ISSN 1648-8776
JAUNŲJŲ MOKSLININKŲ DARBAI. Nr. 1 (45). 2016 Doi: 10.21277/jmd.v1i45.40

PASLĖPTŲJŲ MARKOVO MODELIŲ METODO TYRIMAS IR TAIKYMAS BALSO ĮRAŠAMS STENOGRAFUOTI

Jūratė Vaičiulytė, Gražvydas Felinskas
Šiaulių universiteto Informatikos, matematikos ir e. studijų institutas

Įvadas

Kalba – tai natūraliausia ir lengviausia žmonių bendravimo forma. Automatinės šnekos atpažinimo (toliau AŠA) sistemos apdoroja žmogaus tariamus žodžius ir konvertuoja juos į tekstinį atitikmenį. Šios sistemos naudojamos įvairiose srityse: pavyzdžiui, balsu valdomos vartotojo sąsajos, balsu įvedami duomenys, atliekamas automatinis telefono skambučių apdorojimas, stenografavimas ir kt. [1], [10]. AŠA sistemos pritaikytos plačiai vartojamoms kalboms. Lietuvių kalbos struktūra yra gana sudėtinga, jai netinka kitoms kalboms sukurti akustiniai modeliai. Todėl svarbu kurti lietuvių kalbos atpažinimo sistemas, kad šios kalbos vartojimas nebūtų išstumtas iš moderniųjų technologijų [3], [4], [7]. AŠA sistemas būtų galima suskirstyti į dvi pagrindines kategorijas: programos, galinčios interpretuoti atpažinimo rezultatus (komandų ir kontrolės programos), ir programos, neinterpretuojančios atpažinto teksto (diktavimo programos). Balso įrašų stenografavimo įrankis priklausytų antrajai kategorijai. Tokius įrankius dažnai naudoja gydytojai, žurnalistai, kai reikia išsaugoti balsu sakomas pastabas [6].

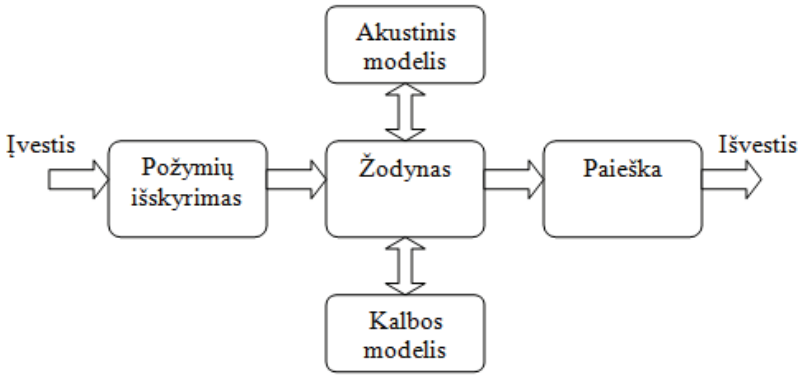
Tyrimo tikslas – išanalizuoti paslėptųjų Markovo modelių metodą ir algoritmus, taikomus siekiant automatinio šnekos atpažinimo, ir pritaikyti juos kuriant programinio įrankio prototipą, skirtą garso įrašams stenografuoti.

***Uždaviniai*:**

- Išanalizuoti paslėptųjų Markovo modelių metodą ir mokslo darbus apie lietuvių kalbos automatinį šnekos atpažinimą.
- Suprojektuoti automatinio šnekos atpažinimo modelį, taikant paslėptųjų Markovo modelių metodą.
- Apmokyti akustinius modelius ir sukurti programinio įrankio prototipą, skirtą garso įrašams stenografuoti.

Automatinio šnekos atpažinimo sistemos veikimo principas

Automatinio šnekos atpažinimo sistemos gautą įvestį – garso signalą – apdoroja ir išveda tekstinį įvesties atitikmenį (žr. 1 pav.). Automatinis šnekos atpažinimas paprastai susideda iš dviejų etapų – pirminio apdorojimo ir galutinio apdorojimo. Pirminis apdorojimas apima požymių išskyrimą, o galutinis apdorojimas sudaro šnekos atpažinimo variklį, turintį akustinį modelį, žodyną ir gramatiką. Jei visos šios dalys yra korektiškos, šnekos atpažinimo variklis identifikuoja labiausiai tikėtiną gautos įvesties atitiktį ir grąžina atpažintus žodžius tekstu. Tinkamų požymių išskyrimo ir šnekos atpažinimo metodų parinkimas turi didelės įtakos atpažinimo sistemos tikslumui [1], [9], [10].



1 pav. Šnekos atpažinimo sistemos struktūra

Šnekos požymių išskyrimo metodai

Skaitmeninis signalas *Y* į automatinio šnekos atpažinimo sistemą dažniausiai įvedamas iš failo arba mikrofono. Šiame garso signale, be ištartų žodžių, gali būti daug kitos informacijos, pavyzdžiui, aplinkos triukšmas, akcentas, intonacija ir kt. Požymių išskyri-

mo dalies užduotis yra transformuoti garso signalą *Y* į požymių seką *X*. Požymių išskyrimas dažniausiai remiasi dažnine signalų analize. Daroma prielaida, kad šnekos signalas yra stacionarus trumpuoju laiko intervalu, tad signalas *Y* skaidomas mažais susiliejančiais langais (kadrais). Dažniausiai naudojamas nuo 15 ms iki 30 ms trukmės langas, o požymių išskyrimo metu langas slenkamas po 10–20 ms. MFCC (angl. *Mel Frequency Cepstral Coefficients*) – požymių išskyrimo metodas, dažnai naudojamas šnekai atpažinti. Šis metodas remiasi trumpalaike analize [2], [10].

