vietos ir užsienio įvykių pranešimų nubrėžti negalima, bet sporto pranešimai gali būti lengvai atskirti nuo garsyno.

Tam tikrų problemų sukėlė užsienietiško tarimo vardai ir sporto įvykių pranešimai. Garso įrašuose yra nemažai užsienietiško tarimo vardų. Pastarieji kėlė problemų žodyno transkribavimo metu, nes užsienietiško tarimo vardai reikalauja naudoti kitų fonetinių sistemų elementus. Keliuose skirtinguose įrašuose randami keli to paties vardo tarimo būdai. Didelių problemų kėlė ir sporto įvykių pranešimai, nes diktorių kalbėjimo greitis yra per aukštas ir tai sukėlė daugelį neteisingų žodžių ištarimų. Sporto įvykiuose ypatingai dominuoja užsienietiškos tarties vardai.

Garso įrašuose be žodžių yra ir įkvėpimo, tylos, pauzių[11], neteisingo žodžių tarimo. Įkvėpimui, tylai ir pauzei žymėti įvedami trys žymenys, atitinkamai *_įkvepimas*, *_tyla*, *_pauze*. Neteisingi žodžių ištarimai žymimi prieš juos dedant ženklą „_“.

[11] Tyla ir pauzė skiriasi trukme. Tyla dažniausiai yra sakinio pradžioje ir pabaigoje, pauzė – tarp žodžių.

Garsyno struktūra. Abu garsynai buvo padalinti į mokymo (*training*), testavimo-tobulinimo (*development*) ir testavimo-įvertinimo (*evaluation*) duomenų rinkinius. Skirtinguose garsynuose skiriasi šių trijų imčių dydžiai, diktorių parinkimas į skirtingas aibes. Jie pateikti 4.1 lentelėje.

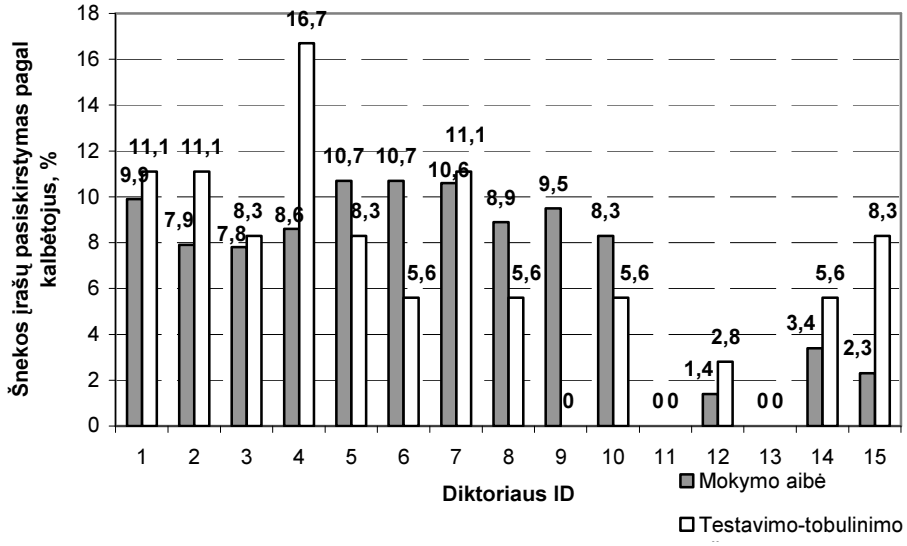
4.1 lentelė. *Garsynų LRN0 ir LRN0.1 mokymo, testavimo-tobulinimo ir testavimo-įvertinimo aibių kiekybinės charakteristikos*

Duomenų aibės pavadinimas	Trukmė (val:min)		Sakinių skaičius		Diktorių skaičius	
	LRN0	**LRN0.1**	**LRN0**	**LRN0.1**	**LRN0**	**LRN0.1**
Mokymo aibė	9:58	15:26	6 566	11 127	23	13
Testavimo-tobulinimo aibė	0:02	0:37	50	736	12	13
Testavimo-įvertinimo aibė	0:16	1:18	360	1 098	13	18
Iš viso	10:16	17:23	6 976	12 961	23	31

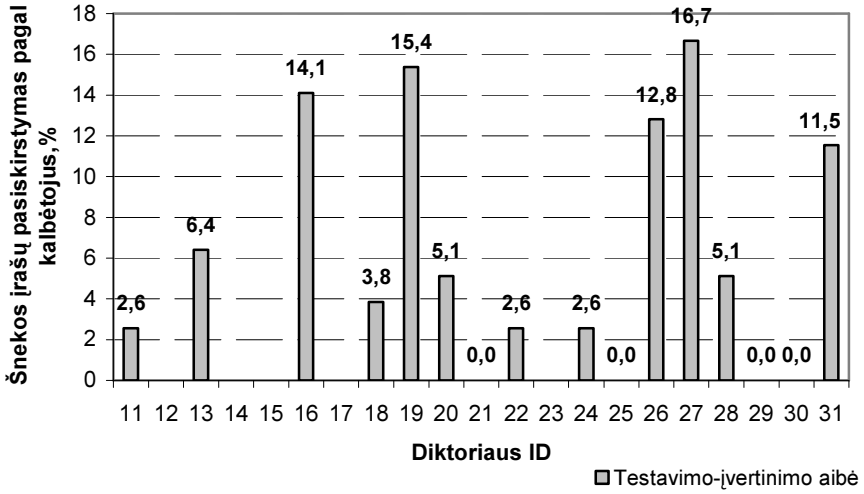
Garsyne LRN0 palyginti su vėlesne jo versija LRN0.1 testavimo duomenų aibių imtys yra mažos, duomenų atrinkimas į skirtingas aibes grindžiamas geresnės kokybės ir trumpesnio signalo atrinkimu į testavimo aibes, nekreipiant dėmesio į diktorių kriterijų.

Garsyno versija LRN0.1 nuo versijos LRN0 skiriasi ~7 valandomis didesne garso įrašų aibe, tolygesniais testavimo imčių dydžiais. Specialus diktorių parinkimas į skirtingas aibes buvo atliekamas siekiant realizuoti šnekos atpažinime nepriklausomumo nuo kalbėtojo idėją, t. y. mokymo, testavimo-tobulinimo ir testavimo-įvertinimo aibės sudarytos iš nepersikertančių kalbėtojų garso įrašų aibių. Garso įrašų pasiskirstymą pagal kalbėtojus trijose duomenų aibėse vaizduoja 4.3 ir 4.4 paveikslai.

Testavimo aibė garsyne LRN0.1 padalinta į dvi beveik lygias dalis. Viena aibė sudaryta atrenkant garso įrašus pagal jų sakinio lygio transkripcijos savybes: transkripcijos ilgumą, minimizuojant asmenvardžių ir neteisingų ištarimų naudojimą. Kita testavimo aibės dalis sudaryta iš likusių garso įrašų, kurių transkripcijoms nebuvo keliami reikalavimai, t. y. šių transkripcijų savybės artimos esančioms mokymo aibėje. Testavimo-tobulinimo aibė LRN0.1 versijoje yra testavimo aibė iš LRN0 versijos. Dar reikia pabrėžti, kad minėtų 10 kalbėtojų įrašai, esantys garsyno LRN0.1 versijoje ir sudarantys 89 % visų garso įrašų, naudojami tik mokymo ir testavimo-tobulinimo tikslams. Likę įrašai naudojami testavimui-įvertinimui.

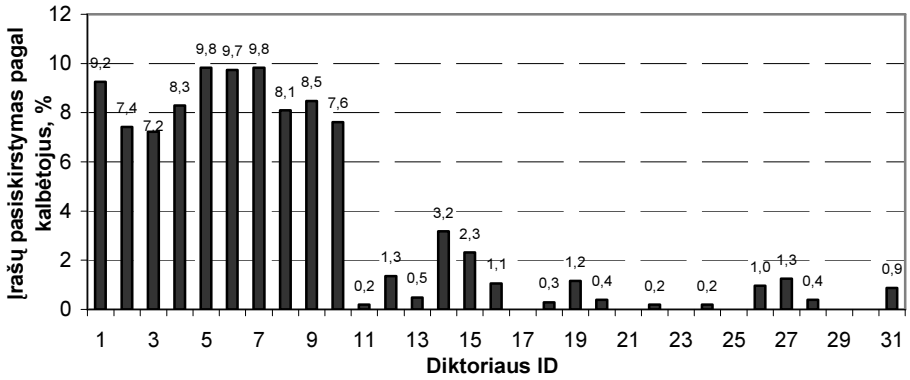


4.3 pav. *Garso įrašų pasiskirstymas pagal kalbėtojus mokymo ir testavimo-tobulinimo aibėse, garsyne LRN0.1*



4.4 pav. *Garso įrašų pasiskirstymas pagal kalbėtojus testavimo-įvertinimo aibėje, garsyne LRN0.1*

Diktorių charakteristikos. Garsyne LRN0 įrašyti 23 abiejų lyčių Lietuvos radijo pranešėjai, garsyne LRN0.1 – 31. Garsyne LRN0.1 pagrindinių dešimties pranešėjų (4 moterys ir 6 vyrai) duomenys sudaro 89 % visų garsyno įrašų. Tai galima matyti 4.5 paveiksle.

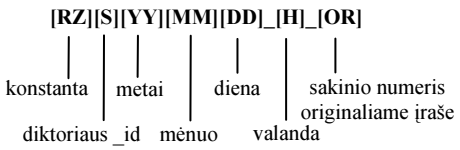


4.5 pav. *Diktorių pasiskirstymas garsyne LRN0.1*

Analogiška situacija yra ir garsyne LRN0, nes abiejose garsyno versijose žinias skaito tie patys diktoriai.

Failų formatai. Skirtingi duomenų tipai garsyne atskiriami pagal failų plėtinius. Plėtinys *.wav* žymi garso įrašą, o plėtinys **.txt* – sakinio anotaciją žodžių lygmeniu. Visi failai, susiję su tuo pačiu ištarimu turi tokį patį pavadinimą. Pavadinimo formatas yra:

[sakinio_id].[praplėtimas], čia **[sakinio_id]** yra formos, pavaizduotos 4.6 paveiksle.



4.6 pav. *Garsyno failo pavadinimo sandara*

Garsyno versija LRN0.1 nėra baigtinė. Kiekvienais metais garsynas papildomas 3–4 garso įrašų valandomis. Ankstesnėmis garsyno versijomis besinaudojantys vartotojai nuo painiavos apsaugomi. Vėliau pridėti duomenys gali būti lengvai atskiriami pagal failo pavadinime esančios konstantos **RZ**

pakitusią dalį **Z**. Kiekvienas naujas papildymas keičia šią dalį abėcėlės tvarka, pradedant nuo **A**.

Fonemų sistema. Daugelis šnekos atpažinimo sistemų yra grįstos fonemų atpažinimu. Tam reikia atlikti garsyno žodyno transkribavimą. Prieš atliekant transkribavimą pasirenkama fonemų aibė. Fonemų aibė garsynui LRN0 buvo pasirinkta atlikus esamų sistemų analizę (Pakerys 2003, Vaitkevičiūtė 2001, Girdenis 2003, Raškinis *et al.* 2003b). Visos fonemų sistemos yra panašios ir skiriasi vos keliomis fonemomis. Kadangi garsynas buvo kuriamas šnekos atpažinimo tikslams, pasinaudota fonetine sistema SAMPA-LT (Raškinis *et al,* 2003b). LRN0 garsyno fonetinė aibė susideda iš 11 balsių (5 trumpi, 6 ilgi), 45 priebalsių (6 poros sprogstamųjų, 4 poros afrikačių, 7 poros pučiamųjų ir 5 poros + 1 sklandžiųjų priebalsių), 7 dvibalsių ir 16 mišriųjų dvigarsių. Tuo būdu naudotoje fonetinėje sistemoje atsispindi priebalsių minkštumas ir balsių ilgumas. Papildoma žymė yra kirtis. Dėl šios žymės naudojimo fonemų ar greičiau fonetinių vienetų skaičius išsiplečia. Išsamus fonetinių vienetų sąrašas su pavyzdžiais pateikiamas priede A.

4.3. Izoliuotų žodžių garsynas

Šis garsynas apima 30 min. garso įrašo. Garsynas sudarytas iš 31 kalbėtojo (18 vyrų ir 13 moterų) 50 skirtingų žodžių, pakartotų 20 kartų, įrašų. Kiekvienas iš 50 žodžių turi 620 tarimo variantų. Kiekvienas žodis įrašytas į atskirą garso failą be tylos žodžio pradžioje ir pabaigoje (žodžiai buvo iškerpami automatiniu žodžių segmentavimo įrankiu WordCut, sukurtu doktorantų M. Skripkausko ir T. Lyguto). Kiekvienas garso įrašas turi jam atitinkantį žodžio lygio anotacijos failą.

Pagrindinės garsyno įrašų charakteristikos: diskretizavimo dažnis – 11 kHz, kanalai – mono, kvantavimo skyra – 16 bitų.

Garsyno žodžiai atrinkti pagal dažninį dabartinės rašomosios lietuvių kalbos žodyną (Grumadienė 1997).

4.4. Ketvirtojo skyriaus rezultatai ir išvados

1. Empiriniais tikslais – siekiant modeliuoti ištisinės šnekos atpažinimą – buvo sukurtos ištisinės šnekos LRN garsyno LRN0 ir LRN0.1 versijos. Darbo autorė buvo atsakinga už: garsyno kūrimo proceso sudarymą; garsyno struktūros, pagrindinių garsyno charakteristikų ir fonemų sistemos parinkimą; darbo užduočių paruošimą ir paskirstymą darbo grupės nariams (tarp kurių buvo ir pati); atliktų darbų tikrinimą ir garsyno aprašymą.

- Garsyną LRN0 sudaro ~10 val. trunkantys garso įrašai su žodžių lygmens anotacijomis. Šalia garsyno pateikiamas žodynas, kuriame kiekvienas žodis transkribuojamas foneminiu lygmeniu.
- Garsyną LRN0.1 sudaro ~17 val. trunkantys garso įrašai su žodžių lygmens anotacijomis. Šalia garsyno pateikiamas ~18 000 žodžių žodynas, kuriame kiekvienas žodis transkribuojamas foneminiu lygmeniu.

2. Empiriniuose tyrimuose naudotas izoliuotų žodžių garsynas jau buvo sukurtas, bet darbo autorės buvo paruoštas naudojimui, atliekant pradinių garso įrašų apdorojimą – šalinant tylą.

5

Akustinio modeliavimo tyrimai ir jų rezultatai

Antrame skyriuje buvo suformuluota akustinio modeliavimo problema, susijusi su kalbos vienetų parinkimu AM kūrimui, t. y. esant didelei kalbos vienetų įvairovei, tyrėjai, atlikdami tyrimus, pasirenka kurį nors vieną kalbos vienetų tipą ir jam kuria AM. Taigi, trūksta tyrimų, susijusių su skirtingų kalbos vienetų panaudojimu šnekai atpažinti naudojant vieną garsyną. Tai atlikus būtų galima teikti rekomendacijas kalbos vienetų parinkimui ir AM aibės sudarymo būdams. Buvo minėta, kad šiame darbe bus nagrinėjami tik lingvistiniu keliu gaunami kabos vienetai. Šis pasirinkimas yra grindžiamas argumentu, kad darbo pradžioje lietuvių šnekai skirta automatinio kalbos vienetų radimo teorija, besiremianti klasterizavimu, dar nebuvo išplėtota.

Taigi, tyrimo tikslas yra pagal skirtingiems kalbos vienetams sukurtus akustinius modelius atlikti lyginamuosius šnekos atpažinimo tyrimus, kurie leistų pasiūlyti įvairių kalbos vienetų akustinio modeliavimo technologijas ir įvertinti kalbos vienetų akustinių modelių efektyvumą ir panaudojimo galimybes. Keliama hipotezė, kad atliekant detalesnį kalbos vienetų parinkimą šnekos akustiniam modeliavimui, galima padidinti šnekos atpažinimo tikslumą. Tam reikia:

- pasirinkti kalbos vienetų tipus;
- kiekvienam kalbos vienetų tipui atlikti išsamią analizę, po to kurti akustinius modelius; nesant nusistovėjusiai kurio nors kalbos vienetų tipo naudojimo šnekai atpažinti schemai ją pasiūlyti;
- išnagrinėti kelis šnekos tipus.

Įgyvendinant minėtus uždavinius buvo pasiūlytos akustinio modeliavimo schemos pagal kalbos vienetų ir šnekos tipus (3.4 poskyris), sukurti ir paruošti garsynai (4 skyrius). ASA sistemų modeliavimas atliekamas naudojant HTK programinę įrangą, papildomai darbui su kalbos vienetais naudojamos Exel terpėje Visual Basic programavimo kalba sukurtos programos.

Šiame skyriuje aprašyti kalbos vienetų naudojimo šnekai atpažinti tyrimai. Tyrimų aprašymas pateiktas jų sudėtingumo didėjimo tvarka. Pradedama nuo paprasčiausio šnekos tipo – izoliuotų žodžių. Nuosekliai išnagrinėjami žodžių, skiemenų ir fonemų kalbos vienetų tipai. Vėliau tie patys kalbos vienetų tipai (išskyrus žodžių) bus nagrinėjami ištisinės šnekos atveju, papildomai nagrinėjant kontekstinės fonemos ir kontekstinio skiemens kalbos vienetų tipus. Pristatomi šie tyrimai:

- izoliuotų žodžių žodžiais grįstas atpažinimas;
- izoliuotų žodžių žodžiais, skiemenimis ar fonemomis grįstas atpažinimas;
- ištisinės šnekos fonemomis grįstas atpažinimas;
- ištisinės šnekos fonemomis ir skiemenimis grįstas atpažinimas.

Šiame skyriuje pristatomi tyrimai ir jų rezultatai buvo publikuoti (Laurinčiukaitė 2004, Šilingas *et al.* 2004b, Šilingas *et al.* 2006, Laurinčiukaitė ir Lipeika 2006, Laurinčiukaitė ir Lipeika 2007).

5.1. Izoliuotų žodžių žodžiais grįsto atpažinimo tyrimai

5.1.1. Izoliuotų žodžių žodžiais grįsto atpažinimo tikslas ir uždaviniai

Visuose iki šiol Lietuvoje atliktuose tyrimuose buvo naudoti paprastesni šnekos atpažinimo metodai, tokie kaip DTW. Svarbu ištirti, kaip atpažinimo rezultatai pasikeičia naudojant sudėtingesnį šnekos atpažinimo metodą. Palyginamieji tyrimų rezultatai nėra įmanomi dėl skirtingų duomenų bazių naudojimo. Orientaciniais rezultatais laikomi rezultatai, gauti naudojant DTW šnekos atpažinimo metodą ir nedidelį garsyną, susidedantį iš 12 žodžių (Lipeika *et al.* 2002). Geriausi ASA sistemos rezultatai – 99,17 % atpažinimo tikslumo nuo kalbėtojo priklausomu šnekos atpažinimo atveju ir 98,06 % – nuo kalbėtojo nepriklausomu šnekos atpažinimo atveju. Tyrimo metu sukurta ASA sistema ir AM aibė. AM kurti tikrinti tokiems įtakos šnekos atpažinimui faktoriams, kaip: priklausomumas/nepriklausomumas nuo kalbėtojo, mokymo ir testavimo imčių dydžių parinkimas. Naudojant skirtingus AM

pastebima šnekos atpažinimo kokybės priklausomybė nuo mokymo etape pasirinktos akustinių modelių formavimo tendencijos. Ši tendencija įžvelgiama ir kituose šio darbo tyrimuose.

Siekiant tikslų reikia sukurti ir testuoti šiuos AM:

- Pagal vieno kalbėtojo (identifikacinis numeris – P1, vyras) duomenis. Atpažinimui pateikiami šie duomenys: to paties kalbėtojo ir keturių pašalinių kalbėtojų šnekos pavyzdžiai (P5, P6 ir P2, P7 – tai du vyrai ir dvi moterys). Šiuos AM naudojanti ASA sistema vadinama priklausoma nuo kalbėtojo ir gauna kodinį pavadinimą – IZ_PNK (trumpinys šifruojamas taip: izoliuoti žodžiai, priklausoma nuo kalbėtojo). Pagal tai, kokio dydžio yra mokymo ir testinės imtys, išskiriamos penkios AM grupės.
- Pagal 25 kalbėtojų duomenis. Atpažinimui pateikiami mokyme nenaudoti penkių kalbėtojų (P1, P2, P5, P6 ir P7) šnekos pavyzdžiai, naudoti IZ_PNK sistemos testavimui. Šiuos AM naudojanti sistema vadinama nuo kalbėtojo nepriklausoma ASA sistema. Kodinis sistemos pavadinimas IZ_NNK (trumpinys šifruojamas taip: izoliuoti žodžiai, nepriklausoma nuo kalbėtojo).

Atpažinimo etape AM vertinami skaičiuojant atpažinimo tikslumą (ZAT).

Siekiant palyginti netiesioginį (garso įrašų) ir tiesioginį (gyvai vykstantį) šnekos atpažinimą, išbandytas gyvai vykstantis šnekos atpažinimas (5 kalbėtojai). Kodinis sistemos pavadinimas – IZ_RS (trumpinys šifruojamas taip: izoliuoti žodžiai, realios sąlygos).

5.1.2. Tyrimo eigos aprašymas

Izoliuotų žodžių žodžiais grįsto atpažinimo tyrimas atliktas pagal 3.5.1 skyrelyje pateiktą izoliuotų žodžių žodžiais grįsto atpažinimo sistemos modeliavimo schemą, nusakančią objektus, kuriais operuojama, veiksmų seką.

Akustinių modelių aibių kūrimas. Pirmoji – IZ_PNK sistema. IZ_PNK sistemos pagrindas – kalbėtojo P1 šnekos pavyzdžiais apmokyta akustinių modelių aibė. Iš pradžių reikėjo nustatyti mokymo ir testinių duomenų imčių, atskirų modelių struktūrą, o tik paskui pradėti modelių mokymo etapą.

Mokymo ir testavimo imtys. Kalbėtojo P1 garso įrašų aibė – 50 sesijų, t. y. kiekvienas iš penkiasdešimties žodžių kartojamas 50 kartų. Siekiant surasti geriausią garso įrašų dalinimo variantą ir kartu patikrinti mokymo duomenų imties dydžio įtaką modelių apmokymo kokybei (kuri tikrinama atpažinimo metu), atlikti penki garso įrašų aibės suskaidymai, kiekvienam suskaidymo būdui kuriant akustinių modelių aibę. Garso įrašų aibės padalinimo būdai ir juos atitinkančios akustinių modelių aibės vaizduojamos 5.1 lentelėje. Lentelė taip pat matyti, kiek žodžio egzempliorių naudojama mokymui ir kiek testavimui.

Kita testinių duomenų imtis buvo sudaryta iš keturių pašalinių kalbėtojų garso įrašų. Kiekvienas kalbėtojas pateikė po 20 sesijų atpažinimui, t. y. po 20 kiekvieno žodžio pavyzdžių kiekvienai iš penkių akustinių modelių aibių. Šių kalbėtojų testavimo imtys vaizduojamos 5.1 lentelėje.

Modelių dydžiai. Atskiri žodžiai yra skirtingo ilgio, todėl sudarant tuos žodžius atitinkančių pirminių modelių šablonus (kurie vėliau mokomi), tikslinga į tai atsižvelgti ir modelio dydį parinkti pagal žodžio ilgį. Žodžio ilgio matavimo vienetu pasirinkta fonema, t.y. jų skaičius žodyje turi nusakyti žodžio ilgį. Akustiniame modelyje fonemos ilgis nusakomas trimis būsenomis. Žodį atitinkančio modelio dydis nustatomas suskaičiuojant fonemas ir jų skaičių padauginant iš trijų.

5.1 lentelė. *IZ_PNK sistemos 5 akustinių modelių aibės*

AM aibės pavadinimas	**Garso įrašų aibės padalinimo būdas (sesijų sk.)**		**Pašalinių kalbėtojų testavimo duomenų imtis (sesijų sk.)**	**Akustinių modelių aibės apibūdinimas**
	Mokymui	**Testavimui**		
AM_10_40	10	40	4×20	Vieno kalbėtojo 50-ies izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 10 mokymo vienetų ir skirti testuoti 40 testinių vienetų.
AM_20_30	20	30	4×20	Vieno kalbėtojo 50-ies izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 20 mokymo vienetų ir skirti testuoti 30 testinių vienetų.
AM_25_25	25	25	4×20	Vieno kalbėtojo 50-ies izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 25 mokymo vienetus ir skirti testuoti 25 testinius vienetus.
AM_30_20	30	20	4×20	Vieno kalbėtojo 50-ies izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 30 mokymo vienetų ir skirti testuoti 20 testinių vienetų.
AM_40_10	40	10	4×20	Vieno kalbėtojo 50-ies izoliuotų žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 40 mokymo vienetų ir skirti testuoti 10 testinių vienetų.

