



**KAUNO TECHNOLOGIJOS UNIVERSITETAS  
ELEKTROS IR ELEKTRONIKOS FAKULTETAS**

**Remigijus Račys**

**SKAIČIŲ IR VARDŲ GARSYNŲ ATPAŽINIMO,  
NAUDOJANT FONEMOMIS GRĮSTUS PASLĖPTUOSIUS  
MARKOVO MODELIUS, TYRIMAS**

Baigiamasis magistro projektas

**Vadovas**

Doc. dr. Kastytis Ratkevičius

**KAUNAS, 2017**

**KAUNO TECHNOLOGIJOS UNIVERSITETAS**  
**ELEKTROS IR ELEKTRONIKOS FAKULTETAS**  
**AUTOMATIKOS KATEDRA**

**SKAIČIŲ IR VARDŲ GARSYNŲ ATPAŽINIMO,  
NAUDOJANT FONEMOMIS GRĮSTUS PASLĖPTUOSIUS  
MARKOVO MODELIOUS, TYRIMAS**

Baigiamasis magistro projektas  
Valdymo technologijos (kodas 621H66001)

**Vadovas**

Doc. dr. Kastytis Ratkevičius

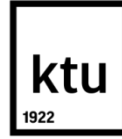
**Recenzentas**

Doc. dr. Vytautas Gargasas

**Projektą atliko**

Remigijus Račys

**KAUNAS, 2017**



KAUNO TECHNOLOGIJOS UNIVERSITETAS

Elektros ir elektronikos fakultetas

(Fakultetas)

Remigijus Račys

(Studento vardas, pavardė)

Valdymo technologijos, 621H66001

(Studijų programos pavadinimas, kodas)

Baigiamojo projekto „Skaičių ir vardų garsynų atpažinimo, naudojant fonemomis grįstus paslėptuosius Markovo modelius, tyrimas“

### AKADEMINIO SAŽININGUMO DEKLARACIJA

20 \_\_\_\_ m. \_\_\_\_\_ d.

Kaunas

Patvirtinu, kad mano **Remigijaus Račio** baigiamasis projektas tema „Skaičių ir vardų garsynų atpažinimo, naudojant fonemomis grįstus paslėptuosius Markovo modelius, tyrimas“ yra parašytas visiškai savarankiškai, o visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

\_\_\_\_\_  
(vardą ir pavardę įrašyti ranka)

\_\_\_\_\_  
(parašas)

Račys, Remigijus. Skaičių ir vardų garsynų atpažinimo, naudojant fonemomis grįstus paslėptuosius Markovo modelius, tyrimas. *Valdymo sistemų magistro* baigiamasis projektas / vadovas doc. dr. Kastytis Ratkevičius; Kauno technologijos universitetas, Elektros ir elektronikos fakultetas, Automatikos katedra.

Mokslo kryptis ir sritis: Inžinerijos mokslų magistras, Valdymo technologijos

Reikšminiai žodžiai: garsynas, HTK, paslėptieji Markovo modeliai, anotavimas.

Kaunas, 2017. 55 p.

## **SANTRAUKA**

Šiame darbe pateikiami atskirų komandų (skaičių ir vardų) atpažinimo eksperimentų rezultatai, kalbos atpažinimo srityje, pritaikant – fonemomis grįstus – paslėptuosius Markovo modelius. Eksperimentiniams tyrimams paruoštos garsyno gramatikos fonemų lygmenyje. Taip pat „rankiniu“ būdu daliai garsyno sukurtos anotacijos. Nagrinėjama atpažinimo tikslumo priklausomybė nuo naudojamų fonemų rinkinio ir nuo naudojamų HTK paketo įrankių. Mokymo metu naudojamos įvairios fonemų aibės ir tikrinama jų įtaka atpažinimo tikslumui. Aiškinamasi, kokia įtaką turi anotuotas garsynas – tikslumo tyrimams.

Ištyrus izoliuotų vardų ir skaičių komandų atpažinimo tikslumą, gautas geriausias rezultatas – 95,36%, panaudojant išplėstinį SAMPA–LT fonemų rinkinį.

Siekiant patikrinti anotuoto garsyno įtaką atpažinimui, atlikti tyrimai skaičių garsyno tikslumo įvertinimui. Mokymo etape paruoštas anotuotas garsynas šešiams diktoriaus. Atpažinimo tikrinimo metu atlikti tyrimai, testavimui naudojant 24 diktorius. Su minimaliu fonemų rinkiniu, tikslumas gautas – 99,13%, o su SAMPA–LT fonemų rinkiniu – 98%. Sukurtos kontekstinės fonemos (trifonai) su kuriomis, atpažinimo tikslumas naudojant minimalų rinkinį – 99,10%, komandų atpažinimo tikslumas atpažįstant trifonus, sukurtus iš SAMPA–LT rinkinio – 98%.

Račys, Remigijus. Investigation of Digits and Names Corpora Recognition by Phones Based Hidden Markov Models: *Control systems master's degree / supervisor assoc. doc. dr. Kastytis Ratkevičius*; Kaunas University of Technology, Faculty of Electrical and Electronics Engineering, department of Automation.

Research area and field: Master of Engineering Sciences, Control Technologies

Key words: Corpus, HTK, Hidden Markov Models, Annotation.

Kaunas, 2017. 55 p.

## **SUMMARY**

In this paper the detection experiments of separate commands (digits and names) are presented, in the field of speech recognition, adaptation–phonemes–based hidden Markov models. Experimental research of the corpus level grammatical phonemes are prepared. For a part of the corpus annotations were created “manually”. The research recognition accuracy depends on the used set of phonemes and the use of HTK package tools. Experimenting with various phoneme sets was made in order to find their impact on the detection accuracy.

Examination of names and numbers of the isolated command recognition accuracy were 95.36%. This was achieved by using the expanded SAMPA–LT phoneme set.

In order to check the influence of recognition annotated corpus, a number of tests were made to assess the accuracy of the number corpus. An annotated corpus for 6 announcers was made in the training stage. Testing was made in the recognition inspection stage while using 24 announcers. With a minimum set of phonemes, the accuracy obtained – 99.13%, and using SAMPA–LT phoneme set the result obtained was 98%. Contextual phonemes (triphones) were created with which the detection accuracy of the minimum set was 99.10%, while the recognition accuracy using triphones created from SAMPA–LT set – 98%.

# TURINYS

ĮVADAS .....	7
1 KALBOS ATPAŽINIMAS .....	9
1.1 Veiksniai įtakojantys kalbos atpažinimo sudėtingumą .....	9
1.2 Kalbos atpažinimo būdai.....	10
1.3 Paslėptieji Markovo modeliai .....	11
1.4 HTK programinis įrankis .....	13
2 METODINĖ DALIS .....	15
2.1 Vardų ir skaičių garsynas .....	15
2.1.1 Garsyno paruošimas tyrimams.....	15
2.2 Hidden Markov Model Toolkit įsisavinimas .....	16
2.3 Garsyno duomenų anotavimas .....	27
2.4 Anotuoto garsyno tyrimas.....	29
2.5 Kontekstinių fonemų (trifonų) tyrimas .....	31
3 TYRIMO REZULTATAI .....	33
3.1 Atpažinimas naudojant minimalų fonemų rinkinį .....	33
3.2 SAMPA–LT fonemų rinkinys.....	34
3.3 Išplėsto fonemų rinkinio atpažinimo tikslumas .....	38
3.4 Anotuoto garsyno tyrimo rezultatai .....	41
3.4.1 Minimalaus fonemų rinkinio atpažinimas .....	41
3.4.2 Minimalaus anotuoto fonemų rinkinio atpažinimo tikslumas .....	42
3.4.3 SAMPA–LT anotuoto fonemų rinkinio atpažinimo tikslumas.....	45
3.4.4 Trifonų atpažinimo tyrimo tikslumas .....	47
3.5 Anotuoto skaičių garsyno atpažinimo tyrimo rezultatų apibendrinimas .....	50
3.6 Apmokytų fonemų įterpimas į skaičių ir vardų garsyno tyrimą .....	52
IŠVADOS IR REZULTATAI .....	53
LITERATŪROS SĄRAŠAS .....	54

## IVADAS

Kalba yra mums labiausiai žinoma ir taikoma bendravimo priemonė norint perduoti savo mintis. Kalbantysis koduoja mintis į garsus (rašantysis koduoja į raides), o klausantysis dekoduoja garsus (skaitantysis dekoduoja raides) į mintis. Svarbiausias bendravimo aspektas – tai tarpusavio informacijos keitimosi suvokimas. Plečiantis technologijų naudojimosi galimybėms siekiama, kad technologijos taip pat būtų pritaikytos įvairiais lygiais skaitmeninant kalbą. Vienas iš jų – automatinis šnekos atpažinimo sistemų kūrimas ir tobulinimas. Technologijų valdymo balsu sąsajos sukūrimas taptų neatsiejama gyvenimo dalimi. Todėl pagrindinis valdymo balsu sistemų kūrimo ir tobulinimo tikslas – sukurti technologines sąsajas, kurios galėtų priimti girdimą informaciją, teisingai ją suprasti bei atsakyti ar atlikti veiksmus pagal balsu gautą informaciją.

Kalbos atpažinimo panaudojimas informacinėse sistemose yra neatsiejamas nuo technologijų tobulėjimo. Su šnekamosios kalbos atpažinimo sistemų taikymu jau dirba tokios kompanijos kaip – *Apple* (Siri), *Samsung* (S – Voice), *Windows* (Cortana) ir *Google* (Google Voice Search). Šių kompanijų programos atlieka paprastas žodžiu atpažįstamas komandas, kaip: naudotis naršykle, siųsti trumpąsias žinutes ar atlikti skambučius, tik tada kai naudojamos jų produktais. Sėkmingas technologijų taikymas šnekamosios kalbos atpažinimo srityje reikalauja didelių finansinių išteklių ir didelės apimties duomenų kiekio, apimančio įvairias balso komandas. Lietuvoje „Lietuvių šneka valdoma paslauga (LIEPA)“ atveria vartus į naują aplinką, kurioje galima dirbti, bendrauti su kompiuteriais šnekant lietuviškai. Šis projektas teikia 7 paslaugas, leidžiančias dirbti su kompiuteriu [1]. Kauno technologijos universitete buvo vykdomas projektas „Hibridinė atpažinimo technologija balso sąsajai (INFOBALSAS)“, kurio tikslas – sukurti hibridinę kompiuterinio medicininių ir farmacinių lietuvių šnekos terminų atpažinimo sąsają, t.y. – panaudoti plačiai paplitusių didelių kalbų atpažintuvų elementus derinant juos su specifiniais lietuvių šneikai orientuotais moduliais. Laukiama, kad sukurtos balso sąsajos galės būti naudojamos kuriant praktines paslaugas [2].

Taigi šiame darbe bus naudojamas bendras skaičių ir vardų garsynas tiriant tikslumo atpažinimą fonemų lygmenyje. Taip pat tiriama anotuotų skaičių garsyno atpažinimo tikslumas. Nors ir žinoma, kad automatinio lietuvių kalbos atpažintuvo tikslumas nėra tobulas ir yra erdvės, kurioje galima tobulėti.

**Darbo tikslas** – ištirti skaičių ir vardų garsynų atpažinimo galimybes naudojant fonemomis grįstus paslėptuosius Markovo modelius (PMM), segmentuoti dalį garsyno fonemų

lygmenyje ir patikrinti segmentavimo įtaką atpažinimo tikslumui naudojant fonemomis ir kontekstinėmis fonemomis grįstus PMM.

#### **Darbo uždaviniai:**

1. atlikti literatūros šaltinių analizę garsyno gramatikų sudarymo, įsigilinti į paslėptųjų Markovo modelių sudarymą, įsisavinti HTK programinio paketo veikimą;
2. nustatyti jungtinio skaičių ir vardų garsyno atpažinimo tikslumą naudojant fonemomis grįstus PMM ir minimalų bei SAMPA–LT fonemų rinkinius;
3. pabandyti pagerinti garsyno atpažinimo tikslumą įvedant naujas fonemas;
4. įsisavinti garsyno segmentavimo fonemų lygmenyje bei gautų žymių pritaikymo HTK paketui metodiką;
5. išanalizuoti gautus darbo rezultatus ir pateikti išvadas.

**Darbo objektas ir metodai.** Skaičių ir vardų daugiadiktorinis garsynas, paslėptieji Markovo modeliai, HTK programinių įrankių rinkinys. Tyrimų rezultatai bus gaunami naudojant HTK programinį įrankį pritaikant fonemų modeliavimo transkripcijas.

**Problema.** Ruošiant ligų pavadinimų atpažinimo pagal jų kodus (juos sudaro raidė ir skaitmenys) programą, reikia turėti jungtinio vardų ir skaičių garsyno atpažinimo tikslumo tyrimų rezultatus bei PMM (paslėptųjų Markovo modelių) akustinius modelius.

Iš S. Laurinčiukaitės daktaro disertacijos seka, kad izoliuotų komandų atpažinimui geriau tinka žodžiais grįsti PMM lyginant su fonemomis grįstais PMM [3].

Kol kas atlikti atskirų garsynų atpažinimo tikslumo tyrimai naudojant fonemomis grįstus PMM nenaudojant garsyno anotavimo ir gauti blogesni atskirų garsynų atpažinimo tikslumo rezultatai lyginant su žodžiais grįstais PMM.

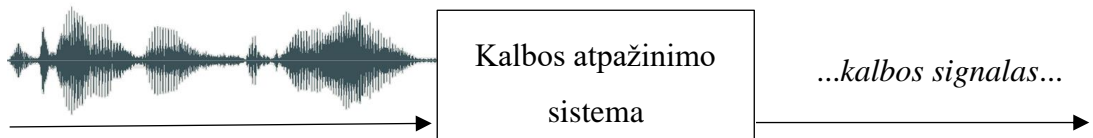
Literatūros analizė ir projekto „INFOBALSAS“ rezultatai rodo, kad galima gauti labai gerus izoliuotų komandų atpažinimo rezultatus naudojant kontekstinėmis fonemomis (trifonais) grįstus PMM ir garsyno segmentavimą fonemų lygmenyje.

**Temos aktualumas.** Automatinis šnekos atpažinimas turintis platesnį žodyną, šiuo atveju skaičių ir vardų, būtų labai aktualus, tai pritaikant, norint užkoduoti kokią nors skaičių ir raidžių seką. Toks informacijos kodavimas paspartintų informacijos perdavimą.



# 1 KALBOS ATPAŽINIMAS

Pagrindinis kalbos atpažinimo sistemos tikslas yra suprasti, girdėti, kalbėti ir veikti pagal balsu gautos informacijos užduotį. Kalbos atpažinimo procesą, galima išskaidyti į tris pagrindinius etapus: duomenų įvedimą į sistemą, požymių išskyrimą ir atpažinimą. Pirmiausia, kad garsą galėtume tirti, jį turime paversti skaitmeniniu signalu, tai padaroma naudojantis mikrofonu ir įrašymo programa.



1.1 pav. Kalbos atpažinimo procesas

Kalbos signalą galima apibūdinti kaip laike kintantį nepriklausomą signalą, kurį apibūdina trys dydžiai – laikas, dažnis ir amplitudė. Antrame etape fiksuotais laiko momentais imami tam tikro ilgio kalbos signalo kadrai, kuriuose skaičiuojami tam tikrų dažnio juostų energiją nusakantys parametrai. Pagal šiuos parametrus – trečiame atpažinimo (klasifikacijos) etape, nustatoma, kokios fonemos yra kalbos signale. Kadangi kalbos atpažinimas yra sudėtingas uždavinys, reikia išsiaiškinti, kas nulemia sudėtingumą ir kokie specifiniai kalbos atpažinimo metodai yra taikomi [4].

## 1.1 Veiksniai įtakoję kalbos atpažinimo sudėtingumą

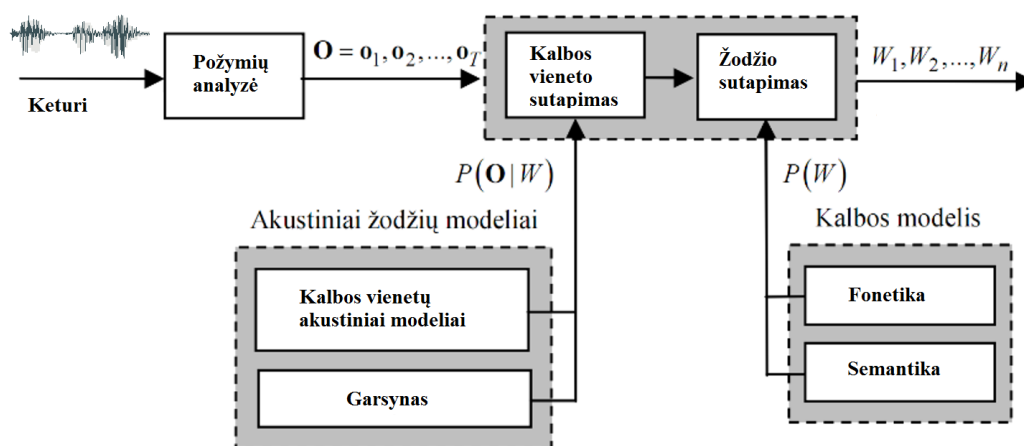
Kalbos atpažinimas yra labai sudėtingas uždavinys reikalaujantis suprasti sudėtingą laike kintantį fizinių procesą. Taip pat galima išskirti esminius trikdžius, kurie kalbos atpažinimą paverčia sunkiu uždaviniu:

1. tos pačios fonemos akustinė realizacija gali labai skirtis, net jei ją ištarė tas pats diktorius;
2. kalbėjimo greitis kinta labai netiesiškai, todėl keičiantis kalbėtojo intonacijai keičiasi kiekvienos fonemos realizacijos ilgis;
3. fonemų akustinė realizacija priklauso nuo gretimų fonemų. Šis reiškinys vadinamas koartikuliacija;
4. kalbėjimo sraute nėra aiškiai apibrėžtų ribų tarp fonemų;

5. kiekvieno žmogaus dikcija yra skirtinga, todėl reikalinga sistema paruošta su kuo didesniu diktorių skaičiumi;
6. jei kuriama atpažinimo sistema paremta fonemų atpažinimu, gali būti pernelyg didelis fonemų skaičius;
7. kalbėjimo fone gali būti ir nekalbinių fragmentų (pvz., triukšmas, kosulys), kuriuos reikia atskirti ir pašalinti;
8. taikant praktikoje, papildomų problemų sukelia foninis triukšmas;
9. regioniniai ir socialiniai dialektai turi skirtingus kalbai būdingus požymius;
10. lyties įtaka balsuose. Moterų trumpesnis vokalas ir pats balso tonas bent du kartus aukštesnis nei vyrų [5].

## 1.2 Kalbos atpažinimo būdai

Kalbos atpažinimo būdai skirstomi pagal taikymo sritį ir poreikį. Jei reikalingas kalbos signalo vertimas į tekstą, naudojamos *kalbos atpažinimo* technologijos. Jei norimas tekstas verčiamas į garsinius signalus, tai reikalingos *kalbos sintezės* technologijos. Sudėtingiausias kalbos atpažinimo uždavinys yra atpažinti *ištisinę arba natūralią kalbą*. Šiomis technologijomis ir paremtos visos programos gebančios išskirti ir atlikti užduotis [3]. Šio darbo užduotis izoliuotų kalbos komandų atpažinimas fonemų lygmenyje.



1.2.1 pav. Izoliuotos komandos atpažinimo sistemos schema [6]

Pagrindinės automatinės izoliuotos komandos atpažinimo sistemos dalys:

- požymių išskyrimas iš kalbos signalo;

- garsinę informaciją atitinkančio modelio formavimas ir lyginimas su garsyne esančiais modeliais;
- kalbos fonetinio fragmento sutapimo atveju testavime pateikiamas atitinkama komanda.

Fonologija nagrinėja fonetinius kalbos elementus, kuriuos panaudojant kuriamos kalbos analizės sistemos. Fonologija atsako į klausimą, kiek ir kokių fonetinių vienetų yra kalbos vienetė, o akustinė fonetika pateikia šių vienetų charakteristikas spektriniais metodais.

Būtina tinkamai pasirinkti tinkamą komandos segmentų skaičių, kuris lemia tinkamą kalbos atpažinimo sistemos kūrimą. Pagrindiniai naudojami kalbos tyrimo vienetai yra:

- *fonema (phone)* minimalus kalbos garsas iš kurio susideda tariamas žodis;
- *kontekstinė fonema (triphone)* minimalių fonemų junginys, susietas eiliškumo grandine.