Akustinis modelis

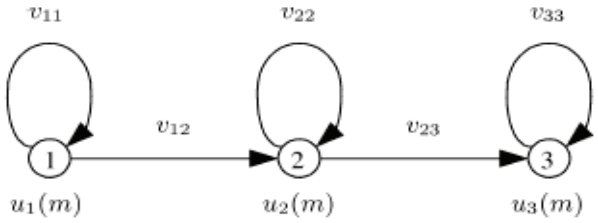
Akustinio modelio užduotis yra įvertinti žodžių sekos tikimybę. Teoriškai būtų galima surinkti daug žodžio *w* garso pavyzdžių ir taip sudaryti tikimybinių požymių vektorių pasiskirstymo priklausomybę nuo žodžio, bet praktiškai tai sunkiai įgyvendinama, nes didelio žodyno atveju daug žodžių retai pasitaiko mokymo duomenyse arba iš viso net neturima žodį atitinkančio garso įrašo. Todėl dažniausiai požymių vektoriaus *X* skirstinys modeliuojamas pagal mažesnius negu žodis fonetinius vienetus – fonemas, kontekstines fonemas arba skiemenis. Šiam skirstiniui modeliuoti naudojami paslėptieji Markovo modeliai (toliau PMM). PMM galima įsivaizduoti kaip atsitiktinį procesą, keliaujantį per būsenų aibę *S* ir generuojantį požymių vektorius *X*. Kiekvienas akustinis vienetas modeliuojamas vienu PMM, sudarytu iš kelių būsenų. Dažniausiai naudojami trijų būsenų (garso pradžia, vidurys ir pabaiga) PMM. Gausaus žodyno atveju formuojamas statinis arba dinaminis žodžių tinklas, sudarytas iš daugelio PMM, tinkle ieškoma tokios būsenų sekos *S*, kuri sugeneruotų požymių vektorius *X* su didžiausia tikimybe. Geriausiai sekai surasti naudojamas Viterbio (Viterbi) algoritmas [2], [9].

Kalbos modelis

Negausaus žodyno atskirai tariamų žodžių atpažinimo atveju kartais kalbos modelis (toliau KM) nenaudojamas ir pasikliaunama tik akustiniu modeliu, bet rišlios šnekos atpažinimo sistemose KM būtinas. Šnekoje egzistuoja taisyklės (gramatika), pavyzdžiui, žodžiai jungiami vienas su kitu, ir šių žinių panaudojimas leidžia labai sumažinti klaidingai atpažįstamų žodžių kiekį. KM kaupiamos lingvistinės žinios, jis susideda iš dviejų dalių: žodyno ir gramatikos taisyklių. KM gramatika gali būti aprašoma formaliomis lingvistinėmis taisyklėmis, pavyzdžiui, bekontekstėmis gramatikomis, arba sudaroma remiantis statistiniais metodais, t. y. naudojant statistinius KM, modeliuojančius žodžio pasirodymo tekste tikimybinius skirstinius pagal gretimus žodžius. Bekontekstės gramatikos gali būti naudojamos tik specializuotiems šnekos atpažinimo uždaviniams spręsti. Statistiniai KM plačiai naudojami rišlios šnekos atpažinimo sistemose dėl jų paprastumo. Kalbos modeliavimo užduotis yra parinkti kuo geresnę istorijos klasifikavimo funkciją. Funkcija turi leisti ir gerai prognozuoti būsimą žodį, klasės turi būti gana dažnos, kad būtų galima patikimai įvertinti tikimybes. Populiariausi KM yra n-gramos. Daroma prielaida, kad žodis priklauso tik nuo $n-1$ ankstesnio žodžio. Nors ir paprastos, tačiau n-gramos sugeba įvertinti ir sintaksinius, ir prasminius žodžių ryšius, todėl yra vienos iš geriausių modelių. Tyrimuose n-gramos paprastai laikomos atskaitos modeliais ir bet koks modeliavimo patobulinimas įvertinamas lyginant su n-gramomis. Bigramos ($n = 2$) ir trigramos ($n = 3$) atsižvelgia į žodžio istoriją [9].

Paslėptaisiais Markovo modeliais grįstas šnekos atpažinimas

Paslėptuosiuose Markovo modeliuose nagrinėjamas dvigubas atsitiktinis procesas, t. y. perėjimas iš vienos būsenos į kitą ir būsenos stebėjimas. Dažniausiai naudojami pirmos eilės Markovo modeliai, kuriuose perėjimo tikimybė priklauso tik nuo ankstesnės būsenos. Paslėptuoju vadinamas antrasis procesas, kadangi jis vykdomas per pirmąjį procesą ir tiesiogiai nestebimas. Perėjimus nusako tikimybių v_{ij} aibė, *m*-ojo simbolio b_m stebėjimą būsenos *i* – tikimybių $u_i(b_m)$ aibė (žr. 2 pav.) [8].