Akustinių modelių aibių kūrimas. Antroji – IZ_NNK sistema. Antrosios – IZ_NNK sistemos pagrindas yra akustiniai modeliai, kurių mokymui naudoti 25 kalbėtojų garso įrašai.

Mokymo ir testavimo imtys. Mokymo imtis apima visą 25 kalbėtojų turimų garso įrašų aibę, t. y. 25 kalbėtojų visos 20 sesijų panaudojamos mokymui. Sudaryta viena akustinių modelių aibė AM_20_20 (25 kalbėtojų 50-ies izoliuotų

žodžių akustiniai modeliai, sukurti modelio mokymui naudojant 20×25 mokymo vienetų ir skirti testuoti 20 testinių vienetų). Testavimo imtį sudarė penkių pašalinių kalbėtojų garso įrašai, t. y. kiekvienam iš penkių kalbėtojų po 20 sesijų (5.2 lentelė).

5.2 lentelė. *IZ_NNK sistemos akustinių modelių aibė*

Akustinių modelių aibės pavadinimas	Garso įrašų aibės padalinimo būdas		Pašalinių kalbėtojų testavimo duomenų imtis (sesijų sk.)
	Mokymui (sesijų sk.)	Testavimui (sesijų sk.)	
AM_20_20	20×25	0	5×20

Modelių dydžiai išliko tapatūs IZ_PNK sistemos atvejui, t. y. modelio dydis parenkamas pagal žodžio ilgį, o mokant modelius naudojamos tikslios žodžių ribos.

Atpažinimo etapas. IZ_PNK sistema. Sistemos IZ_PNK testavimo rezultatai pateikti 5.3 lentelėje. Izoliuotų žodžių atpažinimas vyko dviem etapais:

- testuojami to paties kalbėtojo P1, kurio garso įrašai buvo panaudoti modelių mokymui, garso įrašų pavyzdžiai su skirtingomis testinėmis imtimis (5.3 lentelė, (2) skiltis);
- testuojami keturių kalbėtojų P2, P5, P6, P7 garso įrašai, kiekvieno po 20 sesijų (5.3 lentelė, (4) skiltis). 5.3 lentelės (3) skiltyje pavaizduotas bendras keturių kalbėtojų garso įrašų ZAT.

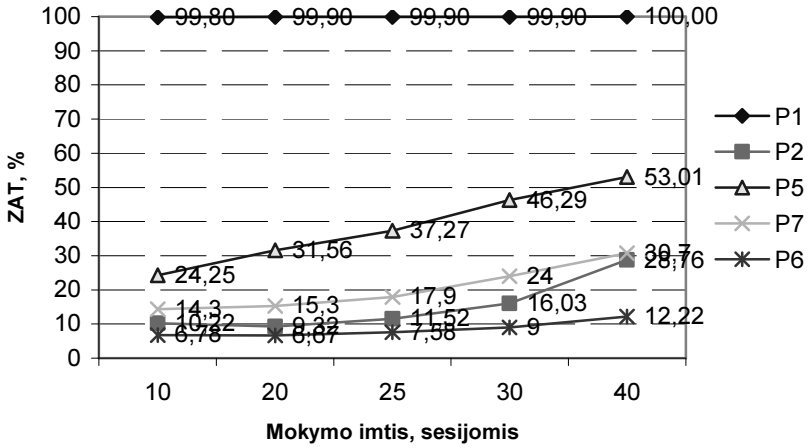
5.3 lentelė. *IZ_PNK sistemos šnekos atpažinimo rezultatai, pateikiant ZAT procentais. Ties kalbėtojo identifikaciniu kodu nurodoma kalbėtojo lytis (m – moteris, v – vyras)*

(1) Akustinių modelių aibės pavadinimas	(2) P1 kalbėtojo ZAT, %	(3) Bendri pašalinių kalbėtojų ZAT, %	(4) Atskiri kalbėtojų ZAT, % (po 20 sesijų)				
			(4a) P1 (v)	(4b) P2 (m)	(4c) P5 (v)	(4d) P6 (v)	(4e) P7 (m)
AM_10_40	99,40 ±0,28	13,89±8,88	99,80	10,22	24,25	6,78	14,30
AM_20_30	99,87 ±0,15	15,71±13,12	99,90	9,32	31,56	6,67	15,30
AM_25_25	99,92 ±0,13	18,57±15,48	99,90	11,52	37,27	7,58	17,90
AM_30_20	99,90 ±0,16	23,83±19,01	99,90	16,03	46,29	9,00	24,00
AM_40_10	100	31,17±19,69	100	28,76	53,01	12,22	30,70

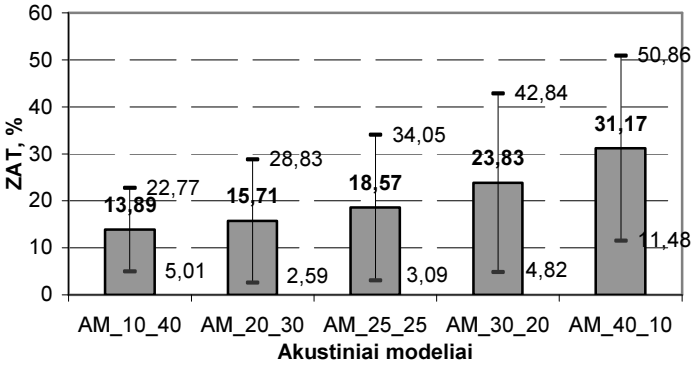
5.1 paveiksle vaizduojami rezultatai tiek vienų, tiek kitų testinių duomenų atveju parodo ZAT priklausomybę nuo mokymo duomenų imties. Antrųjų

testinių duomenų atveju patvirtinamas faktas, kad pašalinių kalbėtojų garso įrašų ZAT, naudojant priklausomą nuo kalbėtojo ASA sistemą, nukenčia. Didinant mokymo imties dydį, ZAT didėja, tačiau net kai P1 kalbėtojo ZAT pasiekia 100 %, bendras keturių likusių kalbėtojų ZAT siekia vos 31,17 %.

Nepastebėtas skirtumas tarp moterų ir vyrų garso įrašų ZAT: vieno vyro garso įrašai atpažįstami geriausiai, o antrojo – blogiausiai.



5.1 pav. *IZ_PNK sistemos ZAT tam pačiam P1 kalbėtojui ir 4 pašaliniams P2, P5, P6, P7*



5.2 pav. *IZ_PNK sistemos bendras ZAT 4 pašaliniams kalbėtojams P2, P5, P6, P7 su 95 % patikimumo intervalais*

Minėta, kad pagal vieno kalbėtojo mokymo duomenis suformuoti akustiniai modeliai buvo panaudoti atpažįstant mokyme nedalyvavusių kalbėtojų duomenis.

Reikia pastebėti, kad tokiu atveju gaunami nenuspėjami, nestabilūs rezultatai, tai rodo pasikliautinųjų intervalų dydžiai 5.2 paveiksle.

Atpažinimo etapas. IZ_NNK sistema. Sistemos IZ_NNK testavimo rezultatai pateikti 5.4 lentelėje. Lyginant tos pačios keturių kalbėtojų testinės garso įrašų imties ZAT pagal dvi skirtingas akustinių modelių aibes, matyti, kad naudojant IZ_NNK sistemą ZAT pakyla. Kalbėtojui P1 ZAT beveik visada (išskyrus sistemos IZ_PNK akustinių modelių aibę AM_10_40) yra didesnis naudojant vien tik jo duomenimis apmokytus akustinius modelius, taigi ir sistemą IZ_PNK.

5.4 lentelė. *IZ_NNK sistemos šnekos atpažinimo rezultatai, pateikiant bendrąjį ir individualius ZAT procentus ir paskaičiuojant 95 % pasikliautinuosius intervalus. Ties kalbėtojo identifikaciniu kodu pateikiama kalbėtojo lytis*

Akustinių modelių aibės pavadinimas	**Bendras kalbėtojų ZAT, %**	**Atskiri kalbėtojų ZAT, % (po 20 sesijų)**				
		P1 (v)	**P2 (m)**	**P5 (v)**	**P6 (v)**	**P7 (m)**
AM_20_20	99,44 ±0,48	99,70	99,60	99,90	99,40	98,60

IZ_RS sistema. Šios sistemos konstravimui naudoti penkių kalbėtojų garso įrašų duomenys – 99 žodžiai. Skirtingi kalbėtojai įrašė skirtingą sesijų skaičių[12], pateiktą 5.5 lentelėje.

5.5 lentelė. *Kalbėtojų P1, P2, P3, P4, P5 įrašų skaičius sesijomis*

Kalbėtojo ID	**Įrašų skaičius (sesijomis)**
P1	20
P2	23
P3	15
P4	14
P5	16

Modeliuojant šią sistemą buvo sukonstruotos trys akustinių modelių aibės, o viena aibė adaptuota naujam kalbėtojui. Šios aibės apibūdinamos 5.6 lentelėje. Akustinių modelių sandara ir mokymo algoritmas yra analogiški IZ_PNK ir IZ_NNK sistemoms. Mokymo imtį sudarė visų turimų garso įrašų aibė. Tikslesnės imtys pateiktos 5.6 lentelėje. Kiekviena akustinių modelių aibė testuota vieno kalbėtojo (P2) realiomis sąlygomis, kiekvieną iš 99 žodžių pakartojant 10 kartų.

[12] Šis tyrimas yra ankstesnis, nei ASA sistemų IZ_PNK ir IZ_NNK modeliavimas, todėl kalbėtojų P1, P2, P3, P4, P5 naudojamų įrašų skaičius šiame ir kituose tyrimuose gali skirtis.

Iš tyrimų rezultatų 5.7 lentelės matyti, kad naudojant visų kalbėtojų duomenimis apmokytą akustinių modelių aibę AM_5k, gaunami geriausi ZAT rezultatai 97,50 ±0,82 %. Atpažinimo rezultatai, gauti naudojant akustinių modelių aibę, apmokytą vieno kalbėtojo duomenimis AM_1k, turėjo būti geresni (realiai buvo atpažinta 92,20 ±1,40 %). Darant prielaidą, kad apmokant šiuos žodžius atitinkančius modelius pritrūko mokymo duomenų (modeliai apmokomi 23 ištarimais), mokymo imtis buvo padvigubinta iki 46-ių ištarimų atrinktiems vienuolikai žodžių, kurių atpažinimas buvo blogiausias. Buvo sukurta nauja AM aibė AM_1k+, kurioje kai kurie žodžiai turėjo didesnes mokymo imtis. Atlikus pakartotinį atpažinimą gauta, kad klaidų sumažėjo, ZAT buvo 97,60 ± 0,80 %.

5.6 lentelė. *IZ_RS sistemos akustinių modelių aibių apibūdinimas*

Akustinių modelių aibės pavadinimas	Apibūdinimas	Mokymo imtis, sesijomis	Testavimo imtis, sesijomis
AM_1k	Apmokyta vieno kalbėtojo (P2) duomenimis	23	10
AM_5k	Apmokyta penkių kalbėtojų (P1, P2, P3, P4, P5) duomenimis	88	10
AM_4k	Apmokyta keturių (P1, P3, P4, P5) kalbėtojų duomenimis	65	10
AM_4k_ad	Adaptuota AM_4k_	65+	10

Lyginant atpažinimo tyrimų rezultatus, gautus dviem AM aibėms: apmokytai keturių kalbėtojų duomenimis be testuotojo duomenų AM_4k, ir naujajam kalbėtojui adaptuotai keturių kalbėtojų aibei AM_4k_ad, kaip ir buvo tikėtasi, AM_4k_ad atpažinimo tyrimų rezultatai buvo geresni, t. y. pagerėjo nuo 87 ±1,76 % iki 92,10 ±1,41 %.

5.7 lentelė. *Atpažinimo realiomis sąlygomis tyrimų rezultatai, skaičiuojant 95 % pasikliautinuosius intervalus*

Akustinių modelių aibės pavadinimas	ZAT, %
AM_1k	92,20 ±1,40
AM_1k +	97,60 ±0,80
AM_5k	97,50 ±0,82
AM_4k	87,00 ±1,76
AM_4k_ad	92,10 ±1,41

Pagrindinė kliūtis, su kuria susiduriama tokio tipo atpažinime, – atpažinimo terpės nepastovumas. Kadangi esančios sąlygos kalbos/tylos detektoriumi matuojamos tik vieną sykį iš pat pradžių, tai vėliau atsiradę nauji nuolatiniai (mašinos, oro kondicionieriaus ūžimas); staigūs, spontaniški (durų trinktelėjimas, telefono skambutis) ar paties kalbėtojo sukelti (užkimimas) trikdžiai daro įtaką

atpažinimui. Jei nuolatinį triukšmą galima eliminuoti dar kartą tikslinant aplinkos sąlygas kalbos/tylos detektoriumi, tai momentinio triukšmo taip lengvai „apeiti“ negalima. Pasaulinėje praktikoje tokie trikdžiai ne „apeinami“, bet specialiai ieškomi, konstruojant jiems skirtus specialius modelius.

Teoriniai šaltiniai pateikia dar vieną svarbų faktą – atpažinimą turėtų lemti garso plokštės ir mikrofono charakteristikos. Tačiau šia kryptimi tyrimai nebuvo atliekami.

5.1.3. Pirmojo tyrimo išvados ir rezultatai

Ištyrus izoliuotų žodžių atpažinimą, konstruojant akustinius modelius žodžiams ir modeliuojant dvi šnekos atpažinimo sistemas – priklausomą nuo kalbėtojo (IZ_PNK), ir nepriklausomą nuo kalbėtojo (IZ_NNK) – daromos šios išvados:

1. Pagal vieno kalbėtojo duomenis sukurti akustiniai modeliai turi būti naudojami tik to kalbėtojo šnekos atpažinimui (galima pasiekti iki 100 % šnekos atpažinimo tikslumą). Atpažįstant pašalinių kalbėtojų šneką gaunami nestabilūs, nepatikimi ir žemi rezultatai, iki 31 ±19,7 % (pateikiamas keturių kalbėtojų gautų rezultatų vidurkis).
2. Pagal daugelio kalbėtojų duomenis (tirti 25 kalbėtojai) sukurti akustiniai modeliai gali būti naudojami įvairių kalbėtojų šnekos atpažinimui (pasiekiant 99 ±0,5 % šnekos atpažinimo tikslumą).
3. Šnekos atpažinimo tikslumui įtakos turi modelių mokymo aibės dydis. IZ_PNK sistemos atveju didėjant mokymo imčiai atpažinimo tikslumas didėja nuo 99 ±0,3 % iki 100 % kalbėtojo P1 atveju ir nuo 14 ±8,9 % iki 31 ±19,7 % keturių pašalinių kalbėtojų atveju (pateikiamas keturių kalbėtojų gautų rezultatų vidurkis).
4. Gyvos šnekos atpažinimo tikslumas priklauso nuo atpažinimo terpės stabilumo.
5. Prieš kuriant akustinius modelius žodžio kalbos vienetų tipui rekomenduojama:
   - Apsibrėžti kalbėtojų skaičių, kurių šneką siekiama atpažinti, ir pagal tai parinkti akustinių modelių imties turinį.
   - Parenkant akustinių modelių mokymo aibės dydį reikia atsižvelgti į akustinių modelių aibės dydį, kuris atliekant tyrimą didėjo nuo 20 iki 30 mokymo egzempliorių vienam akustiniam modeliui apmokyti, akustinių modelių aibės dydžiui didėjant nuo 50 iki 99 elementų.
   - Akustinių modelių mokymo aibės dydis turi didėti panašiai skambančių žodžių akustiniams modeliams

5.2. Izoliuotų žodžių žodžiais, skiemenimis ar fonemomis grįsto atpažinimo tyrimai

Šiame etape toliau tirtas izoliuotų žodžių atpažinimas, bet grįstas atskirai fonemomis, skiemenimis, žodžiais. Minėtų kalbos vienetų tipų naudojimas buvo tiriamas plačiau, bet prieš pradedant didesnės apimties tyrimus buvo atliktas preliminarus lyginamasis tyrimas. Tyrimo rezultatai yra paskelbti (Laurinčiukaitė 2004).

Naudotas 31 kalbėtojo izoliuotų žodžių garsynas, aprašytas garsynų dalyje.

5.2.1. Izoliuotų žodžių žodžiais, skiemenimis ar fonemomis grįsto atpažinimo tikslas ir uždaviniai

Pagrindinis tikslas – trijų kalbos vienetų tipų palyginimas konstruojant AM ir AM naudojant šnekai atpažinti. Modeliuojama nepriklausoma nuo kalbėtojų ASA sistema izoliuotų žodžių atpažinimui žodžių, skiemenų ir fonemų pagrindu. Šios sistemos kodinis pavadinimas IZ_NNK_ZSF (izoliuoti žodžiai, nepriklausoma nuo kalbėtojo, žodžiai-skiemenys-fonemos). Iš tikrųjų sukurta ir testuota 30 akustinių modelių aibių, kiekvienam kalbos vienetų tipui po 10. Dėl paprastumo IZ_NNK_ZSF suskaidoma į tris ASA sistemas: IZ_NNK_Z, IZ_NNK_S ir IZ_NNK_F, kurių pagrindas – skirtingiems kalbos vienetų tipams sukurti AM. Vėliau kiekvienos iš sistemų 10 rezultatų redukuojami į tris pagrindinius skaičius, rodančius skirtingų segmentinių vienetų efektyvumą šnekos atpažinimo procese. Šie rezultatai gali būti laikomi preliminariais rezultatais, leidžiančiais prognozuoti tolimesnių sudėtingų tyrimų baigtį.

Tyrimui pasirinkta kryžminio tikrinimo (*cross-validation*) struktūra, kai vieno kalbėtojo duomenys naudojami testavimui, o likę duomenys – mokymui. Tuo būdu kiekvienam kalbėtojui sukuriama po akustinių modelių aibę. Toks atpažinimo būdas ne tik leidžia gauti bendrą šnekos atpažinimo rezultatą, bet ir leidžia palyginti atskirų kalbėtojų šnekos atpažinimą.

Tyrime išskiriami mokymo ir testavimo etapai. Mokymo etape siekiama sukurti po tris akustinių modelių aibes (reprezentuojančias skirtingus segmentinius vienetus: fonemas, skiemenis ir žodžius) kiekvienam iš atrinktų kalbėtojų. Kiekvienai aibei nustatyti modelių parametrai. Uždaviniai:

- išskirti mokymo ir testavimo aibes nepriklausomų nuo kalbėtojo akustinių modelių kūrimui ir testavimui, paruošti pagalbinius sąrašus mokymo procesui;
- apibrėžti žodžių, skiemenų ir fonemų segmentinių vienetų aibes;
- parinkti kiekvienam iš trijų segmentinių tipų AM parametrus;
- paruošti kiekvieno segmentinio tipo modelių mokymo procedūrą;
- kiekvienam kalbėtojui parinkti AM mokymo ir testavimo imtis;

- atlikti mokymą, testavimą.

5.2.2. Tyrimo eiga

Tyrime AM kurti žodžiams, skiemenimis ir fonemoms, todėl buvo taikytos dvi ASA sistemos modeliavimo schemos. Pirmoji schema, skirta darbui su žodžių AM, yra pateikiama 3.5.1 skyrelyje, o schema, skirta darbui su skiemenų ir fonemų AM, yra pateikta 3.5.2 skyrelyje.

Modeliuojant nepriklausomą nuo kalbėtojo ASA sistemą, 30 kalbėtojų garso įrašų duomenys buvo naudojami mokymui, o likusio vieno – testavimui. Ši iteracija pakartota dešimčiai kalbėtojų. Kiekvienam iš jų daromi atskiri tyrimai trijų tipų akustinių modelių aibių sukūrimui ir tų modelių testavimui šnekai atpažinti.

Mokymo ir testavimo imčių skyrimas. Tyrime naudotas 31 kalbėtojo garsynas, kuriame kiekvienas kalbėtojas turi po 20 penkiasdešimties žodžių ištarimo sesijų. Pagal kryžminio tikrinimo strategiją, vieno kalbėtojo duomenis paliekant testavimui, o likusius naudojant mokymui, mokymo ir testavimo imtys yra tos pačios – $30\,kalbėtojų \times 20\,sesijų$ mokyti ir 20 sesijų testuoti. Tai tik garsyno skyrimas į mokymo ir testavimo imtis. Iš tikrųjų kiekvienas segmentinis vienetas turi vis kitą mokymo imtį. Kalbant apie tokius segmentinius vienetus kaip skiemuo ir fonema, juos atitinkančių modelių mokymo imtys didėja dėl to paties segmentinio vieneto pasikartojimo kitame žodyje. Dėl šios priežasties kiekvienas iš šių segmentinių vienetų turi tik jam būdingą mokymo imties dydį. Žodžių segmentinių vienetų mokymo imtis visiems žodžiams yra vienoda ir sutampa su garsyno skyrimu į mokymo ir testavimo dalis.

Žodžio, skiemens ir fonemos akustinių modelių parametrų parinkimas ir mokymo procesas. Žodžiui skirto akustinio modelio struktūra yra analogiška naudotai IZ_PNK sistemoje: būsenų modelyje yra $3 \times$ *fonemų skaičius žodyne* ir vienas Gausinis mišinys būsenoje. Mokymo procesas taip pat analogiškas – dvi mokymo iteracijos.

Skiemeniui skirto akustinio modelio struktūra: devynios būsenos ir vienas Gausinis mišinys būsenoje. Fonemai skirto akustinio modelio struktūra: penkios būsenos ir vienas Gausinis mišinys būsenoje. Šie segmentiniai vienetai yra panašūs tuo, kad tikslios jų ribos žodyje nėra žinomos. Tuo būdu naudojamas kitas – iteratyvus algoritmas modelių mokymui.

Žodyno analizė. Naudotas garsynas turėjo 50 žodžių žodyną. Jis buvo analizuojamas kiekvienam segmentinių vienetų tipui, surandant segmentinių vienetų aibes, pagal kurias turi būti kuriami AM, ir tikrinant kiekvieno kuriamo AM mokymo imties dydį.

Kadangi garsyne buvo 50-ies žodžių garso įrašai, tai žodžių segmentinių vienetų irgi 50. Skiemenų ir fonemų aibės buvo gautos atlikus garsyno žodyno

žodžių transkribavimą skiemenimis ir fonemomis. Fonemų aibė neperdengė lietuvių kalbos fonemų aibės (kurioje yra 56 fonemos) dėl garsyno žodyno dydžio. Nustačius, kiek kiekvienoje iš trijų segmentinių vienetų aibių yra elementų, pagal mokymo ir testavimo imties dydžius nustatyti minimalūs ir maksimalūs akustinių modelių mokymo imčių dydžiai. 5.8 lentelėje pateikiami skirtingų segmentinių vienetų akustinių modelių aibių dydžiai ir jų elementų mokymo imtys.

5.8 lentelė. *ASA sistemos IZ_NNK_ZSF akustinių modelių mokymo imtys*

AM aibės pavadinimas	Aibės dydis (elementų skaičius)	Elementų mokymo imties dydis (vienetais)
Žodžių AM	50	600
Skiemenų AM	91	600 – 6 600
Fonemų AM	47	600 – 14 400

Iš 5.8 lentelės matyti, kad akustinių modelių aibės pagal jos dydį išsirikiavo tokia tvarka: fonemų, žodžių, skiemenų AM. Nors fonemų akustinių modelių aibės dydis yra mažiausias, tačiau kai kurie modeliai turi dideles mokymo imtis.