Pasirinkus automatinę kalbos atpažinimo sistemą, kuri išskiria tokias smulkesnes žodžio dalis, kaip kontekstinės fonemos ar fonemos, galima sukurti sistemą, kuri atskirtų fonemų junginius ir atpažintų fonemų kombinacijas atitinkantį žodį [7].

### 1.3 Paslėptieji Markovo modeliai

Paslėptųjų Markovo modelių (angl. *Hidden Markov Models (HMM)*) metodas gali aprašyti tam tikro laike kintančio parametrinio atsitiktinio proceso savybių kitimą ir to kitimo statistines charakteristikas. Paslėptasis Markovo modelis modeliuoja atsitiktinius procesus būsenų sekomis susietomis perėjimo tikimybėmis. Atsitiktinis procesas juda būsenų seka, kiekvienoje būsenoje generuodamas kokį nors įvykį. Pagal įvykių seką negalima pasakyti kokia yra būsenos seka. Čia įvykis yra tikimybinės būsenos funkcija, nes viena būseną atitinka kelis įvykius, pasikartojančius kitose būsenose [8]. Izoliuotų komandų atpažinime, sukuriami atskirų fonemų paslėptieji Markovo modeliai. Skaičiuojant izoliuotos komandos sudarančių fonemų sekos tikimybes, visame PMM fonemų rinkinyje ieškoma geriausio kelios atitinkančio fonemų perėjimo būsenų sekos, toks procesas yra atsitiktinių įvykių seka, aprašoma galimų įvykių aibe tačiau negalima nusakyti būsenos sekos. Todėl ir vadinamas paslėptasis Markovo modelis. Šnekos atpažinime nagrinėjama fonetinių vienetų seka yra atsitiktinis procesas, kuris negali būti stebimas tiesiogiai: atpažinimo sistema fiksuoja požymių vektorių sekas, kurios pačios yra atsitiktinis procesas ir stebimą atsitiktinį procesą turi susieti su tiesiogiai matomais fonetiniais vienetais.

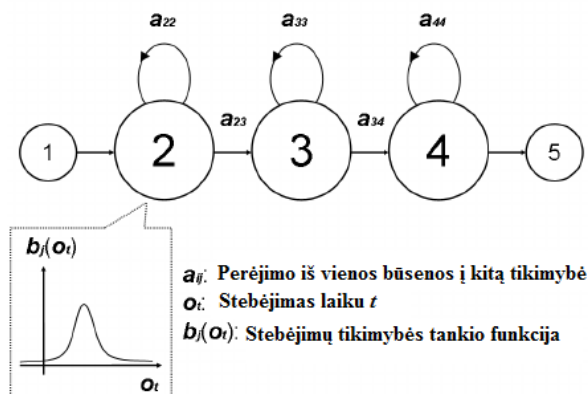
Paslėptasis Markovo modelis aprašomas nusakant penkis dydžius:

- būsenų  $Q = (q_1, q_2, \dots, q_n)$  skaičius  $N$ .

- skirtingų stebėjimų būsenoje skaičius  $D$  – diskrečiu atveju arba stebėjimų  $O = (o_1, o_2, \dots, o_T)$  pasiskirstymo tankis – tolydžiu atveju.
- perėjimo iš vienos būsenos į kitą tikimybių matricą  

$$A = \{a_{ij}\}, i, j = 1, \dots, N, \forall i, j \ a_{ij} \geq 0 \text{ ir } \forall i \sum_{j=1}^N a_{ij} = 1.$$
- stebėjimų tikimybinių skirstinių  $B = \{b_j(k)\}, k = 1, \dots, D, j = 1, \dots, N$  – diskrečiuoju atveju arba stebėjimų tikimybės tankio funkcija  $B = \{b_j(o_t)\}, j = 1, \dots, N, t = 1, \dots, T$  – tolydžiu atveju.
- pradinio (initial) buvimo būsenoje tikimybė  $\pi_i = P(q_1 = i), i = 1, \dots, N$ . Šnekos atpažinimo modeliavime pradine būsena laikoma pirmoji paslėptojo Markovo modelio būsena  $\pi = 1$ .

Šie 5 dydžiai apibūdina konkretų paslėptojo Markovo modelį kiekvienai fonemai [3].



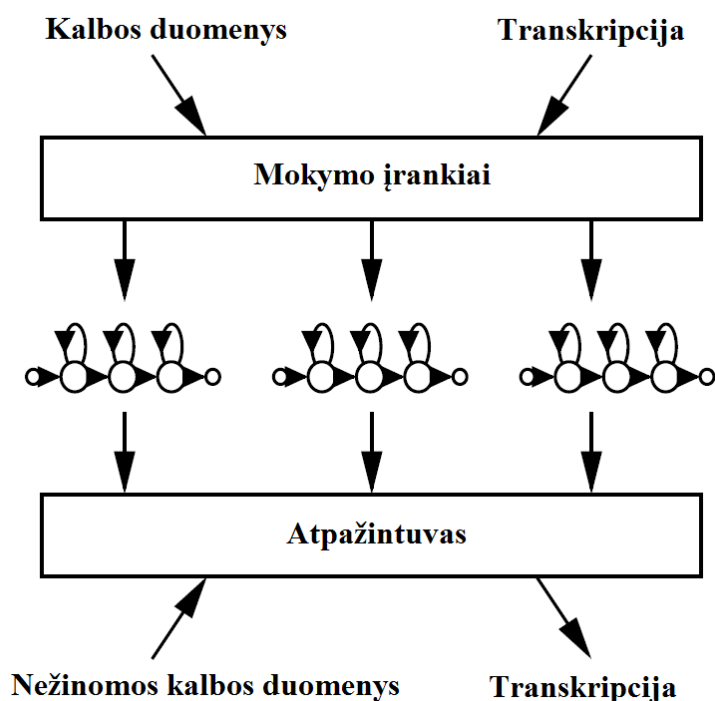
**1.3.1 pav.** Iš kairės į dešinę nukreiptas paslėptasis Markovo modelis

Siekiant realizuoti izoliuotų žodžių atpažinimą fonemų lygmenyje, kiekvienai fonemai sukonstruojama paslėpto Markovo modelio būsena. Mokymo metu kiekvieną žodį sudarančių fonemų mokymo seka segmentuojama į būsenas ir tada studijuojamos vektorių savybės, taip suformuojant kiekvienos būsenos stebėjimų aibę, aprašomą skirstiniu. Kalbant plačiau, prieš mokymą atliekamas modelio įverčių (perėjimo tikimybių, *Gausinių* stebėjimų skirstinių) inicializavimas. Jis atliekamas *Viterbi* algoritmu, kuris iteracijų būdu ieško kalbos vektorių priklausymo vienai ar kitai būsenai didžiausios tikimybės – ieškoma labiausiai tikėtina būsenų seka, atitinkančia mokomąjį pavyzdį. Po inicializavimo atliekamas tikrasis mokymas – jau esamų įverčių patikslinimas. Jis atliekamas *Baum–Welch* algoritmu, kuris jau inicializuotas reikšmes dar kartą tikslina. Atpažinimo metu daroma prielaida, kad nežinomą komandą atstovaujanti tiriamų kalbos vektorių seka yra gaunama iš paslėptojo Markovo modelio. Skaičiuojamos tos komandos fonemų atitikimo kiekvienam modeliui tikimybės ir labiausiai tikėtinas modelis identifikuoja

komandą. Atpažinimo procesas grindžiamas *Viterbi* algoritmu. Šis algoritmas gali būti vizualizuotas kaip geriausio kelio paieška matricoje, kurioje vertikaliai išdėstomos paslėptosios būsenos, o horizontaliai – kalbos kadrai (laikas, vektoriai).

#### 1.4 HTK programinis įrankis

HTK (angl. *Hidden Markov Model Toolkit*) paketą galima laikyti, įrankiu, skirtu kurti ir manipuluoti paslėptuosius Markovo modelius. Iš principo HTK paketas sukurtas manipuluoti bet kokius laike kintančius procesus, tačiau pagrindine naudojamas kalbos tyrimo realizavimui.



1.4.1 pav. HTK paketo struktūrinis panaudojimas

Kaip parodyta 1.4.1 paveikslėlyje yra du pagrindiniai apdorojimo etapai. Pirmiausia HTK mokymo įrankiai naudojami paslėptųjų Markovo modelių parametrų įvertinimui, panaudojant mokymo kalbos duomenis (garso failus) ir jiems priskirtas transkripcijas. Antrame etape nežinomiems kalbos duomenims yra priskiriamos transkripcijos naudojantis atpažintuvo įrankiais.

Vykdam atpažinimą su HTK paketu, kiekvienam etapui atlikti reikia panaudoti po atskirą programą iš HTK atpažinimo programų rinkinio. Bendruoju atveju galima įvardinti 4 pagrindinius etapus atliekant bandymus su HTK [9]:

- žodyno sudarymas. Pirmiausia reikalingi kalbos duomenys su kuriais bus mokomi PMM. Taip pat reikalingas žodynas, apibūdinantis galiojančias

komandas ir tų komandų akustinių vienetų (fonemų) kombinacijas, kurios reikalingos atpažinimui.

- požymių išskyrimas. Atpažinimo sistemoje naudojami ne patys įrašai, o iš kalbos signalo išskirtos parametrizuotos požymių vektorių sekos pasinaudojant Melų dažnių skalės kepstiniais koeficientais (MFCC).
- akustinių modelių mokymas. Mokymo etape visoms žodyno komandoms pagal parinktas transkripcijas sudaromi tikimybiniai akustiniai modeliai ir apskaičiuojami kiekvieno akustinio modelio tikėtimumo įverčiai pasitelkiant paslėptųjų Markovo modelių metodiką.
- testavimo duomenų atpažinimas. Testavimui naudojami kiti kalbos duomenys, kurie nebuvo naudojami mokymo procese. Sistema palygina žinoma komandą su nežinomu kalbos elementu ir pateikia tą komandą, kurią reprezentuoja didžiausią tikėtimumo įvertį turintis modelis [10].

HTK programinio paketo redagavimui ir tobulinimui naudojamos papildomos programos („*Notepad++*“ ir „*Far Manager 3*“), kurios padidina darbo našumą ir suteikia galimybę matyti esamas klaidas atliekant tyrimus.

HTK programinis paketas naudoja rašytinį komandinį failą „*batch*“ („*\*.bat*“). Šių failų paskirtis – aprašyti atliekamų komandų nuorodas, kurios įvykdomos su kiekvienam failui priskirtu „*exe*“ failu.

## 2 METODINĖ DALIS

### 2.1 Vardų ir skaičių garsynas

Garsyną galima laikyti signalų rinkiniu, skirtu šnekos atpažinimui, sintezei, diktoriaus identifikavimui, verifikavimui ar kitiems uždaviniams, reikalaujantiems šnekos signalo apdorojimo. Garsynai dažnai naudojami kartu su papildoma programine įranga (HTK), skirta šnekos signalų rinkinių apdorojimui, duomenų tvarkymui ir pateikimui. Šnekos duomenų bazių – garsynų – yra daug ir labai įvairių, o tai suteikia galimybę taikyti kalbos technologijas įvairiose sferose [11].

Garsynams yra keliami reikalavimai, kad juos efektyviai galėtume koreguoti norėdami atlikti tyrimus. Garsyno duomenų bazė privalo būti lengvai prieinama ir apdorojama, garsyno pateikiama akustinė/ fonetinė medžiaga turi būti išsami, atvaizduoti diktorių įvairovę ir aktualų žodyno turinį. Struktūrizuota garsų įrašų aibė turi turėti savo žodyno atitikmenis skirtingais transkribavimo lygio vienetais (žodžiais, skiemenims ar fonemomis), kurių reikalauja tyrimo paskirtis.

TIMIT (angl. *Texas Instruments/Massachusetts Institute of Technology*) šnekos duomenų bazė laikoma klasikinio garsyno pavyzdžiu. TIMIT garsynas yra hierarchinės duomenų bazės pavyzdys, kur svarbų vaidmenį atlieka failų sistemos hierarchinė struktūra ir ten sudėta informacija. TIMIT modelio direktorių struktūroje užkoduojama pagrindinė garsyno organizacinė informacija (\*.wav – garsiniai audio failai), o tekstiniuose failuose (\*.txt) pateikiama papildoma informacija apie kalbos signalus, diktorius ir jų anotacijas žodžio lygmenyje ir viskas aprašoma ASCII formatu.

#### 2.1.1 Garsyno paruošimas tyrimams

Skaičių ir vardų balso komandos buvo sudiktuotos ir įrašytos naudojant MS DOS operacinės sistemos aplinkoje veikiančią programą „inp\_sr16.exe“.

Daugiadiktorinį vardų garsyną sudaro 21 diktoriaus balso įrašai tariantys vardus. Katalogas „Vardai“ talpina vardų garsyną, kuriame yra balso komandos. Pavyzdžiui, failo pavadinimas „0A00.wav“ nurodo, kad 0 – tai diktoriaus eilės numeris, A – lotyniškai raidei priskirtas vardas „Austėja“, o pavadinimas „20Z19.wav“ reiškia 20 – diktoriaus eilės numeris, Z

– lotyniškai raidei priskirtas vardas „*Zacharijus*“. Daugiadiktorinį garsyną sudaro 12 moterų ir 9 vyrų balso įrašų komandos. Kiekvienas diktorius tą pačią komandą ištaria po 20 kartų. Vardų garsyną sudaro 26 skirtingų vardų balso komandos – atitikmenys kiekvienai lotyniškos abėcėlės raidei. Vardų garsyno apimtį sudaro:  $Vardų\ garsyno\ apimtis = diktorių\ skaičius \times ištarių\ skaičius \times skirtingų\ komandų\ skaičius = 21 \times 20 \times 26 = 10920$ .

Skaičių garsyną sudaro arabiški skaitmenys nuo 0 iki 9, kiekvienas diktorius tą pačią komandą ištaria 20 kartų. Daugiadiktorinį skaičių garsyną sudaro 23 moterų ir 7 vyrų balso įrašų komandos. „*Skaičiai*“ katalogas talpina skaičių garsyną, kuriame yra balso komandos. Pavyzdžiui, failo pavadinimas „*FVAIVAI000.wav*“ tai, F ( angl. *Female*) – moteris, „VAI...VAI...“ – pirmos trys diktoriaus vardo ir pavardės raidės, ...0... – tai skaičiaus „*nulis*“ balso komanda, ...00 – kelintą kartą yra ištartas tas skaičius. Dar vienos komandos failo pavyzdys, „*MRIMAPA919.wav*“ M ( angl. *Male*) – vyras, tariantis skaičių „*devyni*“ devynioliktąjį kartą.  $Skaičių\ garsyno\ apimtis = diktorių\ skaičius \times ištarių\ skaičius \times skirtingų\ komandų\ skaičius = 30 \times 20 \times 10 = 6000$ .

Visą bendrą garsyną sudaro 16920 vardų ir skaičių „*\*.wav*“ failų duomenų bazė. Audio formato „*\*.wav*“ parametrai: 16000Hz, 16 bitų monofoninis garso failas. Tokie parametrai būtini, kad tiktų HTK programiniam paketui.

## 2.2 Hidden Markov Model Toolkit įsisavinimas

Šiame skyriuje yra aprašoma darbo eiga, vardų ir skaičių komandų atpažinimo tikslumo tyrimams atlikti. Naudojamas HTK ( angl. *Hidden Markov Model Toolkit*) programinis paketas, sukurtas Kembridžo universiteto inžinerijos departamente. Taigi, šiame skyriuje aptariamas praktinis HTK programinio paketo panaudojimas ir įsisavinimas, fonemomis grįstais paslėptaisiais Markovo modeliais bei reikalingų duomenų paruošimas izoliuotų komandų atpažinimui. HTK paketo modeliavimo metodiką sudaro šie žingsniai [12]:

**1 žingsnis** – žodyno gramatikos „*gram*“ kūrimas, kuriame aprašomos naudojamos lietuviškų vardų ir skaičių komandos, kurios faile atrodo taip:

---

```
$word=AUSTEJA|BOLESLOVAS|CECILIJA|DONATAS|EIMANTAS|FAUSTA|GRAZVYDAS|HANSAS|IZ  
AOKAS|JONAS|KAROLIS|LAIMA|MARTYNAS|NOJUS|OSKARAS|PATRIKAS|KJU|RICARDAS|SANDRA  
|TEODORAS|ULIJONA|VACYS|WASHINGTON|IKSAS|YGREKAS|ZACHARIJUS|VIENAS|DU|TRYS|KE  
TURI|PENKI|SHESHI|SEPTYNI|ASHTUONI|DEVYNI|NULIS;  
($word)
```

---



Panaudojant HTK programinį įrankį – „Hparse.bat“, kurio komandinė eilutė „HParse gram wdnet“, kuris naudoja „gram“ – žodžių sąrašo failą, automatiškai sukuria žodžių tinklo gardelę „wdnet.txt“.

**2 žingsnis** – fonemų modelių žodyno paruošimas. Failai – „dict.txt“, „monophones1.txt“, „monophones0.txt“ paruošiami rankiniu būdu.

Faile „dict“ pateikiamas skaičių ir vardų komandų žodynas, suskaidytas į smulkesnius komandų segmentus – fonemas. Fonema „sp“ – trumpą pauzę reiškianti fonema, kuri prirašoma prie kiekvienos komandos tam, kad būtų pavaizduota komandos pabaigą nurodanti pauzė. „SENT-END [] sil“ ir „SENT-START [] sil“ yra naudojamos aprašyti tylos modeliui, kuris yra komandos pradžioje ir pabaigoje. Failo „dict“ turinys atrodo taip:

---

AUSTEJA	au sm tm Ek j a	sp
BOLESLOVAS	b o lm ek s l o v a s	sp
CECILIJA	cm em cmi ik lm ik j a	sp
DONATAS	d o n ak t a s	sp
EIMANTAS	ei m a nn t a s	sp
FAUSTA	f aku s t af	sp
GRAZVYDAS	g r ak Zm vm y d a s	sp
HANSAS	h akm nn s a s	sp
IZAOKAS	i z a ok k a s	sp
JONAS	j oo n a s	sp
KAROLIS	k ak r o lm i s	sp
LAIMA	l aki aki m a	sp
MARTYNAS	m a rrm tm ii n a s	sp
NOJUS	n oo j u s	sp
OSKARAS	oks s k a r a s	sp
PATRIKAS	p ak tm rm i k a s	sp
KJU	km j uk	sp
RICARDAS	rm ik C a rr d a s	sp
SANDRA	s akm nn d r a	sp
TEODORAS	tm e o d ok r a s	sp
ULIJONA	u lm i j oo n a	sp
VACYS	v a cm ii s	sp
WASHINGTON	v ak shm i ng g t o n	sp
IKSAS	ikk ikk kx s a s	sp
YGREKAS	ii gm rm e k a s	sp
ZACHARIJUS	z a x ak rm i j u s	sp
ASHTUONI	a sh t uo nm ik	sp
DEVYNI	dm e vm ii nm ik	sp
DU	d ud ud	sp
KETURI	km et t u ri ir	sp
NULIS	n uk lm i s	sp
PENKI	pm en nk km ik	sp
SEPTYNI	sem e pp tm y nm ik	sp
SHESHI	shm esh shm ish	sp
TRYS	tm rm y s	sp
VIENAS	vm i ek n a s	sp
SENT-END	[]	sil
SENT-START	[]	sil

---

Iš failo „dict“ išrenkamos visos pasikartojančios fonemos ir surašomos į „monophones1“ failą įtraukiant ir trumpos pauzės „sp“ bei tylos „sil“ fonemas. Taip pat sukuriamas failas „monophones0“, kuris neturi pauzės fonemos „sp“. Failo „monophones1“ turinys atrodo taip:

---

au	a	s
sm	b	l
tm	o	v
Ek	lm	cme
j	ek	em

cm	z	pm
e	ok	uo
ik	k	nk
i	oo	sh
d	aki	nm
n	rrm	dm
ak	ii	esh
t	u	ish
ei	oks	en
m	p	et
nn	rm	ri
f	km	ir
aku	uk	ud
af	C	kx
g	rr	cmi
r	shm	sem
Zm	ng	pp
vm	ikk	sp
y	gm	sil
h	x	
akm	ie	

**3 žingsnis** – MLF „*Master label file*“ kūrimas. Failai „*words.mlf*“ ir „*mkphones.led*“ paruošiami „rankiniu“ būdu. „*words.mlf*“ faile yra surašomas visų komandų rinkinys. „*words.mlf*“ paruošiamas pagal „*dict*“ faile esančias komandas ir duomenų, skirtų apmokymui, failų pavadinimus. „*words.mlf*“ failo pradžios ir pabaigos fragmentas atrodo taip:

```

#!MLF!#
"*0A00.lab"          "*" /MVYGVAI917.lab"
AUSTEJA              DEVYNI
.
"*0A01.lab"          "*" /MVYGVAI918.lab"
AUSTEJA              DEVYNI
.
"*0A02.lab"          "*" /MVYGVAI919.lab"
AUSTEJA              DEVYNI
.