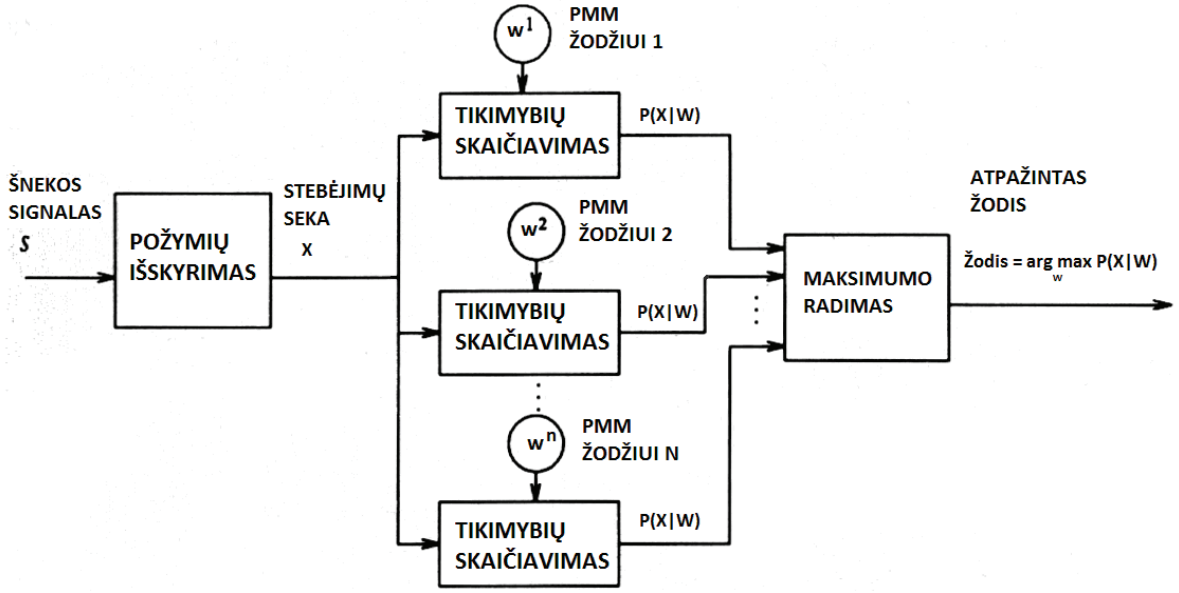


2 pav. Trijų būsenų Markovo modelis

PMM nusako penki parametrai: *M* – būsenos stebėjimo simbolių skaičius, *T* – būsenų skaičius, *V* – perėjimo tikimybių pasiskirstymas, *U* – būsenos stebėjimų tikimybių pasiskirstymas ir π – pradinės būsenos tikimybių pasiskirstymas. Tarkime, kad turime *V* šnekos pavyzdžių, kuriems atpažinti norime pritaikyti PMM metodą. Pirmasis žingsnis – žodyno sukūrimas. Kiekvienam iš *V* pavyzdžių sukuriame modelį λ. Modelio tikimybiniai parametrai *V, U* ir π nustatomi taikant įvertinimo procedūras iš apmokymui pateiktų pavyzdžių. Nagrinėdami nežinomą kalbos pavyzdį, mes atliekame signalo analizę ir taip gauname stebėjimų seką *X*. Atpažintuoju pavyzdžiu paskelbiamas etaloninis pavyzdys, kurio modelis geriausiai atitinka nagrinėjamą stebėjimų seką [8] (žr. 3 pav.):

$$Z = \arg\max_{1<k<V} P(X \mid \lambda_k) \quad (1)$$

Taigi, norint sėkmingai naudoti PMM kalbos signalams atpažinti, reikia išspręsti tris uždavinius: (a) įvertinimo uždavinys: turint stebėjimų seką $X=(x_1, x_2, ..., x_T)$ ir grandinę aprašančio modelio parametrus $\lambda=(V,U,\pi)$, reikia apskaičiuoti tikimybę $P(X|\lambda)$, kad nagrinėjamoji stebėjimų seka buvo sugeneruota nagrinėjamo modelio; (b) paslėptųjų būsenų nustatymo uždavinys: turint stebėjimų seką $X=(x_1, x_2, ..., x_T)$, reikia nustatyti būsenų seką, optimalią tam tikro pasirinkto prasmingo kriterijaus prasme; (c) apmokymo uždavinys: kaip parinkti modelio parametrus, kad būtų padidinama tikimybė $P(O|\lambda)$ [8].



3 pav. Automatinis šnekos atpažinimas naudojant PMM

Automatinio šnekos atpažinimo sistemų klasifikavimas

AŠA sistemos gali būti klasifikuojamos pagal tai, koks yra šnekėjimo pobūdis, kalbėtojo modelio tipas, žodyno tipas. Pagal *šnekėjimo pobūdį* išskiriamos šios grupės: pavienių žodžių atpažinimas, rišlių frazių atpažinimas (minimalios pauzės tarp žodžių), ištisinės šnekos atpažinimas (beveik natūrali kalba, klaidos tikimybė didėja gausėjant žodynui), spontaniškos šnekos atpažinimas (gali apdoroti natūralią šneką, turinčią netgi beprasmių garsų, pauzių, su tempo ar balso kitimu). Pagal *kalbėtojo modelio tipą* galima išskirti šias grupes: nuo kalbėtojo priklausomos ir nuo kalbėtojo nepriklausomos (nereikia individualaus apmokymo) grupės. AŠA sistemų *naudojami žodynai* gali būti maži (dešimtys žodžių), vidutiniai (šimtai žodžių), dideli (tūkstančiai žodžių), labai dideli (dešimtys tūkstančių žodžių ir daugiau) [1], [5], [10].