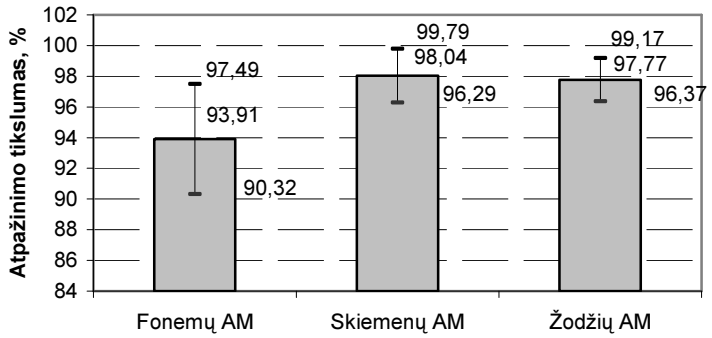
5.9 lentelė. *Individualūs ir bendri kiekvieno kalbėtojo ASA sistemos IZ_NNK_ZSF atpažinimo ZAT rezultatai, paskaičiuojant 95 % pasikliautinuosius intervalus*

Kalbėtojo kodas	Lytis	Fonemų AM ZAT, %	Skiemenų AM, ZAT, %	Žodžių AM, ZAT, %
D01	*m*	79,20 ±2,11	90,30 ±1,54	94,80 ±1,15
D03	*v*	97,20 ±0,86	98,90 ±0,54	98,00 ±0,73
D04	*m*	87,39 ±1,73	95,60 ±1,07	92,20 ±1,39
D06	*m*	97,80 ±0,76	99,60 ±0,33	98,80 ±0,57
D07	*m*	93,90 ±1,24	99,70 ±0,28	98,50 ±0,63
P1	*v*	94,00 ±1,24	97,90 ±0,75	98,20 ±0,69
P2	*m*	94,39 ±1,20	99,10 ±0,49	99,60 ±0,33
P3	*m*	98,60 ±0,71	99,80 ±0,20	98,10 ±0,82
P4	*m*	98,40 ±0,78	99,80 ±0,20	99,70 ±0,30
P5	*v*	98,20 ±0,73	99,70 ±0,30	99,80 ±0,20
Bendrai, %:		**93,91 ±3,59**	**98,04 ±1,75**	**97,77 ±1,40**

Atpažinimo etapas. Sistemos IZ_NNK_ZSF testavimo įvairiomis modelių aibėmis rezultatai pateikti 5.9 lentelėje. Kiekvienam iš dešimties kalbėtojų pateikiami treji izoliuotų žodžių atpažinimo rezultatai, testuojant tris akustinių modelių aibes: žodžių, skiemenų ir fonemų.

Pagal apibendrintus rezultatus, atpažinimo etape naudojant skiemenų AM, pasiekiamas didžiausias ZAT. Artimas rezultatas pasiekiamas naudojant žodžių

AM. Fonemų AM rezultatai akivaizdžiai blogesni. Grafinis izoliuotų žodžių atpažinimo rezultatų pasiskirstymas pagal naudojamų akustinių modelių aibės tipą pateiktas 5.3 paveiksle.



5.3 pav. *Sistemos IZ_NNK_ZSF testavimo rezultatai pagal akustinių modelių aibės tipą*

5.2.3. Antrojo tyrimo išvados ir rezultatai

Ištyrus nepriklausomos nuo kalbėtojo automatinio šnekos atpažinimo sistemos IZ_NNK_ZSF izoliuotų žodžių atpažinimui tris variantus: atpažinimą grindžiant žodžio, skiemens ir fonemos kalbos vienetais, daromos šios išvados:

1. Izoliuotų žodžių atpažinimui labiausiai tinka skiemenų ir žodžių kalbos vienetų tipai (galintys pasiekti 98 ±1,8 % ir 98 ±1,4 % atpažinimo tikslumą);
2. Vertinant fonemų, skiemenų ir žodžių akustinio modeliavimo sudėtingumą, modeliavimas paprasčiau atliekamas žodžiams;
3. Siekiant atpažinti izoliuotus žodžius ir renkantis akustiniam modeliavimui kalbos vienetą iš fonemų, skiemenų ir žodžių kalbos vienetų tipų, rekomenduojama rinktis žodžių kalbos vienetų tipą, nes didesnis šnekos atpažinimo tikslumas ir paprastesnis akustinis modeliavimas.

5.3. Ištisinės šnekos fonemomis grįsto atpažinimo tyrimai

Šioje dalyje aprašomas garsyno LRN0 pagrindu atliktas ištisinės lietuvių šnekos atpažinimo tyrimas[13]. Pagrindinis tikslas – išanalizuoti kelių modifikuotų

[13] Šis tyrimas buvo atliktas kartu su dr. D. Šilingu. Eksperimentinio tyrimo apraše akcentuojamos tos eksperimento dalys, kurias atliko šio darbo autorė (papildomų AM įkvėpimui, tylai, pauzei ir

lietuvių kalbos fonemų aibių įtaką šnekos atpažinimo tikslumui. Standartinėje fonetinėje bazėje buvo naudojamos kirčio, priebalsių minkštumo žymės. Tyrime analizuota jų įtaka. Taip pat skirtas dėmesys dvigarsio reprezentacijai: ar verta dvigarsį traktuoti kaip neskaidomą kalbos vienetą, ar jį skaidyti į du garsus. Iš visų modifikuotų fonemų aibių išrinktos kelios, kurias naudojant gauti geriausi atpažinimo rezultatai. Pagal atrinktas fonemų aibes suformuotos kontekstinių fonemų aibės. Tyrimų rezultatai yra skelbti (Šilingas *et al.* 2004b, Šilingas *et al.* 2006).

5.3.1. Ištisinės šnekos fonemomis grįsto atpažinimo tikslas ir uždaviniai

Visos kalbos turi kalbininkų apibrėžtas fonemų aibes, todėl fonema ir yra šnekai atpažinti labiausiai paplitęs segmentinis vienetas. Apibrėžtosios aibės turi būti išanalizuotos ir pritaikytos šnekai atpažinti, nes šiame procese garsinė informacija turi lemiamą reikšmę. Ištisinės šnekos fonemomis grįsto atpažinimo tyrimo tikslas – sumodeliuoti ASA sistemą IS_PNK_F (ištisinė šneka, priklausoma nuo kalbėtojų, fonemos) fonemų pagrindu ir išsiaiškinti fonemų aibės, reikalingos AM kūrimui, sudarymo ypatumus. Buvo siekiama peržvelgti, kokio laipsnio įtakos šnekos atpažinimui turi tokie kalbos elementai kaip: kirtis, priebalsių minkštumas ir ilgumas. Kadangi kuriant garsyną jau atsižvelgta į minėtus elementus, tai pradinė – bazinė fonemų aibė yra fonemų aibė su priebalsių minkštumu, balsių ilgumais ir kirčiais.

Tyrimais buvo siekta surasti tiksliausią lietuvių šnekos fonetinei erdvei fonemų aibę; surasti pagrindinius šnekos elementus, lemiančius geresnį sistemos darbingumą, ir pateikti rekomendacijas. Gauti rezultatai yra priklausomi nuo tyrimų bazės, tačiau pasitelkus žinias apie lietuvių kalbą, galima daryti apibendrinimus.

Kitas etapas, kuriame atpažinimo sistemos darbingumas pagerinamas, yra susietas su kontekstinių fonemų sąvoka. Kontekstinių fonemų naudojimas atpažinime didina sistemos darbo tikslumą, nes palyginti su fonema, kontekstinėje fonemoje labiau atsižvelgiama į koartikuliaciją. Atliekant tyrimą buvo kelios geriausios fonemų aibės performuotos į kontekstinių fonemų aibes, sukurti AM, iš naujo sumodeliuotos atpažinimo sistemos ir išrinkta geriausia kontekstinių fonemų aibė.

Tyrimui iškelti uždaviniai:

- išskirti fonemų aibes pagal šiuos fonemų požymius: kirtį, priebalsių minkštumą, mišriųjų dvigarsių skaidymą;

---

dvibalsiams modeliavimas; klasterizavimo medžių kontekstinėms fonemoms klasterizuoti paruošimas).

- pagal išskirtas fonemų aibes sukurti AM, atliekant PMM parametrų parinkimą;
- pagal atrinktas didžiausią atpažinimo tikslumą (ZT) teikiančias fonemų aibes sukonstruoti kontekstines fonemų aibes ir sukurti jų AM.

5.3.2. Tyrimo eiga

Tyrime naudota ištisinės šnekos fonemomis ar skiemenimis grįstos atpažinimo sistemos modeliavimo schema (3.5.2 skyrelis 3.7 paveikslas). Ji taikytina visiems šiame darbe atliktiems tyrimams, kuriuose kalbos vienetas yra ne žodis, o skiemuo ar fonema. Kuriant kontekstinius AM buvo vadovaujamasi 3.5.2 skyrelyje 3.8 paveiksle pateikta mokymo ir testavimo schema.

Buvo išskirtos sistemos dvi IS_PNK_F modeliavimo stadijos: mokymo ir testavimo. Mokymo procesą galima vadinti paruošiamuoju, nes prieš AM mokymą atliekami papildomi darbai, susiję su fonemų aibių formavimu, žodynų pagal jas ruošimu, AM parametrų parinkimu.

Fonemų aibių išskyrimas ir žodynų paruošimas. Išskiriant analizuotinas fonemų aibes, pateiktas 5.10 lentelėje, pirmine ir pagrindine, iš kurios kildinamos kitos, laikoma garsyno LRN0 fonemų aibė. Joje yra priebalsių minkštumo, kirčio ženklo žymenys, dvigarsiai laikomi vienu kalbos vienetu. Garsyno fonemų aibė buvo parinkta išanalizavus kalbininkų ir jau esamų garsynų naudojamas fonemų aibes. Ji atitinka minėtas fonemų aibes, bet yra papildyta vienu fonemų požymių – kirčiu. Tikimasi, kad didesnis fonemos požymių komplektas geriau diferencijuos garsus.

Antroji fonemų aibė, kurioje pašalintas priebalsių minkštumas, parinkta keliant hipotezę, kad kontekstinės fonemos atveju minkštumo požymis tampa nereikalingas, nes minkštumą atspindi fonemos kaimynės.

Trečioji fonemų aibė, kurioje pašalinti priebalsių minkštumo ir kirčio žymenys priartina šią aibę prie grafemų – rašto ženklų; norima patikrinti, ar atmetus papildomus diferencinius požymius ir supaprastinus fonemų aibę negalima gauti geresnių atpažinimo rezultatų.

Ketvirtoji fonemų aibė, kurioje nelieka kirčio ženklų, atspindi tipinę, labiausiai naudojamą fonemų aibę.

Penktoje fonemų aibėje keičiamos afrikatos ir skaidomi dvigarsiai atsižvelgiant į mažus afrikatų pasikartojimo dažnius mokymo duomenyse ir prielaidą, kad išskaidžius dvigarsius bus sumažintas modelių skaičius, bet fonemų aibės reprezentatyvumas dėl to nesumažės.

Visoms penkioms fonemų aibėms buvo paruošti žodynai. Kaip buvo minėta, pagrindinis garsyno žodynas atspindi pirmąją smulkiausiai fonemas aprašančią fonemų aibę, o kiti žodynai buvo gauti transformacijų būdu iš pagrindinio.

5.10 lentelė. *Fonemos segmentinio vieneto aibių tipai ir juos atitinkančios akustinių modelių aibės*

AM aibės pavadinimas	Fonemos segmentinio vieneto aibių apibūdinimai
AM_MKD	Fonemų aibė **su** priebalsių minkštumu ir **su** kirčiais;
AM_KD	Fonemų aibė **be** priebalsių minkštumo žymių, bet **su** kirčiais;
AM_D	Fonemų aibė **be** priebalsių minkštumo žymių ir **be** kirčių;
AM_MD	Fonemų aibė **su** priebalsių minkštumu ir **be** kirčių;
AM_MK	Fonemų aibė **su** priebalsių minkštumu ir **su** kirčiais, bet mišriuosius dvigarsius išskaidžius į atskiras komponentes;

Akustinių modelių parametrų parinkimas ir modelių mokymas. Kiekvienai aibių (5.10 lentelė) fonemai kuriamas atskiras modelis. Kiekvieno fonemos modelio būsenų skaičius – 5, Gausinių mišinių skaičius būsenoje – 4. Mokymas vyko keliomis stadijomis, didinant parametrų skaičių ir 3 kartus vykdant mokymo iteraciją.

Papildomi modeliai buvo sukurti tylai, įkvėpimui ir pauzei, todėl nustatant būsenų skaičių šiuose modeliuose buvo atlikti tyrimai. Taip pat buvo ieškomas geriausiai dvigarsius reprezentuojantis modelis.

Atpažįstant anglų šneką, tylos ir pauzės modeliai sujungiami, t. y. pirmiausia sukonstruojamas 5 būsenų tylos modelis, o iš jo centrinė būsena perkeliama į pauzės modelį. Tylos modelio perėjimų paeiliui iš vienos būsenos į kitą matricoje įvedami du papildomi perėjimai, leidžiantys peršokti iš pirmos būsenos iš karto į paskutinę. Sieti tylos ir pauzės modelius logiška – jų charakteristikos tos pačios, nes tyla ir pauzė atitinka tą patį signalą, skiriasi tik trukmė.

Duotoje tyrimų bazėje garso signale jau buvo įdėtos tylos, pauzės ir įkvėpimo žymės, todėl atsirado galimybė patikslinti minėtų modelių topologijas. Pradiniai tyrimai atlikti naudojant AM_MKD modelių aibę (5.10 lentelė) tris kartus padidinus mišinių būsenose skaičių. Deriniai: *1_derinys*, *2_derinys*, *3_derinys*, tirti naudojant garsyno LRN0 testavimo-tobulinimo aibę, o deriniai: *2.1_derinys*, *2.2_derinys*, tirti naudojant testavimo aibę. 5.11 lentelėje pateikti rezultatai su modifikuotomis įkvėpimo, tylos ir pauzės modelių būsenomis.

Iš lentelės matyti, kad didžiausias atpažinimo tikslumas pasiekiamas tylos ir įkvėpimo modeliams turint po 5 būsenas, o pauzės – 3 būsenas. Tylos ir pauzės modeliai nebuvo siejami. Kaip matyti, įkvėpimo modelio būsenų skaičius nebuvo keičiamas. Modelis su 5 būsenomis visiškai pasiteisino, nes peržiūrėjus atpažinimo rezultatus, įkvėpimas buvo atpažintas 100 %. Būtent tie modeliai, parodę geriausius rezultatus, buvo naudoti tolesniems tyrimams.

Dar viena modelių topologijos paieška yra susijusi su dvigarsio reprezentacijos klausimu. Pagrindinis klausimas buvo toks: dvigarsio modelyje turi būti 5 ar 8 būsenos. Kadangi vienai fonemai reprezentuoti skiriamos 5 būsenos, tai dvigarsiui, sudarytam iš 2 fonemų, reprezentuoti, atrodo, reikėtų 8. Tačiau yra manoma, jog dvigarsis greičiau yra ne dviejų fonemų samplaika, o

atskiras garsas. Norint tai pasitikslinti, atlikti keli tyrimai, kurių rezultatai pateikti 5.11 lentelėje. Tyrimas atliktas analogiškomis sąlygomis, kaip ir ieškant modelių tylai, pauzei ir įkvėpimui. Rezultatai patvirtino, kad dvigarsį reikia laikyti vienu garsu, o ne fonemų junginiu, todėl jį geriau modeliuoti 5 būsenomis, panašiai kaip ir vienos fonemos atveju.

5.11 lentelė. *Modelių topologijos paieškos rezultatai tylos, pauzės, įkvėpimo ir dvigarsių modeliams*

AM modelių su įvairiu būsenų skaičiumi deriniais	AM būsenų skaičius				ZT, %
	Įkvėpimas	Pauzė	Tyla	Dvigarsiai	
1_derinys	5	5	5	5	46,27
2_derinys	**5**	**3**	**5**	**5**	**47,01**
3_derinys	5	3 (iš tylos)	5	5	45,83
2.1_derinys	**5**	**3**	**5**	**5**	**58,71**
2.2_derinys	5	3	5	8	56,47

Kontekstinės fonemos. Fonemos šnekai atpažinti naudojamos rečiau už kontekstines fonemas. Pagal atrinktas AM_MKD, AM_MK ir AM_MD fonemų aibes sukonstruotos kontekstinių fonemų aibės ir jų AM. Didelių kontekstinių fonemų aibių modelių parametrus įprasta sieti. Tai atliekama naudojant sprendimų medį. Modeliuojant sistemą IS_PNK_F, reikėjo sukurti sprendimų medžius kontekstinių fonemų klasterizavimui. Kuriant šiuos medžius buvo analizuoti šie sprendimų medžiai: TIMIT garsyno ir Lietuvos tyrėjų sukurtieji (Raškinis, Raškinienė 2003b, 2004). Kiekvienai kontekstinių fonemų aibei buvo sukonstruotas atskiras sprendimų medis. Klausimai šiems sprendimų medžiams buvo ruošiami remiantis lietuvių kalbos fonemų akustinių savybių skirtumais (Pakerys 2003).

Geriausi atpažinimo rezultatai buvo gauti naudojant AM_MKD ir AM_MK fonemų aibes. Jas modifikavus suformuotos dar dvi. Galutinis kontekstinių fonemų akustinių modelių sąrašas pateiktas 5.12 lentelėje.

5.12 lentelė. *Kontekstinių fonemų aibių tipai ir juos atitinkančios akustinių modelių aibės*

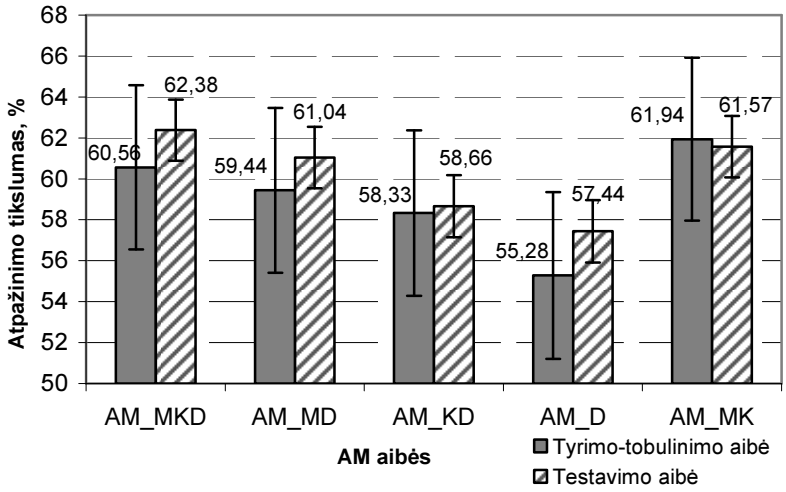
AM aibės pavadinimas	Fonemos segmentinio vieneto aibių apibūdinimai
T_AM_MKD	Kontekstinių fonemų aibė, suformuota pagal AM_MKD
T_AM_KD	Kontekstinių fonemų aibė **be** priebalsių minkštumo žymių, bet **su** kirčiais, suformuota pagal AM_MKD
T_AM_MK	Kontekstinių fonemų aibė, suformuota pagal AM_MK
T_AM_K	Kontekstinių fonemų aibė **su** kirčiais, suformuota pagal AM_MK

5.3.3. Rezultatai

Šnekos atpažinimo rezultatai gaunami dviem aibėms – testavimo-tobulinimo ir testinei. Kadangi testavimo-tobulinimo aibė yra nedidelės trukmės, išvados daromos pagal testinę aibę. Fonemomis grįsto atpažinimo penkiomis AM aibėmis rezultatai pateikiami 5.13 lentelėje ir 5.4 paveiksle.

5.13 lentelė. *Šnekos atpažinimo rezultatai fonemų akustinių modelių aibėms. Rezultatai pateikti ZT testavimo-tobulinimo aibės 1, 2, 3, 4 mišiniams ir testavimo aibei 4 mišiniams, paskaičiuojant 95 % pasikliautinuosius intervalus*

AM aibės pavadinimas	Gausinių mišinių skaičius / ZT, %				
	1	2	3	4	4
AM_MKD	41,39	55,56	57,50	60,56 ±4,01	62,38 ±1,49
AM_MD	43,33	54,44	57,50	59,44 ±4,03	61,04 ±1,50
AM_KD	38,89	50,83	54,17	58,33 ±4,04	58,66 ±1,52
AM_D	37,22	48,89	52,78	55,28 ±4,08	57,44 ±1,52
AM_MK	41,94	54,17	55,56	61,94 ±3,98	61,57 ±1,50

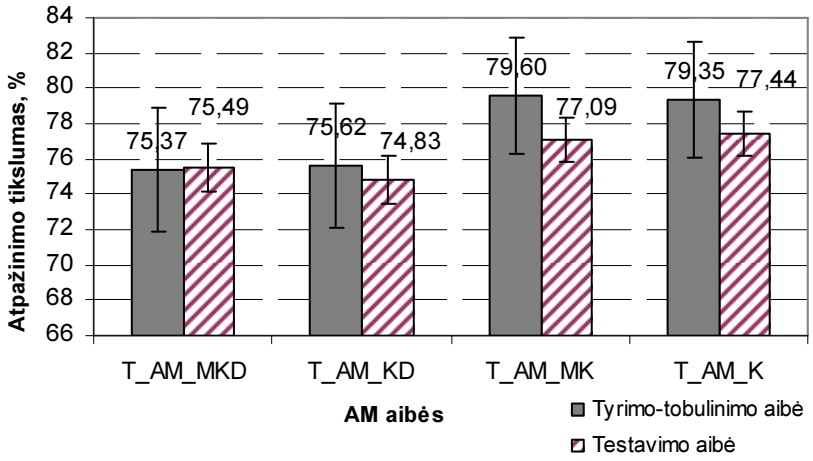


5.4 pav. *Atpažinimo rezultatai ZT prasme, gauti panaudojus pagal 5 fonemų aibes suformuotus AM testavimo-tobulinimo ir testavimo aibėms*

Kaip matyti iš paveikslo, geriausias atpažinimo rezultatas buvo gautas naudojant pirmąją fonemų aibę, kuri greta fonemų turi papildomos informacijos apie kirčius ir priebalsių minkštumą. Antrasis rezultatas priklauso 5-tai fonemų aibei, kurioje egzistuoja tiek kirčio, tiek priebalsių minkštumo žymės, bet mišrieji dvigarsiai išskaidyti į dvi komponentes. Pasižiūrėjus į kitus rezultatus matyti, kad

priebalsių minkštumas – informacija apie kontekstą turi didesnę įtaką, nei kirčių buvimas.

Šnekos atpažinimo rezultatai, gauti kontekstinių fonemų AM, pateikiami 5.14 lentelėje ir 5.5 paveiksle. Matyti, kad geriausias atpažinimo rezultatas pasiektas naudojant kontekstinių fonemų aibę, kurioje yra žymės apie kirčius, bet nėra priebalsių minkštumo žymių, o mišrieji dvigarsiai išskaidyti į dvi komponentes. Tokią kontekstinių fonemų aibę rekomenduotina naudoti tolesniuose tyrimuose.



5.5 pav. *Atpažinimo rezultatai pagal 4 kontekstinių fonemų aibes testavimo-tobulinimo ir testavimo aibėms*

5.14 lentelė. *Šnekos atpažinimo rezultatai kontekstinių fonemų akustinių modelių aibėms. Rezultatai pateikti ZT testavimo-tobulinimo aibės 1, 2, 3, 4 mišiniams ir testavimo aibės 4 mišiniams, paskaičiuojant 95 % pasikliautinuosius intervalus*

AM aibės pavadinimas	Gausinių mišinių skaičius / ZT, %				
	1	2	3	4	4
T_AM_MKD	69,15	75,62	74,13	75,37 ±3,53	75,49 ±1,32
T_AM_KD	67,91	74,38	74,13	75,62 ±3,52	74,83 ±1,34
T_AM_MK	72,64	79,10	79,35	79,60 ±3,31	77,09 ±1,29
T_AM_K	71,14	76,37	80,35	79,35 ±3,32	77,44 ±1,29

5.3.4. Trečiojo tyrimo išvados ir rezultatai

Atlikus ištisinės šnekos fonemomis grįsto atpažinimo tyrimus daromos šios išvados:

1. Iš penkių ištirtų fonemų rinkinių geriausi, nes pasiekia didžiausią šnekos atpažinimo tikslumą (62 ±1,5 % ir 62 ±1,5 %), yra tie rinkiniai, kurie greta fonemos naudoja priebalsių minkštumo ir kirčio žymes, o dvigarsių neskaido (pirmame rinkinyje) arba skaido (antrame rinkinyje) į dvi komponentes.
2. Iš keturių ištirtų kontekstinių fonemų rinkinių geriausi, nes pasiekia didžiausią šnekos atpažinimo tikslumą (77 ±1,3 % ir 77 ±1,3 %), yra tie rinkiniai, kurie greta kontekstinės fonemos, pirmuoju atveju, naudoja tik kirčio žymes, o dvigarsius skaido į dvi komponentes; antruoju atveju, naudoja priebalsių minkštumo ir kirčio žymes, dvigarsius taip pat skaido į dvi komponentes.
3. Nors naudojant kontekstinių fonemų akustinius modelius pasiekiamas didesnis šnekos atpažinimo tikslumas (77 ±1,3 %) nei naudojant paprastų fonemų akustinius modelius (62 ±1,5 %), kontekstinių fonemų atveju gali atsirasti problemų šnekos atpažinimui pateikiant mokymo metu žodyne nebuvusį žodį.
4. Šnekos atpažinimui naudojant fonemų ar kontekstinių fonemų modelius rekomenduojama:
   - Renkantis tarp paprastos fonemos ir kontekstinės fonemos atpažinimo tikslumo atžvilgiu rinktis kontekstinę fonemą, o modeliavimo paprastumo atžvilgiu – paprastą fonemą.
   - Naudoti fonemų aibę be (arba su) minkštumo žymių (-ėmis), su kirčio žymėmis ir išskaidžius dvigarsius į atskiras komponentes.
   - Modeliuojant dvigarsius jiems parinkti fonemos modelio topologiją.

