

**KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS**

Donatas Sipavičius

**„GOOGLE“ LIETUVIŲ ŠNEKOS ATPAŽINTUVO EFEKTYVUMO
ĮVERTINIMO TYRIMAI**

Baigiamasis magistro projektas

Vadovas

Doc. dr. Rytis Maskeliūnas

KAUNAS, 2016

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

**„GOOGLE“ LIETUVIŲ ŠNEKOS ATPAŽINTUVO EFEKTYVUMO
ĮVERTINIMO TYRIMAI**

Baigiamasis magistro projektas
Informatikos studijų programa (kodas 621I10003)

Vadovas

Doc. dr. Rytis Maskeliūnas

Recenzentas

Doc. dr. Kastytis Ratkevičius

Projektą atliko

Donatas Sipavičius

KAUNAS, 2016



KAUNO TECHNOLOGIJOS UNIVERSITETAS

Informatikos fakultetas

(Fakultetas)

Donatas Sipavičius

(Studento vardas, pavardė)

Informatikos studijų programa 621I10003

(Studijų programos pavadinimas, kodas)

Baigiamojo projekto „Google“ lietuvių šnekos atpažintuvo efektyvumo įvertinimo tyrimai“

AKADEMINIO SAŽININGUMO DEKLARACIJA

20 16 m. gegužės 20 d.
Kaunas

Patvirtinu, kad mano, **Donato Sipavičiaus**, baigiamasis projektas tema „Google“ lietuvių šnekos atpažintuvo efektyvumo įvertinimo tyrimai“ yra parašytas visiškai savarankiškai ir 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)

TURINYS

Terminų ir santrumpų žodynas	6
Įžanga.....	7
1. Probleminės srities analizė.....	9
1.1. Tyrimo objektas, sritis ir problema	9
1.2. Tyrimo objekto analizė	9
1.3. Tyrimo objekto naudotojų analizė	17
1.4. Esamų problemos sprendimo metodų analizė (Lietuvos ir tarptautiniu mastu).....	17
1.5. Darbo tikslas, uždaviniai ir siekiami privalumai	20
1.6. Siekiamo sprendimo apibrėžimas	20
1.7. Analizės išvados.....	20
2. Sprendimo reikalavimų specifikacija ir projektas, formalus aprašas.....	21
2.1. Reikalavimų specifikacija	21
2.2. Tyrimo metodo aprašas	22
2.3. Sistemos loginė architektūra	23
2.4. Sistemos elgsenos modelis.....	23
2.5. Duomenų modelis	42
3. Sprendimo realizacija ir testavimas	44
3.1. Sprendimo realizacijos ir veikimo aprašas.....	44
3.2. Testavimo modelis, duomenys, rezultatai.....	49
4. Eksperimentinis sprendimo tyrimas.....	50
4.1. Eksperimento planas	50
4.2. Eksperimento duomenys	50
4.3. Eksperimento rezultatai.....	50
4.4. Sprendimo veikimo ir savybių analizė, kokybės kriterijų įvertinimas.....	57
4.5. Sprendimo taikymo rekomendacijos.....	57
5. Išvados	59
6. Literatūra.....	60
7. PRIEDAI.....	62
7.1. Panaudos atvejų aprašas.....	62
7.2. Dalykinės srities esybių klasių modelis	64
7.3. Testavimo atvejai	65
7.4. Detalūs šnekos atpažinimo rezultatai.....	68

Donatas Sipavičius. „Google“ lietuvių šnekos atpažintuvo efektyvumo įvertinimo tyrimai“. Magistro baigiamasis projektas / vadovas doc. dr. Rytis Maskeliūnas; Kauno technologijos universitetas, Informatikos fakultetas.

Mokslo kryptis ir sritis: Informatika (09P), fizinių mokslų sritis

Reikšminiai žodžiai: „Google“ lietuvių šnekos atpažintuvas, šnekos atpažinimas, klaidingų žodžių dažnis (WER), balso ir triukšmo signalų santykis (SNR).

Kaunas, 2016. 77 p.

SANTRAUKA

Šiame darbe atlikti „Google“ lietuvių šnekos atpažintuvo efektyvumo tyrimai. Eksperimentui atlikti buvo pasirinktas metodas, sudarytas iš trijų etapų: pirmiausiai apdoroti visus balso įrašus be triukšmo; apdoroti visus balso įrašus su kelių skirtingų tipų triukšmais, modifikuotais taip, kad būtų gauti tam tikri balso įrašo ir triukšmo signalų santykiai (SNR); po 1 mėn. pakartotinai apdoroti visus balso įrašus be triukšmo ir įvertinti, ar pagerėjo šnekos atpažinimo kokybė. Šnekos atpažinimo kokybei vertinti pasirinkta metrika WER – klaidingų žodžių dažnis.