```

Faile „*mkphones0.led*“ surašomos komandos, kurios bus taikomos aktyvavus programą „*HLEd.bat*“. Failo „*mkphones0.led*“ turinys atrodo taip:

```

EX
IS sil
DE sp

```

Komanda „*EX*“ pakeičia kiekvieną ištariamą „*words.mlf*“ faile pagal atitinkamą fonemą iš fonemų sąrašo aprašyto žodyno „*dict*“ faile.

Komanda „*IS*“ įterpia tylos fonemą „*sil*“ kiekvieno ištaramo pradžioje ir pabaigoje.

Komanda „*DE*“ ištrina visas trumpų pauzių – „*sp*“ žymes.

Paruošus „*words.mlf*“ ir „*mkphones0.led*“ failus aktyvuojamas HTK programinis įrankis – „*HLEd.bat*“, kurio komandinė eilutė „*HLEd -l features -d dict -i phones0.mlf mkphones0.led words.mlf*“, kuris aktyvuojamas kartu su failais „*words.mlf*“ ir „*mkphones0.led*“ išplečia komandų

transkripcijas į fonemų modelio transkripcijas – sukuria failą „*phones0.mlf*“, kurio fragmentas atrodo taip:

---

```
#!MLF!#
"features/*0A00.lab"
sil
au
sm
tm
Ek
j
a
sil
.
"features/*0A01.lab"
sil
au
sm
tm
Ek
j
a
sil
.
```

---

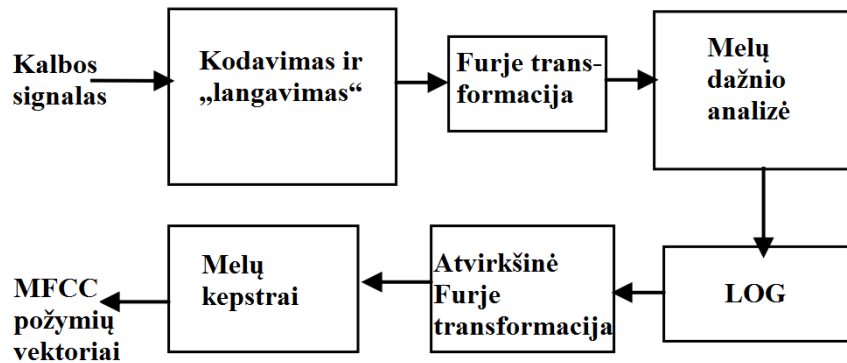
Po to aktyvuojama programa „*HLEd2.bat*“, kuri atitinkamai aktyvuojamas kartu su failuose „*words.mlf*“ ir „*mkphones1.led*“ esančia informacija sukuria failą „*phones1.mlf*“.

„*phones0.mlf*“ ir „*phones1.mlf*“ failai yra beveik identiški, tik „*phones0.mlf*“ ištarimų transkripcijos yra be trumpųjų pauzių fonemų „*sp*“.

**4 žingsnis** – duomenų kodavimas.

Garsynų duomenys „\*.wav“ formatu yra laikomi duomenų kataloge „*data*“, o šiame etape sukurtus požymius laikysime kataloge – „*features*“.

Duomenų kodavimas vyks aktyvavus HTK programinį įrankį „*HCopy.bat*“, kurio komandinė eilutė „*HCopy -A -D -T 1 -C config -S codetr.scf*“, kuris automatiškai pagal nurodytus parametrus „*config*“ faile, apdoroja kalbos signalą ir išskiria MFCC požymio vektorius. MFCC (angl. *Mel Frequency Cepstral Coefficients*) – Melų dažnių kepstriniai koeficientai, Melu dažnio skalėje išreikšti logaritminiai energijos spektrai išreikšti iš *Furjė* transformacijos. Kepstrinis šnekos spektro atvaizdavimas išskiria geras signalo savybes kiekvienam lango kadru. MFCC apskaičiuojami iš kalbos signalo iškerpant 25ms trukmės Hemingo langą, kuris atnaujinamas kas 10ms. Šiam langui atliekama Furjė transformacija ir gautas amplitudinis spektras išskaidomas į 22 trikampus filtrus ir jiems atliekama logaritminė moduliacija. Šiems trikampių filtrų logaritminėms reikšmėms atliekama atvirkštinė Furjė transformacija ir gaunami Melų dažnių skalės kepstriniai koeficientai (MFCC).



2.2.1 pav. Kalbos signalo požymių išskyrimas MFCC metodu [13]

Duomenų kodavimui reikalingus tekstinius „*config*“ ir „*codetrscp*“ failus sukuriame „rankiniu“ būdu. Parametrai „*config*“ faile aprašomi taip:

---

```

SOURCEFORMAT=WAVE
HIFREQ=8000
TARGETKIND=MFCC_E
TARGETRATE=100000
SAVECOMPRESSED=T
SAVEWITHCRC=T
WINDOWSIZE=250000
USEHAMMING=T
PREEMCOEF=0.97
NUMCHANS=26
CEPLIFTER=22
NUMCEPS=12
  
```

---

Parametrai nekeičiami, nes tokie nurodymai yra HTK programinio paketo [14].

Faile „*codetr.scf*“ surašome visų programos „*HCOPY.bat*“ naudojamų šaltinio failų direktorijas ir nurodoma direktoriją į kurią reikia perkelti sukuriamus MFCC vektorių koeficientus. Toks parametrų konvertavimas reikalingas, nes HTK paketu negalime tiesiogiai tirti audio failų. Failo „*codetr.scf*“ turinio fragmentas atrodo taip:

---

```

data\0A00.wav features\0A00.mfc
data\0A01.wav features\0A01.mfc
data\0A02.wav features\0A02.mfc
.....
data\MVYGVAI917.wav features\MVYGVAI917.mfc
data\MVYGVAI918.wav features\MVYGVAI918.mfc
data\MVYGVAI919.wav features\MVYGVAI919.mfc
  
```

---

Paruošus „*config*“ ir „*codetr.scf*“ failus aktyvuojamas „*HCOPY.bat*“, kuris apdoroja duomenis pagal nurodytus konvertavimo parametrus ir sukuria foneminių atpažinimo modelių požymių vektorių failus skirtus apmokymui ir patalpina juos kataloge „*features*“.

5 žingsnis – foneminių modelių kūrimas. Sukuriami du tekstiniai failai „*train.scf*“ ir „*config6*“. Failas „*config6*“ skiriasi nuo „*config*“ tik Melų dažnių skalės koeficiento požymiu,

kuris turi didesnę reikšmę 39 (ši reikšmė nurodo fonemos modelio vektoriaus dydį). Vektoriaus dydis 39 gaunamas prie MFCC parametrų vektoriaus ilgio (13) pridėjus delta koeficientus (+13) ir pridėjus akceleracijos koeficientus (+13). Failo „*config6*“ turinys atrodo taip:

---

```
TARGETKIND=MFCC_E_D_A_Z
TARGETRATE=100000
SAVECOMPRESSED=T
SAVEWITHCRC=T
WINDOWSIZE=250000
USEHAMMING=T
PREEMCOEF=0.97
NUMCHANS=26
CEPLIFTER=22
NUMCEPS=12
ALLOWXWRDEXP=T
```

---

Taip pat sukuriamas „*train.scp*“ failas, kuriame yra nuorodos, kad programa „*HCompv.bat*“, kurios komandinė eilutė „*HCompV -A -D -T 1 -C config6 -f 0.01 -m -S train.scp -M hmm0 proto*“ pasiimtų jai reikalingus komandų požymių failus. Failo „*train.scp*“ fragmentas atrodo taip:

---

```
features\0A00.mfc
features\0A01.mfc
features\0A02.mfc
.....
features\MVYGVAI917.mfc
features\MVYGVAI918.mfc
features\MVYGVAI919.mfc
```

---

Turint paruoštus „*train.scp*“ ir „*config6*“ failus kuriamas katalogas su pavadinimu „*hmm0*“, po to aktyvuojame HTK programinį įrankį „*HCompv.bat*“. Programai reikalingas yra dar vienas pradinis modelio failas „*proto*“, paimtas iš HTK paketo, kuriame yra išdėstyti pradiniai modelio vektoriaus reikšmių koeficientai. Pradinio „*proto*“ failo turinys atrodo taip:

---

```
~o <VecSize> 39 <MFCC_E_D_A_Z>
~h "proto"
<BeginHMM>
<NumStates> 5
<State> 2
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State> 3
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State> 4
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
```

---

```

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<Transp> 5
0.0 1.0 0.0 0.0 0.0
0.0 0.6 0.4 0.0 0.0
0.0 0.0 0.6 0.4 0.0
0.0 0.0 0.0 0.7 0.3
0.0 0.0 0.0 0.0 0.0
<EndHMM>

```

`~o` – pradžios simbolis;

`<VECSIZE> 39<NULLD><MFCC_E_D_A_Z>` – šioje eilutėje nurodomas vektoriaus dydis – tie patys 39 požymiai, `MFCC_E_D_A_Z` reiškia, kad bus naudojami MFCC požymis ir `E` – kalbos signalo energijos koeficientas, `D` – delta koeficientas, `A` – pagreičio koeficientas ir `Z` – nulinės reikšmės, naudojamas vektoriaus normalizavimui, požymio koeficientas [14];

`~h "proto"` – tai modelio failo pavadinimas;

`<BEGINHMM>` – modelio pradžios žyma;

`<NUMSTATES>` – nurodomas modelių faile naudojamas būsenų skaičius;

`<STATE> 2` – antrosios būsenos numeris;

`<MEAN> 39` – požymių vidurkiai, kurie modelių mokymo metu bus perskaičiuojami automatiškai;

`<VARIANCE> 39` – požymių dispersijos, kurios apmokymo metu bus perskaičiuotos automatiškai;

`<TRANSP>` – „*transponētē*“ – paskutinė būsena. Šioje būsenoje yra sudaroma tikimybių matrica;

`<ENDHMM>` – modelio pabaigos žyma;

Programa „*HCompv.bat*“ išanalizuoja požymio vektorius ir suskaičiuoja vidurkį bei kovariaciją. Šie duomenys pakeičia pirminiame modelyje esančias nulines reikšmes ir kataloge „*hmm0*“ sukuria du naujus failus „*proto*“ (naujame proto atsiranda vektoriaus koeficientų reikšmės pagal naudojamus duomenis) ir „*vFloors*“.

Kataloge „*hmm0*“ reikia susikurti du naujus failus – „*macros*“ ir „*hmmdefs*“. Failas „*macros*“ sukuriamas prie „*vFloors*“ faile esančios informacijos pridėdant kelias papildomas parametrų eilutes:

```

~o
<STREAMINFO> 1 39
<VECSIZE> 39<NULLD><MFCC_E_D_A_Z><DIAGC>
~v varFloor1
<Variance> 39
6.676555e-001 4.758350e-001 4.600532e-001 6.292701e-001 5.724778e-001 4.315946e-001 4.618952e-
001 3.731705e-001 3.511576e-001 3.265702e-001 3.147837e-001 2.186281e-001 1.010098e-003
2.138309e-002 1.999862e-002 1.948002e-002 2.603166e-002 2.719533e-002 2.570571e-002 2.634540e-002

```

2.615026e-002 2.325648e-002 2.304726e-002 2.121749e-002 1.704650e-002 2.238667e-005 2.932143e-003  
2.874293e-003 2.931769e-003 3.860753e-003 4.484176e-003 4.342295e-003 4.584453e-003 4.661352e-003  
4.134187e-003 4.097667e-003 3.754075e-003 3.134713e-003 2.875983e-006

Naujai sukurtas „*proto*“ failas pridamas prie kiekvienos fonemos naudojamos garsyne, taip sukuriamas failas „*hmmdefs*“. Prie „*monophones0*“ failo esančių fonemų sąrašo, kiekvienai fonemai pridama „~h“ žymė ir „*hmm0*“ esančiame failo „*proto*“, su perskaičiuotomis reikšmėmis, fragmentą prasidedantį <*BEGINHMM*> ir pasibaigiantį <*ENDHMM*>.

```
~h "au"
<BEGINHMM>
<NUMSTATES> 5
<STATE> 2
<MEAN> 39
8.616629e-009 3.657269e-009 1.885546e-009 2.590139e-009 2.164106e-009 6.823547e-010 1.482190e-009 2.671190e-009 5.252258e-010
9.635480e-010 -1.899006e-010 -4.288125e-010 5.085518e-001 -3.810956e-002 -1.362875e-002 -8.100101e-003 1.296204e-003 7.181698e-
003 7.608147e-003 1.141536e-002 1.056800e-002 7.810549e-003 8.757939e-003 5.014183e-003 4.401922e-003 -3.264264e-004 -9.556224e-
004 -1.139633e-003 -5.760615e-004 -1.524245e-004 5.534238e-004 3.708863e-004 6.497892e-004 8.818322e-004 8.973945e-004 5.504883e-
004 5.288412e-004 4.615969e-004 -4.664228e-005
<VARIANCE> 39
6.676556e+001 4.758350e+001 4.600533e+001 6.292701e+001 5.724778e+001 4.315947e+001 4.618952e+001 3.731705e+001 3.511576e+001
3.265702e+001 3.147837e+001 2.186281e+001 1.010098e-001 2.138309e+000 1.999862e+000 1.948002e+000 2.603166e+000 2.719533e+000
2.570571e+000 2.634540e+000 2.615026e+000 2.325648e+000 2.304726e+000 2.121749e+000 1.704650e+000 2.238667e-003 2.932143e-001
2.874293e-001 2.931769e-001 3.860753e-001 4.484176e-001 4.342295e-001 4.584453e-001 4.661352e-001 4.134188e-001 4.097667e-001
3.754075e-001 3.134713e-001 2.875983e-004
<GCONST> 9.817532e+001
<STATE> 3
<MEAN> 39
8.616629e-009 3.657269e-009 1.885546e-009 2.590139e-009 2.164106e-009 6.823547e-010 1.482190e-009 2.671190e-009 5.252258e-010
9.635480e-010 -1.899006e-010 -4.288125e-010 5.085518e-001 -3.810956e-002 -1.362875e-002 -8.100101e-003 1.296204e-003 7.181698e-
003 7.608147e-003 1.141536e-002 1.056800e-002 7.810549e-003 8.757939e-003 5.014183e-003 4.401922e-003 -3.264264e-004 -9.556224e-
004 -1.139633e-003 -5.760615e-004 -1.524245e-004 5.534238e-004 3.708863e-004 6.497892e-004 8.818322e-004 8.973945e-004 5.504883e-
004 5.288412e-004 4.615969e-004 -4.664228e-005
<VARIANCE> 39
6.676556e+001 4.758350e+001 4.600533e+001 6.292701e+001 5.724778e+001 4.315947e+001 4.618952e+001 3.731705e+001 3.511576e+001
3.265702e+001 3.147837e+001 2.186281e+001 1.010098e-001 2.138309e+000 1.999862e+000 1.948002e+000 2.603166e+000 2.719533e+000
2.570571e+000 2.634540e+000 2.615026e+000 2.325648e+000 2.304726e+000 2.121749e+000 1.704650e+000 2.238667e-003 2.932143e-001
2.874293e-001 2.931769e-001 3.860753e-001 4.484176e-001 4.342295e-001 4.584453e-001 4.661352e-001 4.134188e-001 4.097667e-001
3.754075e-001 3.134713e-001 2.875983e-004
<GCONST> 9.817532e+001
<STATE> 4
<MEAN> 39
8.616629e-009 3.657269e-009 1.885546e-009 2.590139e-009 2.164106e-009 6.823547e-010 1.482190e-009 2.671190e-009 5.252258e-010
9.635480e-010 -1.899006e-010 -4.288125e-010 5.085518e-001 -3.810956e-002 -1.362875e-002 -8.100101e-003 1.296204e-003 7.181698e-
003 7.608147e-003 1.141536e-002 1.056800e-002 7.810549e-003 8.757939e-003 5.014183e-003 4.401922e-003 -3.264264e-004 -9.556224e-
004 -1.139633e-003 -5.760615e-004 -1.524245e-004 5.534238e-004 3.708863e-004 6.497892e-004 8.818322e-004 8.973945e-004 5.504883e-
004 5.288412e-004 4.615969e-004 -4.664228e-005
<VARIANCE> 39
6.676556e+001 4.758350e+001 4.600533e+001 6.292701e+001 5.724778e+001 4.315947e+001 4.618952e+001 3.731705e+001 3.511576e+001
3.265702e+001 3.147837e+001 2.186281e+001 1.010098e-001 2.138309e+000 1.999862e+000 1.948002e+000 2.603166e+000 2.719533e+000
2.570571e+000 2.634540e+000 2.615026e+000 2.325648e+000 2.304726e+000 2.121749e+000 1.704650e+000 2.238667e-003 2.932143e-001
2.874293e-001 2.931769e-001 3.860753e-001 4.484176e-001 4.342295e-001 4.584453e-001 4.661352e-001 4.134188e-001 4.097667e-001
3.754075e-001 3.134713e-001 2.875983e-004
<GCONST> 9.817532e+001
<TRANS> 5
0.000000e+000 1.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000
0.000000e+000 6.000000e-001 4.000000e-001 0.000000e+000 0.000000e+000
0.000000e+000 0.000000e+000 6.000000e-001 4.000000e-001 0.000000e+000
0.000000e+000 0.000000e+000 0.000000e+000 7.000000e-001 3.000000e-001
0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000
<ENDHMM>
```

Paslėptųjų *Markovo* modelių apmokymas naudojant HTK programinį įrankį „*HERest.bat*“ „*HERest -A -D -T 1 -C config6 -I phones0.mlf -t 250.0 150.0 1000.0 -S train.scf -H hmm0/macros -H hmm0/hmmdefs -M hmml monophones0*“, kuris, taikydamas *Baum–Welch* algoritimą, patikslina sukurto modelio reikšmes. „*HERest.bat*“ įrankis bus aktyvuojamas tris kartus sukurtų fonemų modelių esančių „*phones0.mlf*“ perskaičiavimui, atsižvelgiant į požymio vektorius.

Pirmajam sukurtų fonemų modelių perskaičiavimui sukuriamas naujas katalogas ir pavadinamas „*hmml*“. Tada aktyvuojamas programinis įrankis „*HERest.bat*“, kuris naudodamasis „*hmm0*“ kataloge esančiais pirminiais „*hmmdefs*“ ir „*macros*“ failais, „*hmml*“ kataloge sukuria naujus „*hmmdefs*“ ir „*macros*“ failus.

Antrajam sukurtų fonemų modelių perskaičiavimui sukuriamas naujas katalogas ir pavadinamas „*hmm2*“. Tada aktyvuojamas programinis įrankis „*HERest2.bat*“, kuris naudodamas „*hmm1*“ kataloge esančiais „*hmmdefs*“ ir „*macros*“ failais, „*hmm2*“ kataloge sukuria naujus „*hmmdefs*“ ir „*macros*“ failus.

Trečiajam sukurtų fonemų modelių perskaičiavimui sukuriamas naujas katalogas ir pavadinamas „*hmm3*“. Tada aktyvuojamas programinis įrankis „*HERest3.bat*“, kuris naudodamas „*hmm2*“ kataloge esančiais „*hmmdefs*“ ir „*macros*“ failais, „*hmm3*“ kataloge sukuria naujus „*hmmdefs*“ ir „*macros*“ failus.