5.4. Ištisinės šnekos fonemomis ir skiemenimis grįsto atpažinimo tyrimai

ASA sistemose, skirtose lietuvių šnekai, dažniausiai naudojamas fonemomis ar kontekstinėmis fonemomis grįstas šnekos atpažinimas. Skiemenų tipo tyrimai yra fragmentiški ir neišbaigti. Šiuo tyrimu siekiama nuosekliai išnagrinėti fonemomis ir skiemenimis grįstą šnekos atpažinimą. Tyrimo rezultatai yra skelbti (Laurinčiukaitė, Lipeika 2006, Laurinčiukaitė, Lipeika 2007).

Tyrime naudojamas garsynas LRN0.

5.4.1. Ištisinės šnekos fonemomis ir skiemenimis grįsto atpažinimo tikslas ir uždaviniai

Tikslas: sumodeliavus ASA sistemą IS_PNK_S (ištisinė šneka, priklausoma nuo kalbėtojo, skiemenys) išsiaiškinti ir pasiūlyti skiemenų aibės sudarymo būdą.

Tyrimo metu buvo pasiūlyta nauja struktūra, tuo pačiu ir metodika, skiemenų ir fonemų kalbos vienetų aibės formavimui (3.5.3 skyrelis). Ši metodika tapo pagrindu atliekant skiemenimis ir fonemomis grįstą šnekos atpažinimą.

Suformuoti fonemų ir skiemenų atrinkimo metodiką leidusių uždavinių sprendimas:

- duotam garsynui rasti žodžių skaidinius skiemenimis;
- atlikti skiemenų atranką, nuspręsti, kaip elgtis su likusia skiemenų gausybe;
- paruošus fonemų ir skiemenų aibę bei jos – žodyną, sukurti AM;
- surasti papildomą būdą be testinių duomenų atpažinimo rezultatų lyginimo vertinti sukurtus AM.

5.4.2. Tyrimo eiga

Tyrime taikytos 3.5.2 skyrelyje 3.7 ir 3.8 paveiksluose pateiktos sistemos modeliavimo schemos. Tyrimo metu buvo naudoti du skirtingi mokymo ir testinės imčių formavimo būdai:

- Taikant kryžminio tikrinimo principą 10 valandų garsynas buvo padalintas į 10 panašaus dydžio dalių; akustiniai modeliai buvo mokomi 9 dalimis, o viena garsyno dalimi testuojama.
- Garsynas buvo padalintas į 3 dalis: 9 val. 58 min. skiriant mokymui, 2 min. – testavimui-tobulinimui ir 16 min. testavimui.

Pirmuoju variantu gauti šnekos atpažinimo rezultatai laikomi patikimais dėl kryžminio tikrinimo principo – mokymo ir testinių imčių skirtingumo, dydžio. Tačiau tiriant šiuo tyrimo principu sugaišta daug laiko, ir darbas tapo neefektyvus. Ši priežastis lėmė perėjimą prie antrojo mokymo ir testavimo imčių formavimo varianto, o kryžminio tikrinimo principas buvo taikytas tik iš pat pradžių.

Tyrime vadovautasi 3.5.3 skyrelyje išdėstyta skiemenų ir fonemų aibės formavimo metodika, buvo atlikti metodikos tikslinimo darbai. Tyrimo aprašas pateikiamas prisilaikant 3.9 paveiksle pateiktos schemos.

Žodyno skaidymas skiemenimis (1 blokas)

Žodyno žodžiai buvo išskaidyti skiemenimis[14], pagal aprašytą skiemenavimo algoritmą (Kasparaitis 2004), kai kiekvienam skiemeniui yra taikoma bendra struktūra, apibūdinama formule STRARTSK. Šioje formulėje kiekvieną raidę atitinka tam tikros raidinių simbolių aibės, pavaizduotos 5.6 paveiksle.

[14] Nors sakoma, kad žodžiai buvo skaidomi skiemenimis, iš tikrųjų šiai kategorijai priklauso ir fonemos, gautos skaidymo metu.

S	T	R	A	R	T	S	K
s	b	j		j	b	s	k
š	d	l		l	d	š	t
z	g	m		m	g	z	
ž	k	n		n	k	ž	
	p	r		r	p		
	t	v		v	t		
	c				c		
	č				č		
	dz				dz		
	dž				dž		
	ch				ch		
	h				h		
	f				f		

Bet kokia balsė ar balsių grupė

5.6 pav. *Bendrą skiemens struktūrą sudarančių aibių rinkiniai (Kasparaitis 2004)*

Skiemenavimo procesą iliustruoja 5.7 paveikslas. Būtina pabrėžti, jog žodis skaidomas skiemenimis pradedant nuo jo pabaigos. Imama viena raidė ar raidžių kombinacija (5.7 paveikslo pirmajame stulpelyje tiriamoji raidė pabraukta), po vieną skiemens struktūros elementai, pradedant nuo struktūros pabaigos (kaip parodyta 5.7 paveikslo antrajame stulpelyje) ir tikrinama, kuriam elementui priklauso nagrinėjamoji raidė. Kai randamas toks elementas, kaip parodyta 5.7 paveikslo trečiajame stulpelyje: $s \in S$, $a \in A$, $m \in R$, $n \in R$ atvejais, pereinama prie kitos žodžio raidės, bet struktūros formulė nagrinėjama toliau. Kai prieinama žodžio ar skiemens struktūros pabaiga, konstatuojama, kad šis trūkis yra skiemens riba, o skiemenį sudaro tos raidės, kurios gavo patvirtinimą apie jų priklausymą skiemens struktūros elementui, kaip parodyta 5.7 paveikslo trečiajame stulpelyje. Išnagrinėta žodžio dalis atskiriama kaip skiemuo. Likusiai žodžio daliai taikoma ta pati procedūra, jei reikia kartojama kelis kartus.

Žodžių skaidymo skiemenimis algoritmas nėra išbaigtas. Nepatikslinti išimtiniai atvejai, kai skaidant žodį skiemenimis atsižvelgiama į žodžio sandarą. Jei žodis sudarytas iš kelių žodžių ar yra priešdėlis, lietuvių kalbos gramatikos taisyklės reikalauja, kad priešdėlis būtų atskiras skiemuo, o skaidant skiemenimis sudurtinį žodį, būtų prisilaikoma atskirų žodžių ribų. Šios išimtys reikalauja atlikti žodžio sandaros analizę ir susiję su išsamesnėmis ir specifinėmis kalbos žiniomis.

Buvo dvejojama, kaip elgtis išimtinais atvejais: ar 1) atsižvelgti į gramatikos taisykles ir peržiūrint tokius žodžius pataisyti, ar 2) palikti tokius, kaip pateikė algoritmas. Svarbiausia yra kuo labiau automatizuoti visus darbus, todėl pasirinktas antras variantas.

Žodžio raidė	Skiemens struktūros elementas	Patvirtinimas	Žodžio pabaiga	Skiemens struktūros pabaiga
namas	STRARTSK	s∉K	ne	ne
namas	STRARTSK	s∈S	ne	ne
namas	STRARTSK	a∉T	ne	ne
namas	STRARTSK	a∉R	ne	ne
namas	STRARTSK	a∈A	ne	ne
namas	STRARTSK	m∈R	ne	ne
namas	STRARTSK	a∉T	ne	ne
namas	STRARTSK	a∉S	ne	taip
	Skiemuo	**mas**		
namas	STRARTSK	a∉K	ne	ne
namas	STRARTSK	a∉S	ne	ne
namas	STRARTSK	a∉T	ne	ne
namas	STRARTSK	a∉R	ne	ne
namas	STRARTSK	a∈A	ne	ne
namas	STRARTSK	n∈R	taip	ne
	Skiemuo	**na**		
	Suskiemenuotas žodis	**na-mas**		

5.7 pav. *Žodžio namas skiemenavimo procesas*

Probleminiai žodžių skaidymo skiemenimis pavyzdžiai. Dalis priešdėlinių žodžių yra skaidomi pagal gramatikos taisykles, dalis – ne. Priešdėlinių žodžių skiemenavimo klaidos nebuvo taisomos dėl tokių žodžių gausos. 5.15 lentelėje pateiktas algoritmiškai gautas rezultatas, kuris galutiniame skaidinyje toks ir liko.

5.15 lentelė. *Neteisingų priešdėlinių žodžių skaidinių skiemenimis pavyzdžiai*

Priešdėlis	Neteisingi skaidiniai	
ap-	a-plan-ky-ti	a-plin-ko-sau-ga
iš-	i-šau-gin-tas	i-šda-vi-mą
at-	a-tme-tė	a-tlei-stu
už-	u-žda-riu-si	u-žge-sin-ti

Tam tikrų problemų kilo susidūrus dviem balsiams, priklausantiems skirtingiems skiemenims. Tokios klaidos buvo taisomos. Pavyzdžiai pateikti 5.16 lentelėje. Sudurtiniuose ir tarptautiniuose žodžiuose sutinkamos klaidos taip pat buvo taisomos. Pavyzdžiai pateikti 5.17 lentelėje.

5.16 lentelė. *Neteisingų su dviem skirtingiems skiemenims priklausančiais balsiais žodžių skaidinių skiemenimis pavyzdžiai*

Žodis	Neteisingas skaidinys	Pataisytas variantas
neišleido	nei-š-lei-do	ne-iš-lei-do
neuždengtais	neu-žden-tais	ne-už-den-gtais
suimti	suim-ti	su-im-ti

5.17 lentelė. *Neteisingų sudurtinių ir tarptautinių žodžių skaidinių skiemenimis pavyzdžiai*

Žodis	Neteisingas skaidinys	Pataisytas variantas
pusfinalio	pu-sfi-na-lio	pus-fi-na-lio
pusmečio	pu-sme-čio	pus-me-čio
specialistai	spe-cia-li-stai	spe-ci-a-li-stai

Ištaisius dviejų skirtingiems skiemenims priklausančių balsių sandūros bei sudurtinių ir tarptautinių žodžių skiemenavimo klaidas, žodžių skaidiniuose skiemenimis pažeidžiama tik priešdėlio laikymo vientisu nedalomu skiemeniu taisyklė, kuri esminė kalbotyroje, bet tyrėjų manymu, ne tokia svarbi šnekai atpažinti.

Skiemenų koregavimas (2 blokas)

Algoritme atsižvelgiama į skiemens kontekstą žodžio viduje. Kontekstas daro įtaką skiemeniui suteikdamas priebalsiams minkštumą, juos duslindamas ar skardindamas. Papildomos konteksto įtakos žymės buvo suteikiamos tik skiemenų pabaigoms. Pavyzdžiai pateikiami 5.18 lentelėje.

5.18 lentelė. *Skiemens konteksto žodžio viduje įtakos skiemenims pavyzdžiai*

Modifikuotas skaidinys	Žodis	Paaiškinimas
a*b*-*g*a-din-tas	a*pg*adintas	priebalsio *p* suskardėjimas dėl *g*
a-pli*w*-*k*o-sau-ga	apli*nk*osauga	priebalsio *n* pakitimas dėl *k*
re-mia*n*‘-*č*iai	remia*nč*iai	priebalsio *n* suminkštėjimas dėl *i*

Nepasikartojantys, skirtingi skiemenavimo algoritmu gauti suskaidyti vienetai toliau buvo analizuojami atsiejus juos nuo konkrečių žodžių, t. y. traktuojami kaip skiemenų ir fonemų sąrašas. Bendras gautų suskaidytų vienetų skaičius buvo 3 037. Skiemenavimo metu tokie raidiniai simboliai, kaip: *į* ir *y*, *ų* ir *ū*, buvo traktuojami kaip skirtingi objektai, nors tai yra viena ir ta pati fonema, ta pati akustinė realizacija, todėl teisingausia būtų įvesti vieningą žymėjimą. Tai leistų sumažinti skiemenų skaičių. Įvestas bendras žymėjimas vieną fonemą reiškiantiems raidiniams simboliams pateiktas 5.19 lentelėje[15]. Tuo pačiu patikrinama ir asimiliacija skiemens viduje, 5.20 lentelė.

5.19 lentelė. *Įvesti nauji žymėjimai*

Įvestas žymėjimas	Žymimieji raidiniai simboliai	Skiemenų pavyzdžiai
u:	ū, ų	šū, šų tampa šu:
i:	į, y	dį, dy tampa di:

[15] Šie žymėjimai sutampa su pateiktais priede A.

5.20 lentelė. *Skiemenų modifikacijos atsižvelgus į asimiliaciją pavyzdžiai*

Originalus skiemuo	Atsižvelgus į asimiliaciją pakeistas skiemuo	Pavyzdys	Transkripcija žodyne
sda	zda	įve*sda*ma	į-ve-*zda*-ma
šda	žda	pareik*šda*mas	pa-reik-*z2da*-mas
žtai	štai	grie*žtai*	grie-*s2tai*
žto	što	grie*žto*jo	grie-*s2to*-jo

Atliktos modifikacijos leido sumažinti suskaidytų vienetų skaičių iki 2 959, t. y. 78 vienetais.

Statistinė skiemenų analizė (3–8 blokai)

Žodynui, kurio apimtis 18 075 žodžių, gauta ~ 3000 skiemenų ir fonemų. Kitas žingsnis yra patikrinti ar šiems elementams užtenka mokymo duomenų. Čia galimi du veikimo būdai:

- surasti skiemenų pasikartojimo skaičių žodyne;
- surasti skiemenų pasikartojimo skaičių mokymo duomenyse.

Pirmuoju būdu gauti duomenys rodys skiemens panaudojimo lygį skiemenavimo procese. Tačiau neatsižvelgiant į mokymo duomenis, skiemenų ir fonemų aibė tampa priklausoma nuo žodyno ir atsiranda galimybė, kad į bazinę aibę gali patekti skiemuo, esantis tik testavimo duomenyse, o jo mokymo imtis yra maža arba apskritai neegzistuoja. Saugesnis yra antrasis būdas. Čia gautos skiemenų ir fonemų aibės elementai turės pakankamas mokymo imtis, bet į žodyno žodžių skaidinių ilgį atsižvelgiama nebus. Pirmoji skiemenų ir fonemų aibė pavadinama H_1 (H – hibridinė kalbos vienetų aibė, 1 – aibės numeris), antroji – H_2. Toliau aprašomi abiejų aibių sudarymo būdai[16].

Atlikus skiemenų mokymo imčių analizę pagal mokymo duomenis, buvo gauti tokie rezultatai: iš ~3000 elementų 22 % skiemenų buvo vieno egzemplioriaus, 10 % turėjo du mokymo pavyzdžius, 6 % – tris egzempliorius. Skiemenų mokymo imtys pagal žodyną buvo dar mažesnės, o norint sukurti reprezentatyvius skiemenų modelius, turi būti pakankamos mokymo imtys.

Aibės H_1 formavimas. Atvejai, kai mokymo imtyje yra mažiau nei 10 mokymo egzempliorių formuojant aibę H_1 yra visiškai nepriimtini, o mažiau 50 – nepageidautini (Šilingas *et al.* 2004b). Atsižvelgiant į tai, kad skiemuo yra ilgesnės trukmės segmentinis vienetas, jam reikėtų didesnės mokymo imties. Išnagrinėti keli slenksčiai – 50, 60 ir 70 egzempliorių pasikartojimo ribos.

Pradinėje stadijoje užfiksuota 50 mokymo pavyzdžių riba tam, kad skiemuo patektų į bazinę skiemenų ir fonemų aibę H_1. 234 skiemenys ir fonemos, kurių dažniai virš 50, traktuojami galutinės skiemenų ir fonemų aibės objektais. Šioje aibėje yra beveik visa lietuvių kalbos fonemų aibė – 56 fonemos (trūksta fonemų

[16] Aibių H_1 ir H_2 formavimo būdai yra pateikti turint fiksuotas mokymo ir testavimo imtis. Pagal kryžminio tikrinimo principą reikėjo suformuoti vieną H_1 aibę, ir net dešimt H_2 aibių.

dz ir *o:*, kurių atsisakyta dėl mažo *dz* pasikartojimo ir nepakankamų žinių *o:* žymėjimui). H_1 aibėje yra 293 vienetai, iš kurių 227 – skiemenys, o 63 – fonemos ir dvigarsiai, 3 akustiniai modeliai kuriami pašalinių garsų atpažinimui (tyla, pauzė, įkvepimas).

Suformavus bazinį skiemenų ir fonemų sąrašą H_1, reikia žodyno pradinius skaidinius skiemenimis transformuoti į skaidinius baziniais H_1 elementais. Kyla klausimas, kaip elgtis su likusiais nebaziniais skiemenimis. Galimi du būdai: 1) likusius skiemenis išskaidyti fonemomis, 2) likusius skiemenis išskaidyti skiemenimis iš bazinės aibės ir fonemomis. Antrasis būdas skiemenį kaip vientisą vienetą iškraipo, o tuo atžvilgiu pirmasis būdas yra korektiškesnis.

Likusių skiemenų skaidymui fonemomis buvo panaudota automatinė transkribavimo programa. Skaidant likusius skiemenis baziniais skiemenimis ir fonemomis buvo parašyta programa kiekvienam iš ~ 3000 skiemenų parenkanti tinkamiausius skiemenis iš bazinių skiemenų aibės. Vėliau gautą sąrašą peržiūri žmogus, išrenka geriausią skiemenų ir fonemų kombinaciją. Pirmenybė teikiama ilgesniems skiemenims, o paskui – dažnesniems.

5.21 lentelė. *Žodžių skaidiniai skiemenimis ir fonemomis panaudojus bazinę aibę H_1*

Žodis	Skiemenavimo algoritmu gautas skaidinys	Modifikuotas skaidinys	Pastabos
vaidmenį	vaid‘-me-nį	vai-d'-me-ni:	*vaid‘* išskaidytas į *vai* ir *d‘*
vairuotojas	vai-ruo-to-jas	vai-ruo-to-jas	Nepakito
minimalūs	mi-ni-ma-lūs	mi-ni-ma-l-u:-s	*lūs* išskaidyta į *lu:* ir *s*
geriausią	ge-riau-sią	ge-r'-eu-s'-ą	*riau* išskaidyta į *r‘* ir *eu*, *sią* į *s‘* ir *ą*

Tuo būdu skiemenų ir fonemų aibę H_1 atitinka du žodynai: H_1SP (skaidant likusius skiemenis naudoti skiemenys ir fonemos) ir H_1P (skaidant likusius skiemenis naudotos fonemos). 5.21 lentelėje pateikiami keli žodyno po skaidymo skiemenimis ir fonemomis pagal bazinę aibę H_1 pavyzdžiai.

Aibės H_2 formavimas. Analogiškai aibei H_1, suformuojama skiemenų ir fonemų bazinė aibė H_2. Paskaičiavus po pradinio žodžių skiemenavimo gautų skiemenų pasikartojimo dažnius mokymo imtyje, iš 2 959 skiemenų sąrašo atrenkama 234 skiemenų ir fonemų aibė H_2. Pridėjus likusias fonemas, aibės H_2 dydis yra 289 elementai (223 skiemenys, 63 fonemos, 3 modeliai pašaliniams garsams). Aibės H_2 dydis derinamas su aibės H_1 dydžiu, kad būtų galima atlikti šnekos atpažinimo rezultatų palyginimą. H_1 ir H_2 aibių dydžiai skiriasi keliais kalbos vienetais, bet kokybinis skirtumas – 45 skiemenys.

Kaip ir aibėje H_1, sudaromi du žodyno variantai: H_2P ir H_2SP.

Fonemų aibės modifikacijos

Fonemų aibę reikia apibūdinti atskirai. Atliekant tyrimus iš esmės buvo laikomasi (Raškinis *et al.* 2003b) straipsnyje išdėstytos lietuvių kalbos garsų erdvės skirstymo struktūros, tačiau yra tam tikrų skirtumų, t. y.:

- atsisakyta žymėti balsių ilgumą, išskyrus atvejį, kai ilgumas yra žymimas nosine raide, tuo būdu išnyko ilgasis balsis *o:* ;
- nėra afrikatos *dz* (išskaidyta į *d* ir *z*), dėl jos vienintelio egzemplioriaus duomenyse;
- prie 6 dvibalsių prijungtas nelietuviškos kilmės dvibalsis *eu*.

Tad jei SAMPA nurodo 58 fonemų ir 6 (+3) dvibalsių grupę, šioje grupėje yra 56 fonemos ir 7 dvibalsiai.

Čia nagrinėjama aibės H_1 modifikacija ir H_1SF tipo žodyno transkripcijos. Analizuojant skiemenų-fonemų modelių aibės vertinimo rezultatus *Pirminių šnekos atpažinimo rezultatų* dalyje, pastebėta, kad blogiausi ZT prasme yra fonemų modeliai. Nuspręsta sumažinti fonemų modelių skaičių juos apjungiant. Kadangi apjungimo laipsnis gali būti labai įvairus, o atliktuose tyrimuose vyravo vienas iš žemesnių – fonemas išskaidant į kuo daugiau grupių; pabandytas vienas iš aukštesnių apjungimo laipsnių – formuojant 5 fonemų grupes, tad 5 jų modelius (balsiai, afrikatos, sprogstamieji, pučiamieji ir sklandieji priebalsiai). Šių grupių fonemų sąrašas pateikiamas 5.22 lentelėje. H_1 aibės fonemas pakeitus apjungtų grupių žymenimis, gautoji aibė pažymėta H_1_K (_K žymi atliktą klasterizavimą).

Kitas fonemų apjungimo būdas – atmesti priebalsių minkštumo požymį. Pašalintos skiemenų pabaigoje esančios minkštumo žymės, tačiau skiemens viduje – kontekste – slypintis minkštumas išliko. Atlikus minėtą modifikaciją turimai 2 959 skiemenų ir fonemų aibei, ji sumažėjo 366 skiemenimis iki 2 593. Po rūšiavimo pagal mokymo imties dydį, atrinkti 244 skiemenys ir fonemos, kurie paskelbti baziniai. Likę (pagal anksčiau taikytą analogiją) išreiškiami šių skiemenų ir fonemų skaidiniais. Naujai skiemenų ir fonemų aibei suteiktas pavadinimas H_1_M (_M žymi minkštumo neutralizavimą). Galutinėje H_1_M skiemenų-fonemų aibėje yra 233 skiemenys ir 47 fonemos – iš viso 280 elementų.

5.22 lentelė. *Jungtinių fonemų grupių elementai*

Apjungtos grupės pavadinimas	Grupės elementai
Balsiai	a, a:, e, e:, e3:, i, i:, o, u, u:, h, h‘, j‘, ai, au, ei, eu, ie, ui, uo
Afrikatos	c2, c2‘, c, c‘, dz, dz2, dz2‘
Sprogstamieji	k, k‘, t, t‘, r, r‘, l, l‘, p‘, p‘, g, g‘, d, d‘, b, b‘, v, v‘
Pučiamieji	s, s‘, s2, s2‘, z2, z2‘, z, z‘, f, f‘, ch, ch‘
Sklandieji	n, n‘, w, w‘, m, m‘

Ganapathiraju (2001) aprašo fonemų modelių mokymą kitais duomenimis nei mokomi skiemenų modeliai. Tai buvo pritaikyta ir šiame tyrime. Fonemų modeliai buvo mokomi pagal tą patį garsyną, bet kitas mokymo imtis (naudojamas žodynas, kuriame žodžių transkripcijos sudarytos tik iš fonemų). Tokiu būdu gauti fonemų modeliai pakeičia modelius, gautus mokymo procese kartu su skiemenimis, t. y. fonemų modelius anksčiau gautoje aibėje H_1. Nauja modelių aibė pavadinama H_1dPm.

Atskiru tyrimu formuojant fonemų modelius, buvo sukurti ir kontekstinių fonemų modeliai. Jie pakeitė fonemų modelius aibėje H_1. Ši modelių aibė pavadinta H_1dTm. Gautos modelių aibės testuotos atpažįstant testavimo-tobulinimo aibę.