Analizuojant eksperimento rezultatus pastebėta, kad didžiausią įtaką šnekos atpažinimo kokybei turi balso įrašo ir triukšmo signalo santykis (SNR) bei šnekos tipas (geriausiai atpažįstamos pavienės balso komandos, prasčiausiai – spontaniška šneka). Tuo tarpu tokios charakteristikos, kaip kalbėtojo lytis, kalbėjimo sklandumas, kalbėjimo greitis, kalbėjimo garsumas šnekos atpažinimo kokybei ženklios įtakos neturi.

Sipavičius, Donatas. „Google“ Lithuanian Speech Recognition Efficiency Evaluation Research: Master's thesis in Informatics / supervisor assoc. prof. Rytis Maskeliūnas. The Faculty of Informatics, Kaunas University of Technology.

Research area and field: Informatics (09P), physical sciences

Key words: „Google“ Lithuanian speech recognition, speech recognition, WER (Word Error Rate), signal-to-noise ratio (SNR).

Kaunas, 2016. 77 p.

SUMMARY

This study presents „Google“ Lithuanian speech recognition efficiency evaluation research. For the experiment it was chosen method that consists of three parts: first of all to process all voice records without adding any noise, secondly process all voice records with several different types of noise, modified so as to get some predefined signal to noise ratios (SNR) and at last after one month reprocess all voice records without any additional noise and to assess improvements in the quality of the speech recognition. It was chosen WER (Word Error Rate) metrics for speech recognition quality assessment.

Analyzing the results of the experiment it was observed that the greatest impact on the quality of speech recognition has a signal-to-noise ratio (SNR) and speech type (most recognizable is isolated words, the worst - spontaneous speech). Meanwhile, characteristics such as the gender of the speaker, smooth speech, speech speed, speech volume does not make any significant influence on speech recognition quality.

TERMINŲ IR SANTRUMPŲ ŽODYNAS

ASR	–	automatinis šnekos atpažinimas (angl. <i>Automatic Speech Recognition</i>)
DB	–	duomenų bazė
DNN	–	gilieji neuroniniai tinklai (angl. <i>deep neural networks</i>)
DTW	–	dinaminis laiko skalės kraipymas (angl. <i>Dynamic Time Warping</i>)
GMM	–	Gauso mišinių modeliai (angl. <i>Gaussian Mixture Models</i>)
HMM	–	paslėptieji Markovo modeliai (angl. <i>Hidden Markov Models</i>)
PA	–	panaudos atvejis
PĮ	–	programinė įranga
Sistema	–	eksperimento programinė įranga
SNR	–	signalų / triukšmo santykio matas (angl. <i>signal-to-noise ratio</i>)
WER	–	klaidingų žodžių dažnis (angl. <i>Word Error Rate</i>)

IŽANGA

Šnekos atpažinimo technologijos taikomos daugelyje sričių: įvairiems įrenginiams valdyti, kalbiniam dialogui su kompiuteriu, gyvam (angl. *live*) pokalbiui internetu tarp asmenų, kurie kalba skirtingomis kalbomis (kalbos vertimas iš vienos kalbos į kitą), balso įrašams stenografuoti, kompiuterinėms programoms valdyti ir pan. Šios taikymo sritys yra aktualios ne tik neįgaliesiems, negalintiems fiziškai atlikti tam tikrų veiksmų.

Šnekamosios kalbos efektyvumas yra neginčijamas: nepatyręs naudotojas per minutę gali įvesti maždaug 20 žodžių naudodamasis klaviatūra, 24 užrašyti ranka, o padiktuoti apie 150 [1]. Šnekos technologijų panaudojimas balsu valdomose sąsajose (angl. *voice user interface*) padeda tokiose situacijose, kada naudotojas negali skirti pakankamai dėmesio valdymo įrenginių manipuliavimui, tekstinės ar grafinės informacijos suvokimui arba naudotojo sąsaja yra nepatogi miniatiūrinuose ar pramoniniuose įrenginiuose [1].

Šnekos atpažinimo technologijų ištakos siekia dar 1952 metus, kai „Bell Labs“ pristatė pirmą sėkmingą šnekos atpažintuvą. Tai buvo nuo vieno kalbėtojo priklausomas techninis įrenginys, kuris gebėjo atpažinti skaičius nuo nulio iki devynių. Buvo teigiama, kad šio atpažintuvo tikslumas yra 98 %. Bėgant laikui, šnekos atpažintuvų tikslumas beveik nekito. Toks tikslumo rodiklis buvo pasiektas bandymuose dalyvaujant apmokytiems kalbėtojams ar patiems atpažintuvų kūrėjams, kurie žinojo, ką sakyti atpažintuvui bandymų metu [2].