Lietuvių kalbos automatinis šnekos atpažinimas

Lietuvių kalbos šnekai atpažinti sukurtos kelios programos. Programa *Atpažinimas* yra priklausoma nuo kalbėtojo, grįsta dinaminiu laiko skalės kraipymu, joje atliekamas nežinomo ištarimo palyginimas su etalonais. Programa *Žodžių atpažintuvas* yra nepriklausoma nuo kalbėtojo, atpažinimo procesui naudoja PMM, turi menką žodyną, naudoja nuo konteksto nepriklausomus fonemų akustinius modelius. Ja remiantis buvo sukurtas interneto naršyklės valdymo balsu sistemos prototipas [4]. Sukurta ir pavienių žodžių atpažinimo ir segmentavimo sistema KAS [8]. Lietuvos mokslininkai atliko lietuvių kalbos šnekos atpažinimo tyrimus, naudodami kitų kalbų atpažinimo variklius. Tyrimai atlikti skaičiams ir balso komandoms atpažinti. Aukščiausi rezultatai gauti naudojant ispanų kalbai skirtos šnekos atpažinimo variklį [3].

Lietuvių kalbos automatinio šnekos atpažinimo kalbos ir akustiniai modeliai

Norint automatiškai atpažinti lietuvišką šneką, reikia:

- kalbos modelio: žodyno (žodžių sąrašas) ir gramatikos (žodžių kombinacijų rinkiniai);
- akustinio modelio, galinčio statistiškai aprašyti skirtingus garsus, kuriuos sukelia kiekvienas vartojamo kalbos modelio žodis. Norint apmokyti akustinius modelius, reikia sudaryti garsyną;
- kalbos atpažinimo variklio, žmogaus ištartą garsą susiejančio su akustiniame modelyje esančiais garsais, o po to randančio tokią fonemų seką gramatikoje.

Šiam darbui reikėjo *frazių garsyno*, kuris buvo sudarytas iš LRT mediatekoje esančių žinių laidų įrašų. Įrašyti šnekos signalai yra aukštos kokybės, tarimas aiškus ir taisyklingas. Buvo atrinkta 50 įrašų (trukmė – 2 valandos ir 20 minučių), kurie suskaidyti į sakinius ir anotuoti pagal juos atitinkančius tekstus žodžių lygiu. Taip buvo sudarytas garsyne esančių žodžių sąrašas (5 859 žodžiai). Be to, reikėjo *izoliuotų žodžių garsyno*, kuris buvo sudarytas iš 10 kalbėtojų 150 skirtingų žodžių, pakartotų 10 kartų, įrašų, kurių trukmė iš viso siekė 6 valandas. Kiekvienas žodis įrašytas į atskirą garso failą, turintį jį atitinkantį žodžio lygio anotacijos failą. Įrašai sukurti su *Audacity* programa. Su turimais paruoštais duomenimis kalbos ir akustinis modeliai apmokomi naudojant HTK įrankius.

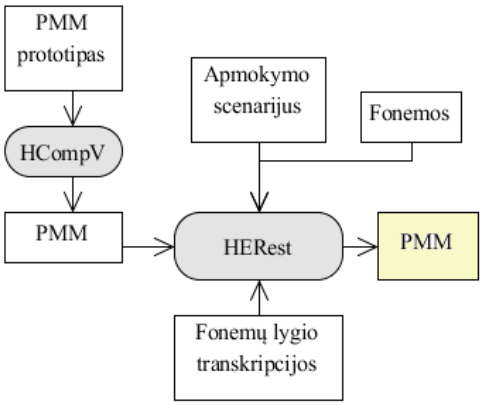
Transkripcijų žodynas

Norint sukurti lietuvių šnekos akustinį modelį, reikia sudaryti tarimo žodyną ir fonemų transkripcijas, įrašyti garsyną ir konvertuoti jį į formatą, tinkamą požymio vektoriams išskirti. Pagal išskirtus požymio vektorius reikia apmokyti PMM. Norint gauti tikslesnius rezultatus, reikia turėti frazių garsyną su atitinkamomis tekstinėmis frazėmis ir jų transkripcijomis. Visiems garsyne panaudotiems žodžiams sudarytas transkripcijų žodynas. Norint atlikti transkribavimą, reikia pasirinkti, pagal kokią fonetinę sistemą bus transkribuojama. Šiame darbe fonetinė sistema, kurioje yra 85 fonemos, sudaryta remiantis lietuvių kalbos tarties žodynu, kuriame atsižvelgiama į kirčiavimą, priebalsių minkštumą ir kietumą, balsių ilgumą ir trumpumą.