Skiemenų ir fonemų modelių parametrų parinkimas

Modelio sąvoka apibrėžiamas paslėptasis Markovo modelis, kurio struktūriniai elementai skirtingoms kalbos vieneto reikšmėms nekinta. Pagrindiniai struktūriniai elementai, nustatomi tyrėjo, yra: būsenų skaičius ir mišinių skaičius būsenoje. Kalbos vieneto reikšmės kitimui jautriausias dydis yra būsenų skaičius. Pagrindinis skirtumas tarp kalbos vienetų – jų trukmė. Konstruojant modelį ši savybė yra svarbi parenkant būsenų skaičių – kuo ilgesnės trukmės kalbos vienetas, tuo daugiau jo savybių galima dirbtinai atspindėti modeliuose. Fonemos trukmė išreiškiama 5 būsenomis, iš kurių tik trys tėra aktyvios. Ilgesnės trukmės kalbos vienetų modelio būsenų skaičius nustatomas padauginus kalbos vienetą sudarančių fonemų skaičių iš 3 ir pridedant 2.

Šiame tyrime elgtasi analogiškai: skiemenų ir fonemų būsenų skaičius modelyje priklausė nuo fonemų skaičiaus. Vienintelis pauzę reprezentuojantis modelis turėjo tris būsenas, visi fonemų modeliai – po 5, o skiemenų modeliai 8–11, tad būsenų skaičius įvairiuose kalbos vienetuose buvo nuo 3 iki 11.

Kito parametro – mišinių skaičiaus būsenoje – parinkimas yra sudėtingesnis. Mišinių didinimo savybe naudojamasi tada, kai kuriamam kalbos vieneto modeliui yra daug ir įvairių mokymo duomenų: skirtingi diktoriai, skirtingai ištariami tie patys garsai. Pastaroji savybė leidžia toje pačioje būsenoje atspindėti skirtingus to paties garso ar jo dalies tarimo variantus. Galima kelti hipotezę, kad mišinių skaičiaus didinimas būsenoje susijęs su mokymo imties dydžiu.

Modelių mokymo schema remiantis anksčiau atliktais tyrimais (Šilingas *et al.* 2004b) buvo: mišinių skaičiaus visuose modeliuose didinamas iki 4 nuosekliai po kiekvieno padalinimo atliekant trijų iteracijų mokymo etapą.

Pirminiai šnekos atpažinimo rezultatai

Prieš pateikiant pirminius šnekos atpažinimo rezultatus, 5.23 lentelėje pateikiamos visos tiriamos skiemenų ir fonemų aibės, žodynai.

Mokymo imčių skaičiavimo (4 blokas) ir ne pagrindinėje aibėje esančių skiemenų skaidymo (7 blokas) būdų palyginimas. Skiemenų ir fonemų aibėms H_1 ir H_2 sukurti AM ir tikrinama, kuriuo būdu suformuota aibė tinka šnekos atpažinimui pagal ZT. Tolesnės transformacijos taikytos geriausiai aibei. Kartu buvo patikrintas ir nebazinių skiemenų skaidymo būdas, naudojant žodynus H_1P, H_2P ir H_1SP, H_2SP. Šnekos atpažinimo rezultatai taikant kryžminio tikrinimo principą pateikti 5.24 lentelėje, o fiksuojant mokymo ir testinę aibes – 5.25 lentelėje.

5.23 lentelė. *Tiriamų skiemenų ir fonemų aibių, žodynų aprašai*

Skiemenų ir fonemų aibė/ Žodynas	Aprašas
Skiemenų ir fonemų aibės formavimo būdas	
H_1	Skiemenų ir fonemų iš 293 elementų (227 skiemenys, 63 fonemos ir dvibalsiai, 3 pašalinių garsų žymenys) aibė, suformuota pagal elementų dažnį (> 50) žodyne.
H_2	Skiemenų ir fonemų aibė iš 289 elementų, išrenkant 223 skiemenis iš sąrašo pagal dažnį mokymo duomenyse. Skirtumas nuo H_1 – 45 skiemenys.
H_1_K	Skiemenų ir fonemų aibė iš 235 elementų (227 skiemenys, 5 fonemų ir dvibalsių grupės, 3 pašalinių garsų žymenys). Skiemenys išlieka kaip ir bazinėje aibėje H_1 , o fonemos ir dvibalsiai suskirstomi į grupes.
H_1_M	Skiemenų ir fonemų aibė iš 283 elementų (233 skiemenys, 47 fonemos ir dvibalsiai, 3 pašalinių garsų žymenys). Aibėje panaikintas priebalsių minkštumo žymuo.
H_1dPM	Skiemenų ir fonemų aibė, kurioje skiemenys imami iš bazinės aibės H_1, o fonemų modeliai formuojami atskiru tyrimu.
H_1dTM	Skiemenų ir fonemų aibė kurioje skiemenys imami iš bazinės aibės H_1, fonemų modeliai formuojami atskiru tyrimu, prijungiamos kontekstinės fonemos.
Žodyno formavimo būdas	
H_1P, H_2P	Žodynai, suformuoti pagal skiemenų ir fonemų bazines aibes H_1 ir H_2, nebazinius skiemenis keičiant fonemomis.
H_1SP, H_2SP	Žodynai, suformuoti pagal skiemenų ir fonemų bazines aibes H_1 ir H_2, nebazinius skiemenis keičiant skiemenų ir fonemų seka.

ZT rezultatai, pasiekti atpažįstant testavimo-tobulinimo ir testavimo aibes 4 mišinių AM, yra skirtingi. To priežastis gali būti nepakankamas testavimo-tobulinimo aibės dydis. Šiame etape testavimo-tobulinimo aibės pakeisti neįmanoma, nes nebūtų galima atlikti tyrimų lyginimo. Dėl to testavimo-tobulinimo aibė lieka, tačiau lemiamas sprendimas priimamas remiantis testavimo aibės gautais rezultatais. Pagal 5.25 lentelėje pateiktus testinės aibės atpažinimo rezultatus, geresnis skiemenų ir fonemų aibės sudarymo būdas

remiasi skiemenų atrinkimu pagal jų pasikartojimą mokymo duomenyse, t. y. naudojant bazinę aibę H_2.

5.24 lentelė. *Šnekos atpažinimo rezultatai taikant kryžminio tikrinimo principą skiemenų ir fonemų aibėms H_1 ir H_2. Rezultatai pateikti apskaičiavus dešimties ZT vidurkį. Naudojami 4 mišiniai*

Skiemenų ir fonemų aibė	Žodynas	ZT vidurkis, esant 95 % patikimumo intervalams
H_1	H_1P	46,96 ±0,37
	H_1SP	48,69 ±0,42
H_2	H_2P	53,08 ±0,22
	H_2SP	56,67 ±0,33

Kalbant apie žodynų formavimo būdą, geresnis yra grįstas nebazinių skiemenų skaidymu skiemenimis ir fonemomis. Tai patvirtina dviem aibėms H_1 ir H_2 suformuoti žodynai H_1SP ir H_2SP. Jei 2.1.1 skyrelyje išskiriami 2 galimi kalbos vienetų tipai: 1) paremtas automatiniu ir 2) lingvistiniu kriterijais, tai naujasis būdas yra antrojo būdo modifikacija. Šiuo atveju pradedama nuo kalbos vienetų, gautų žodynui taikant lingvistinį kriterijų, tačiau modifikuojant žodyną šis kriterijus keičiamas pirmenybę teikiant žodžio skaidymui į kuo ilgesnės trukmės kalbos vienetus. Gautajame žodyne žodžių skaidiniai yra greičiau pseudo-skiemenys ir fonemos, nei skiemenys ir fonemos.

5.25 lentelė. *Šnekos atpažinimo rezultatai fiksuojant mokymo ir testines aibes skiemenų ir fonemų aibėms H_1 ir H_2. Rezultatai pateikti ZT testavimo-tobulinimo aibės 1, 2, 3, 4 mišiniams ir testavimo aibei 4 mišiniams, paskaičiuojant 95 % pasikliautinuosius intervalus*

Skiemenų ir fonemų aibė	Žodynas	ZT (testavimo-tobulinimo aibė)				ZT (testavimo aibė)
		1	2	3	4	4
H_1	H_1P	45,27	56,47	61,44	65,67	60,94 ±1,50
	H_1SP	52,24	62,94	64,93	70,15	63,81 ±1,48
H_2	H_2P	46,02	60,70	60,95	63,93	61,92 ±1,49
	H_2SP	52,24	61,44	64,93	66,42	**65,38 ±1,46**

Tolesniuose tyrimuose naudojamas naujas modifikuotas kalbos vienetų žodynui konstruoti būdas ir naudojami H_1SP ir H_2SP tipo žodynai.

Fonemų modelių modifikacijų palyginimas. Fonemų aibės ir jos elementų modeliavimui, buvo tirtos 5 skirtingos aibės. Skiemenų aibė sutampa su H_1 aibės skiemenimis, o fonemų atveju buvo naudota: 1) 56 fonemos (aibė H_1), 2) 5 fonemų grupės (aibė H_1_K), 3) 45 fonemos, kurių aibėje nėra

minkštumo žymens (H_1_M), 4) 56 fonemos, kurių modeliai gauti atskiru tyrimu (aibė H_1dPm), 5) aibės H_1dTm fonemos, prie kurių pridedamos kontekstinės fonemos. Gauti rezultatai pateikiami 5.26 lentelėje.

Atlikus tyrimus nustatyta, kad atskiras fonemų modelių formavimas nėra efektyvus ir rekomenduojama naudoti kartu su skiemenų modeliais gautus fonemų modelius.

5.26 lentelė. *Skiemenų ir fonemų modelių aibių atpažinimo rezultatai testavimo-tobulinimo aibei, naudojant 4 mišinius būsenose, paskaičiuojant 95 % pasikliautinuosius intervalus*

Skiemenų ir fonemų aibės	ZT, %
H_1_K	49,50 ±4,10
H_1_M	66,42 ±3,87
H_1	70,15 ±3,75
H_1dPm	68,66 ±3,81
H_1dTm	65,67 ±3,90

Skiemenų atrankos į PA slenksčio parinkimo rezultatai (6.2 blokas). Tolesniame etape buvo keičiamas skiemenų patekimo į bazines aibes H_1 ir H_2 slenkstis (pradinis slenkstis buvo 50) ir taip mažinant ar didinant pradinių aibių dydį formuojamos naujos skiemenų ir fonemų aibės. Aibei H_1 keisti slenksčiai yra 50, 60 ir 70 mokymo egzempliorių. Aibėje H_2 bandyta išlaikyti tą patį kaip ir aibei H_1 tapatų kalbos vienetų skaičių, nors skiemenų patekimo slenksčiai yra kiti. Dėl patogumo vietoje aibės H_2 atveju naudotų slenksčių pateikiami aibės H_1 slenksčius atitinkantys skaičiai. 5.27 lentelėje tikrieji skiemenų pateikimo į PA slenksčiai yra pateikiami stulpelyje „Tikrasis slenkstis“, o aibės H_2 atveju stulpelyje „Simbolinis slenkstis“ esantys slenksčiai yra perimti iš aibės H_1 ir susieti su „Tikraisiais slenksčiais“ pagal aibių H_1 ir H_2 dydžius. Pagal simbolinius slenksčius atitinkamų H_1 ir H_2 aibių dydžiai nėra vienodi dėl to, kad skiemenys į PA buvo atrenkami kartu su fonemomis. Neatsižvelgus į šių kalbos vienetų proporcijas pateko skirtingas skiemenų ir fonemų skaičius. Vėliau PA buvo papildytos likusiomis fonemomis (fonemų aibės dydis yra fiksuotas). Taigi H_1 ir H_2 PA dydžių skirtumas slenksčio 50 atveju atsirado neatsižvelgus į pradines kalbos vienetų proporcijas.

Iš 5.27 lentelėje pateiktų rezultatų matyti, kad keičiant skiemenų patekimo į PA slenkstį, didėja PA ir pakyla šnekos atpažinimo tikslumas. Stebint šnekos atpažinimo tikslumą kintant skiemenų patekimo į H_2 slenksčiui, pastebėta, kad yra tokia slenksčio reikšmė (H_2 aibės atveju slenksčiui įgyjant reikšmes 20 ir 30), kai atpažinimo tikslumas didėja nežymiai. Garsyne LRN0 iš ~3000 skiemenų ir fonemų į H_2 atrinkus 11 % elementų, pasiekiamas geriausias

ZT – 67,38 ±1,44 %. Tiesa, užtektų atrinkti ir 10 % elementų, nes ZT kitimas yra nedidelis.

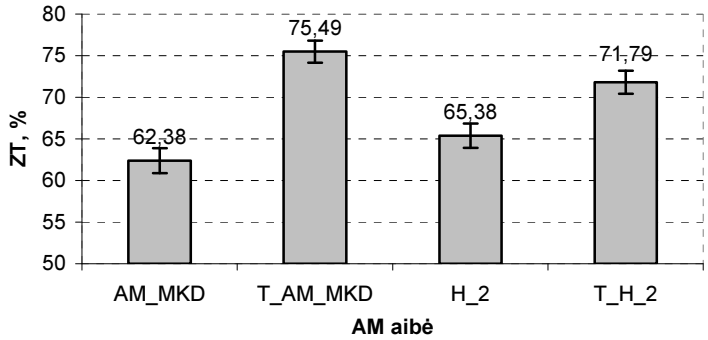
5.27 lentelė. *Šnekos atpažinimo rezultatai skiemenų ir fonemų aibėms H_1 ir H_2, keičiant skiemenų patekimo į bazinę aibę slenkstį (aibei H_1 – 50, 60, 70 ir H_2 – 20, 30, 40, 50, 60, 70). Rezultatai pateikti ZT testavimo-tobulinimo aibės 1, 2, 3, 4 mišiniams ir testavimo aibei 4 mišiniams, paskaičiuojant 95 % pasikliautinuosius intervalus*

Skiemenų ir fonemų aibė	Simbolinis slenkstis	Tikrasis slenkstis	Aibės dydis	ZT, % (testavimo-tobulinimo aibė)				ZT, % (testavimo aibė)
				1	2	3	4	4
H_1	50	50	293	52,24	62,94	64,93	70,15	63,81±1,48
	60	60	261	53,23	62,94	66,17	69,40	64,30±1,47
	70	70	232	52,74	60,70	64,68	67,91	64,26±1,47
H_2	20	133	381	53,48	64,43	67,16	70,40	**67,38±1,44**
	30	161	352	51,74	63,18	65,67	70,65	67,24±1,44
	40	195	319	50,25	63,68	65,17	68,41	66,57±1,45
	50	224	289	52,24	61,44	64,93	66,42	65,38±1,46
	60	260	261	52,49	61,94	65,17	67,41	64,68±1,47
	70	266	239	53,73	61,94	66,17	67,77	64,37±1,47

Kontekstiniai skiemenų modeliai. 5.3.2 skyrelyje analizuojant fonemų rinkinius, buvo sukonstruoti, apmokyti ir testuoti tiek fonemas, tiek kontekstines fonemas reprezentuojantys modeliai. Kontekstinių fonemų modelių naudojimas padidino šnekos atpažinimo tikslumą nuo 61,57 % (AM_MK) iki 77,09 % (T_AM_MK). Analizuojant skiemenų ir fonemų aibes, viena aibė buvo išplėsta iki kontekstinių skiemenų ir fonemų modelių. Išplėtimui pasirinkta parodžiusi geriausią atpažinimo tikslumą (65,38 ±1,44 %) aibė H_2 su žodynu H_2SP ir skiemenų atrinkimo į pagrindinę aibę slenksčiu 50.

Kontekstinių skiemenų modelių konstravimo ir mokymo schema yra analogiška standartiškai naudojamai kontekstinių fonemų schemai (Young 2003), skiemenų atvejui atliekant modifikaciją (Šilingas 2005), kurios esmė – kurti modelius kontekstiniams skiemenims atliekant skiemens kairės ir dešinės kaimynių apribojimą iki fonemų, pavyzdžiui, kontekstinis skiemuo *pa-mai+na* bus modifikuotas į *a-mai+n*. Po kontekstinių skiemenų ir fonemų modelių sukūrimo ir apmokymo buvo atliktas jų testavimas atpažįstant garsyno LRN0 testinę aibę. Naujai sukurta kontekstinių AM aibė T_H_2 pasiekia 71,79 ±1,38 % atpažinimo tikslumą ir palyginti su AM aibe H_2 jį padidina. Vis dėlto geresni išlieka AM rinkinio T_AM_MKD rezultatai – 75,49 ±1,32 %.

Lyginamieji fonemų ir skiemenų-fonemų AM efektyvumo rezultatai pateikti 5.8 paveiksle.



5.8 pav. *Fonemų AM rinkinio AM_MKD, pagal jį išplėsto kontekstinių fonemų AM rinkinio T_AM_MKD, skiemenų ir fonemų AM rinkinio H_2 ir jį atitinkančio kontekstinių AM rinkinio T_H_2 lyginamieji garsyno LRN0 testinės aibės atpažinimo rezultatai. Visi AM testuoti esant 4 mišiniams būsenose*

Skiemenų ir fonemų modelių vertinimas (6.1 blokas)

Siekiant geriau suprasti atpažinimo klaidų priežastis ir keliant hipotezę, kad vienas iš faktorių – blogi skiemenų ir fonemų modeliai, buvo atlikta H_1 modelių aibės analizė, kurios tikslai buvo:

- surasti kriterijų, kuriuo būtų vertinami modeliai;
- surasti slenkstį, kurį peržengus modelis laikomas nebeatitinkantis atpažinimo tikslus;
- išskirti modelių grupę, nepatenkinančią keliamų reikalavimų;
- pasiūlyti būdų, kaip būtų galima taisyti netinkamus modelius.

Duomenų, reikalingų vertinimui, išskyrimas. Kadangi modelių vertinimas remiantis tik testinės aibės atpažinimo rezultatais yra gana paviršutiniškas ir neinformatyvus, siekta rasti kitą būdą. Pasirinktas toks mechanizmas: apmokytus H_1 skiemenų-fonemų aibės modelius panaudoti tos pačios mokymo aibės atpažinimui, bet atmetus žodyną, t. y. atpažinimo metu tik pagal signalo akustines savybes jam bus priskiriama skiemenų-fonemų seka (taip atsisakoma visų apribojimų, koncentruojantis į fonemų modelių gebėjimą pagal signalo akustines savybes surasti skiemenį ar fonemą atspindintį segmentą). Etalonas, su kuriuo lyginami gauti rezultatai, gaunamas tokiu būdu: pagal žodžių lygio transkripcijas ir žodyną panaudojant modelius, skiemenų-fonemų vienetų segmentai surandami automatiškai naudojant taip vadinamą Viterbi išlygiavimą. Šis etalonas nėra idealus, kadangi tikslių skiemenų-fonemų vienetų ribų šiuo metu esančiomis technologijomis nėra įmanoma surasti (tai atlieka specialiai mokytas žmogus). Gautų rezultatų ir etalonų lyginimo metu, generuotos

papildomos charakteristikos, padedančios vertinti skiemenų-fonemų modelius (nurodoma, koks yra bendras jo pasikartojimų skaičius etalone *V*, kiek buvo atpažinta teisingai *T*).

Modelių vertinimo kriterijus. Galimi keli modelių vertinimo kriterijai:

- Bendra modelio klaidos tikimybė, kaip dviejų klaidos tikimybių (klaidingą kito modelio pasirinkimo vietoje savo ir klaidingą šio modelio pasirinkimo vietoje tikrojo) suma. Paskaičiavus abi tikimybes, paaiškėja, kad skirtumas tarp jų yra didelis ir matyti, jog antroji tikimybė praktiškai nedaro įtakos sumai.
- Kiek geriau suvokiamas kriterijus – skaičiuoti teisingai atpažintų elementų skaičių – atpažinimo tikimybę:

$$AT = \frac{T}{V} \times 100\,\%, \tag{5.1}$$

čia V – modelio pasikartojimų skaičius etalone, T – atpažintų teisingai modelių skaičius, AT – modelio atpažinimo tikimybė.

Skiemenų modelių analizė. Pagal kiekvieno skiemens atpažinimo tikimybę bandyta surasti skiemenų, turinčių didesnę tikimybę, išskirtinius požymius. Pradinė aibės peržvalga leido išskirti 2 dideles aibes: skiemenų aibę, turinčią skiemenyje 2 fonemas, ir aibę, turinčią 3 fonemas. Pastarosios aibės atpažinimo tikimybės yra didesnės, todėl apsiribota skiemenų aibės, skiemenyje turinčios po 2 fonemas, tyrimu. 5.28 lentelėje pateikiama tirta skiemenų aibė (padalinta į segmentus, išrikiuotus pagal abėcėlės tvarką ir atpažinimo tikimybės dydį). Toks rikiavimo būdas leidžia palyginti antrojo elemento – balsio įtaką. Šios analizės metu siekta rasti išskirtinius skiemenų požymius, lemiančius tikslų atpažinimą ir pabandyti suformuluoti reikalavimus naujos aibės formavimui.

Analizuojant šią aibę, pastebėti 2 dalykai: 1) skiemenys, prasidedantys pučiamosiomis priebalsėmis *š, ž, z*, afrikate *c2,* turi aukštą atpažinimo tikimybę, 2) bet kokio garso kombinacija su garsu *i* (išskyrus *bi, fi,*) *di, gi, ki, li, mi, ni, pi, ri, si, ti, vi* turi žemą atpažinimo tikimybę. Formuojant skiemenų-fonemų aibę galimos rekomendacijos: 1) rinktis skiemenis, prasidedančius pučiamąja priebalse, 2) sukurti bendrą modelį bet kokio garso kombinacijos su *i* tipo skiemenims.

Tie skiemenų modeliai, kurių atpažinimo tikimybės nesiekia 50 % slenksčio, buvo analizuoti detaliau. Pastebėta, kad pvz.: *le* ir *lia*, *re* ir *ria*, *se* ir *sia,* nesiekia šios ribos. Nors po ankstesnių tyrimų atsisakyta tokio tipo skiemenų porų jungimo į vieną modelį, vis dėlto šioms trims skiemenų poroms rekomenduojama sukurti po vieną modelį.

Nagrinėjant skiemenų grupes pastebėta, kad skiemenys, prasidedantys priebalsiu *j*, nesiekia 50 % ribos, todėl siūloma jų atsisakyti.

Naujoji skiemenų-fonemų aibė formuota iš senosios tokiu būdu:

- 6 modeliai, atitinkantys skiemenis *di*, *gi*, *ki*, *pi*, *ti*, *bi* sujungiami į vieną *sp_i*;
- skiemenų poroms *sia* ir *se*, *lia* ir *le*, *ria* ir *re* sukurta po modelį, pavadinimais *se*, *le*, *re*;
- pašalinti 7 modeliai, atitinkantys skiemenis, prasidedančius garsu *j*;
- pašalinti 8 skiemenys, prasidedantys balse *at*, *at'*, *is2*, *is2'*, *ap*, *ap'*, *in'*, *ar'*;
- vietoje dingusių 24 modelių, prijungti 23 nauji (+ naujai sukurtas modelis *sp_i*), išrinkti iš pagal dažnius išrikiuotų po 3 fonemas skiemenyje turinčių skiemenų (naujai prijungtų skiemenų sąrašas pateikiamas vėliau).

Naujoji aibė, pavadinta H_1_1P (_1P – pirmas pakeitimas), yra tokio pat dydžio, kaip ir H_1 – 293 skiemenų-fonemų.