**6 žingsnis** – pauzės fonemos taisymas. Šiame etape sukuriamas naujas katalogas ir pavadinamas „*hmm4*“. Į šį katalogą nukopijuojami failai „*hmmdefs*“ ir „*macros*“ iš „*hmm3*“ katalogo. Į kataloge „*hmm4*“ esantį failą „*hmmdefs*“ įtraukiamas trumpų pauzių „*sp*“ modelis, gautas iš tylos „*sil*“ modelio pašalinant antrą (2) ir ketvirtą (4) būsenas, paliekant tik trečią (3) būseną ir pakeičiant „*transp*“ pradinę 5x5 matricą į 3x3 matricą. Papildyto, trumpų pauzių fonemos modelio „*sp*“, failo „*hmmdefs*“, turinys atrodo taip:

---

```

~h "sp"
<BEGINHMM>
<NUMSTATES> 3
<STATE> 2
<MEAN> 39
-3.583859e+000 1.947640e+000 -9.582480e-001 4.062145e+000 3.073115e+000 2.215589e+000
2.804457e+000 8.581215e-001 9.912094e-001 1.087161e+000 1.122483e+000 3.181509e-001 1.837713e-001
-2.237545e-001 2.202104e-002 8.705210e-002 1.970315e-003 -6.545750e-002 -4.281981e-002 -
7.388708e-002 -1.083715e-001 -1.105275e-001 -8.702906e-002 -9.363876e-002 -6.907409e-002
1.433468e-002 1.184308e-001 -1.234493e-002 9.083836e-002 -1.118520e-001 -1.172332e-001 -
8.405770e-002 -1.007509e-001 -7.694263e-002 -7.850000e-002 -5.159383e-002 -3.231612e-002 -
9.536900e-003 1.126397e-002
<VARIANCE> 39
6.959997e+001 1.991402e+001 2.382462e+001 1.989569e+001 2.093261e+001 2.160735e+001
2.114713e+001 2.237295e+001 2.210080e+001 2.013294e+001 1.914513e+001 1.471322e+001 3.867243e-002
3.894617e+000 1.022980e+000 1.149515e+000 1.105702e+000 1.359647e+000 1.495765e+000 1.640249e+000
1.778079e+000 1.787622e+000 1.673895e+000 1.579106e+000 1.362201e+000 1.064076e-003 6.800699e-001
2.123577e-001 2.622249e-001 2.626384e-001 2.968138e-001 3.103075e-001 3.301210e-001 3.385828e-001
3.380930e-001 3.254366e-001 2.986770e-001 2.515035e-001 3.302395e-004
<GCONST> 8.225308e+001
<TRANSP> 3
0.0 1.0 0.0
0.0 0.9 0.1
0.0 0.0 0.0
<ENDHMM>

```

---

Toliau sukuriamas naujas katalogas ir pavadinamas „*hmm5*“. Tada aktyvuojamas „*HHEd.bat*“ failas, kuris prideda reikalingų papildomų perėjimų perkeltiant trumpų pauzių „*sp*“ modelį į tylos modelio „*sil*“ centrą. Programa „*HHEd.bat*“ „*HHEd -A -D -T 1 -H hmm4/macros -H hmm4/hmmdefs -M hmm5 sil.hed monophones1*“ veikia panašiai kaip „*HLEd.bat*“, naudodama paruoštus duomenis perskaičiuoja fonemų paslėptuosius Markovo modelius. „*HHEd.bat*“ programa naudoja failus „*hmmdefs*“ ir „*macros*“ iš katalogo „*hmm4*“ ir dar vieną failą „*sil.hed*“, kuriame esanti informacija yra nurodyta HTK pakete. Failo „*sil.hed*“ turinys atrodo taip:



---

```
AT 2 4 0.2 {sil.transP}
AT 4 2 0.2 {sil.transP}
AT 1 3 0.3 {sp.transP}
TI silst {sil.state[3],sp.state[2]}
```

---

Aktyvavus „*HHEd.bat*“ programą, kataloge „*hmm5*“ sukuriami „*hmmdefs*“ ir „*macros*“ failai.

Toliau tęsiamas paslėptųjų Markovo modelių apmokymas naudojant HTK programinį įrankį „*HERest.bat*“. Šiame etape „*HERest.bat*“ įrankis bus aktyvuojamas du kartus faile „*hmmdefs*“ sukurtų fonemų modelių, papildytų trumpų pauzių modeliu „*sp*“ perskaičiavimui, atsižvelgiant į požymio vektorius.

Pirmajam sukurtų fonemų modelių, papildytų trumpų pauzių modeliu „*sp*“, perskaičiavimui sukuriamas naujas katalogas ir pavadinamas „*hmm6*“. Tada aktyvuojamas programinis įrankis „*HERest4.bat*“, kuris naudodamas „*hmm5*“ kataloge esančius „*hmmdefs*“ ir „*macros*“ failus, „*hmm6*“ kataloge sukuria naujus „*hmmdefs*“ ir „*macros*“ failus.

Antrajam sukurtų fonemų modelių, papildytų trumpų pauzių modeliu „*sp*“, perskaičiavimui sukuriamas naujas katalogas ir pavadinamas „*hmm7*“. Tada aktyvuojamas programinis įrankis „*HERest5.bat*“, kuris naudodamas „*hmm6*“ kataloge esančius „*hmmdefs*“ ir „*macros*“ failus, „*hmm7*“ kataloge sukuria naujus „*hmmdefs*“ ir „*macros*“ failus.

**7 žingsnis** – fonemų modelių mokymas. Fonemų modelių apmokymui naudojama programa „*HVite.bat*“, kurios komandinė eilutė „*HVite -A -D -T 1 -l features -o SWT -b SENT-END -C config6 -a -H hmm7/macros -H hmm7/hmmdefs -i aligned.mlf -m -t 250.0 150.0 1000.0 -y lab -a -I words.mlf -S train.scf dict monophones1*“. Pagal jau sukurtus fonemų modelius sukuria naujas komandų transkripcijas fonemų modelių apmokymui balso komandoms atpažinti. „*HVite.bat*“ programa naudoja fonemų paslėptuosius Markovo modelius failuose „*hmmdefs*“ ir „*macros*“, kurie yra kataloge „*hmm7*“, perskaičiuojant „*words.mlf*“ faile surašytą pilną visų komandų transkripcijų rinkinį, naudojant „*dict.txt*“ faile esančius komandų foneminius modelius, į naujas transkripcijas, kurioms programa sukuria failą „*aligned.mlf*“. Failo „*aligned.mlf*“ turinys yra lygiai toks pat kaip failo „*phones1.mlf*“.

Toliau tęsiamas paslėptųjų Markovo modelių apmokymas naudojant HTK programinį įrankį „*HERest.bat*“. Šiame etape „*HERest.exe*“ įrankis bus aktyvuojamas du kartus sukurtų fonemų modelių esančių „*aligned.mlf*“ perskaičiavimui, atsižvelgiant į požymio vektorius.

Šiame žingsnyje, naujai sumodeliuotų fonemų modelių perskaičiavimui sukuriamas naujas katalogas ir pavadinamas „*hmm8*“. Tada aktyvuojamas programinis įrankis „*HERest6.bat*“, kuris

naudodamas „*hmm7*“ kataloge esančius „*hmmdefs*“ ir „*macros*“ failus, „*hmm8*“ kataloge sukuria naujus „*hmmdefs*“ ir „*macros*“ failus.

Pakartotiniam sumodeliuotų fonemų modelių perskaičiavimui sukuriamas naujas katalogas ir pavadinamas „*hmm9*“. Tada aktyvuojamas programinis įrankis „*HERest7.bat*“, kuris naudodamasis „*hmm8*“ kataloge esančiais „*hmmdefs*“ ir „*macros*“ failais „*hmm9*“ kataloge sukuria naujus „*hmmdefs*“ ir „*macros*“ failus. Kataloge „*hmm9*“ esantys fonemų paslėptieji Markovo modeliai failuose „*hmmdefs*“ ir „*macros*“ yra naudojami balso komandų atpažinimo testavimui pagal fonemų modelius.

**8 žingsnis** – atpažinimo testavimas pagal fonemų modelius. Šiame etape pirmiausia sukuriamas naujas katalogas ir pavadinamas „*Testavimas*“. Kataloge „*hmm9*“ esantys apmokyti fonemų paslėptieji Markovo modeliai failuose „*hmmdefs*“ ir „*macros*“ yra naudojami balso komandų atpažinimo testavimui pagal fonemų modelius, todėl failai „*hmmdefs*“ ir „*macros*“ yra nukopijuojami į naujai sukurta „*Testavimas*“ katalogą. Atpažinimo testavimo etape, apmokytų fonemų modelių atpažinimo tikslumo testavimui naudojami kiti garsyno duomenys. Katalogas „*data*“ su naujais testavimui skirtais duomenis įkeliamas į katalogą „*Testavimas*“. Naujų duomenų kodavimui reikalingi tekstiniai „*config*“ ir „*codetr.scp*“ failai, kur surašomi programos „*HCopy.bat*“ aktyvavimui reikalingi parametrai ir informacija. Parametrai „*config*“ faile atrodo taip:

---

```
SOURCEFORMAT=WAVE
TARGETKIND=MFCC_E_D_A_Z
TARGETRATE=100000
SAVECOMPRESSED=T
SAVEWITHCRC=T
WINDOWSIZE=250000
USEHAMMING=T
PREEMCOEF=0.97
NUMCHANS=26
CEPLIFTER=22
NUMCEPS=12
FORCECXP=EXP=T
ALLOWXWRDEXP=F
```

---

Testavimui naudojamas „*codetr.scp*“ failas, kuriame nurodomos testavimui naudojamų duomenų direktorijos.

Paruošus „*config.txt*“ ir „*codetr.scp*“ failus aktyvuojamas „*HCopy.bat*“ duomenų kodavimo programinis įrankis, kuris sukuria testavimui naudojamų požymių failus ir patalpina juos naujai sukurtame kataloge „*features*“.

Fonemų modelių testavimui naudojama „*HVite.bat*“, kurios komandinė eilutė „*HVite -H hmm9/macros -H hmm9/hmmdefs -S test.scp -l features -i recout.mlf -w wdnnet -p 0.0 -s 5.0 dict monophones1*“ programa. „*HVite.bat*“ programa naudoja fonemomis grįstus paslėptuosius

Markovo modelius kataloge „*hmm9*“ esančiuose failuose „*hmmdefs*“ ir „*macros*“. Paleidus programą sukuriamos naujos transkripcijos, kurios perkeliamos į naują failą „*recout.mlf*“.

Testavimo etape „*HVite.bat*“ programa naudoja papildomus failus. Failas „*test.scf*“ kuriamas taip pat, kaip „*train.scf*“ failas, failas „*testref.mlf*“ kuriamas, kaip „*words.mlf*“ failas. Taip pat reikalingi failai yra „*wdnet*“, „*dict*“ ir „*monophones1*“, kurie yra jau paruošti pirmuose modelių apmokymo paruošimo etapuose ir nukopijuojami į katalogą „*Testavimas*“.

Paslėptųjų Markovo modelių, gristomis fonemomis, testavimo rezultatams gauti naudojamas HTK programinis įrankis „*HResults.bat*“ „*HResults -p -I testref.mlf dict recout.mlf >>rez*“. Jį aktyvavus, remiantis failuose „*testref.mlf*“, „*dict*“, „*recout.mlf*“ paruošta informacija, sukuriama atpažinimo tikslumo matrica pateikta faile „*rez*“ [15]. Pavyzdį galima pamatyti 3.1.1 paveiksle.

### 2.3 Garsyno duomenų anotavimas

Apie kalbos vienetų išskaidymą minimi terminai anotavimas (angl. *annotation*), „markiravimas“ (angl. *mark-up*), „transkribavimas“ (angl. *transcription*). Jie visi, vienaip ar kitaip, susiję su šnekos signalo arba kitų pirminių kalbos duomenų papildymu meta-duomenimis.

S. Bird ir M. Lieberman siūlo anotavimą, „markiravimą“ ir „transkribavimą“ suprasti, kaip tą patį procesą, nes iš tiesų, kalbant apie lingvistinius duomenis, procesas yra tas pats – suteikiama papildomos, simboliais išreikštos, reikšmės tam tikrai lingvistinei duomenų porcijai.

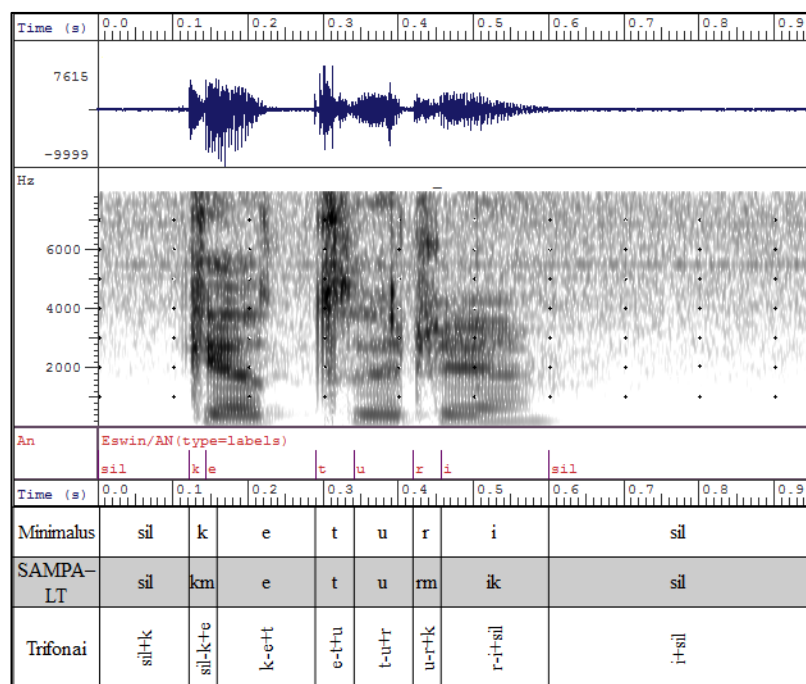
Šnekos signalo anotavimas yra svarbus, nes anotacijos duoda papildomą informaciją, be kurios šnekos atpažinimo ar sintezės algoritmų kūrimas būtų labai komplikotas arba netikslus [11].

Šiuo metu prikurta labai daug anotavimo modelių. Tačiau rankiniam anotavimui bus naudojama „*Speech Filing System*“ (*SFS*) programa, kuri yra šnekos signalų apdorojimo ir analizės sistema, galinti savo šnekos duomenų failus formatuoti ir apdoroti. Svarbiausia tai, kad ji yra suderinta su HTK paketu, o patogia sąsaja galima analizuoti, įrašyti tirti ir apdoroti garso failus. Tai padeda spręsti šnekos atpažinimo uždavinius paslėptųjų Markovo modelių teorijos metodais [16].

SFS naudoja duomenų modelius, leidžiančius viename faile aprašyti daugiau failo lygmenų ir juos susieti į hierarchines struktūras. Viename faile saugoma informacija apie visus šnekos signalo anotavimo lygmenis. Be ortografinės informacijos saugoma ir kita lingvistinė informacija apie sintaksę bei nurodomas kiekvienos fonemos pradžios ir pabaigos laikas.

Rankinio segmentavimo metu SFS programoje atidaromas pasirinktas garso failas ( žr. 2.3.1 pav. „*FAGNGRA400.wav*“). Programos lange matoma oscilograma, kur nustatomos tylos

ribos „sil“ ir iš kalbos įrašo virpesių pokyčių galima pastebėti, kur prasideda ir kur baigiasi kalbos garso virpesiai. Pagal spektrogramą ir perklausant signalą priskiriamos kiekvienos fonemos ribos.



2.3.1 pav. „SFS“ programos anotavimo langas

Atlikus anotavimą suformuojame failą, tinkantį tolesniems tyrimams su HTK paketu. Šiuo atveju sukuriamas „FAGNGRA400.lab“ Taip atrodo failo viduje atrodo sukurtos anotavimo ribos susiejančios kalbos duomenų elementus su atitinkamais fonemų pradžios ir pabaigos laiko momentais:

---

```

0 979100 sil
979100 1245100 k
1245100 2788200 e
2788200 3139400 t
3139400 3852400 u
3852400 4315300 r
4315300 5863700 i
5863700 9843700 sil

```

---

Anotacijos pirmu atveju sukurtos su minimaliu fonemų rinkiniu. Vėlesniuose tyrimuose visų fonemų transkripcijos perrašytos, kad tiktų tyrimams su SAMPA-LT fonemų rinkiniu. Garsyno anotavimas atliktas 6 skaičių komandų diktoriais. Viso anotavimo metu sukurtos 1200 komandų anotavimo žymės.

## 2.4 Anotuoto garsyno tyrimas

Turint anotuotą garsyną galima tiksliau atlikti tyrimą, nes turimos kiekvienos fonemos žymės kiekvienoje komandoje. Dėl to į anksčiau aptarta tyrimo seką įterpiamas papildomas žingsnis.

**5 žingsnis** – anotuotų foneminių modelių kūrimas. Reikalingi failai „*train.scp*“ ir „*config6*“. Visi turimo anotuoto garsyno failai perkeliama į „*labels*“ direktoriją. Taip pat sukuriama papildomas katalogas „*model*“, kuriame aprašomos visos tyrime naudojamos fonemos. Fonemos „*a*“ modelis atrodo taip:

---

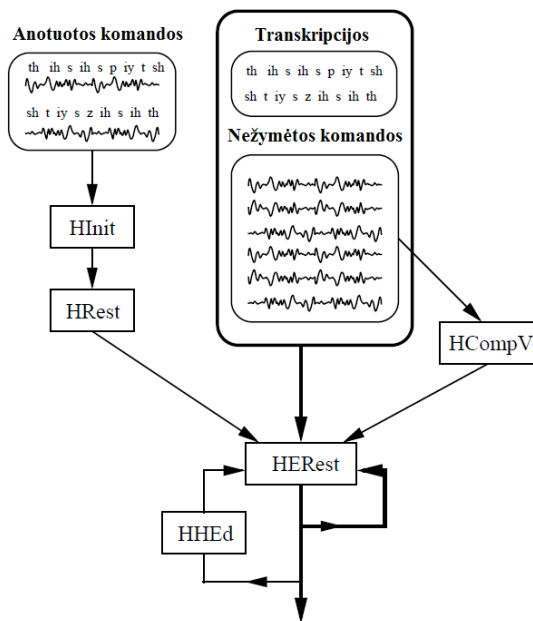
```
~o <VecSize> 39 <MFCC_E_D_A_Z>
~h "hmm_a"
<BeginHMM>
<NumStates> 5
<State> 2
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State> 3
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State> 4
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<Transp> 5
0.0 1.0 0.0 0.0 0.0
0.0 0.6 0.4 0.0 0.0
0.0 0.0 0.6 0.4 0.0
0.0 0.0 0.0 0.7 0.3
0.0 0.0 0.0 0.0 0.0
<EndHMM>
```

---

Programa „*HInit.bat*“, kurios komandinė eilutė „*HInit.exe -A -D -T 1 -C config6 -S train.scp -H model/hmm\_a -M hmm0 -l a -L labels hmm0/hmm\_a*“ perskaičiuoja pradines parametrų vertes, perskaito visus apmokymui skirtus duomenis ir iškerpa tik reikalingų fonemų intarpus. Pirmajame iteracijų cikle apmokymo duomenys sujungiami su segmentuoto modelio būsenos grandine.

Programa „*HRest.bat*“, kurios komandinė eilutė „*HRest.exe -A -D -T 1 -C config6 -S train.scp -H model/hmm\_a -M hmm0 -l a -L labels hmm0/hmm\_a*“ perskaičiuoja kiekvienos

fonemos požymio vektorius, kurie sudaryti iš požymių vidurkių, požymių dispersijos ir tikimybių matricos. Šie duomenys pakeičia pirminiame modelyje esančias nulines reikšmes ir kataloge „*hmm0*“ sukuria naują failą „*hmm\_a*“ (naujuose failuose atsiranda duomenys apie vektoriaus koeficientus). Tolesniam paslėptųjų *Markovo* modelių mokymui naudojamas HTK įrankis „*HERest.bat*“



2.4.1 pav. Anotuoto garsyno mokymas PMM [15]

Žemiau pateiktoje 2.4.1 lentelėje nurodomas HTK paketo naudojimas turint kalbos komandų anotuotus failus. Modeliavimo metodika nelabai kuo skiriasi nuo anksčiau aptartos. Tačiau šiuo atveju prisideda 5 žingsnis, kuriame tiksliai perskaičiuojami kiekvienos fonemos požymio vektoriai. Taip pat 7 žingsnyje parengiant tylos fonemą reikia nukopijuoti „*hmmdfes*“ faile esančią fonemą „*sil*“ ir iš jos sukurti pauzės fonemą „*sp*“. Visi kiti žingsniai lieka tokie patys.