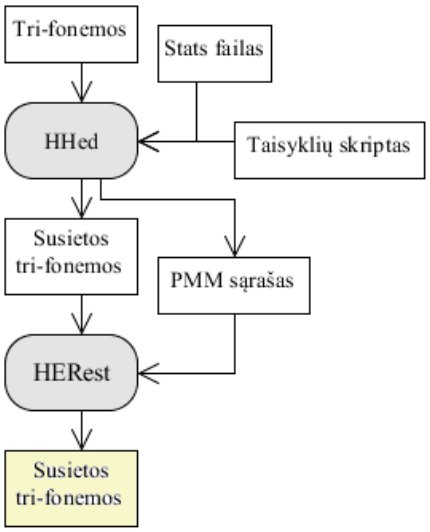
Paslėptųjų Markovo modelių apmokymas

Kuriant akustinius modelius buvo pasirinktas HTK įrankis, su kuriuo gali būti kuriami tiek pavieniams žodžiams, tiek ištisinei šnekai skirti akustiniai modeliai, grįsti PMM metodu. Jį sudaro priemonės duomenims rengti, akustiniams ir kalbos modeliams mokyti ir testuoti, akustiniams modeliams pritaikyti konkrečiam kalbėtojui, rezultatams analizuoti. Galima modeliuoti pavienių žodžių ir ištisinės šnekos atpažinimo sistemas. Paslėptųjų Markovo modelių apmokymas susideda iš kelių žingsnių. Apmokant buvo naudojami garso įrašų failai, jau sukurti transkripcijų failai, fonemos. Pirmiausia **HLEd** įrankiu žodžių lygio transkripcijos išplėstos į fonemų lygio transkripcijas. **HCopy** įrankiu iš visų garsyno garso failų išskiriami požymių vektoriai ir sukuriami MFCC formato failai, t. y. požymių failai. Scenarijaus faile nurodoma, kokius failus reikia konvertuoti ir kur išsaugoti požymių failus.

PMM modeliams apmokyti sukurtas prototipo modelis, aprašantis 25 MFCC požymių vektorius. Standartiškai prototipe požymių vektoriams naudojama iki 39 MFCC. Tačiau turimai apmokymo duomenų imčiai ir apmokymo įrašų kokybei (netriukšmingiems duomenims) pakanka 25 MFCC, kadangi atpažinimo tikslumas išlieka toks pat. Prototipo sukūrimo tikslas – apibrėžti modelio struktūrą. Įrankis **HCompV** peržiūri visus MFCC formato failus, suskaičiuoja vidurkį ir kovariaciją, o rezultatus priskiria duotam PMM prototipo modeliui. Prototipas pritaikomas kiekvienai fonemai. Įrankiu **HERest** fonemos iš naujo įvertinamos pagal sukurtą prototipą ir MFCC požymių vektorius (žr. 4 pav.). Sukurtieji PMM modeliai buvo koreguojami – jiems pridėti tylos modeliai *sp*, aprašantys trumpas pauzes, galinčias atsirasti ištisinėje šnekoje tarp ištariamų žodžių. Modelyje jau buvo sukurtas tylos modelis *sil*, kuris yra ilgesnis už *sp* ir naudojamas sakinio pradžioje ir pabaigoje. Tylos modelis *sp* sukuriamas **HHEd** įrankiu panaudojant *sil* tylos modelį. PMM vėl įvertinami **HERest** įrankiu jau su *sp* tylos modeliais. Šiame etape sukurti PMM, panaudojant fonemas, gali būti naudojami šnekai atpažinti.



4 pav. PMM apmokymas



5 pav. PMM apmokymas trifonemomis

PMM apmokymas trifonemomis

Trifonema – tai trijų fonemų grupė, sudaryta pagal formą $L - X + R$ (L – kairėje esanti fonema, X – vidurinė fonema, R – dešinėje esanti fonema). Tą patį žodį galima sudaryti iš fonemų (pavyzdžiui, EITI → `e i t i`) ir iš trifonemų (pavyzdžiui, EITI → `e+i i-t+i t+i`). Trifonemos naudojamos norint tiksliau atpažinti šneką, kadangi tuomet žiūrima į fonemos kontekstą (bandoma atpažinti iš eilės einančius tris garsus, o ne vieną, kaip daroma fonemų atveju). Trifonemos sumažina tikimybę, kad vienas garsas bus supainiotas su kitu. Trifonemos sudaromos konvertuojant fonemas su **HLEd** įrankiu. PMM modeliai iš naujo įvertinami įrankiu **HERest**, remiantis sukurtomis trifonemomis. Sukuriamas ir skriptų failas *tree.hed*, kuriame pateikiamos taisyklės galimam fonemų klasterizavimui atlikti. Remiantis sudarytomis taisyklėmis, **HHed** įrankiu trifonemos „surišamos“ jas klasterizuojant. PMM vėl įvertinami įrankiu **HERest** (žr. 5 pav.).

Apmokytų PMM testavimas

Šnekos atpažinimo tikslumas buvo vertinamas skaičiuojant žodžių atpažinimo procentą. Eksperimentų metu tiriama, kaip kinta frazių ir pavienių žodžių atpažinimo tikslumas, kai naudojami fonemomis ir trifonemomis apmokyti PMM. Frazėms atpažinti testuojant parinkti 100 sakinių iš radijo žinių įrašų, kurių trukmė siekė 20 minučių. Pavieniams žodžiams atpažinti testuojant buvo įrašyta 740 pavienių žodžių, kurių bendra trukmė siekė 20 minučių. Įrankiu **HBuild** sukurtas žodžių tinklas, nusakantis, kad gali būti ištartas bet kuris žodyne esantis žodis. Testuojama su **HVite** įrankiu. Žodžių atpažinimo tikslumui įvertinti naudojamas įrankis **HResults**. ***Pirmajame tyrimo etape*** buvo testuojamas frazių atpažinimas. Testuojant su PMM, apmokytais fonemomis, atpažintų žodžių procentas ir žodžių atpažinimo tikslumas buvo neaukšti, o testuojant tuos pačius duomenis su trifonemomis apmokytais PMM atpažinimo tikslumas ženkliai padidėjo. Frazių atpažinimo fonemomis (A) ir trifonemomis (B) testavimo rezultatai pateikti 1 lentelėje.