5.28 lentelė. *Skiemenų, turinčių 2 fonemas, aibė, išdėstyta abėcėlės ir AT tvarka*

Skiemens AM	AT, %	Skiemens AM	AT, %	Skiemens AM	AT, %	Skiemens AM	AT, %	Skiemens AM	AT, %
at':	**50,26**	**is2:**	**59,12**	me:	57,06	s2i:	52,16	si:	62,83
at:	46,37	is2':	54,40	me3:	54,39	**sa1:**	**69,04**	su:	58,88
bu:	**71,90**	**ju:**	**47,35**	mu:	53,35	siu:	67,74	so:	56,01
bu:	68,02	jo:	46,86	ma1:	46,89	si:	62,83	sio:	55,26
bi:	65,48	je:	35,56	**ni:**	**60,71**	su:	58,88	se3:	52,91
ba:	64,82	ji:	27,82	nio:	59,95	so:	56,01	sa:	51,94
bo:	55,75	ju:	21,61	nu:	57,47	sio:	55,26	se:	30,88
be3:	48,87	ja1:	21,58	na1:	57,32	se3:	52,91	si:	27,14
be:	45,05	ja:	17,34	niu:	57,25	sa:	51,94	sia:	26,50
c2iu:	**70,05**	**ku:**	**68,83**	no:	53,93	se:	30,88	**tu:**	**60,42**
c2io:	64,88	ko:	65,99	ne:	52,68	si:	27,14	tu:	59,43
c2ia:	50,43	ku:	61,87	nu:	49,96	sia:	26,50	to:	51,12
ce:	**62,04**	ki:	61,69	ne1:	39,41	**pa:**	**65,56**	te:	48,62
ci:	48,39	ke:	59,69	ni:	38,58	pu:	64,38	ti:	42,54
du:	**70,39**	ka:	57,06	ne3:	38,36	po:	54,80	ta:	39,67
da:	65,60	ka1:	55,91	**pa:**	**65,56**	pe:	50,52	ti:	35,77
di:	55,22	ke3:	48,24	pu:	64,38	pi:	50,04	ta1:	32,00
do:	53,89	ki:	46,49	po:	54,80	**ru:**	**67,84**	te3:	29,96
de3:	52,30	**lio:**	**70,16**	pe:	50,52	rio:	63,74	**vu:**	**72,22**

Skiemens AM	AT, %	Skiemens AM	AT, %	Skiemens AM	AT, %	Skiemens AM	AT, %	Skiemens AM	AT, %
de	51,74	lu	64,69	pi	50,04	ro	61,06	vo	60,25
di	45,81	liu	64,52	**ru**	**67,84**	ru	57,81	vi	56,85
fe	**74,63**	la	59,28	rio	63,74	ri	53,56	va	44,59
fi	63,50	liu	57,09	ro	61,06	ra	41,91	ve	32,98
gu	**78,45**	lo	54,99	ru	57,81	ria	40,24	ve3	31,08
gi	68,23	le	47,44	ri	53,56	re3	37,12	vi	30,50
gu	66,97	li	40,45	ra	41,91	ri	24,97	**z2u**	**90,14**
go	66,86	le3	40,31	ria	40,24	re	24,06	z2a	87,54
ge	61,36	lia	36,07	re3	37,12	**s2a**	**90,36**	z2i	82,93
ga1	61,19	li	35,11	ri	24,97	s2i	80,27	z2i	81,60
ge3	59,43	**mo**	**63,82**	re	24,06	s2e	74,68	**za**	**76,76**
gia	58,87	mu	61,12	**s2a**	**90,36**	s2i	52,16	zi	74,65
ga	58,83	mi	60,38	s2i	80,27	**sa1**	**69,04**		
gi	47,36	ma	59,80	s2e	74,68	siu	67,74		

Šios aibės modelių įvertinimas buvo atliktas analogiškai aibei H_1. Kartu atliktas ir testavimo-tobulinimo, ir testavimo aibių atpažinimo tyrimai. Gauti atpažinimo rezultatai pateikti 5.29 lentelėje. Iš rezultatų matyti, kad aibė H_1_1P, nors ir formuota stengiantis panaudoti tas skiemenų savybes, kurios turėtų būti svarbios skiemens atpažinimui, nepateisino lūkesčių: atpažinimo rezultatai yra blogesni palyginti su atpažinimo rezultatais, gautais naudojant aibę H_1.

5.29 lentelė. *Lyginamieji atpažinimo rezultatai, gauti naudojant aibes H_1 ir H_1_1P bei testavimo-tobulinimo ir testavimo aibes, skaičiuojant 95 % pasikliautinuosius intervalus*

Skiemenų ir fonemų aibė	ZT (testavimo-tobulinimo aibė)				ZT (testavimo aibė)
	1	2	3	4	4
H_1	52,24	62,94	64,93	70,15	63,81 ±1,48
H_1_1P	46,52	60,20	65,42	69,15	63,25 ±1,48

Šnekos atpažinimo rezultatų, gautų atpažinimą grindžiant fonemomis ir skiemenimis-fonemomis, palyginimas

5.3 poskyryje buvo aprašyti šnekos atpažinimo tyrimai, siekiant išsiaiškinti fonemų aibių su skirtingomis žymėmis (kirčio, minkštumo žymės, mišriųjų dvigarsių skaidymo) naudojimo efektyvumą. Buvo tirtos 5 aibės: 1) fonemų aibė

su kirčio ir minkštumo požymiais, 2) fonemų aibė tik su minkštumo požymiu, 3) fonemų aibė tik su kirčio požymiu, 4) fonemų aibė be kirčio ir minkštumo požymių, 5) fonemų aibė su kirčio ir minkštumo požymiais ir išskaidytais mišriaisiais dvigarsiais. Atpažinimo tyrimai buvo atlikti su testavimo-tobulinimo ir testavimo aibėmis.

5.30 lentelė. *Skiemenų, esančių H_1_1P ir nesančių H_1, AM atpažinimo tikslumas AT*

Skiemens AM	AT, %	Skiemens AM	AT, %	Skiemens AM	AT, %	Skiemens AM	AT, %
kra	85,54	par	91,67	z2ai	88,80	for	97,37
kan'	97,80	niu	41,92	vas	86,05	dau	85,56
tro	89,77	mie	70,00	sian'	98,31	c2iai	83,69
sku	89,31	kuo	79,70	rius	81,43	tri	82,24
sias	73,60	jau	68,37	pla	95,21	stas	93,66
ras	83,05	dvi	84,57	z2ei	89,52		

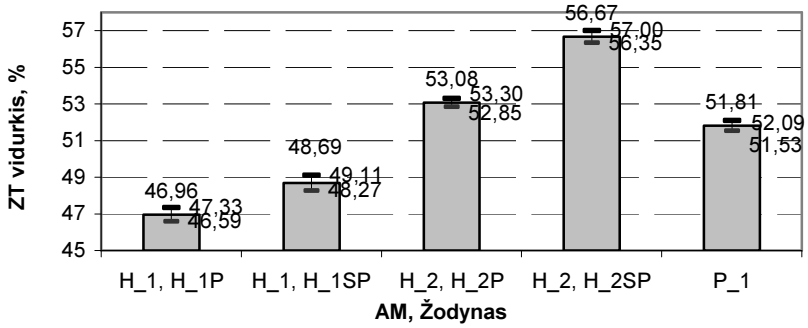
Šnekos atpažinimo rezultatų lyginimas buvo atliktas dviem būdais: 1) palyginti skirtingų tyrimų gauti rezultatai, kai mokymo ir testinės aibės yra fiksuotos, 2) išrinkus geriausią šnekai atpažinti fonemų aibę ir pritaikius kryžminio tikrinimo principą gauti nauji rezultatai. Pirmojo atvejo atpažinimo rezultatai yra 5.31 lentelėje, o antrojo – 5.9 paveiksle.

5.31 lentelė. *Skirtingais kalbos vienetais grįsto šnekos atpažinimo lyginamieji rezultatai testavimo aibei, paskaičiuojant 95 % pasikliautinuosius intervalus*

Nr.	Skiemenų ir fonemų aibė (dydis)	ZT, %
1.	AM_MKD(229)	62,38 ±1,49
2.	AM_MD (139)	61,04 ±1,50
3.	AM_KD(140)	58,66 ±1,52
4.	AM_D (86)	57,44 ±1,52
5.	AM_MK (87)	61,57 ±1,50
6.	H_1 (232)	64,26 ±1,47
7.	H_2 (239)	**64,37 ±1,47**

Taikant kryžminio tikrinimo principą testuota 5.3.2 poskyryje išrinkta geriausia fonemų aibė AM_MKD, čia žymima P_1.

Iš šnekos atpažinimo rezultatų palyginimo keliais būdais matyti, kad rezultatai yra analogiški, t. y. skiemenimis ir fonemomis grįstas atpažinimas pateikia geresnius rezultatus, nei vien tik fonemomis grįstas atpažinimas, nepriklausomai nuo fonemų rinkinio.



5.9 pav. *Šnekos atpažinimo rezultatai skiemenų ir fonemų aibėms H_1, H_2, žodynams H_1SP, H_2SP ir fonemų aibei P_1. Rezultatai pateikti paskaičiavus dešimties ZT vidurkį ir 95 % patikimumo intervalus. Mišinių skaičius akustinių modelių būsenose – 4*

5.4.3. Ketvirtojo tyrimo išvados ir rezultatai

Ištyrus ištisinės šnekos atpažinimą skiemenų ir fonemų pagrindu, daromos šios išvados:

1. Skiemenų ir fonemų akustinio modeliavimo metodikos lietuvių šnekai darbo pradžioje nebuvo, o pasirodžiusi vėliau buvo nepakankama dėl mažai skiriamo dėmesio skiemens akustiniam modeliavimui;
2. Pagal naujai pasiūlytą skiemenų ir fonemų atrinkimo ir žodyno sudarymo metodiką atrinkus skiemenis ir fonemas ir jiems sukūrus akustinius modelius, pasiektas šnekos atpažinimo tikslumas kryžminio tikrinimo principu (57 ±0,3 %) yra didesnis, nei pasiekiamas naudojant fonemų akustinius modelius (52 ±0,3 %);
3. Atskirų metodikos blokų analizė leido padidinti šnekos atpažinimo tikslumą:
   - Akustinių modelių mokymo imtį formuojant pagal mokymo duomenis ir akustinių modelių kūrimui nenaudojamų skiemenų specialiu skaidymo būdu šnekos atpažinimo tikslumą pavyko padidinti nuo 61 ±1,5 % iki 65 ±1,5 %.
   - Analizuojant skiemenų atrinkimo į aibę, pagal kurią kuriami akustiniai modeliai, slenkstį, šnekos atpažinimo tikslumas padidėjo nuo 65 ±1,5 % iki 67 ±1,4 %.
   - Pasiūlytas papildomo kriterijaus (pagrindinis – skiemenų dažniai) skiemenų-fonemų aibės sudarymui – atpažinimo tikslumo kiekvienam modeliui kriterijus. Analizuojant skiemenų atrinkimo į aibę, pagal kurią kuriami akustiniai modeliai, kokybinį kriterijų, besiremiantį skiemenį sudarančių fonemų diskriminatyvumu, atpažinimo tikslumas nukrito

nuo 64 ±1,5 % iki 63 ±1,5 %. Analogiški rezultatai gauti atlikus fonemų AM modifikacijas.

4. Kontekstinių skiemenų ir fonemų akustiniais modeliais pasiektas šnekos atpažinimo tikslumas atpažįstant LRN0 testinę aibę (72 ±1,4 %) yra mažesnis nei atpažinimo tikslumas, pasiektas kontekstinių fonemų modeliais (75 ±1,3 %). Akustiniame modeliavime pereinant nuo fonemų prie kontekstinių fonemų šnekos atpažinimo tikslumas didėja labiau nei pereinant nuo skiemenų-fonemų prie kontekstinių skiemenų-fonemų;
5. Akustinio modeliavimo sudėtingumo ir naujo žodžio įtraukimo į žodyną aspektais paprasčiau modeliuoti fonemų ir skiemenų-fonemų kalbos vienetus. Šnekos atpažinimo tikslumui iš šių kalbos vienetų naudingiau modeliuoti skiemenų-fonemų kalbos vienetus.

5.5. Penktojo skyriaus rezultatai ir išvados

1. Ištyrus izoliuotų žodžių atpažinimą, akustinius modelius konstruojant žodžiams ir modeliuojant dvi šnekos atpažinimo sistemas – priklausomą nuo kalbėtojo (IZ_PNK) ir nepriklausomą nuo kalbėtojo (IZ_NNK), galima teigti, kad:
   - Akustinius modelius sukūrus pagal vieno kalbėtojo duomenis, šie akustiniai modeliai turi būti naudojami tik to kalbėtojo šnekos atpažinimui, atpažįstant pašalinių kalbėtojų šneką gaunami netikslūs rezultatai.
   - Akustinius modelius sukūrus pagal daugelio kalbėtojų duomenis, šie akustiniai modeliai gali būti naudojami įvairių kalbėtojų šnekos atpažinimui.
   - Šnekos atpažinimo tikslumui įtakos turi modelių mokymo aibės dydis.
2. Ištyrus nepriklausomos nuo kalbėtojo automatinio šnekos atpažinimo sistemos izoliuotų žodžių atpažinimui tris variantus (atpažinimą grindžiant žodžio, skiemens ir fonemos kalbos vienetais), padarytos šios išvados:
   - Izoliuotų žodžių atpažinimui labiausiai tinka skiemenų ir žodžių kalbos vienetų tipai.
   - Vertinant fonemų, skiemenų ir žodžių akustinio modeliavimo sudėtingumą, modeliavimas paprasčiau atliekamas žodžiams.
3. Atlikus ištisinės šnekos atpažinimo tyrimus fonemų kalbos vieneto pagrindu, daromos šios išvados:
   - Iš penkių ištirtų fonemų rinkinių geriausi fonemų rinkiniai, pasiekiantys didžiausią šnekos atpažinimo tikslumą (62 ±1,5 % ir 62 ±1,5 %), greta fonemos naudoja priebalsių minkštumo ir kirčio žymes, o dvigarsių neskaido (pirmame rinkinyje) arba skaido (antrame rinkinyje) į dvi komponentes.

- Iš keturių ištirtų kontekstinių fonemų rinkinių geriausi kontekstinių fonemų rinkiniai, pasiekiantys didžiausią šnekos atpažinimo tikslumą (77 ±1,3 % ir 77 ±1,3 %), greta kontekstinės fonemos, pirmuoju atveju, naudoja tik kirčio žymes, o dvigarsius skaido į dvi komponentes; antruoju atveju, naudoja priebalsių minkštumo ir kirčio žymes, dvigarsius taip pat skaido į dvi komponentes.
- Nors naudojant kontekstinių fonemų akustinius modelius pasiekiamas didesnis šnekos atpažinimo tikslumas (77 ±1,3 %) nei naudojant paprastų fonemų akustinius modelius (62 ±1,5 %), kontekstinių fonemų atveju gali atsirasti problemų šnekos atpažinimui pateikiant mokymo metu žodyne nesantį žodį.

4. Ištyrus ištisinės šnekos atpažinimą skiemenų ir fonemų kalbos vienetų tipų pagrindu, daromos šios išvados:
   - Atskirų pasiūlytos metodikos blokų analizė leido pakelti šnekos atpažinimo tikslumą arba parodė, kad tikslumo padidinti negalima.
   - Kontekstinių skiemenų-fonemų akustiniais modeliais pasiektas šnekos atpažinimo tikslumas atpažįstant LRN0 testinę aibę (72 ±1,4 %) yra mažesnis, nei atpažinimo tikslumas, pasiektas kontekstinių fonemų modeliais (75 ±1,3 %). Akustiniame modeliavime pereinant nuo fonemų prie kontekstinių fonemų šnekos atpažinimo tikslumas didėja daugiau, nei pereinant nuo skiemenų-fonemų prie kontekstinių skiemenų-fonemų.
   - Akustinio modeliavimo sudėtingumo ir naujo žodžio įtraukimo į žodyną aspektais paprasčiau modeliuoti fonemų ir skiemenų-fonemų kalbos vienetus. Šnekos atpažinimo tikslumo prasme iš šių kalbos vienetų naudingiau modeliuoti skiemenų-fonemų kalbos vienetus.

6

Disertacijos rezultatai ir išvados

Disertacijoje ištirtas žodžiais, fonemomis, skiemenimis, kontekstinėmis fonemomis ir kontekstiniais skiemenimis grįstas lietuvių šnekos atpažinimas. Atliekant lietuvių šnekos akustinį modeliavimą ir šnekos atpažinimo tyrimus gauti šie rezultatai:

1. Modeliuojant skiemenimis ir fonemomis grįstą ištisinės šnekos atpažinimo sistemą, pasiūlyta metodika skiemenų ir fonemų akustinių modelių aibės ir žodyno sudarymui. Sukurti šią metodiką realizuojantys įrankiai. Pagrindinis naujumas metodikoje:
   - Pasiūlytas akustinio modeliavimo kiekybinis kriterijus skiemenų ir fonemų atrinkimui.
   - Pasiūlytas naujas kalbos vienetas – pseudo-skiemuo, didinantis šnekos atpažinimo tikslumą.
2. Buvo pateiktos trys automatinio šnekos atpažinimo sistemų modeliavimo schemos pagal šnekos ir kalbos vienetų tipus:
   - Žodžiais grįsta izoliuotų žodžių atpažinimo sistemos modeliavimo schema.
   - Fonemomis ar skiemenimis grįsta ištisinės šnekos atpažinimo sistemos modeliavimo schema.
   - Kontekstinėmis fonemomis ar kontekstiniais skiemenimis grįstos ištisinės šnekos atpažinimo sistemos modeliavimo schema.
3. Sukurtos ištisinės šnekos LRN garsyno LRN0 ir LRN0.1 versijos. Darbo autorė buvo atsakinga už: garsyno kūrimo proceso sudarymą; garsyno struktūros, pagrindinių garsyno charakteristikų ir fonemų sistemos parinkimą; darbo užduočių paruošimą ir paskirstymą darbo grupės nariams (tarp kurių buvo ir pati); atliktų darbų tikrinimą ir garsyno aprašymą.

Gauti rezultatai ir atlikti lyginamieji šnekos atpažinimo tyrimai leidžia daryti šias išvadas:

1. Atlikta analitinė kalbos vienetų akustinio modeliavimo apžvalga parodė, kad lietuvių šnekos akustiniam modeliavimui trūksta ne tik įvairesnių kalbos vienetų akustinio modeliavimo analizės (kai kuriems kalbos vienetams ji nėra atlikta), bet ir neįmanoma atlikti šių modeliavimų efektyvumo palyginimo dėl testavimui naudojamų skirtingų garsynų. Atlikus visų kalbos vienetų akustinio modeliavimo analizę būtų galima pasiūlyti kiekvieno kalbos vieneto akustinio modeliavimo technologiją su efektyvumo įverčiu.
2. Pasiūlyta skiemenų ir fonemų atrinkimo akustiniam modeliavimui metodika sudaro sąlygas atlikti akustinį modeliavimą, dažniau grindžiamą skiemenimis nei fonemomis. Atskirų metodikos ir akustinio modeliavimo etapų analizė didina ištisinės šnekos atpažinimo tikslumą (nuo 61 ±1,5 % iki 72 ±1,4 %).
3. Tyrimo rezultatai parodė, kad įvedus naują kalbos vienetą – pseudo-skiemenį, palyginanti su standartiškai apibrėžiamais skiemenimis padidėja ištisinės šnekos atpažinimo tikslumas (nuo 61 ±1,5 % iki 65 ±1,5 %).
4. Atlikti izoliuotų žodžių ir ištisinės šnekos atpažinimo tyrimai teikia technologijas atskirų kalbos vienetų akustiniam modeliavimui, naudojant sudarytas automatinio šnekos atpažinimo sistemų modeliavimo schemas, rekomendacijas ir sukurtus įrankius.
5. Akustinio modeliavimo tyrimai pateikia šį atskirų kalbos vienetų akustinio modeliavimo efektyvumą:
   - Tyrimų rezultatai patvirtino, kad dažniausiai naudojamas kontekstinėmis fonemomis grindžiamas ištisinės šnekos atpažinimas teikia didesnį atpažinimo tikslumą (76 ±1,3 %) palyginanti su kitais kalbos vienetais grindžiamu ištisinės šnekos atpažinimu, tačiau neišsprendžia naujo žodžio įtraukimo į žodyną problemos.
   - Tyrimu buvo įrodyta, kad skiemenimis ir fonemomis grindžiamas ištisinės šnekos atpažinimas yra pranašesnis (67 ±1,4 %) už fonemomis grindžiamą ištisinės šnekos atpažinimą (62 ±1,5 %) ir todėl rekomenduojamas naudoti.
   - Tyrimų rezultatai parodė, kad izoliuotų žodžių atpažinimui labiau tinka žodžiais grįstas šnekos atpažinimas (atpažinimo tikslumas iki 100 %).
6. Sudarytas garsynas, tyrimuose sukurti akustiniai modeliai yra universalūs ir gali būti taikomi įvairiose šnekos atpažinimo sistemose. Sukurtieji akustiniai modeliai gali padidinti šnekos atpažinimo tikslumą.

Literatūros sąrašas

[1] Ambrazas, V.; Garšva, K.; Girdenis, A. 2005. *Dabartinės lietuvių kalbos gramatika*. Vilnius: Mokslo ir enciklopedijų leidybos institutas, 742 p.

[2] Atal, B. S. 1983. „Efficient coding of LPC parameters by temporal decomposition“, in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP'83* 8: 81–84.

[3] Bacchiani, M.; Ostendorf, M.; Sagisaka, Y.; Paliwal, K. 1996. „Design of a speech recognition system based on acoustically derived segmental units“, in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP'96* 1: 443–446.

[4] Bahl, L. R.; Jelinek, F.; Mercer, R. L. 1983. „A maximum likelihood approach to continuous speech recognition“, in *IEEE Trans. on Pattern Analysis, Machine Intelligence* 5: 179–190.

[5] Bahl, L. R.; Bellegarda, J. R.; de Souza, P. V.; Gopalakrishnan, P. S.; Nahamoo, D.; Picheny, M. A. 1993. „Multonic Markov word models for large vocabulary continuous speech recognition“, in *IEEE Trans. on Speech and Audio Processing* 1(3): 334–344.

[6] Bahl, L. R.; Bellegarda, J. R; de Souza, P. V.; Gopalakrishnan, P. S.; Nahamoo, D.; Picheny, M. A. 1996. „A new class of fenonic Markov word models for large vocabulary continuous speech recognition“, in *Proc. of Int. Conf. on Spoken Language* 2: 1077–1080.

[7] Bahl, L. R.; Brown, P. F.; de Souza, P. V.; Mercer, R. L.; Picheny, M. A. 1993a, „A method for the construction of acoustic Markov models for words“, in *IEEE Trans. on Speech and Audio Processing*1(4): 443–452.

[8] Baum, L. E.; Petrie, T.; Soules, G.; Weiss, N. 1970. „A maximisation technique occuring in the statistical analysis of probabilistic functions of Markov chains“, in *Annals of Mathematical Statistics* 41: 164–171.

[9] Cole, R.; Mariani, J.; Uszkoreit, H.; Zaenen, A.; Zue, V.; Varile, G. B.; Zampolli, A. 1998. *Survey of the state-of-the-art in human language technology*. Cambridge: Cambridge University Press, 533 p.

[10] CSLU. <http://cslu.cse.ogi.edu/toolkit/ >, [Žiūrėta 2007 01 16].

[11] Černocký, J. 2002. „Units for automatic language independent speech processing“, in *Proc. of LREC'02*, Las Palmas: ELRA, 7–13.

[12] Černocký, J. 1998. *Speech processing using automatically derived segmental units: applications to very low rate coding and speaker verification*, daktaro disertacija, Paris, 100 p.

[13] Davis, S. B.; Mermelstein, P. 1980. „Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences“ in *IEEE Trans. on Acoustics, Speech and Signal Processing* 28(4): 357–366.

[14] Deligne, S.; Bimbot, F. 1997. „Inference of variable-length acoustic units for continuous speech recognition“, in *Proc. of Int. Conf. on Acoustic, Speech and Signal Processing* 3: 1731–1734.

[15] Deller, J. H.; Hansen, J. H. L.; Proakis, J. G. 1993. *Discrete-time processing of speech signals*. New York: Macmillan., 988 p.

[16] Driaunys, K.; Rudžionis, V.; Žvinys, P. 2005. „Hierarchine fonemų struktūra grindžiamo LTDIGITS fonemų klasifikavimo tyrimas“, Konferencijos „Informacinės technologijos 2005“ pranešimų medžiagoje, Kaunas, 283–288 psl.

[17] Fant, G. 1973. *Speech sounds and features*. MIT Press, 221 p.

[18] Filipovič, M. 2005. *Lithuanian isolated word recognition using neural networks and hidden Markov models approach*, daktaro disertacija, Vilnius, 138 p.

[19] Filipovič, M.; Lipeika, A. 2004. „Development of HMM/Neural Network-based medium-vocabulary isolated-word Lithuanian speech recognition system“, *Informatica* 15(4): 465–474.

[20] Fosler-Lussier, E.; Greenberg, S.; Morgan, N. 1999. „Incorporating contextual phonetics into automatic speech recognition“, in *Proc. of ICPHS'99* 1: 611-614.

[21] Fotonija 1996. „Skiemuo“. <http://www.fotonija.lt/Products/Skiemuo.aspx> [Žiūrėta 2005 05 05].

[22] Fukada, T.; Bacchiani, M.; Paliwal, K. K.; Sagisaka, Y. 1996. „Speech recognition based on acoustically derived segment units“, in *Proc. of Int. Conf. on Spoken Language* 2: 1077–1080.

[23] Ganapathiraju, A.; Hamaker, J.; Picone, J.; Ordowski, M.; Doddington, G. R. 2001. „Syllable-based large vocabulary continuous speech recognition“, in *IEEE Trans. on Speech and Audio Processing* 9(4): 358–366.