2.4.1 lentelė. Fonemomis grįsto modelio metodo taikymas [15]

		Pavadinimas	Programinis įrankis	Programos aktyvavimo „*.bat“ failai	Reikalingi failai	
Žingsniai	1	Gramatikos failo kūrimas	<i>Hparse.exe</i>	HParse gram wdnnet	<i>gram.dict</i>	Mokymas

2	Fonemų modelių sudarymui reikalingų failų paruošimas, anotuotų failų paruošimas			<i>dict, words.mlf, monophones0 labs, labels, models</i>
3	Požymių kūrimas	<b>HLEd.exe</b>	HLEd -l features -d dict -i phones0.mlf mkphones0.led words.mlf; HLEd -l features -d dict -i phones1.mlf mkphones1.led words.mlf	<i>words.mlf, mkphones0.led,</i>
4	Duomenų kodavimas	<b>HCopy.exe</b>	HCopy -A -D -T 1 -C CONFIG -S codetr.scp	<i>CONFIG, codetr.scp</i>
5	Anotacijų priskyrimas fonemoms	<b>HInit.exe, HRest.exe</b>	HInit -A -D -T 1 -C config6 -S train.scp -H model/hmm_v -M hmm0 -l v -L labels hmm0/hmm_v; HRest -A -D -T 1 -C config6 -S train.scp -H model/hmm_v -M hmm0 -l v -L labels hmm0/hmm_v	<i>proto, config6, train.scp, phones0.mlf, macros, hmmdefs, monophones0 labels hmm0/hmm_v</i>
6	PMM modelių parengimas	<b>HCompv.exe, HRest.exe</b>	HCompV -A -D -T 1 -C config6 -f 0.01 -m -S train.scp -M hmm0 proto; 3x HRest -A -D -T 1 -C config6 -I phones0.mlf -t 250.0 150.0 1000.0 -S train.scp -H hmm0/macros -H hmm0/hmmdefs -M hmm1 monophones0	<i>proto, config6, train.scp, phones0.mlf, macros, hmmdefs, monophones0</i>
7	Tylos fonemos parengimas	<b>HHEd.exe; HRest.exe</b>	HHEd -A -D -T 1 -H hmm4/macros -H hmm4/hmmdefs -M hmm5 sil.hed monophones1; 2x HRest -A -D -T 1 -C config6 -I phones1.mlf -t 250.0 150.0 1000.0 -S train.scp -H hmm5/macros -H hmm5/hmmdefs -M hmm6 monophones1	<i>macros, hmmdefs, config6, train.scp, phones0.mlf, monophones1</i>
8	Modelių apmokymas	<b>HVite.exe, HRest.exe</b>	HVite -A -D -T 1 -l features -o SWT -b SENT-END -C config6 -a -H hmm7/macros -H hmm7/hmmdefs -i aligned.mlf -m -t 250.0 150.0 1000.0 -y lab -a -I words.mlf -S train.scp dict monophones1; 2x HRest -A -D -T 1 -C config6 -I aligned.mlf -t 250.0 150.0 3000.0 -S train.scp -H hmm7/macros -H hmm7/hmmdefs -M hmm8 monophones1	<i>config6, macros, hmmdefs, words.mlf, train.scp, dict, monophones1, aligned.mlf</i>
9	Modelių testavimas	<b>HCopy.exe, HVite.exe, HResults.exe</b>	HCopy -A -D -T 1 -C CONFIG -S codetr.scp; HVite -H hmm9/macros -H hmm9/hmmdefs -S test.scp -l features -i recout.mlf -w wdnnet -p 0.0 -s 5.0 dict monophones1; HResults -p -I testref.mlf dict recout.mlf>>rez	<i>CONFIG, codetr.scp, macros, hmmdefs, test.scp, recout.mlf, wdnnet, dict, monophones0, testref.mlf</i>
				<b>Testavimas</b>

## 2.5 Kontekstinių fonemų (trifonų) tyrimas

Trifonų sąrašas sudaromas iš elementarių fonemų. Kad sugeneruotume trifoną reikia turėti paprastą į atskirus elementus išskaidytą foneminį kalbos signalą. Trifonui (t.y. trijų atskirų fonemų grupei) sukurti imame kairįjį elementarųjį fonemos fragmentą (pvz. *K*), kuris eina prieš esamą fonemą (pvz. *F*), o po to seka dešinysis elementarus fonemos fragmentas (pvz. *D*). Trifonas šiuo atveju sudaromas – iš kairiojo elemento atimant fonemą, kuri seka po jo bei pridėdant vėliau einantį fonemos fragmentą (t.y. *K-F+D*). Tokios formos trijų fonemų grupė galime pavadinti trifonu. Komandą „Keturi“ galima išskirti fonemomis „*k e t u r i*“. Trifonais komandą „*keturi*“ galima pavaizduoti „*k+e k-e+t e-t+u t-u+r u-r+i r-i*“. Reikia atkreipti dėmesį, kad komandos pradžioje ir pabaigoje yra ne trifonas ir ne fonema, o taip vadinamoji „*dviguba fonema*“ [12].

Tokiu trifonu akustiniame modelyje ieškoma panašios sekos t.y. atskirų fonemų prieš ir po reikiamo fragmento. Taip pagerinamas tikslumas, nes ieškoma ne vieno garso, o trijų atskirų garsų sekos, trifono. Trifonų tyrimas taip sumažina klaidos galimybes, nes ieškoma visos eilės garsų, o ne vieno fragmento. Taip pagerinamas ir viso žodžio išskyrimas, nes žodžio pradžioje ir pabaigoje išskiriamos dvi gretimos fonemos [17].

Kontekstinėmis fonemomis paremtas tyrimai atliekamas tik po anksčiau aptarto fonemomis grįsto modelio taikymo, kai jau turimi apmokytų fonemų požymiai kataloge. Pirmajame žingsnyje fonemos perskaičiuojamos į trifonų transkripcijas. Vėliau sukuriama dvigubos fonemos, kurios apibrėžia komandos pradžios ir pabaigos ribas. Toliau perskaičiuojami požymių vektoriai ir dispersijos tarp fonemų. Kitame žingsnyje sugrupuojamos trifonų grupės su jų perėjimo matricos. Dar kartą yra perskaičiuojami požymiai kurie naudojami patikrinimui.

**2.5.1 lentelė.** Trifonais grįsto modelio taikymas [15]

		Pavadinimas	Programinis įrankis	Programos aktyvavimo „*.bat“ failai	Reikalingi failai	
Žingsniai	1	Požymių kūrimas	<i>HLEd.exe</i>	HLEd -A -D -T 1 -n triphones1 -l features -i wintri.mlf mktri.led aligned.mlf	<i>features, hmm9, mktri.led, aligned.mlf</i>	Mokymas
	2	Trifonų gramatikos kūrimas	„Perl.exe“	perl -w maketrihed monophones1 triphones1	monophones1, triphones1,maketrihed	
	3	Požymių perskaičiavimas	<i>HHEd.exe, HERest.exe</i>	HHEd -A -D -T 1 -H hmm9/macros -H hmm9/hmmdefs -M hmm10 mktri.hed monophones1>>HHed_log; HERest -A -D -T 1 -C config -I wintri.mlf -t 250.0 150.0 3000.0 -S train.scf -H hmm10/macros -H hmm10/hmmdefs -M hmm11 triphones1; HERest -A -D -T 1 -C config -I wintri.mlf -t 250.0 150.0 1000.0 -s stats -S train.scf -H hmm11/macros -H hmm11/hmmdefs -M hmm12 triphones1	<i>hmm9, monophones1, triphones1, train.scf, config</i>	
	4	Surištų būsenų trifonai	<i>HDMAN.exe Perl.exe HHEd.exe HERest.exe</i>	HDMAN -A -D -T 1 -b sp -n fulllist -g global.ded -l flog dict-tri dict; perl -w mkclscript.prl TB 350 monmophones0>>tree.hed; HHEd -A -D -T 1 -H hmm12/macros -H hmm12/hmmdefs -M hmm13 tree.hed triphones1; HHEd -A -D -T 1 -H hmm12/macros -H hmm12/hmmdefs -M hmm13 tree.hed triphones1; 2x HERest -A -D -T 1 -C config -I wintri.mlf -s stats -t 250.0 150.0 3000.0 -S train.scf -H hmm13/macros -H hmm13/hmmdefs -M hmm14 tiedlist	Fulllist, global.ded, mkclscript.prl, dict, monmophones0, triphones1, config, wintri.mlf, train.scf, tiedlist	
	5	Modelių testavimas	<i>HCopy.exe, HVite.exe HResults.exe</i>	HCopy -A -D -T 1 -C CONFIG -S codetr.scf; HVite -H hmm15/macros -H hmm15/hmmdefs -S test.scf -l features -i recout.mlf -w wdnet -p 0.0 -s 5.0 dict tiedlist HResults -p -I testref.mlf dict recout.mlf>>rez	<i>CONFIG, codetr.scf, macros, hmmdefs, test.scf, recout.mlf, wdnet, dict, monophones0, testref.mlf</i>	



### 3 TYRIMO REZULTATAI

Vardų ir skaičių komandų atpažinimo tyrimo metu naudojami diktorių įrašai yra išskaidomi į apmokymui ir testavimui skirtas komandas. Mokymo procese naudojami 18-likos diktorių įrašai tariantys vardus ir 24-rių diktorių įrašai tariantys skaičius. Likęs garsynas su 3-jų vardų komandų diktoriais ir 6-ių skaičių komandų diktoriais naudojamas testavimo žingsnyje.

#### 3.1 Atpažinimas naudojant minimalų fonemų rinkinį

Vardų ir skaičių garsyno atpažinimo tyrimo rezultatai pateikiami trimis etapais. Pirmajame etape tikslumo tyrimas atliktas naudojantis minimaliu fonemų rinkiniu. Minimalų fonemų rinkinį sudaro komandų išskaidymas į minimaliausius elementus raides ir dvigarsius. Minimalų fonemų rinkinį sudaro 36 fonemos. Toks rinkinys atrodo taip:

---

<b>AUSTEJA</b> au stEja	sp	<b>NOJUS</b> nojus	sp	<b>ASHTUONI</b> a sh tu oni	sp
<b>BOLESLOVAS</b> boleslovas	sp	<b>OSKARAS</b> oskaras	sp	<b>DEVYNI</b> devyni	sp
<b>CECILJA</b> cecilija	sp	<b>PATRIKAS</b> patrikas	sp	<b>DU</b> du	sp
<b>DONATAS</b> donatas	sp	<b>KJU</b> k j U	sp	<b>KETURI</b> keturi	sp
<b>EIMANTAS</b> eimantas	sp	<b>RICARDAS</b> riCar das	sp	<b>NULIS</b> nulis	sp
<b>FAUSTA</b> fausta	sp	<b>SANDRA</b> sandra	sp	<b>PENKI</b> penki	sp
<b>GRAZVYDAS</b> graZvy das	sp	<b>TEODORAS</b> teodoras	sp	<b>SEPTYNI</b> septyni	sp
<b>HANSAS</b> hansas	sp	<b>ULIJONA</b> ulijona	sp	<b>SHESHI</b> sheshi	sp
<b>IZAOKAS</b> izaokas	sp	<b>VACYS</b> vacys	sp	<b>TRYS</b> trys	sp
<b>JONAS</b> jonas	sp	<b>WASHINGTON</b> wa shing ton	sp	<b>VIENAS</b> vienas	sp
<b>KAROLIS</b> karolis	sp	<b>IKSAS</b> iksas	sp	<b>SENT-END</b> []	sil
<b>LAIMA</b> laima	sp	<b>YGREKAS</b> ygrekas	sp	<b>SENT-START</b> []	sil
<b>MARTYNAS</b> martynas	sp	<b>ZACHARIJUS</b> za x a ri jus	sp		

---

Su šiuo fonemu rinkiniu atliekamas tikslumo nustatymo tyrimas, kuris aprašytas metodinės dalies 2.2 skyriuje. Visi tikslumo rezultatai yra gaunama tikslumo pasiskirstymo matricos pavidalu, kurį galima matyti 3.1.1 paveiksle.



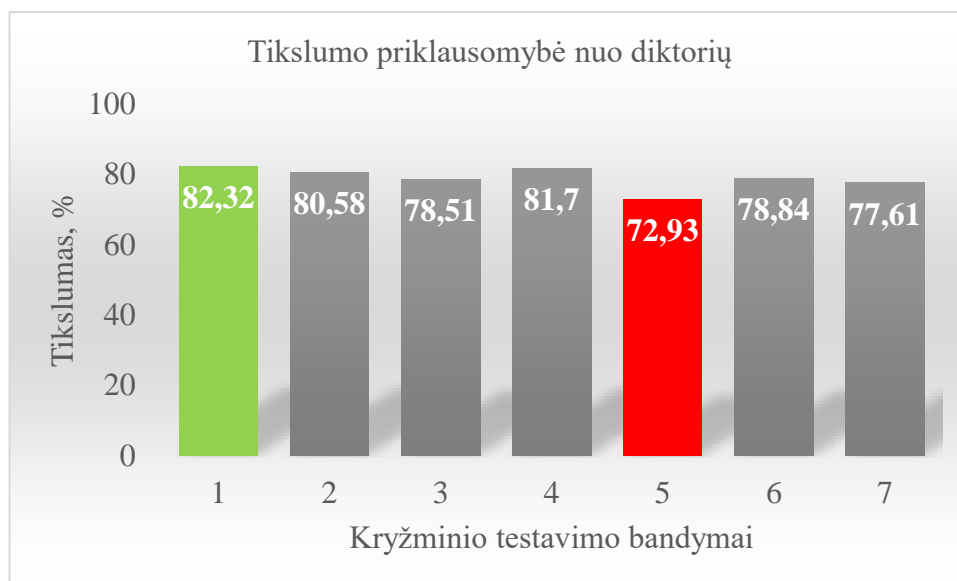
v	Zm	uk
cm	vm	C
e	y	rr
ik	h	shm
i	akm	ng
d	z	gm
n	ok	x
ak	k	ie
t	oo	pm
ei	aki	uo
m	rrm	nk
nn	ii	sh
f	u	nm
aku	p	dm
g	rm	sil
r	km	

Norint patikrinti SAMPA–LT fonemų rinkinio įtaką, atliekamas kryžminis patikrinimas. Kryžminiu patikrinimu galime laikyti testavimo tikslumo priklausomybę nuo mokomų diktorių. Mokymo metu naudojami 18 diktorių, kurie taria vardus ir 24 diktoriai, kurie taria skaičius. Testavimui paliekami 3 vardus ir 6 skaičius tariantys diktoriai. Testavimas su skirtingais diktoriais atliekamas 7 kartus, testavimui paimant vis skirtingus prieš tai testavime nenaudotus vardų ir skaičių diktorius.

Visi, naudojant SAMPA–LT fonemų rinkinį, gauti kryžminio testavimo rezultatai pateikiami 3.2.1 lentelėje. Matome, kad lyginant su minimaliu fonemų rinkiniu, kurio tikslumas 68,88 proc. yra pagerėjęs, tačiau nėra puikus. Geriausias tikslumas gautas panaudojus „0,1,2 ir FAGNGRA, FAGNVIN, FAISIZI, FAISZYM, FAUSNEM, FDAILOI“ diktorius testavimo metu, kada gautas 82,32 proc. tikslumas. Blogiausias 72,93 proc. tikslumas gautas testavimo metu naudojant „11,12,13, MDAIGUS, MEDGVOL, MKAZANU, MMODSLE, MRIMAPA, MVYGVAI“ diktorius.

**3.2.1 lentelė.** Kryžminio testavimo metu naudoti diktoriai ir tikslumo priklausomybė nuo diktorių

Bandymo numeris	1	2	3	4	5	6	7
Vardų diktorių numeris	0,1,2	3,4,5	6,7,10	8, 9,16	11,12,13	14,15,17	18,19,20
Skaičių diktorių numeris	FAGNGRA FAGNVIN FAISIZI FAISZYM FAUSNEM FDAILOI	FGINGED FIEVJUR FIEVSAB FKAMMOS FLAUZET FRAISAV	FRUTNAN FSIMMEI FVAIVAI FVANPEC FVILVAI FVIONAB	FIEVVIS FJUSKIN FUGNBUC FUGNNOV FZIVSTA MLINJUR	MDAIGUS MEDGVOL MKAZANU MMODSLE MRIMAPA MVYGVAI	FAGNGRA FGINGED FRUTNAN FAISIZI FIEVVIS FLAUZET	FAISZYM FIEVSAB FVAIVAI MKAZANU FUGNBUC FRAISAV
Tikslumas	82,32%	80,58%	78,51%	81,70%	72,93%	78,84%	77,61%



**3.2.1 pav.** Tikslumo priklausomybė nuo diktorių naudojant SAMPA–LT fonemų rinkinį

Matoma, kad kryžminio testavimo ( žr. 3.2.1 pav.) metu tiksliausiai komandos atpažintos pirmajame bandyme priskirtais testavimo diktoriais, o blogiausiai atpažintos komandos su penktame bandyme naudotais diktoriais.

Sujungus visus kryžminio testavimo rezultatus gaunamas bendras visų vardų ir skaičių atpažinimo tikslumas, kuris pateikiamas 3.2.3 lentelėje.

Matoma, kad šių komandų: „*Austėja* (88,8%), *Cecilija* (50,24%), *Fausta* (34,52%), *Izaokas* (89,75%), *Laima* (60,25%), *Oskaras* (13,08%), *Kju* (72,85%), *Iksas* (73,55%), *Ygrekas* (65,72%), *Du* (63,8%), *Keturi* (65,55%), *Nulis* (45,11%), *Penki* (42,48%), *Šeši* (35,58%), *Trys* (78,45%)“ tikslumas vis dar nėra pakankamas.

Išvardintos komandos buvo klaidinamos 3.2.2 lentelėje nurodytų komandų.

**3.2.2 lentelė.** Atpažinimo klaidos

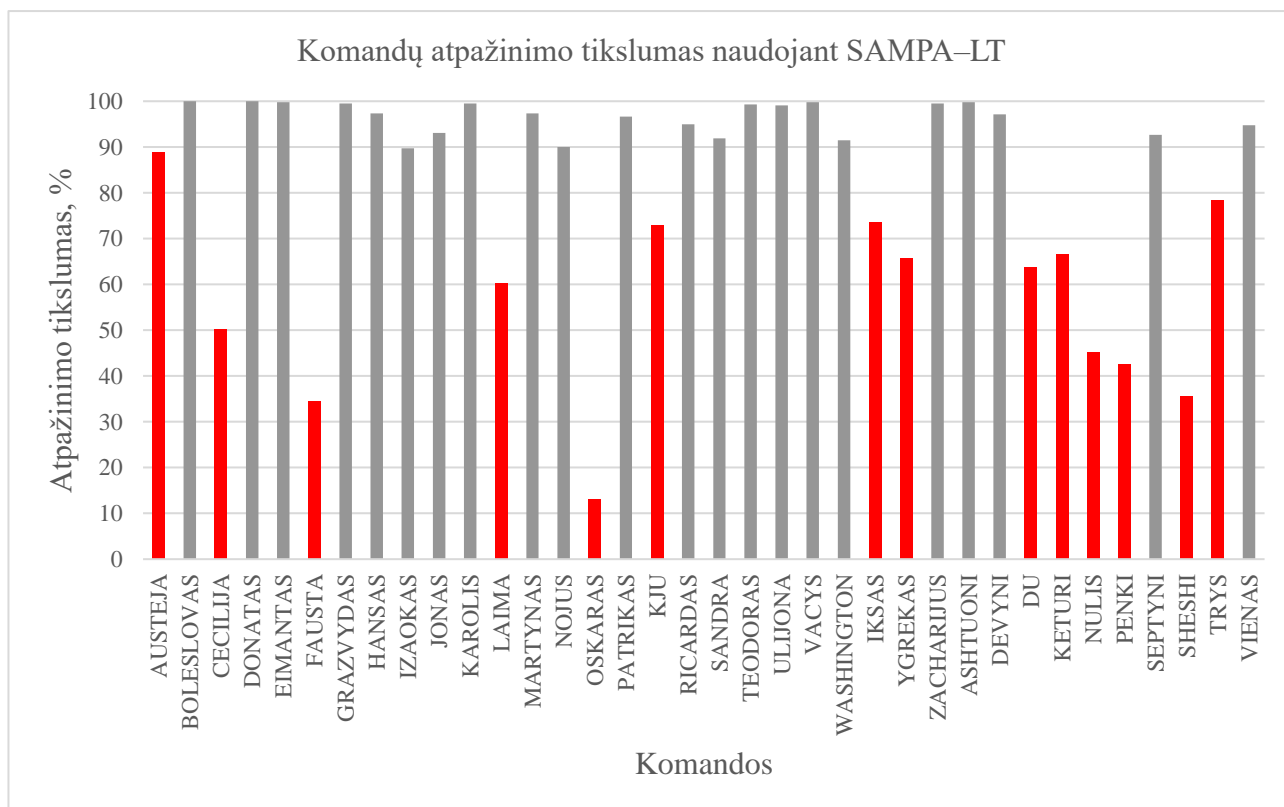
Komandos	Klaidingai atpažino
Austėja	Aštuoni
Cecilija	Devyni, Septyni
Fausta	Izaokas, Aštuoni
Laima	Donatas, Karolis, Aštuoni
Oskaras	Izaokas
Kju	Septyni
Iksas	Ygrekas
Ygrekas	Teodoras

Du	Septyni, Aštuoni
Keturi	Septyni, Aštuoni
Nulis	Karolis, Martynas, Teodoras
Penki	Septyni
Šeši	Keturi, Penki
Trys	Vacys, Martynas, Septyni, Laima

Atpažinimo metu daromos tokios klaidos, kad vardų komandų atpažinimą klaidina skaičių komandos, o skaičiai dažniausiai klysta savo komandas maišant su komanda „*Septyni*“.