2014 m. pabaigoje buvo išleista „Google“ lietuvių šnekos atpažintuvo testinė versija. 2015 m. šis atpažintuvas lietuvių šnekai buvo oficialiai pristatytas. „Google“ teigia, kad jo šnekos atpažintuvo WER rodiklis (klaidingų žodžių dažnis) siekia 8 % [3].

Naudojantis šnekos atpažintuvais dažnai susiduriama su problema, kad dalis žodžių visai neatpažįstami ar atpažįstami neteisingai. Tai lemia šnekos atpažinimo kokybės priklausomybė nuo šių veiksnių:

- Atpažįstamos kalbos ypatumai (pvz., lietuvių kalboje žodžių ryšiai sakinyje reiškiami keičiant galūnes; pakeitus žodžio galūnę, jis įgyja kitokią prasmę);
- Specifinė terminologija, būdinga tik tam tikrai taikymo sričiai;
- Kalbėtojo tarmė, kalbėjimo garsumas, raiškumas, įvairūs šnekos sutrikimai;
- Foninio triukšmo lygis (garsumas) ir jo pobūdis (fabrikas, riedantis traukinys, prekybos centras ir pan.).

Renkantis šnekos atpažintuvą būtina atsižvelgti į jo charakteristikas, tokias kaip: šnekos tipas (atpažįstamos tik pavienės balso komandos, keleto žodžių seka, rišli šneka ar spontaniška šneka), žodyno dydis (galimų atpažinti žodžių kiekis), priklausomumas / nepriklausomumas nuo kalbėtojo (atpažįstamų kalbėtojų skaičius), ryšio kanalai ir aplinka (atsparumas ryšio kanalo iškraipymams ir aplinkos triukšmams).

Šio darbo tyrimo objektas – „Google“ lietuvių šnekos atpažintuvas. Tyrimo sritis – šnekos atpažinimo sistemų efektyvumo įvertinimo metodai.

Šio darbo tikslas – įvertinti „Google“ lietuvių šnekos atpažintuvo efektyvumą.

Darbo uždaviniai:

- Išanalizuoti šnekos atpažinimo sistemų efektyvumo įvertinimo metodus.
- Pasirinkti tinkamą šnekos atpažinimo sistemų efektyvumo įvertinimo metodą.
- Sukurti prototipą pasirinkto metodo efektyvumui įvertinti.
- Atlikti eksperimentinius tyrimus „Google“ šnekos atpažintuvo efektyvumui įvertinti.
- Nustatyti, kokie faktoriai (kalbėtojo savybės, kalbėjimo savybės, foninis triukšmas, šnekos tipas) daro įtaką šnekos atpažinimo kokybei.

Šiame darbe buvo atlikti „Google“ lietuvių šnekos atpažintuvo efektyvumo įvertinimo tyrimai. Eksperimentas buvo atliktas su 63299 balso įrašais, 3318 skirtingų frazių (įgarsintų tekstų). Visus šiuos balso įrašus „Google“ šnekos atpažintuvas apdorojo mažiausiai (neskaičiuojant eksperimento programinės įrangos testavimo) 18 kartų: 1 kartą be triukšmo, 16 kartų su triukšmu (4 skirtingi triukšmai, 4 skirtingi SNR) ir dar 1 kartą be triukšmo praėjus 1 mėnesiui nuo pirmo balso įrašų apdorojimo be triukšmo pabaigos. Eksperimento metu vertinta WER reikšmė ir balso įrašų, kuriems nebuvo gautas šnekos atpažinimo rezultatas, skaičius.

Eksperimentinio tyrimo rezultatai parodė, kad „Google“ šnekos atpažintuvo kokybė dar nėra tokia, kad jį būtų galima taikyti kuriant sistemas ar įvairias programėles, kuriose naudojama šnekos įvestis – vidutinė WER reikšmė, kai „Google“ šnekos atpažintuvas balso įrašus apdorojo be triukšmo, yra 40,74 %. Iš 63299 „Google“ šnekos atpažintuvui pateiktų apdoroti balso įrašų 20784 balso įrašams išvis nebuvo pateikti šnekos atpažinimo rezultatai, o tai sudaro net 32,83 % visų balso įrašų. Pakartojus eksperimentą po 1 mėn., buvo gauti šiek tiek prastesni rezultatai: vidutinė WER reikšmė 0,08 % didesnė ir nebuvo gautas šnekos atpažinimo rezultatas 84 balso įrašais daugiau, nei pirmojo eksperimento etapo metu. Taigi, šnekos atpažinimo kokybė nepagerėjo net ir tuos pačius balso įrašus apdorojus mažiausiai 18 kartų (įgarsintos frazės balso įrašuose pasikartojo dar daugiau kartų).