[24] Gauvain, J. L.; Lamel, L. 1996. „Large vocabulary continuous speech recognition: from laboratory systems toward real-world applications“, in *Trans. of Inst. of Electronics, Information and Communication Engineers* 2: 2005–2021.

[25] Girdenis, A. 2003. *Teoriniai lietuvių fonologijos pagrindai*. Vilnius: Mokslo ir enciklopedijų leidybos institutas, 387 p.

[26] Greenberg, S. 1996. „Understanding speech understanding: towards a unified theory of speech perception“, in *Proc. of ABSP*, 1–8.

[27] Greenberg, S. 1998. „Speaking in Shorthand – a syllable-centric perspective for understanding pronunciation variation“. In *Proc. of MPV,* 47–56.

[28] Greenberg, S.; Chang, S. 2000. „Linguistic dissection of switchboard-corpus automatic speech recognition systems“, in *Proc. of ASR*, 195–202.

[29] Grumadienė, L. 1997. *Dažninis dabartinės rašomosios lietuvių kalbos žodynas: mažėjančio dažnio tvarka*. Vilnius: Lietuvių kalbos institutas, 474 p.

[30] Hain, T.; Woodland, P. C. 2000. „Modelling sub-phone insertations and deletions in continuous speech recognition“, in *Proc. of Int. Conf. on Spoken Language Processing ICSLP'00* 4: 172–176.

[31] Hain, T.; Woodland, P. C. 1999. „Dynamic HMM selection for continuous speech recognition“, in *Proc. of European Conference on Speech Communication and Technology „Eurospeech'99"* 2: 1327–1330.

[32] Holmes, J.; Holmse, W. 2001. *Speech synthesis and recognition*. Second Edition. New York: Taylor & Francis, 298 p.

[33] Hu, Z.; Schalkwyk, J.; Barnard, E.; Cole, R. 1996. „Speech recognition using syllable-like units“, in *Proc. of Int. Conf. on Spoken Language Processing ICSLP'96* 2:1117–1120.

[34] Huang, X.; Acero, A.; Hon, H. W. 2001. *Spoken language processing: a guide to theory, algorithm and system development*. New Jersey: Prentice-Hall, Inc, 980 p.

[35] HTK toolkit. <http://htk.eng.cam.ac.uk/>, [Žiūrėta 2003 10 10].

[36] Hwang, M. Y. 1993. *Subphonetic acoustic modeling for speaker-independent continuous speech recognition*, daktaro disertacija, 176 p.

[37] Hwang, M. Y.; Huang X. 1992. „Subphonetic modeling with Markov states-senone“, in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP'92* 1: 33–36.

[38] Jelinek, F. 1985. „A real-time, isolated-word, speech recognition system for dictation transcription“, in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP'85* 10: 858–861.

[39] Juang, B. H. 1985. „Maximum likelihood estimation for mixture multivariate stochastic observations of Markov chains“, in *AT&T Technical Journal* 64: 1235–1249.

[40] Junqua, J. C.; Haton, J. P. 1996. *Robustness in automatic speech recognition: fundamentals and applications*. USA: Kluwer academic publishers, 437 p.

[41] Jurafsky, D.; Martin, J. H. 2000. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. New Jersey: Prentice-Hall, Inc, 934 p.

[42] Kasparaitis, P. 1999. „Transcribing of the Lithuanian text using formal rules“, *Informatica* 10(4): 367–376.

[43] Kasparaitis, P. 2000. „Automatic stressing of the Lithuanian text on the basis of a dictionary“, *Informatica* 11(1): 19–40.

[44] Kasparaitis, P. 2001. *Lithuanian text-to-speech synthesis*, daktaro disertacija, Vilnius.

[45] Kasparaitis, P. 2004. „Skiemenavimas ir žodžių kėlimas“, paskaitų konspektai. <http://www.mif.vu.lt/~pijus/CL/Skiemen.pdf>, [Žiūrėta 2004 09 29].

[46] Kasparaitis, P. 2005. „Diphone databases for Lithuanian text-to-speech synthesis“, *Informatica* 16(2): 193–202.

[47] Kaukėnas, J.; Navickas G.; Telksnys L. 2006. „Human-computer audiovisual interface“, in *Information Technology and Control* 35(2): 87–93.

[48] Kirchhoff, K. 1996. „Syllable-level desynchronisation of phonetic features for speech recognition“, in *Proc. of Int. Conf. on Spoken Language* 4: 2274–2276.

[49] Levinson, S. E.; Roe, D. B. 1990. „A perspective on speech recognition“, in *IEEE Communications Magazine* 28(1): 28–34.

[50] Lipeika, A.; Tamulevičius G. 2004. „Segmentation of nonstationary signals“, in *Proc. of Int. Conf. on Biomedical Engineering*, 37–40.

[51] Lipeika, A.; Lipeikienė, J. 1992. „Estimation of many change-points in the long autoregressive sequences“, *Informatica* 3(1): 37–46.

[52] Lipeika, A.; Lipeikienė, J. 1993. „Speaker identification“, *Informatica* 4(1–2): 45–56.

[53] Lipeika, A.; Lipeikienė, J. 1995. „Speaker identification using vector quantization“, *Informatica* 6(2): 167–180.

[54] Lipeika, A.; Lipeikienė, J. 1996. „Speaker identification methods based on pseudostationary segments of voiced sounds“, *Informatica* 7(4): 469–484.

[55] Lipeika, A.; Lipeikienė, J. 1999. „Speaker recognition based on the use of vocal tract and residue signal LPC parameters“, *Informatica* 10 (4): 377–388.

[56] Lipeika, A.; Lipeikienė, J. 2003. „Word endpoint detection using dynamic programming“, *Informatica* 14(4): 487–496.

[57] Lipeika, A.; Lipeikienė, J.; Telksnys, L. 2002. „Development of isolated word speech recognition system“, *Informatica* 13(1): 37–46.

[58] Liporace, L. A. 1982. „Maximum likelihood estimation for multivariate observations of Markov sources“, in *IEEE Trans. on Information Theory* 28: 729–734.

[59] Lockwood, P.; Blanchet, M. 1993. „An algorithm for the dynamic inference of hidden Markov models (DIHMM)“, in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP '93* 2: 251–254.

[60] McLennan, C. T.; Luce, P. A.; Luce, J. C. 2003. „Representation of lexical form“, in *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29(4): 539–553.

[61] Murviet, H.; Weintraub, M. 1988. „1000-word speaker-independent continuous-speech recognition using hidden Markov models“, in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP '88* 1: 1115–1118.

[62] Noreika, S.; Rudžionis, A. 1991. „Phoneme-like model of speech signal“, in *Proc. of the XIIth International Congress of Phonetic Sciences* 4: 490–493.

[63] Ostendorf, M.; Digalakis, V.; Kimball, O. A. 1996. „From HMM's to segment models: a unified view of stochastic modeling for speech

recognition“, in *IEEE Trans. on Speech and Audio Processing* 4(5): 360–378.

[64] Pakerys, A. 2003. *Lietuvių bendrinės kalbos fonetika*. Vilnius: Enciklopedija, 241 p.

[65] Picone, J. 1993. „Signal modeling techniques in speech recognition“, in *Proc. of the IEEE* 81(9): 1214–1247.

[66] Rabiner, L. R.; Levinson, S. E. 1981. „Isolated and connected word recognition – theory and selected applications“, in *IEEE Trans. on Communications* 29(5): 621–659.

[67] Rabiner, L. R. 1989. „A tutorial on Hidden Markov Models and selected applications in speech recognition“, in *Proc. of IEEE* 77: 257–286.

[68] Rabiner, L. R.; Juang, B. H. 1993. *Fundamentals of speech recognition*. Prentice-Hall, Inc, 496 p.

[69] Raškinis, A.; Raškinis, G.; Kazlauskienė, A. 2003a. „VDU bendrinės lietuvių šnekos universalus anotuotas garsynas“. Konferencijos „*Informacinės technologijos 2003“* pranešimų medžiagoje, IX–(28–34).

[70] Raškinis, A.; Raškinis, G.; Kazlauskienė, A. 2003b. „Speech assessment methods phonetic alphabet (SAMPA) for encoding transcriptions of Lithuanian speech corpora“, in *Information Technology and Control* 4(29): 52–55.

[71] Raškinis, G.; Raškinienė, D. 2003a. „Lietuvių šnekos atpažinimo sistemos, pagrįstos paslėptais Markovo modeliais, parametrų tyrimas ir optimizacija“. Konferencijos „*Informacinės technologijos 2003“* pranešimų medžiagoje, IX–(41–48).

[72] Raškinis, G.; Raškinienė, D. 2003b. „Building medium-vocabulary isolated-word Lithuanian HMM speech recognition system“, *Informatica* 14(1): 75–84.

[73] Raškinis, G.; Raškinienė, D. 2004. „Trumpų rišlių lietuvių šnekos frazių atpažinimas, naudojant paslėptų Markovo modelių metodiką“. Konferencijos „*Informacinės technologijos 2004“* pranešimų medžiagoje, 33–38.

[74] Reichl, W.; Chou, W. 1999. „A unified approach of incorporating general features in decision tree based acoustic modeling“, in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP'99* 2: 573–576.

[75] Rhys, J. J.; Downey, S.; Mason, J. S. 1997. „Continuous speech recognition using syllables“, in *Proc. of European Conference on Speech Communication and Technology „Eurospeech'97"* 3: 1171–1174.

[76] Rudžionis, A. 1987. „Computer recognition of isolated words in fixed length feature space“, in *Proc. of the XIth International Congress of Phonetic Sciences* 5: 255–258.

[77] Rudžionis, A.; Rudžionis, V. 1999. „Phoneme recognition in fixed context using regularised discriminant anglysis“, in *Proc. of Int. Conf. on Speech Communication and Technology „Eurospeech'99"*, 2745–2748.

[78] Rudžionis, A.; Rudžionis, V. 1995. „Phonetical segmentation and averaging of the utterances in speech recognition“, in *Proc. of COST 250 Speaker Recognition in Telephony (Draft Minutes of 3rd Management Comimitee Meeting)*, 62–65.

[79] Sakoe, H.; Isotani, K.; Yoshida, K. I.; Watanabe, T. 1989. „Speaker-independent word recognition using dynamic programming neural networks“, in *Readings in Speech Recognition,* ed. A. Waibel, K.-F. Lee, 439–442.

[80] Sethy, A.; Narayanan, Sh.; Parthasarthy, S. 2002. „A syllable based approach for improved recognition of spoken names“, in *Proc. of the ISCA Pronuncation Modeling Workshop*, <http://www.clsp.jhu.edu/pmla2002/cd/papers/sethy.pdf>, [Žiūrėta 2005 02 12].

[81] Skripkauskas, M.; Telksnys L. 2006. „Automatic trascription of Lithuanian text using dictionary“, *Informatica* 17(4): 587–600.

[82] Skripkauskas, M. 2006. „Lietuvių šnekos signalų segmentavimas kvazifonemomis“, Konferencijos *„Informacinės technologijos 2006"* pranešimų medžiagoje, 76–80.

[83] Smyth, P. 1997. „Clustering sequences with hidden Markov models“, in *Advances in Neural Information Processing Systems* 9: 648–655.

[84] Sphinx. <http://cmusphinx.sourceforge.net/html/cmusphinx.php>, [Žiūrėta 2007 01 16]

[85] Šilingas, D.; Raškinis, G.; Telksnys, L. 2004a. „Review of Lithuanian speech and language processing“, in *Proc. of Int. Conf on Human Language Technologies – The Baltic Perspective*, 144–150.

[86] Šilingas, D.; Telksnys, L. 2004. „Specifics of hidden Markov model modifications for large vocabulary continuous speech recognition“, *Informatica* 15(1): 93–110.

[87] Šilingas, D.; Laurinčiukaitė, S.; Telksnys, L. 2004b. „Towards acoustic modeling of Lithuanian speech“, in *Proc. of Int. Conf. on Speech and Computer SPECOM'04,* 326–333.

[88] Šilingas, D. 2005. *Choosing acoustic modeling units for Lithuanian continuous speech recognition based on hidden Markov models*, daktaro disertacija, Kaunas, 162 p.

[89] Šilingas, D.; Laurinčiukaitė, S.; Telksnys, L. 2006. „A technique for choosing efficient acoustic modeling units for Lithuanian continuous speech recognition“, in *Proc. of Int. Conf. on Speech and Computer (SPECOM'06)*, 61–66.

[90] Strimaitis, M. 2004. „Perėjimo momentų tarp fonemų akustinis modeliavimas naudojant paslėptus Markovo modelius“. Konferencijos „*Informacinės technologijos 2004"* pranešimų medžiagoje, 51–55.

[91] Sugamura, N.; Shikano, K.; Furui, S. 1983. „Isolated word recognition using phoneme-like templates“, in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP'83* 8: 723–726.

[92] Šveikauskienė, D. 2005. „Graph representation of the syntactic structure of the Lithuanian sentence“, *Informatica* 16(3): 407–418.

[93] Telksnys, L. 1987. „Recognition of nonstationary random processes“, in *Proc. of IFAC* 2: 517.

[94] Vaičiūnas, A.; Raškinis, G. 2003. „Statistinių lietuvių kalbos modelių kūrimas ir pirminis tyrimas“. Konferencijos „*Informacinės technologijos 2003"* pranešimų medžiagoje, IX–(35–40).

[95] Vaičiūnas, A.; Raškinis, G. 2005a. „Review of statistical modeling of highly inflected Lithuanian using very large vocabulary“, in *Proc. of Int. Conf. Interspeech*, 1321–1324.

[96] Vaičiūnas, A.; Raškinis G. 2005b. „Statistinis lietuvių kalbos modeliavimas, grupuojant tekstus į žanrus“. Konferencijos „*Informacinės technologijos 2005“* pranešimų medžiagoje, 309–314.

[97] Vaičiūnas, A.; Raškinis, G.; Kaminskas, V. 2004. „Statistical language models of Lithuanian based on word clustering and morphological decomposition“, *Informatica* 15(4): 565–580.

[98] Vaitkevičiūtė, V. 2001. *Lietuvių kalbos tarties pagrindai ir žodynas*. Vilnius, 1241 p.

[99] Veeravalli, A. G.; Pan, W. D.; Adhami, R.; Cox, P. G. 2005. „A tutorial on using hidden Markov models for phoneme recognition“, in *Proc. of Conf. on System Theory SSST‘05*, 154–157.

[100] Young, S. J.; Russel, N. H.; Thornton, J. H. S. 1989. „Token passing: a conceptual model for connected speech recognition systems“, Technical Report CUED/F-INFENG/TR38, <svr-ftp.eng.cam.ac.uk>.

[101] Young, S.; Evermann, G.; Kershaw, D.; Motore, G.; Odell, J.; Ollason, D.; Valtchev, V.; Woodland, Ph. 2003. *The HTK Book.* Cambridge University Engineering Department Speech Group, 354 p.

[102] Zinkevičius, V. 2000. „Morfologinė analizė su Lemuokliu“, žurnale *Darbai ir dienos* 24: 245–273.

Autorės publikacijų sąrašas disertacijos tema

Straipsniai Mokslinės informacijos instituto pagrindinio sąrašo (*Thomson ISI Master Journal List*) leidiniuose

[1A] Laurinčiukaitė, S.; Lipeika, A. 2007. „Framework for Choosing a Set of Syllables and Phonemes for Lithuanian Speech Recognition“, *Informatica* 18(3): 395–406, ISSN 0868-4952.

Straipsniai Mokslinės informacijos instituto pagrindinio sąrašo (*Thomson Scientific (ISI)*) leidiniuose

[2A] Laurinčiukaitė, S.; Lipeika, A. 2006. „Syllable-Phoneme based Continuous Speech Recognition“, in *Electronics and Electrical Engineering* 6(70): 91–94, ISSN 1392–1215.

Straipsniai *INSPEC* tarptautinėje duomenų bazėje referuojamuose leidiniuose

[3A] Laurinčiukaitė, S.; Šilingas, D.; Skripkauskas, M.; Telksnys, L. 2006. „Lithuanian Continuous Speech Corpus LRN 0.1: Design and Potential Applications“, in *Information Technology and Control* 4: 431–440, ISSN 1392-124X.

Straipsniai kitose tarptautinių ir respublikinių konferencijų medžiagose

[4A] Šilingas, D.; Laurinčiukaitė, S.; Telksnys, L. 2004. „Towards Acoustic Modelling of Lithuanian Speech“, in *Proceedings of International Conference on Speech and Computer „SPECOM 2004“*, held in Sankt Petersburg on 20–22 September 2004 (Tarptautinės konferencijos „SPECOM 2004“, įvykusios Sant Peterburge 2004 m. rugsėjo 20–22 d., medžiaga). Sankt Petersburg: Anatolya, 2004, 326–333, ISBN 5-7452-0110-X.

[5A] Šilingas, D.; Laurinčiukaitė, S.; Telksnys, L. 2006. „A Technique for Choosing Efficient Acoustic Modelling Units for Lithuanian Continuous Speech Recognition“, in *Proceedings of International Conference on Speech and Computer „SPECOM 2006“*, held in Sankt Petersburg on 25–29 June 2006 (Tarptautinės konferencijos „SPECOM 2006“, įvykusios Sant Peterburge 2006 m. birželio 25–29 d., medžiaga). Sankt Petersburg: Anatolya, 2006, 61–66, ISBN 5-7452-0074-X.

[6A] Laurinčiukaitė, S. 2004. „On Different Kinds of Speech Units based Isolated Words Recognition of Lithuanian Language“, in *Proceedings of the First Baltic Conference „Human language technologies – The Baltic Perspective“*, held in Riga on 21–22 April 2004 (Tarptautinės konferencijos „Human language technologies – The Baltic Perspective“, įvykusios Rygoje 2004 m. balandžio 21–22 d., medžiaga). Riga: Data Media Groupp, 2004, 139–143.

A PRIEDAS

Garsyno LRN fonetinė sistema

1A lentelė. *Balsiai*

Raidinis simbolis	Garsyno simbolis	Balsių variantai (paprasti, kirčiuoti)	Transkribuoto žodžio pavyzdys (žodis)*
Balsiai (trumpi) /5/			
a	a	a a0	**a** k' i0 s (**a**kìs) p **a0** s' m' er' k' e3: (p**à**smerkė)
e	e	e e0	d' **e** v' in1 t a: (d**e**viñtą) p a t' **e0** g d a v o: (pat**è**kdavo)
i	i	i i0	e0 t' **i** k o: s (èt**i**kos) a p' **i0** p' l' e3: s2' e3: (ap**ì**plėšė)
o	o	o o0	**o** p **o** z' i0 c' i j' a (**o**p**o**zìcija) a t **o0** m' i n' o: (at**ò**minio)
u	u	u u0	p a k' i0 l **u** s (pakìl**u**s) g r **u0** p' e: (gr**ù**pę)
Balsiai (ilgi) /6/			
a	a:	a: a:1	– **a**:1 k' c' i j' a (**ã**kcija)
ą	a:	a: a:1	a:1 k' c' i j' **a:** (ãkcij**ą**) g' ir t **a:1** v' i m o: (girt**ã**vimo)

* Ilgų nekirčiuotų balsių žymėjimo buvo vengiama, dėl nepakankamų žinių, kaip skirti ilguosius ir trumpuosius balsius. Taip pat ne visi išvardinti simboliai pasitaikė garsyne. Dėl šių priežasčių kai kurios lentelės eilutės yra tuščios.

Raidinis simbolis	Garsyno simbolis	Balsių variantai (paprasti, kirčiuoti)	Transkribuoto žodžio pavyzdys (žodis)*
e	e:	e: e:1	– **e:1** s a m a (**ẽ**sama)
ę	e:	e: e:1 e:2	d' i0 d' e l' **e:** (dìdel**ę**) t' **e:1** s' t' i (t**ę́**sti) b' r' **e:2** s t an' c2' u s (br**ę́**stančius)
ė	e3:	e3: e3:1 e3:2	a0 p t a r' **e3:** (àptar**ė**) s u k' r' **e3:1** t u s' o: (sukr**ė̃**tusio) s' v' **e3:2** r' e3: (sv**ė́**rė)
į	i:	i: i:1 i:2	s' k' r' i:1 d' **i:** (skrỹd**į**) s u g' r' **i:1** z2 u s' i (sugr**į̃**žusi) **i:2** g u l a (**į́**gula)
y	i:	i: i:1 i:2	g' i:2 d' **i:** t o: j' a (gỳd**y**toja) s2 o: v' i n' **i:1** s (šovin**ỹ**s) p u s2' **i:2** n a s (pušy**́**nas)
o	o:	o: o:1 o:2	p a:1 j' a m **o:** s (pãjam**o**s) p r **o:1** t o: (pr**õ**to) p r **o:2** g a (pr**ó**ga)
ū	u:	u: u:1 u:2	n uo d u0 g n **u:** s (nuodugn**ū**s) p a r' ei g **u**:1 n ai (pareig**ū̃**nai) p r a d **u:2** r' e3: (prad**ū́**rė)
ų	u:	u: u:1 u:2	m' i n' e r a:1 l **u:** (minerãl**ų**) p a s' **u:1** s' t' i (pasi**ų̃**sti) –

2A lentelė. *Priebalsiai*

Raidinis simbolis	Garsyno simbolis (paprastas, minkštas)	Transkribuoto žodžio pavyzdys (žodis)
Piebalsiai (sprogstamieji) /6 poros/		
p	p p’	a0 **p** t ar t o: s (à**p**tartos) a **p'** t' i0 k o: (a**p**tìko)
b	b b’	a r a:1 **b** u: (arã**b**ų) ar' **b'** i0 t r a s (ar**b**ìtras)
t	t t’	c' en1 **t** r o: (ceñ**t**ro) c' en' **t'** r' e0 (cen**t**rė)
d	d d’	d' e:1 **d** a (dẽ**d**a) **d'** e3:1 **d'** e3: (**d**ẽ**d**ė)
k	k k’	g r a:1 f' i **k** u s (grãfi**k**us) f r a:1 **k'** c' i j' a (frã**k**cija)
g	g g’	i:2 s t ai **g** a (į́stai**g**a) i: **g'** i:2 t' i (į**g**ýti)

* Ilgų nekirčiuotų balsių žymėjimo buvo vengiama, dėl nepakankamų žinių, kaip skirti ilguosius ir trumpuosius balsius. Taip pat ne visi išvardinti simboliai pasitaikė garsyne. Dėl šių priežasčių kai kurios lentelės eilutės yra tuščios.