3.2.3 lentelė. SAMPA–LT fonemų rinkinio bendras vardų ir skaičių atpažinimo tikslumas

Komandos	Tikslumas, %	Komandos	Tikslumas, %	Komandos	Tikslumas, %
Austėja	88.8	Martynas	97.35	Ygrekas	65.72
Boleslovas	100	Nojus	90	Zacharijus	99.51
Cecilija	50.24	Oskaras	13.08	Aštuoni	99.75
Donatas	100	Patrikas	96.67	Devyni	97.14
Eimantas	99.75	Kju	72.85	Du	63.8
Fausta	34.52	Ričardas	95	Keturi	66.55
Gražvydas	99.52	Sandra	91.9	Nulis	45.11
Hansas	97.37	Teodoras	99.28	Penki	42.48
Izaokas	89.75	Ulijona	99.05	Septyni	92.62
Jonas	93.1	Vacys	99.75	Šeši	35.58
Karolis	99.52	Washington	91.44	Trys	78.45
Laima	60.25	Iksas	73.55	Vienas	94.75



3.2.2 pav. Komandų atpažinimas naudojant SAMPA–LT fonemų rinkinį

Apibendrinant visus bendrus kryžminio testavimo metu gautus tikslumo rezultatus matoma, kad net 14 komandų yra atpažįstamos labai klaidingai. Bendras visų komandų tikslumas siekia 78,92 proc.. Tokio netikslaus atpažinimo priežastis yra per maža fonemų įvairovė. Mokymo metu sudaromi požymių vektoriai yra klaidingi ir netinkami visos fonemoms pasikartojančiose skirtingose komandose, todėl testavimo metu skaičiuojant komandos didžiausią atitikimo tikimybę gaunama klaidingai atpažinta komanda.

### 3.3 Išplėsto fonemų rinkinio atpažinimo tikslumas

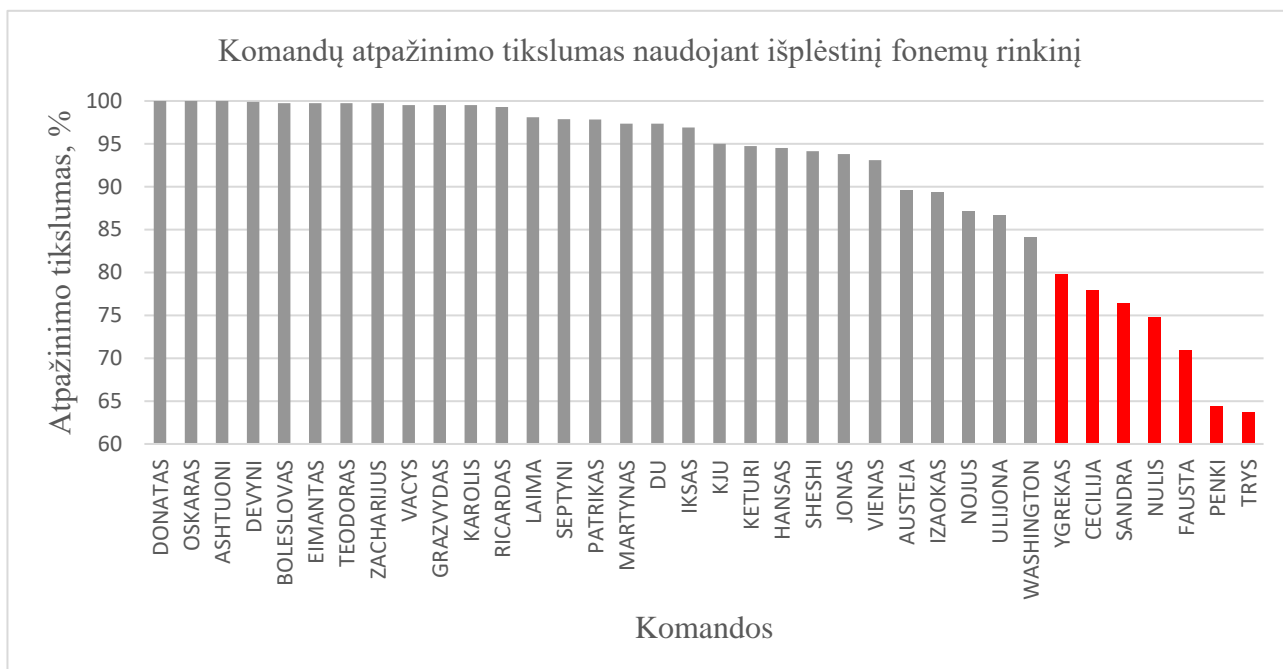
Atlikus tyrimus su SAMPA–LT fonemų rinkiniu matoma, kad 14 komandų atpažinimo tikslumas yra nedidelis. Vis dėl to ne visos komandos yra atpažįstamos klaidingai, todėl fonemų rinkinys gerinamas tik kai kurioms komandoms, kurios aprašytos 3.3.1 lentelėje. Šiuo atveju fonemų skaičius padidėja iki 79.

3.3.1 lentelė. Išplėstinio fonemų rinkinio komandų papildymas

Komanda	SAMPA–LT	Papildyta
Cecilija	cm e cm ik lm i j a sp	cn em cmi ik lm ik j a sp
Fausta	f aku s t a sp	f aku s t af sp
Laima	l aki m a sp	l aki aki m a sp
Oskaras	ok s k a r a s sp	oks s k a r a s sp
Kju	km j uk sp	kmj ju juk sp
Iksas	ik k s a s sp	ikk ikk kx s a s sp
Du	d uk sp	d ud ud sp
Keturi	km e t u rm ik sp	km et t u ri ir sp
Penki	pm e nk km ik sp	pm en nk km ik sp
Septyni	sm e pm tm ii nm ik sp	sem e pp tm y nm ik sp
Šeši	shm e shm ik sp	shm esh shm ish sp
Vienas	vm ie n a s sp	vm i ek n a s



3.3.2 paveiklė pateikiama apibendrinta išplėstinio tyrimo rezultatų lentelė po visų 7 kryžminių tyrimų. Galima matyti, kad blogiausiai atpažįstamos komandos yra „Trys“, „Penki“, „Fausta“, „Nulis“, „Sandra“, „Cecilija“, „Ygrekas“. Šių komandų atpažinimo tikslumas nesiekia nė 80 proc.



**3.3.2 pav.** Komandų atpažinimo tikslumas naudojant išplėstą fonemų rinkinį

Tikslus visų komandų atpažinimo rezultatus matome 3.3.3 lentelėje. Paryškintos yra komandos, kurių atpažinimo tikslumas nesiekia 90 proc.

**3.3.3 lentelė.** Išplėstinio fonemų rinkinio atpažinimo tikslumas

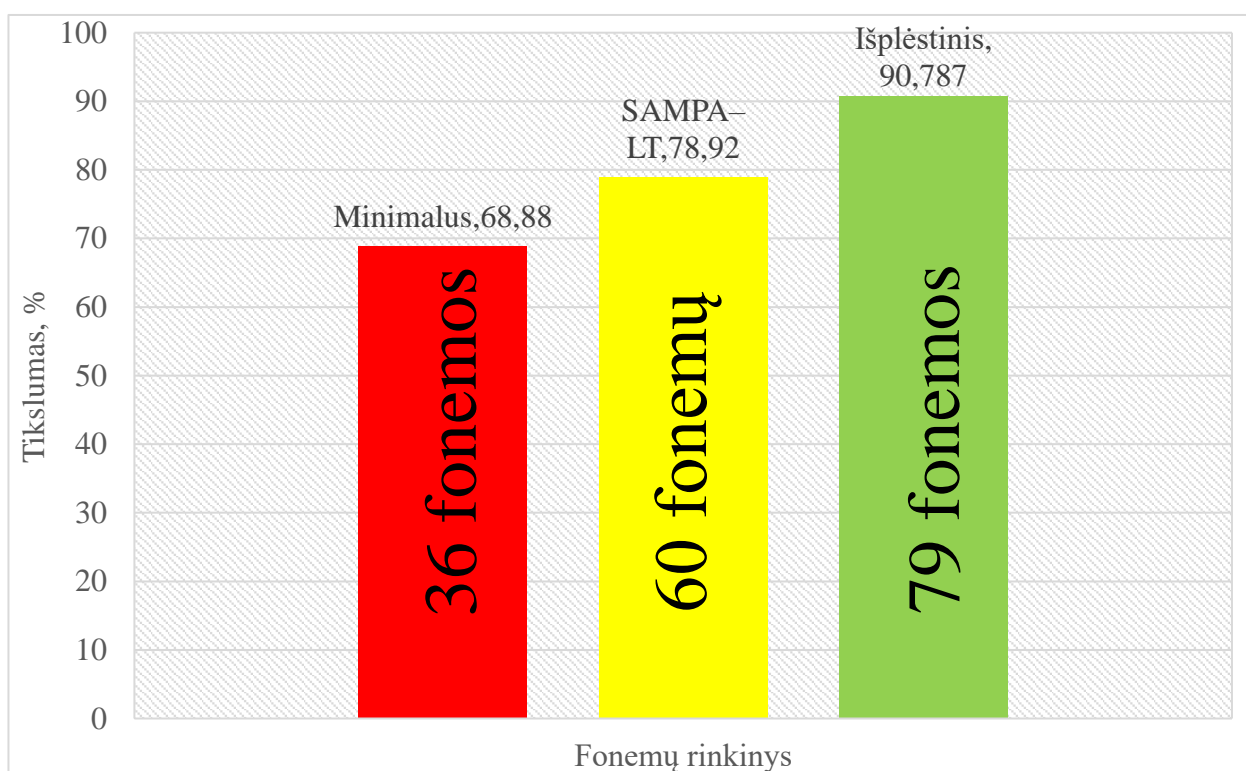
Komandos	Tikslumas, %	Komandos	Tikslumas, %	Komandos	Tikslumas, %
Austėja	89.51	Martynas	97.37	Ygrekas	79.75
Boleslovas	99.75	Nojus	87.14	Zacharijus	99.75
Cecilija	77.87	Oskaras	100	Aštuoni	100
Donatas	100	Patrikas	97.85	Devyni	99.88
Eimantas	99.75	Kju	95	Du	97.34
Fausta	70.94	Ričardas	99.28	Keturi	94.75
Gražvydas	99.51	Sandra	76.42	Nulis	74.75
Hansas	94.51	Teodoras	99.75	Penki	64.4
Izaokas	89.28	Ulijona	86.65	Septyni	97.87
Jonas	93.81	Vacys	99.51	Šeši	94.12
Karolis	99.51	Wašington	84.04	Trys	63.68
Laima	98.08	Iksas	96.91	Vienas	93.1



Išplėstinis fonemų rinkinys panaudotas kryžminiuose tyrimuose kaip ir su SAMPA–LT. Praplėtus fonemų rinkinį gautas minimalus 8,11 proc. tikslumo atpažinimo pagerėjimas ir didžiausias 14,78 proc. tikslumo atpažinimo pagerėjimas. Bendras tikslumas siekia 90,787 proc..

Siekiant pagerinti ir šių komandų atpažinimo tikslumą buvo didinamas fonemų rinkinys, tačiau rezultatai buvo tik blogesni, nes per didelė fonemų įvairovė yra netinkamai priskiriama skaičiuojant požymio vektorius taip suklaidinant HTK atpažintuvą.

Rezultatų apibendrinimą galima matyti 3.3.3 paveiksle. Iš šių rezultatų galima pamatyti, kad automatiname šnekos atpažinime, neturint anotuoto garsyno, didžiausias tikslumas gaunamas 90,787 proc. Toks tikslumas pasiekiamas išplečiant SAMPA–LT fonemų rinkinį.



3.3.3 pav. Atpažinimo tikslumas su skirtingais fonemų rinkiniais

### 3.4 Anotuoto garsyno tyrimo rezultatai

Norint patikrinti anotuoto garsyno įtaką atpažinimo tikslumui buvo paimta 1200 skaičių komandų. Jos sudarytos iš 6 diktorių. Šioms komandoms atliktas – rankinis – anotavimas aprašytas 2.3 skyrelyje. Mokymui, kaip ir testavimui naudojama 1200 anotuoto garsyno failų.

#### 3.4.1 Minimalaus fonemų rinkinio atpažinimas

Pirmi tyrimai atlikti su minimaliu fonemų rinkiniu nenaudojant anotuoto garsyno, kad galėtume patikrinti, kokią įtaką daro anotacijų priskarimas izoliuotų komandų atpažinimui.

```

----- Overall Results -----
SENT: %Correct=59.25 [H=711, S=489, N=1200]
WORD: %Corr=59.25, Acc=59.25 [H=711, D=0, S=489, I=0, N=1200]
----- Confusion Matrix -----
      A  D  D  K  N  P  S  S  T  V
S  E  U  E  U  E  E  E  R  I
T  V  T  L  N  P  S  Y  E
U  Y  U  I  K  T  I  S  N
0  N  R  S  I  Y  A  Del [ %c / %e]
ASTU 119 0 0 0 0 0 1 0 0 0 0 [99.2/0.1]
DEVY 0 104 0 0 0 0 16 0 0 0 0 [86.7/1.3]
DU 0 5 7 92 2 0 13 1 0 0 0 [ 5.8/9.4]
KETU 6 0 0 24 0 0 90 0 0 0 0 [20.0/8.0]
NULI 0 0 0 0 119 0 0 0 1 0 0 [99.2/0.1]
PENK 0 0 0 0 0 27 93 0 0 0 0 [22.5/7.8]
SEPT 0 10 0 0 0 0 110 0 0 0 0 [91.7/0.8]
SESI 46 0 0 0 0 0 0 74 0 0 0 [61.7/3.8]
TRYS 0 0 0 61 1 0 4 0 54 0 0 [45.0/5.5]
VIEN 0 0 0 9 27 0 7 0 4 73 0 [60.8/3.9]
Ins 0 0 0 0 0 0 0 0 0 0 0
-----

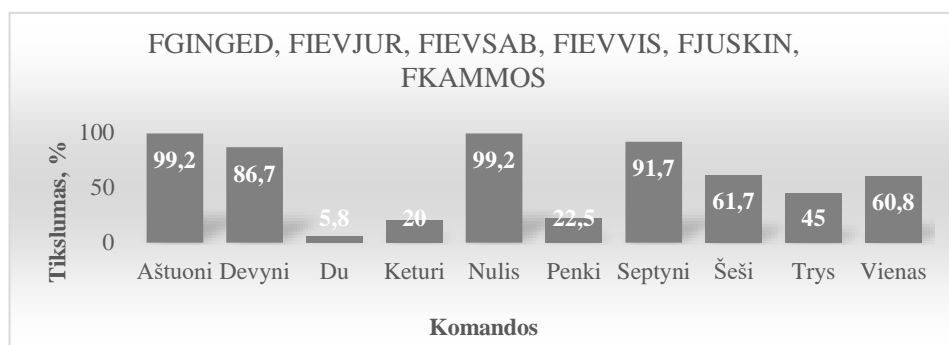
```

### 3.4.1.1 pav. Minimalus fonemų rinkinys (be anotacijų)

Tik skaičių komandoms skirtas minimalus fonemų rinkinys, sudarytas iš 19 fonemų, atrodo taip:

ASTUONI	ashtuoni sp	PENKI	penki sp	SENT-END []	sil
DEVYNI	devyni sp	SEPTYNI	septyni sp	SENT-START []	sil
DU	du sp	SESI	sheshi sp		
KETURI	keturi sp	TRYS	trys sp		
NULIS	nulis sp	VIENAS	vienas sp		

Su minimaliu fonemų rinkiniu gaunamas tik 59,25 proc. atpažinimo tikslumas. Kiekvienos komandos atpažinimo tikslumo rezultatai pateikiami 3.4.1.2 paveiksle. Šį tikslumą lyginsime su anotuoto garsyno atpažinimo tikslumu.



3.4.1.2 pav. Minimalaus fonemų rinkinio tikslumas priklausomai nuo testavimo diktorių (be anotacijų)

### 3.4.2 Minimalaus anotuoto fonemų rinkinio atpažinimo tikslumas

Atlikus anotuoto garsyno tyrimą gauta pasiskirstymo matrica pateikta 3.4.2.1 paveikslėlyje. Čia matomas geriausias gautas atpažinimo tikslumas, kuris yra 100%. Iš

pasiskirstymo matricos galima matyti, kad nė viena komanda atpažinime nubuvo sukklaidinta. Iš 1200 testavimui naudotų komandų atpažintos 1200.

```

----- Overall Results -----
SENT: %Correct=100.00 [H=1200, S=0, N=1200]
WORD: %Corr=100.00, Acc=100.00 [H=1200, D=0, S=0, I=0, N=1200]
----- Confusion Matrix -----
      A  D  D  K  N  P  S  S  T  V
      S  E  U  E  U  E  E  E  R  I
      T  V      T  L  N  P  S  Y  E
      U  Y      U  I  K  T  I  S  N
      O  N      R  S  I  Y      A  Del [ %c / %e]
ASTU 120 0 0 0 0 0 0 0 0 0 0
DEVY 0 120 0 0 0 0 0 0 0 0 0
DU 0 0 120 0 0 0 0 0 0 0 0
KETU 0 0 0 120 0 0 0 0 0 0 0
NULI 0 0 0 0 120 0 0 0 0 0 0
PENK 0 0 0 0 0 120 0 0 0 0 0
SEPT 0 0 0 0 0 0 120 0 0 0 0
SESI 0 0 0 0 0 0 0 120 0 0 0
TRYS 0 0 0 0 0 0 0 0 120 0 0
VIEN 0 0 0 0 0 0 0 0 0 120 0
Ins 0 0 0 0 0 0 0 0 0 0 0
-----

```

**3.4.2.1 pav.** Komandų atpažinimo tikslumas naudojant anotuotą minimalų fonemų rinkinį

Tikslumo vertinimui neatsižvelgiant į diktorius buvo atliktas testavimo tyrimas panaudojant visus turimus skaičių komandų diktorius. Tuomet testavimui buvo naudojami 24 diktoriai ir 4800 komandų.