1 lentelė. ***Testavimo rezultatai***

	Teisingai atpažintų žodžių proc.	**Atpažinimo tikslumas**	**Teisingai atpažintų žodžių sk.**	**Praleistų žodžių sk.**	**Sukeistų žodžių sk.**	**Įterptų žodžių sk.**	**Atpažinimui pateiktų žodžių sk.**
(A)	53,92	44,63	1 010	93	770	174	1 873
(B)	76,67	69,35	1 436	59	378	137	1 873
(C)	29,32	15,68	217	8	515	333	740
(D)	75,14	61,76	556	0	184	99	740
(E)	96,32	94,98	1 804	19	50	25	1 873

Antrajame tyrimo etape buvo testuojamas pavienių žodžių atpažinimas. Vėl išryškėjo trifonemomis apmokytų PMM pranašumas. Pavienių žodžių atpažinimo fonemomis (C) ir trifonemomis (D) testavimo rezultatai pateikti 1 lentelėje.

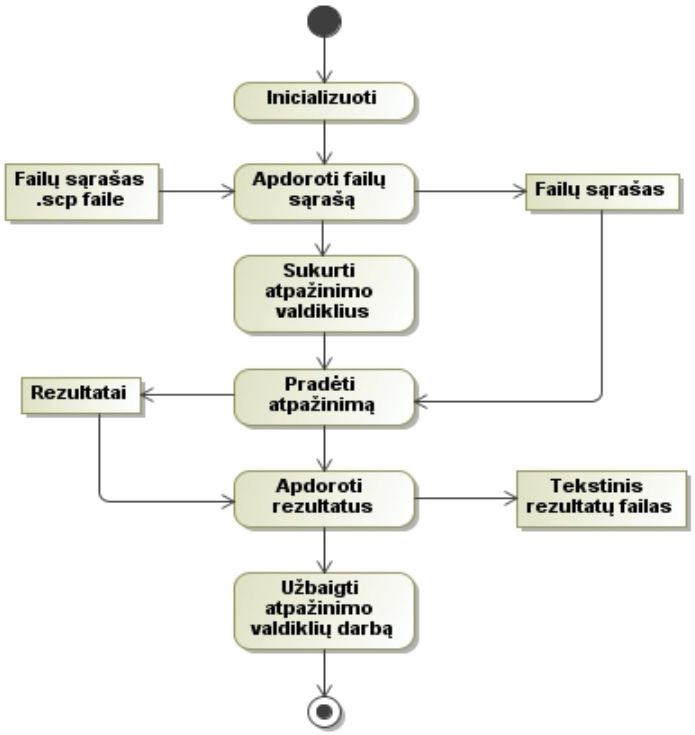
Atlikti pirmieji du tyrimo etapai parodė, kad, naudojant trifonemas, žodžių atpažinimo tikslumas žymiai padidėja. Toliau tyrimai buvo atliekami tik su trifonemomis. ***Trečiasis tyrimo etapas*** atliktas norint patikrinti, ar fonemų, skirtų žodžių transkripcijoms, skaičiaus sumažinimas turės įtakos atpažinimo tikslumui. Buvo sudarytos naujos žodžių fonetinės transkripcijos, kuriose atsisakyta garsų pusilgumą, priebalsių minkštumą ir kirčio ženklus žyminčių fonemų. Naujoji fonetinė sistema sudaryta iš 32 fonemų. Eksperimento rezultatai parodė, kad didelės įtakos atpažinimo tikslumui tai neturėjo – sumažinus fonemų skaičių, žodžių atpažinimo tikslumas padidėjo **2 proc.**, kadangi sumažėjo įterptų žodžių skaičius ($I = 51$). ***Ketvirtasis tyrimo etapas*** atliktas žinant, kad kuriamas stenografavimo įrankis bus skirtas ištisinei šnekai atpažinti, todėl reikėjo padidinti frazių atpažinimo tikslumą. Tai būtų galima padaryti sukurtą gramatiką (žodžių tinklą) praplečiant bigramomis. Bigramos apmokytos iš žinių laidų įrašus (kuriais buvo apmokyti PMM modeliai) atitinkančių sakinių. Tam panaudotas HTK įrankis **HLStats**. Buvo testuojama su frazių atpažinimo įrašų aibe. Žodžių atpažinimo tikslumas žymiai padidėjo ir siekė 94 proc. (duomenys pateikti 1 lentelės (E) eilutėje).