Raidinis simbolis	Garsyno simbolis (paprastas, minkštas)	Transkribuoto žodžio pavyzdys (žodis)
Priebalsiai (afrikatos) /4 poros/		
c	c c’	p r an **c** u:1 z ai (pran**c**ūzai) i m' i t a:1 **c'** i j' a (imitā**c**ija)
č	c2 c2’	g' in1 **c2** o: (giñčo) p' r' i v a **c2'** uo s' e0 (priva**č**iuosè)
dz	dz dz’	– k u **dz'** i: s (Ku**dz**ys)
dž	dz2 dz2’	**dz2** o0 r **dz2** a s (**Dž**òr**dž**as) a t' i **dz2'** u0 s (ati**dž**iùs)
Priebalsiai (pučiamieji) /7 poros/		
f	f f’	in **f** o r m a:1 c' i j' a (in**f**ormãcija) d' el' **f'** i0 n u: (del**f**inų)
s	s s’	in' t' e r' e0 **s** am **s** (interè**s**am**s**) in' t' en' **s'** i: v' i0 (inten**s**yvì)
š	s2 s2’	k a:2r **s2** t o: (kár**š**to) k ar'1 **s2'** t' i s (kar**š**tis)
z	z z’	l a **z** d o: m' i0 s (la**z**domìs) k a **z'** i n o0 (ka**z**inò)
ž	z2 z2’	m a **z2** a m' e0 (ma**ž**amè) m a **z2'** e3:2 j' a (ma**ž**éja)
ch	ch ch’	t' e **ch** n o l o0 g' i j' o: s (te**ch**nològijos) p' s' i0 **ch'** i k o: s (psì**ch**ikos)
h	h h’	al k o **h** o0 l' o: (alko**h**òlio) b e0 k' **h'** e m a s (Bèk**h**emas)
Priebalsiai (sklandieji) /5 poros + 1/		
v	v v’	b u0 **v** o: (bùvo) a0 p' **v'** er' t' e3: (àp**v**erte3)
m	m m’	c' e r' e **m** o0 n' i j' a (cere**m**ònija) ch' e0 **m'** i j' o: s (chè**m**ijos)
n	n n’ w w’	d ai **n** o: m' i0 s (dai**n**omìs) f' i n a:1 l' i **n'** e3: s (finãli**n**ės) v' ie **w** k ar' t' i0 n' i s (vie**n**kartìnis) k o **w'** k' r' e t' i0 (ko**n**kretì)
l	l l’	b a:1 **l** o: (bã**l**o) a p' **l'** iw1 k (ap**l**iñk)
r	r r’	c u0 k **r** au s (cùk**r**aus) b a **r'** e0 (ba**r**è)
j	j'	g al v o: **j'** e0 (galvo**j**è)

3A Lentelė. *Dvigarsiai*

Raidinis simbolis	Garsyno simbolis (paprastas, minkštas)	Dvigarsių variantai (paprasti, kirčiuoti)
Dvibalsiai /7/		
ai	ai	ai (ai), a:2i (ái), ai:1 (aĩ)
au	au	au (au), a:2u (áu), au:1 (aũ),
ei	ei	ei (ei), e:2i (éi), ei:1(eĩ)
Raidinis	**Garsyno**	**Dvigarsių variantai (paprasti, kirčiuoti)**

simbolis	simbolis (paprastas, minkštas)	
eu	eu	eu (eu), e0u (èu)
ui	ui	ui (ui), u0i (ùi), ui:1 (uĩ)
ie	ie	ie (ie), i:2e (íe), ie:1 (iẽ)
uo	uo	uo (uo), u:2o (úo), uo:1 (uõ),
Mišrūs dvigarsiai /16/		
al	al al’	al (al), a:2l (ál), al1 (alĩ) al’ (al’), a:2l’ (ál’), al’1 (alĩ’)
am	am am’	am (am), a:2m (ám), am1 (amm̃) am’ (am’), a:2m’ (ám’), am’1 (amm̃’))
an	an an’ aw aw’	an (an), a:2n (án), an1 (añ) an’ (an’), a:2n’ (án’), an’1 (añ’) aw (aw), a:2w (án), aw1 (añ) aw’ (aw’), a:2w’ (án’), aw’1 (añ’)
ar	ar ar’	ar (ar), a:2r (ár), ar1 (arr̃) ar’ (ar’), a:2r’ (ár’), ar’1 (arr̃’)
el	el el’	el (el), e:2l (él), el1 (elĩ) el’ (el’), e:2l’ (él’), el’1 (elĩ’)
em	em em’	em (em), e:2m (ém), em1 (emm̃) em’ (em’), e:2m’ (ém’), em’1 (emm̃’)
en	en en’ ew ew’	en (en), e:2n (én), en1 (eñ) en’ (en’), e:2n’ (én’), en’1 (eñ’) ew (ew), e:2w (én), ew1 (eñ) ew’ (ew’), e:2w’ (én’), ew’1 (eñ’)
er	er er’	er (er), e:2r (ér), er1 (err̃) er’ (er’), e:2r’ (ér’), er’1 (err̃’)
il	il il’	il (il), i0l (ìl), il1 (ilĩ) il’ (il’), i0l’ (ìl’), il’1 (ilĩ’)
im	im im’	im (im), i0m (ìm), im1 (imm̃) im’ (im’), i0m’ (ìm’), im’1 (imm̃’)
in	in in’ iw iw’	in (in), i0n (ìn), in1 (iñ) in’ (in’), i0n’ (ìn’), in’1 (iñ’) iw (iw), i0w (ìn), iw1 (iñ) iw’ (iw’), i0w’ (ìn’), iw’1 (iñ’)
ir	ir ir’	ir (ir), i0r (ìr), ir1 (irr̃) ir’ (ir’), i0r’ (ìr’), ir’1 (irr̃’)
ul	ul ul’	ul (ul), u0l (ùl), ul1 (ulĩ) ul’ (ul’), u0l’ (ùl’), ul’1 (ulĩ’)
um	um um’	um (um), u0m (ùm), um1 (umm̃) um’ (um’), u0m’ (ùm’), um’1 (umm̃’)
un	un un’ uw uw’	un (un), u0n (ùn), un1 (uñ) un’ (un’), u0n’ (ùn’), un’1 (uñ’) uw (uw), u0w (ùn), uw1 (uñ) uw’ (uw’), u0w’ (ùn’), uw’1 (uñ’)
ur	ur ur’	ur (ur), u0r (ùr), ur1 (urr̃) ur’ (ur’), u0r’ (ùr’), ur’1 (urr̃’)

B PRIEDAS

Skiemenų-fonemų aibės H_1 ir H_2

1B lentelė. *Skiemenų-fonemų aibių H_1 ir H_2 elementai*

H_1	H_2	H_1	H_2	H_1	H_2	H_1	H_2
a	a	ben	be3s	ch'	c2ius	dz2	di:
a1	a1	bi	ben	ci	ch	dz2'	die
ai	ai	bo	bi	d	ch'	dz2ia	do
ap	al'	bu	bo	d'	ci	e	dos
ap'	an	bu:	bos	da	d	e1	du
ar'	ap	c	bu	das	d'	e3	duo
at	at	c'	bu:	de	da	ei	dvi
at'	at'	c2	bus	de3	dar	eu	dz'
au	au	c2'	c	di	das	f	dz2
b	b	c2ia	c'	di:	dau	f'	dz2'
b'	b'	c2io	c2	die	de	fe	e
ba	ba	c2iu:	c2'	do	de3	fi	e1
bai	bai	c2ius	c2em'	du	de3l	g	e3
be	be	ce	c2io	duo	den	g'	ei
be3	be3	ch	c2iu:	dz'	di	ga	eu

H_1	H_2	H_1	H_2	H_1	H_2	H_1	H_2
ga1	f	ka	k'	lu	man	niw	p'
gai	f'	ka1	ka	m	mas	no	pa
gas	fi	kai	kai	m'	me	nos	par
ge	g	kal'	kar	ma	me3	nu	pauze
ge3	g'	kas	kas	ma1	men	nu:	per
gi	ga	kau	kat	mai	mi	nuo	pew'
gi:	ge	ke	kau	mas	mie	o	pi
gia	gi	ke3	ke	me	mis	p	pie
go	gi:	ki	ke3	me3	mo	p'	pir
gos	gia	ki:	ki	men	mos	pa	pir'
gra	go	kie	ki:	mi	mu	pauze	po
gu	gos	klau	kie	mis	mu:	pe	pos
gu:	gu	ko	ko	mo	mus	per	pra
h	h	kon	kos	mos	n	per'	pre
h'	h'	kon'	krep'	mu	n'	pi	pri
i	i	kos	ku	mu:	na	po	prie
i:	i:	kri	ku:	mui	na1	pra	pries2
ie	ie	ku	l	mus	nai	pre	pro
ikvepimas	ikvepimas	ku:	l'	n	nas	pri	pu
in'	ir	kus	la	n'	nau	prie	puo
is2	is2	l	lai	na	ne	pro	r
is2'	j'	l'	lan	na1	ne3	pu	r'
j	ja	la	le	nai	ne3s	puo	ra
ja	ja1	lai	le3	nas	nes	r	ras
ja1	jai	le	lei	nau	ni	r'	re
jai	jas	le3	li	ne	ni:	ra	re3
jan'	jau	lei	li:	ne1	nio	re	rei
jas	je	li	lie	ne3	nis	re3	ri
je	ji	li:	lio	ne3s	niu:	rei	ri:
ji	jo	lia	lis	nes	niw	ri	ria
jo	jok	lie	liu:	ni	no	ri:	rio
jos	jos	lik	lo	ni:	nos	ria	ris
ju	ju	lio	m	niai	nu	rio	riu:
ju:	ju:	lis	m'	nie	nu:	ro	rius
jus	juw	liu	ma	nio	nuo	ru	ro
k	juwk'	liu:	ma1	nis	o	ru:	ru
k'	k	lo	mai	niu:	p	ruo	s

H_1	H_2	H_1	H_2	H_1	H_2	H_1	H_2
s	s'	sius	stan'	tie	tos	ve3	vil'
s'	s2	skai	ste	tis	tre	ver'	vo
s2	s2'	so	sti	tiw	tu	vi	vos
s2'	s2a	sta	sti:	tyla	tu:	vi:	w
s2a	s2e	ste	sto	to	tuo	vie	w'
s2e	s2i	sti	su	tos	tus	vo	z
s2i	s2im	sti:	t	tra	u	vu	z'
s2i:	s2imt	sto	t'	tre	u:	w	z2
s2im	sa	su	ta	tu	ui	w'	z2'
sa	sa1	t	ta1	tu:	uo	z	z2e
sa1	sau	t'	tai	tuo	us2	z'	z2i
sau	se	ta	tar'	tus	v	z2	z2i:
se	se3	ta1	tas	u	v'	z2'	z2mo
se3	sei	tai	tau	u:	va	z2a	z2u
si	sep'	tar'	te	ui	vai	z2i	zi
si:	si	tas	te3	uo	val'	z2i:	
sia	sie	tau	tei	v	vas	z2u	
sie	sios	te	ti	v'	ve	za	
sio	siu:	te3	ti:	va	ve3	zi	
sios	skas	tei	tis	vai	vi		
sis	so	ti	tyla	val'	vi:		
siu:	sta	ti:	to	ve	vie		

C PRIEDAS

Mokymo ir atpažinimo procedūros naudojant PMM

Paslėptojo Markovo modelio parametrų įvertinimas

Mokymas yra iteratyvus PMM parametrų, o tiksliau $\lambda = (\mathbf{A}, \mathbf{B})$, vertinimo procesas. Šnekos atpažinimo procesą nusakančioje formulėje (3.6), naudojant tikimybę $P(\mathbf{O} | M)$, buvo tariama, kad visi modeliai $M \in M^*$ yra apmokyti ir minėta tikimybė nusako stebėjimo duomenų atitikimą vienam ar kitam modeliui. Tikimybėje $P(\mathbf{O} | M)$ modelis tampa nežinomu parametru, kurį reikia įvertinti pagal stebėjimus. Tikimybės išraiška keičiama į tikėtinumo išraišką $L(\mathbf{O} | M)$, kurią reikia maksimizuoti. Atliekamas lokalus maksimizavimas, mokymo procesas yra iteratyvus.

Modelio parametrų vertinimui naudojamas matematinės vilties maksimizavimo metodas. Matematinės vilties maksimizavimą atlieka Baum-Welch algoritmas (Baum *et al.* 1970) – atskiras tikėtinumo maksimizavimo (*Expectation-Maximisation*) algoritmo atvejis. Tikėtinumo maksimizavimas gali būti atliekamas naudojant ir kitus kriterijus, kaip: maksimalios tarpusavio informacijos (*Maximum Mutual Information*), maksimalios aposteriorinės tikimybės (*Maximum A Posteriori*) ir minimalios klasifikavimo klaidos (*Minimum Classsification Error*). Darbe taikytas Baum-Welch algoritmas, todėl toliau pateikiamas trumpas jo aprašymas.

Baum-Welch algoritmas, atlikdamas minėto tikėtinumo $L(\mathbf{O} | M)$ maksimizavimą, suranda modelio *M* parametrus $\lambda = (\mathbf{A}, \mathbf{B})$, kurie maksimizuoja tikėtinumą ankstesnėje iteracijoje surastojo atžvilgiu, t. y.:

$$L\left(\mathbf{O}\,|\,\hat{M}\right) \geq L\left(\mathbf{O}\,|\,M\right), \tag{C.1}$$

čia $\hat{M}$ – modelis su pakitusiais įverčiais $\hat{\lambda} = \left(\hat{\mathbf{A}}, \hat{\mathbf{B}}\right)$.

Prieš pateikiant pastarųjų parametrų perskaičiavimo formules, reikia įvesti tiesiogines-atbulines tikimybes, prastinančias skaičiavimo formules. Tiesioginė tikimybė $\alpha_j(t)$ – tikimybė, kad dalinė stebėjimų seka laiko intervale $[1,t]$ yra $\mathbf{o}_1,...,\mathbf{o}_t$, o modelio būsena momentu t yra j. Šioms tikimybėms skaičiuoti naudojama tokia formulė:

$$\alpha_j(t) = P\left(\mathbf{o}_1,...,\mathbf{o}_t, q_t = j \,|\, M\right) = \left[\sum_{i=2}^{N-1} \alpha_i(t-1)\, a_{ij}\right] b_j\left(\mathbf{o}_t\right). \tag{C.2}$$

Kadangi pirmoji ir paskutinė modelio būsenos yra neemituojančios, atskirai apibrėžiamos pradinės ir galutinės tikimybės:

$$\begin{aligned} &\alpha_1(1) = 1, \\ &\alpha_j(1) = a_{1j} b_j\left(\mathbf{o}_1\right), \quad 1 < j < N, \\ &\alpha_N(T) = \sum_{i=2}^{N-1} \alpha_i(T)\, a_{iN}. \end{aligned} \tag{C.3}$$

Analogiškai skaičiuojama ir atbulinė tikimybė $\beta_j(t)$ – tikimybė, kad dalinė stebėjimų seka laiko intervale $[t;T]$ yra $\mathbf{o}_{t+1},...,\mathbf{o}_T$, o modelio būsena laiko momentu t yra j. Šios tikimybės skaičiuojamos taip:

$$\beta_j(t) = P\left(\mathbf{o}_{t+1},...,\mathbf{o}_T \,|\, q_t = j, M\right) = \sum_{j=2}^{N-1} a_{ij} b_j\left(\mathbf{o}_{t+1}\right) \beta_j(t+1). \tag{C.4}$$

Kadangi pirmoji ir paskutinė modelio būsenos taip pat yra neemituojančios, atskirai apibrėžiamos pradinės ir galutinės tikimybės:

$$\begin{aligned} &\beta_i(T) = a_{iN}, \quad 1 < i < N, \\ &\beta_1(1) = \sum_{j=2}^{N-1} a_{1j} b_j\left(\mathbf{o}_1\right) \beta_j(1). \end{aligned} \tag{C.5}$$

Reikia apibrėžti $\xi_{ij}(t)$ – tikimybę, kad laiko momentu t modelio būsena buvo i, o laiko momentu $t+1$ – būsena j. Šią tikimybę galima išreikšti per tiesioginę ir atbulinę tikimybes:

$$\xi_{ij}(t)=\frac{P(q_t=i,q_{t+1}=j,\mathbf{O}\mid M)}{P(\mathbf{O}\mid M)}=\frac{\alpha_i(t)a_{ij}b_j(\mathbf{o}_{t+1})\beta_j(t+1)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i(t)a_{ij}b_j(\mathbf{o}_{t+1})\beta_j(t+1)}. \quad \text{(C.6)}$$

Įvedamas kintamasis $\gamma_i(t)$ – buvimo laiko momentu *t* modelio būsenoje *i*, o momentu *t+1* būsenoje *j* skaičius su tikimybe $\xi_{ij}(t)$ siejamas taip:

$$\gamma_i(t)=\sum_{j=1}^{N}\xi_{ij}(t). \quad \text{(C.7)}$$

Sumuojant dydžius pagal laiko indeksą *t*, gaunama:

$\sum_{t=1}^{T-1}\gamma_i(t)$ – buvimo būsenoje *i* skaičius;

$\sum_{t=1}^{T-1}\xi_{ij}(t)$ – perėjimų iš būsenos *i* į būseną *j* skaičius.

Naudojant šias išraiškas modelio *M* parametrai $\lambda=(\mathbf{A},\mathbf{B})$ yra perskaičiuojami taip:

$$\hat{a}_{ij}=\frac{\sum_{t=1}^{T-1}\xi_{ij}(t)}{\sum_{t=1}^{T-1}\gamma_i(t)}, \quad \text{(C.8)}$$

$$\hat{b}_j(k)=\frac{\sum_{\substack{t=1\\ o_t=v_k}}^{T}\gamma_j(t)}{\sum_{t=1}^{T}\gamma_j(t)}. \quad \text{(C.9)}$$

Įvedus tikimybę $\mathrm{L}_{jm}(t)=\frac{1}{L(\mathbf{O}\mid M)}\alpha_j(t)\beta_j(t)$, konkrečių parametrų perskaičiavimams taikomos šios formulės (Liporace 1982, Juang 1985):

$$\hat{\mu}_{jm}=\frac{\sum_{t=1}^{T}\mathrm{L}_{jm}(t)\mathbf{o}_t}{\sum_{t=1}^{T}\mathrm{L}_{jm}(t)}, \quad \text{(C.10)}$$

$$\hat{\Sigma}_{jm} = \frac{\sum_{t=1}^{T} \mathrm{L}_{jm}(t)\left(\mathbf{o}_t - \mu_{jm}\right)\left(\mathbf{o}_t - \mu_{jm}\right)^T}{\sum_{t=1}^{T} \mathrm{L}_{jm}(t)}, \tag{C.11}$$

$$\hat{c}_{jm} = \frac{\sum_{t=1}^{T} \mathrm{L}_{jm}(t)}{\sum_{t=1}^{T} \mathrm{L}_{j}(t)}, \tag{C.12}$$

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \alpha_i(t)\, a_{ij} b_j\left(\mathbf{o}_{t+1}\right)\beta_j(t+1)}{\sum_{t=1}^{T-1} \alpha_i(t)\beta_j(t)}, \quad 1 < i, j < N\,. \tag{C.13}$$

čia $\hat{\mu}_{jm}$ – vidurkių vektoriaus įverčiai *j*-tosios būsenos *m*-jam mišiniui, $\hat{\Sigma}_{jm}$ – kovariacijos matricos įverčiai *j*-tosios būsenos *m*-jam mišiniui, $\hat{c}_{jm}$ – *m*-jo mišinio svorinio koeficiento įvertis *j*-jai būsenai, $\hat{a}_{ij}$ – perėjimo iš būsenos *i* į būseną *j* tikimybės įvertis. Perėjimų iš ir į neemituojančias būsenas tikimybės perskaičiuojamos taip:

$$\begin{aligned} \hat{a}_{1j} &= \frac{1}{L(\mathbf{O} \mid M)} \alpha_j(1)\beta_j(1), \\ \hat{a}_{iN} &= \frac{\alpha_j(T)\beta_j(T)}{\sum_{t=1}^{T} \alpha_i(t)\beta_i(t)}. \end{aligned} \tag{C.14}$$

Atpažinimo procedūra naudojant paslėptąjį Markovo modelį

Atpažinimo procedūros (dar vadinamos paieška) tikslas – rasti labiausiai tikėtiną ištarimą, atitinkantį visus akustinius ir lingvistinius apribojimus. Ši paieška gali būti integruoto ar modulinio pobūdžio, t. y.:

- Integruotu požiūriu atpažinimo sprendimas priimamas naudojant visus žinių šaltinius (akustika, žodynas, sintaksė ir semantika) kartu. Tai pasiekiama juos sukompiliavus į vieną baigtinių būsenų tinklą, sudarytą iš akustinių PMM būsenų. Šnekos atpažinimo procedūra tampa

požymių vektorių sekos palyginimo su visomis akustinių būsenų sekomis ir labiausiai tikėtinos žodžių sekos, keliaujant šiuo tinklu, radimo problema. Tai yra šiuo metu labiausiai paplitusi paieškos strategija.

- Modulinei paieškai būdinga nuosekli akustinio ir leksinio atitikimo, sintaksinės ir semantinės analizės seka. Taip kiekvienas modulis gali būti kuriamas, tikrinamas ir optimizuojamas atskirai.

Šiame darbe naudota atpažinimo procedūra yra integruoto pobūdžio.

Ankstesnėje dalyje PMM parametrų įvertinime buvo maksimizuojamas tikėtinumas $L(\mathbf{O}|M)$, t. y. pagal duotą stebėjimų seką $\mathbf{O}$ įvertinti modelio M parametrai $\lambda=(\mathbf{A},\mathbf{B})$, maksimizuojantys $L(\mathbf{O}|M)$. Atpažinimo uždavinyje iš duotų modelių aibės reikia išrinkti vieną modelį, labiausiai atitinkantį stebėjimų seką $\mathbf{O}$. 3.3 poskyryje buvo pateikta išraiška, kaip apskaičiuojama konkretaus modelio tikimybė $P(\mathbf{O}|M)$. Šioje išraiškoje yra visos įmanomos būsenų sekos, o tai sunkina tiesioginį tikimybės skaičiavimą. Dėl to vieno modelio atpažinimą reikia sieti su labiausiai tikėtinos būsenų sekos suradimu pagal stebėjimus. Daugiausiai tam taikomas Viterbi algoritmas (Rabiner 1989), panašus į tiesioginės tikimybės skaičiavimo procedūrą. Viterbi algoritmas dar papildomas sub-optimalia spindulio (*beam search*) paieškos strategija, kuria apribojamas atpažinimo paieškos laukas ir paieška tampa greitesnė.

Tariame, kad modelis M – duotas, stebimų seka yra $\mathbf{O}$. Apibrėžiama labiausiai tikėtina būsenų sekos tikimybė $\varphi_i(t)$ laiko momentu t esant būsenoje i. $\varphi_i(t)$ laiko momentui t ir būsenai q_i yra didžiausias iš visų ir nurodo vieną būsenų seką $q_1,q_2,...,q_i$, kuri geriausiai atitinka duotą stebėjimų seką $\mathbf{o}_1,...,\mathbf{o}_t$:

$$\varphi_i(t)=\max_{q_1,q_2,...,q_{t-1}}\left[q_1,q_2,...,q_{t-1},q_t=i,\mathbf{o}_1,\mathbf{o}_2,...,\mathbf{o}_t\,|\,M\right]. \qquad \text{(C.15)}$$

Įvedamas masyvas $\psi_i(t)$, kuriame įsimenami kiekvieną būsenų trajektoriją sudarančių būsenų numeriai. Procedūra susideda iš:

Inicializacijos:

$$\varphi_i(1)=\pi_i b_i(\mathbf{o}_1),\quad i=1,...,N,$$
$$\psi_i(1)=0.$$

Rekursijos:

$$\varphi_j(t)=\max_{1\le i\le N}\left[\varphi_i(t-1)a_{ij}\right]b_j(\mathbf{o}_t),\quad t=2,...,T,\quad j=1,...,N,$$
$$\psi_j(t)=\arg\max_{1\le i\le N}\left[\varphi_i(t-1)a_{ij}\right],\quad t=2,...,T,\quad j=1,...,N.$$

Pabaigos:

$$P = \max_{1 \le i \le N} \varphi_i(t),$$
$$q_T = \arg\max_{1 \le i \le N} \varphi_i(t).$$

Būsenų sekos suradimo: $q_t = \psi_{t+1}(t+1), \quad t = T-1,...,1.$

Praktiškai $\varphi_i(t)$ skaičiavimas keičiamas logaritminio tikėtinumo skaičiavimu, o vietoje Viterbi algoritmo taikomas alternatyvus žymės perdavimo (*token passing*) algoritmas (Young 1989). Pagal šią schemą kiekviena būsena turi su ja asocijuotą žymę. Šioje žymėje yra laikoma būsenų seka $q_1, q_2, ..., q_i$, kuri buvo naudojama jai pasiekti, ir dalinis sekos tikėtinumas $\varphi_i(t)$. Stebint naują požymių vektorių, šios žymės yra atnaujinamos ir nuosekliai pasiunčiamos visoms modelio būsenoms. Labiausiai tikėtinos žymės parametrai apskaičiuojami paskutinėje modelio būsenoje ir ieškotas maksimalus tikėtinumas $L(\mathbf{O} \mid M)$ yra lygus:

$$\varphi_N(T) = \max_{1 \le i \le N} \left[\varphi_i(T) a_{iN} \right]. \tag{C.16}$$

Žymės perdavimo algoritmas kiekvienoje būsenoje leidžia saugoti *N* žymių. Jei žymių daugiau, paliekamos *N* geriausių. Nagrinėjant ištisinę šneką, būsenų su žymėmis tinklai yra dideli. Mažinant skaičiavimus nagrinėjamos tik tos žymės, kurios turi didžiausią potencialą išlikti, t. y. kiekvienu momentu *t* yra išrenkama geriausia tikimybės atžvilgiu žymė ir pagal jos tikimybę nustatomas spindulys (*beam width*), kuriame esančios žymės bus pagrindinės. Visos likusios žymės nėra nagrinėjamos. Toks žymių atrinkimo mechanizmas vadinamas karpymu (*pruning*). Karpymas atliekamas ne būsenų, o modelių lygiu. Atskiras yra spindulio ilgio parinkimo klausimas. Spinduliui ilgėjant didėja skaičiavimo sąnaudos, trumpėjant – didėja tikimybė, kad įvyks paieškos klaida[17].

[17] Paieškos klaida žymės perdavimo algoritme įvyksta, kai paieškos spindulys yra trumpas, labiausiai tikėtinas kelias bus atmestas anksčiau, nei žymė pasieks ištarimo pabaigą.