3.4.2.2 paveiksle galima matyti, kad iš testavimui panaudotų 4800 komandų buvo suklysta 42. Bendras visų diktorių atpažinimo tikslumas yra 99,13 proc..

```

----- Overall Results -----
SENT: %Correct=99.13 [H=4758, S=42, N=4800]
WORD: %Corr=99.13, Acc=99.13 [H=4758, D=0, S=42, I=0, N=4800]
----- Confusion Matrix -----
      A  D  D  K  N  P  S  S  T  V
      S  E  U  E  U  E  E  E  R  I
      T  V      T  L  N  P  S  Y  E
      U  Y      U  I  K  T  I  S  N
      O  N      R  S  I  Y      A  Del [ %c / %e]
ASTU 480 0 0 0 0 0 0 0 0 0 0
DEVY 0 463 0 2 0 7 7 0 1 0 0 [96.5/0.4]
DU 0 0 473 4 0 1 0 0 2 0 0 [98.5/0.1]
KETU 0 0 0 479 0 0 0 0 1 0 0 [99.8/0.0]
NULI 0 0 2 0 472 0 0 4 2 0 0 [98.3/0.2]
PENK 0 0 0 0 0 479 1 0 0 0 0 [99.8/0.0]
SEPT 0 0 0 2 0 2 476 0 0 0 0 [99.2/0.1]
SESI 0 0 0 0 0 0 0 480 0 0 0
TRYS 0 0 0 0 0 0 0 0 480 0 0
VIEN 0 0 2 0 1 0 0 0 0 1 476 [99.2/0.1]
Ins 0 0 0 0 0 0 0 0 0 0 0
-----

```

**3.4.2.2 pav.** Komandų atpažinimo tikslumas, testavimui panaudojant 24 diktorius

**3.4.2.1 lentelė.** Atskirų diktorių tikslumas su minimaliu fonemų rinkiniu

Komandos Diktoriai	Aštuoni	Devyni	Du	Keturi	Nulis	Penki	Septyni	Šeši	Trys	Vienas	Tikslumas, %	Neatpažintu komandų skaičius
FIEVVIS	100	90	95	100	100	100	100	100	105	100	98,5	3
FSIMMEI	100	95	100	100	100	100	90	100	100	100	98,5	3
FIEVJUR	100	100	90	100	100	100	100	100	100	100	99	2
FVANPEC	100	95	100	100	100	95	100	100	100	100	99	2
FVILVAI	100	90	100	100	100	100	100	100	100	100	99	2
FUGNBUC	100	95	100	100	100	100	100	100	100	100	99,5	1
FKAMMOS	100	95	100	100	100	100	100	100	100	100	99,5	1
FLAUZET	100	95	100	100	100	100	100	100	100	100	99,5	1
FGINGED	100	100	100	100	100	100	100	100	100	100	100	0
FIEVSAB	100	100	100	100	100	100	100	100	100	100	100	0
FJUSKIN	100	100	100	100	100	100	100	100	100	100	100	0
FRAISAV	100	100	100	100	100	100	100	100	100	100	100	0
FRUTNAN	100	100	100	100	100	100	100	100	100	100	100	0
FUGNNOV	100	100	100	100	100	100	100	100	100	100	100	0
FVAIVAI	100	100	100	100	100	100	100	100	100	100	100	0
FVIONAB	100	100	100	100	100	100	100	100	100	100	100	0
FZIVSTA	100	100	100	100	100	100	100	100	100	100	100	0
MRIMAPA	100	95	100	95	70	100	90	100	100	100	95	10
MMODSLE	100	80	95	100	100	100	100	100	100	100	97,5	5
MEDGVOL	100	95	90	100	90	100	100	100	100	100	97,5	5
MDAIGUS	100	100	100	100	100	100	100	100	100	80	98	4
MKAZANU	100	95	95	100	100	100	100	100	100	100	99	2
MLINJUR	100	95	100	100	100	100	100	100	100	100	99,5	1
MVYGVAI	100	100	100	100	100	100	100	100	100	100	100	0
Kiekvienos komandos atpažinimo tikslumas, %	100	96.45	98.54	99.79	98.33	99.79	99.16	100	100	99.16	99.12	42

Atpažinimo tikslumo tyrimo rezultatai pateikiami 3.4.2.1 lentelėje, mokymui buvo naudojami 6 anotuoti diktoriai, o testuojama su kiekvienu diktoriumi atskirai, siekiant pamatyti, kuris diktorius turi didžiausią klaidingumo procentą. Testuojant net su 10 diktorių įmanoma gauti 100 proc. tikslumą.

Mokymo procese naudojami 6 diktoriai yra „*FAGNGRA, FAISZYM, FAISIZI, FDAILOI, FAGNVIN, FAUSNEM*“. Iš garsyno paruošimo tyrimams duomenų aptartą 2.2.1 skyrelyje žinoma, kad mokymo procese yra naudojamos tik moterų ištarimo komandos. Todėl vyrų komandų tikslumo rezultatai yra atskirti brūkšniu. Iš rezultatų matoma, kad vyrų diktorių atpažinimo metu iš 7 diktorių testavimui padarytos 27 klaidos iš 42. Iš tokių duomenų matoma, kad tik 7 diktoriai padarė 64 proc. visų klaidų. Todėl galima teigti, kad turint papildomų anotavimo duomenų su vyrų diktoriais būtų galimas dar tikslesnis atpažinimas.

### 3.4.3 SAMPA–LT anotuoto fonemų rinkinio atpažinimo tikslumas

SAMPA–LT fonemų rinkinys skirtas tik skaičių komandoms, sudarytas iš 28 fonemų, atrodo taip:

ASTUONI	a sh t uo nm ik sp	NULIS	n uk lm i s sp	TRYŠ	tm rm y s sp
DEVYNI	dm e vm ii nm ik sp	PENKI	pm e nk km ik sp	VIENAS	vm ie n a s sp
DU	d uk sp	SEPTYNI	sm e pm tm ii nm ik sp	SENT-END	[] sil
KETURI	km e t u rm ik sp	SESI	shm e shm ik sp	SENT-START	[] sil

Be anotacijų skaičių komandų atpažinimo tikslumas parodytas 3.4.3.1 paveiksle ir yra 80proc..

```

----- Overall Results -----
SENT: %Correct=80.00 [H=960, S=240, N=1200]
WORD: %Corr=80.00, Acc=80.00 [H=960, D=0, S=240, I=0, N=1200]
----- Confusion Matrix -----
      A  D  D  K  N  P  S  S  T  V
      S  E  U  E  U  E  E  E  R  I
      T  V      T  L  N  P  S  Y  E
      U  Y      U  I  K  T  I  S  N
      O  N      R  S  I  Y      A  Del [ %c / %e]
ASTU 120 0 0 0 0 0 0 0 0 0 0
DEVY 0 120 0 0 0 0 0 0 0 0 0
DU 4 0 6 102 0 0 1 0 7 0 0 [ 5.0/9.5]
KETU 0 0 0 102 0 0 18 0 0 0 0 [85.0/1.5]
NULI 0 0 0 0 113 0 0 0 7 0 0 [94.2/0.6]
PENK 0 0 0 0 0 120 0 0 0 0 0
SEPT 0 16 0 0 0 0 104 0 0 0 0 [86.7/1.3]
SESI 39 0 0 1 0 2 5 73 0 0 0 [60.8/3.9]
TRYŠ 0 0 0 4 0 0 2 0 114 0 0 [95.0/0.5]
VIEN 0 0 0 0 0 0 0 0 32 88 0 [73.3/2.7]
Ins 0 0 0 0 0 0 0 0 0 0 0
=====

```

#### 3.4.3.1 pav. SAMPA–LT fonemų rinkinio atpažinimo tikslumas (be anotacijų)

Anotuotas garsynas buvo paruoštas su minimaliu fonemų rinkiniu, todėl fonemų anotacijų reikšmės buvo perrašytos taip, kad tiktų SAMPA–LT fonemų rinkiniui. Tyrimo rezultatai pateikiami 3.4.3.2 paveiksle.

Anotuoto garsyno su SAMPA–LT fonemų rinkiniu geriausias tikslumas 100 proc.. Testavimui naudoti 6 diktoriai „*FGINGED, FJUSKIN, FRUTNAN, FUGNNOV, FVAIVAI, FVANPEC*“.

```

----- Overall Results -----
SENT: %Correct=100.00 [H=1200, S=0, N=1200]
WORD: %Corr=100.00, Acc=100.00 [H=1200, D=0, S=0, I=0, N=1200]
----- Confusion Matrix -----
      A  D  D  K  N  P  S  S  T  V
      S  E  U  E  U  E  E  E  R  I
      T  V      T  L  N  P  S  Y  E
      U  Y      U  I  K  T  I  S  N
      O  N      R  S  I  Y      A  Del [ %c / %e]
ASTU 120 0 0 0 0 0 0 0 0 0 0
DEVY 0 120 0 0 0 0 0 0 0 0 0
DU 0 0 120 0 0 0 0 0 0 0 0
KETU 0 0 0 120 0 0 0 0 0 0 0
NULI 0 0 0 0 120 0 0 0 0 0 0
PENK 0 0 0 0 0 120 0 0 0 0 0
SEPT 0 0 0 0 0 0 120 0 0 0 0
SESI 0 0 0 0 0 0 0 120 0 0 0
TRYS 0 0 0 0 0 0 0 0 120 0 0
VIEN 0 0 0 0 0 0 0 0 0 120 0
Ins 0 0 0 0 0 0 0 0 0 0 0
=====

```

### 3.4.3.2 pav. Anotuoto SAMPA–LT fonemų rinkinio atpažinimo tikslumas

Iš aukščiau pateikiamų duomenų matoma (žr. 3.4.3.2 pav.), kad su šiais diktoriais gautas toks pat 100 proc. tikslumas, kaip naudojant minimalų fonemų rinkinį.

Tikslumo tyrimo su 24 diktoriais rezultatai atvaizduoti 3.4.3.3 paveiksle. Tokio atpažinimo tikslumas 98 proc., kai minimalaus fonemų rinkinio atveju rezultatai buvo 99,13 proc.. Su SAMPA–LT fonemų rinkiniu gautas mažesnis tikslumas. Šiuo atveju suklysta 96, o su minimaliu fonemų rinkiniu suklysta tik 42 kartus.

```

----- Overall Results -----
SENT: %Correct=98.00 [H=4704, S=96, N=4800]
WORD: %Corr=98.00, Acc=98.00 [H=4704, D=0, S=96, I=0, N=4800]
----- Confusion Matrix -----
      A  D  D  K  N  P  S  S  T  V
      S  E  U  E  U  E  E  E  R  I
      T  V      T  L  N  P  S  Y  E
      U  Y      U  I  K  T  I  S  N
      O  N      R  S  I  Y      A  Del [ %c / %e]
ASTU 480 0 0 0 0 0 0 0 0 0 0
DEVY 0 474 0 1 0 3 1 1 0 0 0 [98.8/0.1]
DU 0 0 472 3 0 2 0 0 3 0 0 [98.3/0.2]
KETU 0 0 0 477 0 2 0 0 1 0 0 [99.4/0.1]
NULI 0 0 0 10 0 469 0 0 1 0 0 [97.7/0.2]
PENK 0 0 0 0 0 480 0 0 0 0 0
SEPT 11 0 0 24 0 20 413 11 1 0 0 [86.0/1.4]
SESI 0 0 0 0 0 0 0 480 0 0 0
TRYS 0 0 0 0 0 0 0 0 480 0 0
VIEN 0 0 1 0 0 0 0 0 0 479 0 [99.8/0.0]
Ins 0 0 0 0 0 0 0 0 0 0 0
=====

```

### 3.4.3.3 pav. Anotuoto SAMPA–LT tikslumas naudojant 24 diktorius testavimui

Duomenys apie atpažinimo tikslumo rezultatus pateikiami 3.4.3.1 lentelėje. Testavimo metu tikslumo tyrimas atliktas su kiekvienu diktoriu atskirai. Iš lentelėje pateikiamų duomenų galima įvertinant, kuris diktorius klydo atpažindamas komandas. Testavimo metu daugiausiai neatitikimų turėjo komanda „*Septyni*“. Iš 96 padarytų klaidų 67 klaidos buvo padarytos atpažįstant komandą „*Septyni*“.

### 3.4.3.1 lentelė. Atskirų diktorių tikslumas su SAMPA–LT fonemų rinkiniu

Komandos Diktoriai	Aštuoni	Devyni	Du	Keturi	Nulis	Penki	Septyni	Šeši	Trys	Vienas	Tikslumas, %	Neatpažintu komandų skaičius
FSIMMEI	100	100	95	90	100	100	40	100	100	100	92,5	15
FIEVSAB	100	100	100	100	100	100	70	100	100	100	97	6
FIEVVIS	100	95	95	100	100	100	90	100	100	100	98	4
FKAMMOS	100	100	100	100	100	100	85	100	100	100	98,5	3
FUGNBUC	100	100	100	100	100	100	85	100	100	100	98,5	3
FIEVJUR	100	100	95	100	100	100	100	100	100	100	99,5	1
FLAUZET	100	100	100	100	100	100	95	100	100	100	99,5	1
FRAISAV	100	100	100	100	100	100	95	100	100	100	99,5	1
FGINGED	100	100	100	100	100	100	100	100	100	100	100	0
FJUSKIN	100	100	100	100	100	100	100	100	100	100	100	0
FRUTNAN	100	100	100	100	100	100	100	100	100	100	100	0
FUGNNOV	100	100	100	100	100	100	100	100	100	100	100	0
FVAIVAI	100	100	100	100	100	100	100	100	100	100	100	0
FVANPEC	100	100	100	100	100	100	100	100	100	100	100	0
FVILVAI	100	100	100	100	100	100	100	100	100	100	100	0
FVIONAB	100	100	100	100	100	100	100	100	100	100	100	0
FZIVSTA	100	100	100	100	100	100	100	100	100	100	100	0
MRIMAPA	100	95	100	95	45	100	10	100	100	100	84,5	31
MVYGVAI	100	100	100	100	100	100	45	100	100	100	94,5	11
MEDGVOL	100	95	85	100	90	100	70	100	100	100	95	10
MLINJUR	100	100	100	100	100	100	80	100	100	100	98	4
MMODSLE	100	85	95	100	100	100	100	100	100	100	98	4
MDAIGUS	100	100	100	100	100	100	100	100	100	95	99,5	1
MKAZANU	100	100	95	100	100	100	100	100	100	100	99,5	1
Kiekvienos komandos atpažinimo tikslumas, %	100	98,75	98,33	99,375	97,708	100	86,042	100	100	99,7916	98	96

Iš 3.4.3.1 lentelėje pateiktų duomenų galima matyti, kad testuojant su 9 diktoriais įmanomas 100 proc. tikslumas. Tačiau su SAMPA–LT fonemų rinkiniu atpažinimo metu daugiausia neatitikimu turėjo „MRIMAPA“ diktorius, atpažinimo metu suklysta net 31 kartą. Su šiuo diktoriumi ir minimaliu fonemų rinkiniu buvo padaryta 10 klaidų. Testavimo metu visi vyrų diktoriai padarė klaidų. Iš 7 vyrų diktorių padaryta net 62 klaidos. Taigi, vyrų diktoriai padarė 64,58 proc. klaidų. Tai galima paaiškinti tuo, kad mokymo procese nenaudoti vyrų diktorių anotuoti garso failai.

Taigi SAMPA–LT fonemų rinkinys nepasiteisino tikslumo tyrimuose, su šiuo fonemų rinkiniu padaryta daugiau klaidų, negu su minimaliu fonemų rinkiniu.

### 3.4.4 Trifonų atpažinimo tyrimo tikslumas

Minimalų fonemų rinkinį pavertus kontekstinėmis fonemomis atliktas tikslumo tyrimas, kai testavimo metu pasitelkiami šie 6 diktoriai „FGINGED, FIEVSAB, FKAMMOS, FRAISAV, FRUTNAN, FVAIVAI“. Geriausias atpažinimo rezultatas kaip ir visų anotuotų tyrimų atveju sudarė 100 proc. ( žr. 3.4.4.1 pav.).

```

----- Overall Results -----
SENT: %Correct=100.00 [H=1200, S=0, N=1200]
WORD: %Corr=100.00, Acc=100.00 [H=1200, D=0, S=0, I=0, N=1200]
----- Confusion Matrix -----
      A  D  D  K  N  P  S  S  T  V
      S  E  U  E  U  E  E  E  R  I
      T  V      T  L  N  P  S  Y  E
      U  Y      U  I  K  T  I  S  N
      O  N      R  S  I  Y      A
ASTU 120  0  0  0  0  0  0  0  0  0
DEVY  0 120  0  0  0  0  0  0  0  0
  DU  0  0 120  0  0  0  0  0  0  0
KETU  0  0  0 120  0  0  0  0  0  0
NULI  0  0  0  0 120  0  0  0  0  0
PENK  0  0  0  0  0 120  0  0  0  0
SEPT  0  0  0  0  0  0 120  0  0  0
SESI  0  0  0  0  0  0  0 120  0  0
TRYS  0  0  0  0  0  0  0  0 120  0
VIEN  0  0  0  0  0  0  0  0  0 120
Ins   0  0  0  0  0  0  0  0  0  0
=====

```

### 3.4.4.1 pav. Minimalaus kontekstinių fonemų rinkinio tyrimas su 6 diktoriais

Atliktas kontekstinių fonemų tyrimas su testavimui skirtais 24 diktoriais. Šio tyrimo tikslumas sudarė 99,10 proc., kuomet iš 4800 komandų buvo neteisingai atpažintos 43 ( žr. 3.4.4.2 pav.).

```

----- Overall Results -----
SENT: %Correct=99.10 [H=4757, S=43, N=4800]
WORD: %Corr=99.10, Acc=99.10 [H=4757, D=0, S=43, I=0, N=4800]
----- Confusion Matrix -----
      A  D  D  K  N  P  S  S  T  V
      S  E  U  E  U  E  E  E  R  I
      T  V      T  L  N  P  S  Y  E
      U  Y      U  I  K  T  I  S  N
      O  N      R  S  I  Y      A
ASTU 480  0  0  0  0  0  0  0  0  0
DEVY  0 462  0  1  0  9  7  1  0  0
  DU  0  0 473  4  0  1  0  0  2  0
KETU  0  0  0 480  0  0  0  0  0  0
NULI  0  0  1  0 474  0  0  5  0  0
PENK  0  0  0  0  0 479  1  0  0  0
SEPT  0  0  0  2  0  2 476  0  0  0
SESI  0  0  0  0  0  0  0 480  0  0
TRYS  0  0  0  0  0  0  0  0 480  0
VIEN  0  0  3  0  1  0  0  1  2 473
Ins   0  0  0  0  0  0  0  0  0  0
=====

```

### 3.4.4.2 pav. Minimalaus kontekstinių fonemų rinkinio tyrimas su 24 diktoriais

Siekiant palyginti kokią įtaką fonemų rinkiniai turi kontekstinėms fonemoms, atliktas trifonų sudarymas iš SAMPA–LT fonemų rinkinio. Testavimo metu naudojami šiuos 6 diktoriai „FGINGED, FJUSKIN, FUGNNOV, FVANPEC, FRUTNAN, FVAIVAI“. Testavimo metu gautas 100 proc. tikslumas. ( žr. 3.4.43 pav.).



```

----- Overall Results -----
SENT: %Correct=100.00 [H=1200, S=0, N=1200]
WORD: %Corr=100.00, Acc=100.00 [H=1200, D=0, S=0, I=0, N=1200]
----- Confusion Matrix -----
      A  D  D  K  N  P  S  S  T  V
      S  E  U  E  U  E  E  E  R  I
      T  V      T  L  N  P  S  Y  E
      U  Y      U  I  K  T  I  S  N
      O  N      R  S  I  Y      A
ASTU 120  0  0  0  0  0  0  0  0  0
DEVY  0 120  0  0  0  0  0  0  0  0
  DU  0  0 120  0  0  0  0  0  0  0
KETU  0  0  0 120  0  0  0  0  0  0
NULI  0  0  0  0 120  0  0  0  0  0
PENK  0  0  0  0  0 120  0  0  0  0
SEPT  0  0  0  0  0  0 120  0  0  0
SESI  0  0  0  0  0  0  0 120  0  0
TRYS  0  0  0  0  0  0  0  0 120  0
VIEN  0  0  0  0  0  0  0  0  0 120
Ins   0  0  0  0  0  0  0  0  0  0
-----