Pavienių žodžių atpažinimo tikslumui bigramos įtakos neturėjo. Integravus šią gramatiką į stenografavimo įrankį, jis būtų priklausomas nuo konteksto. Išanalizavus testavimo rezultatus pastebėta, kad dažniausiai neatpažįstami yra panašiai skambantys žodžiai (*įstatymui – įstatymai, projektą – projekte, kalba – kalbą*); daug praleidžiama trumpų žodžių (*į, ir, o, iš*). Todėl reikėtų didesnio skaičiaus duomenų, kuriais būtų galima apmokyti prastai atpažįstamų žodžių akustinius modelius. Praplėtus žodyną ir bi-

gramas apmokius didesniu skaičiumi sakinių, būtų galima pagerinti pavienių žodžių ir trumpų frazių, sudarytų iš dviejų ar trijų žodžių, atpažinimą. Ištisinės šnekos atpažinimo atveju žodžių tinklui išplėsti būtų galima panaudoti trigramas. Trigramoms apmokyti ir joms testuoti būtų galima panaudoti **HDecode** įrankį, kuris yra suderinamas su HTK.

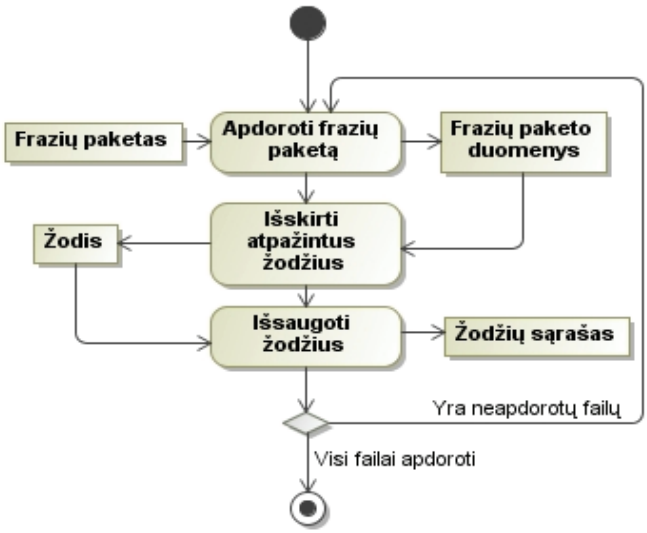
Stenografavimo įrankio prototipo kūrimas

Stenografavimo įrankis kurtas naudojant ATK biblioteką, kuri turi C++ sluoksnį ir yra suderinama su HTK įrankiu sukurtais akustiniais modeliais. Vartotojas programą parenka iš komandinės eilutės, nurodomi parametrai – konfigūracijos failo (aprašančio įrankio šnekos atpažinimo parametrus) ir tekstinio failo (kuriame nurodyti stenografuoti parinkti garso failai) pavadinimai. Įrankio veiksmų diagrama pateikta 6 paveiksle. Pirmiausia patikrinama, ar programos paleidimo metu nurodomi konfigūracijos ir garso įrašų sąrašo failai. Sukuriant atpažinimo valdiklius, inicializuojamas resursų valdiklis *ARMan*, atsakingas už PMM modelių, gramatikos, žodyno integravimą į programą. Duomenys apie PMM modelius, gramatiką ir žodyną gaunami iš konfigūracijos failo.



6 pav. Stenografavimo įrankio veiksmų diagrama

Kitame etape sukuriami vartotojo sąsajos komponentai *AMonitor* atpažinimo eigai stebėti. Į vartotojo sąsajos komponentus įtraukiami *ACode* (atsakingas už požymių vektorių išskyrimą), *ARec* (atsakingas už šnekos atpažinimą), *ASource* (atsakingas už garso signalo įvestį). Atpažinimas vykdomas tik tada, jei yra nurodyti stenografuoti skirti garso failai. Pradėjus atpažinimą pagal konfigūracijos faile nurodytus MFCC parametrus, pradedami apdoroti garso failai. Juos apdorojant (pagal integruotą PMM modelį) renkami *APacket* tipo duomenys apie šnekos atpažinimo paketus ir išskiriami *APhraseData* tipo atpažinimo rezultatus saugantys paketai (žr. 7 pav.). Išskirti duomenys yra konvertuojami į *String* tipo eilutę ir po to perduodami rezultatus apdorojančiai paprogramei, kuri juos įrašo į tekstinį failą. Programos veikimas testuotas su ATK bibliotekų rinkinyje esančiais testiniais duomenimis: žodynu, PMM modeliais, gramatika, garso failais. Programos veikimas yra korektiškas – atpažintus žodžius ji įrašo į tekstinį failą.



7 pav. Atpažinimo rezultatų gavimas

Išvados

1. Tyrimų rezultatai parodė, kad PMM modeliai, apmokyti trifonemomis, leidžia pasiekti geresnių atpažinimo rezultatų nei apmokyti fonemomis. Geresni atpažinimo rezultatai gauti tiek atpažįstant pavienius žodžius (tikslumas siekia iki 61 proc.), tiek frazes (tikslumas siekia iki 69 proc.). Todėl rekomenduojama šnekai atpažinti naudoti trifonemas.
2. Sudarius naują 32 fonemų fonetinę sistemą, kurioje iš pradinės 85 fonemų fonetinės sistemos atsisakyta kirčio ženklų, garsų pusilgumo, minkštumą žyminčių fonemų, ir atlikus fonemų skaičiaus transkripcijose įtakos atpažinimo tikslumui turinčių veiksnių tyrimus, paaiškėjo, kad tai įtakos atpažinimo tikslumui neturėjo.
3. Atlikus tyrimus su bigramomis praplėstu žodžių tinklu buvo nustatyta, kad jos pagerina ištisinės šnekos atpažinimo tikslumą, siekiantį 94 proc., tačiau pavieniams žodžiams atpažinti tai įtakos neturėjo, todėl reikėtų daugiau duomenų, kuriais būtų galima apmokyti prastai atpažįstamų žodžių akustinius modelius.