```

### 3.4.4.3 pav. SAMPA–LT kontekstinių fonemų rinkinio tyrimas su 6 diktoriais

Atliktas ir kontekstinių fonemų tyrimas su testavimui skirtais 24 diktoriais. Šio tyrimo tikslumas 98 proc.. Iš 4800 komandų neteisingai atpažintos 96 komandos. Šio tyrimo tikslumas lygiai toks pat, kaip ir su SAMPA–LT fonemų rinkiniu nenaudojant kontekstinių fonemų. (žr. 3.4.4.4 pav.).

```

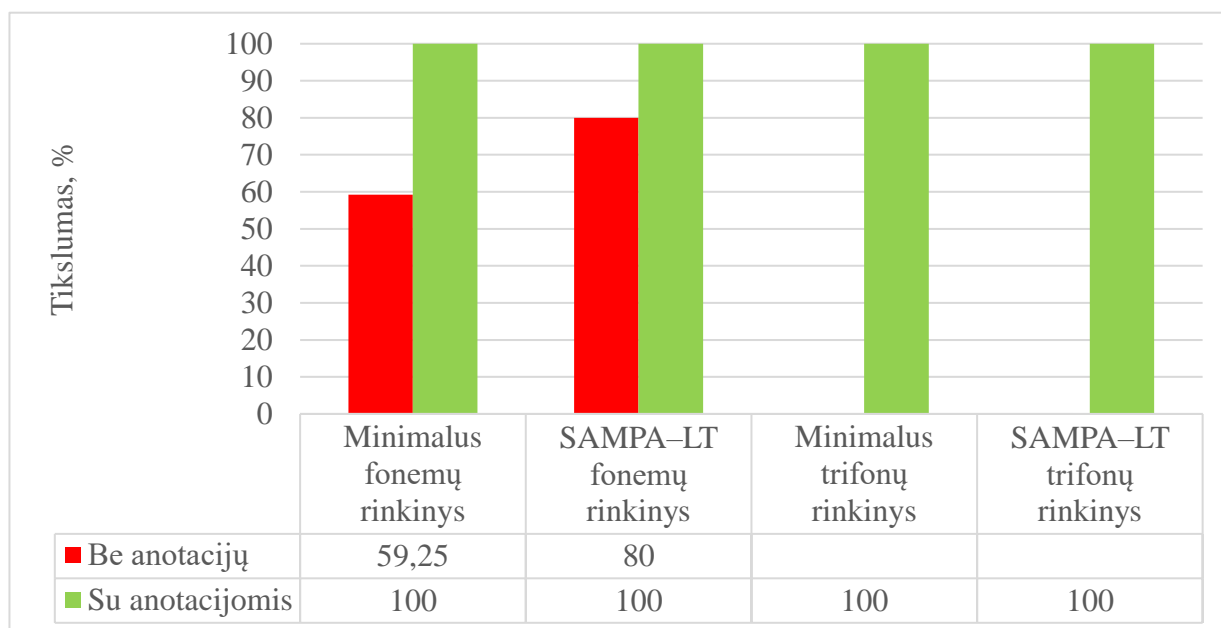
----- Overall Results -----
SENT: %Correct=98.00 [H=4704, S=96, N=4800]
WORD: %Corr=98.00, Acc=98.00 [H=4704, D=0, S=96, I=0, N=4800]
----- Confusion Matrix -----
      A  D  D  K  N  P  S  S  T  V
      S  E  U  E  U  E  E  E  R  I
      T  V      T  L  N  P  S  Y  E
      U  Y      U  I  K  T  I  S  N
      O  N      R  S  I  Y      A
ASTU 480  0  0  0  0  0  0  0  0  0
DEVY  0 474  0  1  0  2  0  1  2  0
  DU  0  0 472  3  0  1  0  0  4  0
KETU  0  0  0 478  0  1  0  0  1  0
NULI  0  0  0 12  0 467  0  0  1  0
PENK  0  0  0  0  0 480  0  0  0  0
SEPT 12  0  0 25  0 16 414 12  1  0
SESI  0  0  0  0  0  0  0 480  0  0
TRYS  0  0  0  0  0  0  0  0 480  0
VIEN  0  0  1  0  0  0  0  0  0 479
Ins   0  0  0  0  0  0  0  0  0  0
-----

```

### 3.4.4.4 pav. SAMPA–LT kontekstinių fonemų rinkinio tyrimas su 24 diktoriais

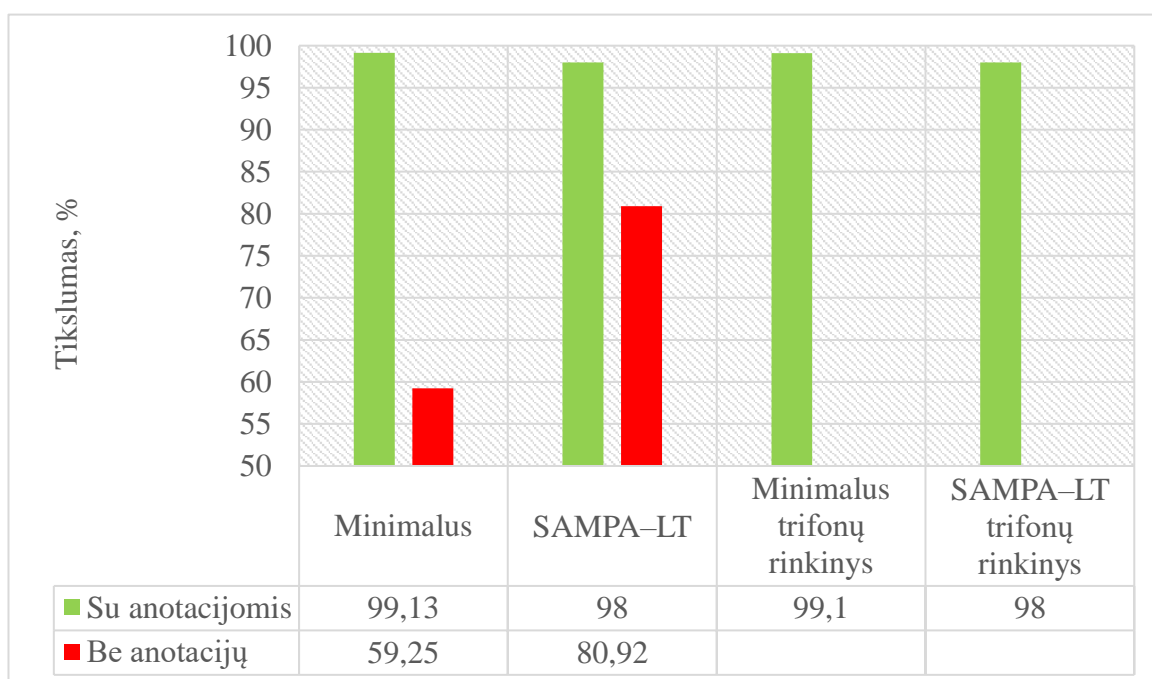
Šie rezultatai tik patvirtina tai, kad turint anotuotą garsyną mokymo procese geriausia tikslumą galima gauti naudojant minimalų fonemų rinkinį.

### 3.5 Anotuoto skaičių garsyno atpažinimo tyrimo rezultatų apibendrinimas



**3.5.1 pav.** Anotuoto garsyno atpažinimo tikslumas testavimui naudojant 6 diktorius

Apibendrinant visus gautus rezultatus, galima pasakyti, kad garsyno duomenų anotavimas ženkliai pagerino atpažinimo tikslumą. Minimalus fonemų rinkinys, testavime naudojant 6 diktorių kalbos signalus nuo 59,25 proc. pakilo iki 100 proc. SAMPA-LT fonemų rinkinio atpažinimas nuo 80 proc. pakilo iki 100 proc. Visais anotuoto garsyno testavimo atvejais galima gauti 100 proc. tikslumą jei naudojama maža testavimo duomenų aibė ( žr. 3.5.1 pav.).



**3.5.2 pav.** Anotuoto garsyno atpažinimo tikslumas testavimui naudojant 24 diktorius

Aktualiausi rezultatai yra tikslumas nepriklausomas nuo diktorių, tam atliktas tyrimas panaudojant visus turimus skaičių komandų diktorius. Matome pagal 3.5.2 paveiksle gautus rezultatus, kad tiksliausias skaičių komandų atpažinimas yra su minimaliu fonemų rinkiniu ir yra 99,13 proc.. Šiuo atveju klaidos tikimybė tik 0,87 proc.. Tikslumas su SAMPA–LT fonemų rinkiniu gavosi blogesnis. Testavimo metu tikslumo tikrinimas atliktas su 4800 komandų. Klaidų skaičius su SAMPA–LT fonemų rinkiniu yra 96 lyginant su 42 klaidomis padarytomis naudojant minimalų fonemų rinkinį. Sukuriant trifonus iš šių fonemų rinkinių nedavė tikslesnių rezultatų.

Tikslumą buvo galima gauti geresnį, jeigu būtų turimas didesnės apimties anotuotas garsynas, nes visi PMM parametrai yra jautrūs mokymo imties dydžiui ir mokymo tikslumas būtų tikslesnis, kuo didesnė mokymo duomenų imtis. Taip pat būtina atsižvelgti į tai, kad mokymo procese nebuvo naudotasi vyrų diktorių balso komandomis. Dėl šios priežasties atpažinimo rezultatuose didesnę klaidų apimtį sudarė vyrų diktorių skaičių ištarimo komandos.

### 3.6 Apmokytų fonemų įterpimas į skaičių ir vardų garsyno tyrimą

Šiame tyrime buvo panaudotos fonemos, kurias mokėme skaičių garsyno tyrime. Minimalus fonemų rinkinys yra sudarytas iš 35 fonemų, jų tarpe naudojamas 18 fonemų apmokytas rinkinys. Likusioms fonemoms yra priskiriami pradiniai modelio vektoriaus reikšmių koeficientai „*proto*“.

```

----- Overall Results -----
SENT: %Correct=96.81 [H=2672, S=88, N=2760]
WORD: %Corr=96.81, Acc=96.81 [H=2672, D=0, S=88, I=0, N=2760]
----- Confusion Matrix -----

```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	K	R	S	T	U	V	W	I	Y	Z	A	D	D	K	N	P	S	S	T	V	Del [ %c / %e]				
AUST	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
BOLE	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
CECI	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
DONA	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
EIMA	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
FAUS	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
GRAZ	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HANS	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
IZAO	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
JONA	0	0	0	0	0	0	0	0	0	59	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[98.3/0.0]	
KARO	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LAIM	0	0	0	0	0	0	0	0	0	0	0	58	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	[96.7/0.1]	
MART	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
NOJU	0	0	0	0	0	0	0	0	0	0	3	0	0	56	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[93.3/0.1]		
OSKA	0	0	0	0	0	8	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[86.7/0.3]		
PATR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
KJU	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[95.0/0.1]		
RICA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
SAND	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
TEOD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ULIJ	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	2	0	0	0	0	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[68.3/0.7]		
VACY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
WASH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
IKSA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
YGRE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ZACH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASHT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	119	0	0	1	0	0	0	0	0	0	0	0	[99.2/0.0]		
DEVY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	114	0	0	0	2	3	0	1	0	0	[95.0/0.2]			
DU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	119	1	0	0	0	0	0	0	0	0	0	[99.2/0.0]			
KETU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	119	0	1	0	0	0	0	0	0	[99.2/0.0]			
NULI	0	0	0	0	0	0	0	0	0	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	83	0	0	0	0	0	0	0	[69.2/1.3]			
PENK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	120	0	0	0	0	0	0	0	0		
SEPT	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	116	0	0	0	0	0	[96.7/0.1]				
SHES	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	120	0	0	0	0	0	0			
TRYS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	120	0	0	0	0	0			
VIEN	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	119	0	[99.2/0.0]			
Ins	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		

3.6.1 pav. Vardų ir skaičių atpažinimo tikslumas panaudojant apmokytas fonemas

Iš 3.6.1 paveiksle parodytos pasiskirstymo matricos galime matyti, kad ir neturint visų apmokytų fonemų reikšmių, galime gauti žymiai geresnius rezultatus (žr. 3.1.1 pav.). Naudojantis minimaliu fonemų rinkiniu gautas 68,88 proc. tikslumas. Neapmokytas fonemas pakeitus mokytomis gautas 96,81 proc. tikslumas. Klaidų skaičius nuo 859 sumažėjo iki 88. Toks didelis tikslumo pagerėjimas patvirtina, kad turint anotuotą garsyną galima gauti žymiai geresnį tikslumą. Dar didesnis tikslumas būtų gautas, jei turėtume ir vardų garsyno anotuotas transkripcijas.

## IŠVADOS IR REZULTATAI

1. Atlikus skaičių ir vardų tikslumo tyrimą su minimaliu fonemų rinkiniu gautas 68,88 proc. tikslumas. Pasinaudojus SAMPA–LT fonemų rinkiniu gautas geresnis 78,92 proc. tikslumas. Siekiant pagerinti tikslumą įvestos papildomos fonemos ir geriausias tikslumas gautas 90,787 proc.. Todėl galima teigti, kad tinkamas fonemų transkripcijų parinkimas padeda pagerinti atpažinimo tikslumą.
2. Baigiamojo darbo metu išsiaiškintas garsyno duomenų anotavimas, ir jo panaudojimas atpažinimo tikslumo tyrimams.
3. Pasinaudojus HTK paketu atliktas anotuoto garsyno tyrimas skaičių komandoms. Minimalaus anotuoto fonemų rinkinio atpažinimo tikslumas 99,13 proc., testuojant su 24 diktoriais, o trifonu, sudarytu iš minimalaus fonemų rinkinio, atpažinimo tyrimo tikslumas 99,10 proc.. SAMPA–LT anotuoto fonemų rinkinio atpažinimo tikslumas 98 proc., testuojant su 24 diktoriais. Tiek pat kontekstiniu fonemų (trifonų), sudarytu iš SAMPA–LT fonemų rinkinio, atpažinimo tyrimo tikslumas 98 proc..
4. Skaičių ir vardų garsyno tyrime pasinaudojus anotuotų fonemų vektorių koeficientų reikšmes tikslumas pagerėjo nuo 68,88 proc. iki 96.81 proc..
5. Palyginus gautus rezultatus galima teigti, kad turint anotuotą garsyną galima gauti žymiai geresnius rezultatus, nei naudojantis tik fonemomis grįžtais paslėptaisiais Markovo modeliais, be anotavimo žymių.

## LITERATŪROS SĄRAŠAS

1. „Apie projektą „LIEPA““. Interaktyvus [žiūrėta 2016-04-13]. Prieiga per internetą: <https://www.xn--ratija-ckb.lt/liepa>
2. Rudžionis, V., Ratkevičius, K., Rudžionis, A., Raškinis, G., Maskeliūnas, R. (2013). “ Hibridinė atpažinimo technologija balso sąsajai (INFOBALSAS)”. Aukštųjų technologijų plėtros 2011–2013 metų programa. Mokslo, inovacijų ir technologijų agentūra. Baigiamoji ataskaita. Kaunas, Lietuva, 80 p.
3. Laurinčiukaitė, S. (2008). Lietuvių šnekos atpažinimo akustinis modeliavimas: daktaro disertacija: technologijos mokslai, informatikos inžinerija 07T / Vilniaus Gedimino technikos universitetas, Vilnius: Technika. 134 p. Interaktyvus [žiūrėta 2016-04-11]. Prieiga per internetą: [www.mii.lt/files/mii\\_dis\\_08\\_laurinciukaite.pdf](http://www.mii.lt/files/mii_dis_08_laurinciukaite.pdf)
4. Kasparaitis, P. (2005). Kompiuterinės lingvistikos paskaitų konspektai. Interaktyvus [žiūrėta 2016-04-11]. Prieiga per internetą: <http://www.mif.vu.lt/~pijus/CL/cl.htm>
5. M. Forsberg „Why is Speech Recognition Difficult?“ (2003). Department of Computing Science, Chalmers University of Technology. Interaktyvus [žiūrėta 2016-04-24]. Prieiga per internetą: [http://www.speech.kth.se/~rolf/gslt\\_papers/MarkusForsberg.pdf](http://www.speech.kth.se/~rolf/gslt_papers/MarkusForsberg.pdf)
6. Vaičiukynas, E. (2009). Balso dialogų technologijos (kompiuteriniai dialogai balso): Informacinių technologijų metodai: Kauno technologijos universitetas (T120D004). Interaktyvus [žiūrėta 2016-05-02]. Prieiga per internetą: [http://www.elen.ktu.lt/studentai/lib/exe/fetch.php?media=balso\\_dialogu\\_tech.pdf](http://www.elen.ktu.lt/studentai/lib/exe/fetch.php?media=balso_dialogu_tech.pdf)
7. Vaičiūnas, A. (2006). Lietuvių kalbos statistinių modelių ir jų taikymo šnekos atpažinimui tyrimas, kai naudojami labai dideli žodynai: daktaro disertacija: fiziniai mokslai, informatika 09P / Vytauto Didžiojo universitetas, Kaunas. Interaktyvus [žiūrėta 2016-11-11]. Prieiga per internetą: [http://donelaitis.vdu.lt/disertacijos/Vaiciunas\\_2006.pdf](http://donelaitis.vdu.lt/disertacijos/Vaiciunas_2006.pdf)
8. Maskeliūnas R. (2009). Lietuviškų balso komandų atpažinimas daugybinių transkripcijų pagrindu: daktaro disertacija: technologijos mokslai, informatikos inžinerija – 07T / Kauno technologijos universitetas. Kaunas. 159 p.
9. B. Resch „Automatic Speech Recognition with HTK“ Interaktyvus [žiūrėta 2016-11-11]. Prieiga per internetą: <https://www.spsc.tugraz.at/system/files/asr.pdf>
10. Laurinčiukaitė, S. (2003). Lietuvių kalbos kompiuteriniai tyrimai (IX sekcija): Atskirai pasakytų Lietuviškų kalbos žodžių atpažinimas, reimiantis paslėptaisiais Markovo modeliais, Matematikos informatikos institutas, Vilnius. Interaktyvus [žiūrėta 2017-03-13]. Prieiga per internetą: [http://www.elibrary.lt/resursai/Konferencijos/KTU\\_01/IT\\_2003/Sekcija09.pdf](http://www.elibrary.lt/resursai/Konferencijos/KTU_01/IT_2003/Sekcija09.pdf)

11. Balvočius, B., Telksnys, L. (2003). Lietuvių kalbos kompiuteriniai tyrimai (IX sekcija): Garsynų duomenų modeliai ir programinės įrangos architektūros / Vytauto Didžiojo universitetas, Matematikos informatikos institutas. Interaktyvus [žiūrėta 2017-03-13]. Prieiga per internetą: [http://www.elibrary.lt/resursai/Konferencijos/KTU\\_01/IT\\_2003/Sekcija09.pdf](http://www.elibrary.lt/resursai/Konferencijos/KTU_01/IT_2003/Sekcija09.pdf)
12. VoxForge. „Tutorial: Create Acoustic Model – Manually“. Interaktyvus [žiūrėta 2016-03-23]. Prieiga per internetą: <http://www.voxforge.org/home/dev/acousticmodels/windows/create/htkjulius/tutorial>
13. Vimala.C., Dr.V.Radha., (2012) „A Review on Speech Recognition Challenges and Approaches“ World of Computer Science and Information Technology Journal (WCSIT). Interaktyvus [žiūrėta 2016-04-18]. Prieiga per internetą: <http://www.wcsit.org/pub/2012/vol.2.no.1/A%20Review%20on%20Speech%20Recognition%20Challenges%20and%20Approaches.pdf>
14. Tiwari. V., (2010). „MFCC and its applications in speaker recognition“. Interaktyvus [žiūrėta 2017-05-05]. Prieiga per internetą: <https://pdfs.semanticscholar.org/b4e9/c14c67b8aa431a40041cce0a3564144e1a2a.pdf>
15. Young S., Evermann G., Kershaw D., Moore G., Odell J., Ollason D., Valtchev V., Woodland P.(2015). The HTK Book (version 3.5). Microsoft Corporation. 433p. Interaktyvus [žiūrėta 2017-03-23]. Prieiga per internetą: <http://htk.eng.cam.ac.uk/ftp/software/htkbook-3.5.alpha-1.pdf>
16. „How To: Use HTK Hidden Markov modelling toolkit with SFS“. Interaktyvus [žiūrėta 2017-02-28]. Prieiga per internetą: <http://www.phon.ucl.ac.uk/resource/sfs/howto/htk.php>
17. Yi J., (1996). „Speech recognition method and system using triphones, diphones and phonemes“. Interaktyvus [žiūrėta 2017-05-08]. Prieiga per internetą: <https://docs.google.com/viewer?url=patentimages.storage.googleapis.com/pdfs/US5502790.pdf>
18. Raškinis, A., Raškinis, G., Kazlauskienė, A. (2003). „SAMPA (Speech assessment methods phonetic alphabet) for encoding transcriptions of Lithuanian speech corpora“. Interaktyvus [žiūrėta 2017-05-02]. Prieiga per internetą: [http://fcim.vdu.lt/~asta\\_kazlauskiene/publikacijos/SAMPA%20for%20encoding%20transcriptions.pdf](http://fcim.vdu.lt/~asta_kazlauskiene/publikacijos/SAMPA%20for%20encoding%20transcriptions.pdf)