Literatūra

1. Anusuya M. A., Katti S. K., 2009, Speech Recognition by Machine: A Review. *International Journal of Computer Science and Information Security*. Prieiga per internetą: <http://arxiv.org/ftp/arxiv/papers/1001/1001.2267.pdf>.
2. ILVB-2006 tutorial, 2006, Automatic Speech Recognition. Prieiga per internetą: <http://isoft.postech.ac.kr/~gblee/Course/CS703AHLT/postech-ASR.ppt>.
3. Bartisiute G., Ratkevicius K., 2012, Speech Server based Lithuanian Voice Commands Recognition. *Elektronika ir elektrotechnika*. Prieiga per internetą: <http://www.eejournal.ktu.lt/index.php/elt/article/view/3061/2125>.
4. Filipovič M., Ringelienė Ž., 2011, Žodžių atpažinimo, grįsto paslėptaisiais Markovo modeliais, vizualizavimo ir analizės programinė įranga. *Informacijos mokslai*. Nr. 56. Prieiga per internetą: <http://www.zurnalai.vu.lt/informacijos-mokslai/article/download/3150/2274/63-72.pdf>.
5. Gaikwad S. K., Gawali B. W., Yannawar P., 2010, A Review on Speech Recognition Technique. *International Journal of Computer Applications (0975–8887)*. Prieiga per internetą: <http://www.ijcaonline.com/volume10/number3/pxc3871976.pdf>.
6. Sinha P., 2010, Speech Processing in Embedded Systems. Prieiga per internetą: <http://link.springer.com/book/10.1007%2F978-0-387-75581-6>.
7. Sledevič T., Stašionis L., 2013, Lietuvių kalbos pavienių žodžių atpažinimo algoritmo įgyvendinimas lauku programuojama logine matrica. *Elektronika ir elektrotechnika*. Prieiga per internetą: <www.mla.vgtu.lt/index.php/mla/article/download/mla.2013.18/pdf>.
8. Tamulevičius G., 2008, Pavienių žodžių atpažinimo sistemų kūrimas. *Daktaro disertacija*. Prieiga per internetą: <http://donelaitis.vdu.lt/disertacijos/Tamulevicius_2008.pdf>.
9. Vaičiūnas A., 2006, Lietuvių kalbos statistinių modelių ir jų taikymo šnekos atpažinimui tyrimas, kai naudojami labai dideli žodynai. *Daktaro disertacija*. Prieiga per internetą: <http://donelaitis.vdu.lt/disertacijos/Vaiciunas_2006.pdf>.
10. Vimala C., Radha V., 2012, A Review on Speech Recognition Challenges and Approaches. *World of Computer Science and Information Technology Journal (WCSIT)*. Prieiga per internetą: <http://www.wcsit.org/media/pub/2012/vol.2.no.1/A%20Review%20on%20Speech%20Recognition%20Challenges%20and%20Approaches.pdf>.

Summary

INVESTIGATION OF HIDDEN MARKOV MODELS AND ADAPTATION FOR TRANSCRIBING VOICE RECORDINGS

J. Vaičiulytė, G. Felinskas

Automatic speech recognition methods have been reviewed in the paper. Hidden Markov models for acoustic modeling were used, a language model and dictionary were created. The models were trained with recorded news, total 5859 words, for 2 hours and 20 minutes. Two phonetic systems (with 85 and 32 phonemes respectively) were developed on the basis of a Lithuanian pronunciation dictionary. The created models were integrated in a stenography app prototype for transcribing voice recordings. The testing results of the acoustic models showed a 94% accuracy for phrase recognition.

Keywords: speech recognition, language models, acoustic models, hidden Markov models.

Santrauka

PASLĖPTŲJŲ MARKOVO MODELIŲ METODO TYRIMAS IR TAIKYMAS BALSO ĮRAŠAMS STENOGRAFUOTI

J. Vaičiulytė, G. Felinskas

Šiame darbe apžvelgti automatinio šnekos atpažinimo metodai. Paslėptųjų Markovo modelių metodu apmokyti akustiniai modeliai, sukurtas kalbos modelis ir žodynas. Akustiniai modeliai apmokyti žinių įrašais, kurių trukmė siekia 2 valandas ir 20 minučių, o žodyno apimtis – 5 859 žodžius. Eksperimentams naudotos dvi fonetinės sistemos, kuriose yra 85 ir 32 fonemos, sudarytos remiantis lietuvių kalbos tarties žodynu. Sukurti modeliai integruoti į stenografavimo įrankio prototipą, skirtą balso įrašams stenografuoti. Remiantis akustinių modelių testavimo rezultatais galima daryti išvadą, kad pasiektas 94 proc. frazių atpažinimo tikslumas.

Prasminiai žodžiai: šnekos atpažinimas, kalbos modeliai, akustiniai modeliai, paslėptieji Markovo modeliai.

Įteikta 2016-02-22
Priimta 2016-05-21