Šis dokumentas sudarytas iš 7 skyrių. Pirmame skyriuje atlikta šnekos atpažintuvų ir galimų metrikų jiems vertinti analizė, pateiktas siekiamo sprendimo aprašas. Antrame skyriuje pateikta kuriamos eksperimento programinės įrangos reikalavimų specifikacija. Trečiame skyriuje aprašyta sukurto sprendimo realizacija, jo testavimas ir testavimo rezultatai. Ketvirtame skyriuje pateiktas eksperimento planas ir pristatomi atlikto eksperimento rezultatai; taip pat pateikta sprendimo veikimo ir savybių analizė, kokybės kriterijų įvertinimas bei sprendimo taikymo rekomendacijos. Darbo išvados pateiktos penktame skyriuje. Šeštame skyriuje pateiktas literatūros sąrašas. Septintame skyriuje pateikiami visi priedai.

1. PROBLEMINĖS SRITIES ANALIZĖ

Analizės tikslas – išsiaiškinti, kokie metodai naudojami vertinant šnekos atpažinimo sistemos efektyvumą ir tinkamumą naudoti, kokios galimos vertinimo metrikos ir gautų rezultatų interpretacijos.

1.1. Tyrimo objektas, sritis ir problema

Magistro darbo tyrimo objektas – „Google“ lietuvių šnekos atpažintuvas.

Tyrimo sritis – šnekos atpažinimo sistemų efektyvumo įvertinimo metodai.

Kaip jau minėta, „Google“ lietuvių šnekos atpažintuvas atsirado visai neseniai. Prieš panaudojant šnekos atpažintuvą tam tikroje taikymo srityje būtina įvertinti jo galimybes ir efektyvumą: kokioje aplinkoje (jokio foninio triukšmo, silpnas foninis triukšmas, stiprus foninis triukšmas), kaip kalbant (rišli kalba, spontaniška kalba ir pan.) atpažintuvas veikia tiksliausiai.

1.2. Tyrimo objekto analizė

Šnekos atpažinimas – balsu tariamų vienetų (frazių, žodžių, jų sekų) automatinis nustatymas, t. y. žmogaus šnekos pavertimas tekstu (akustiniams signalams parenkant žodžių atitikmenis) naudojant kompiuterį. Šnekos atpažinimas tiriamas apie šešiasdešimt metų siekiant panaudoti žmogaus šneką automatizuojant uždavinius, kuriuos žmogus gali atlikti sąveikaudamas su kompiuteriu. Automatinė šnekos atpažinimo technologija suteikė galimybę kompiuteriui „suprasti“ žmogaus šneką ir vykdyti išstartas komandas [4, 5].

Automatinis šnekos atpažinimas yra taikomas įvairiose žmonių veiklose [6]:

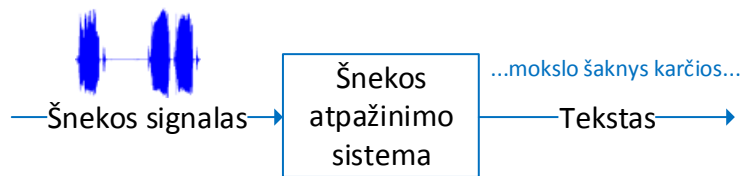
- klientų aptarnavimas didelėse kompanijose – informacijos teikimas, skambučių priėmimų centrų automatizavimas sumažinant laukimo laiką;
- užklausomis grįstoms sistemoms, kurios teikia atnaujintą informaciją apie keliones, akcijų kainų pokyčius, orus;
- automatizuoto teksto rinkimas diktuojant – teksto surinkimas nereikalauja surinkimo įgūdžių ir sunaudojama mažiau žmonių darbo resursų;
- pasinaudodami automatinio šnekos vertimu, žmonės gali susikalbėti ir nemokėdami reikiamos kalbos.

Automatinis šnekos atpažinimas (angl. *Automatic Speech Recognition*, sutr. ASR) gali būti taikomas ir kaip pagalbinė bendravimo priemonė neįgaliems žmonėms, kuriems rinkti tekstą klaviatūra dažniausiai yra sudėtinga, žmonėms, turintiems specifinius gebėjimo skaityti sutrikimus (disleksija), gebėjimo rašyti sutrikimus (disgrafija) ar kitokių sunkumų manipuliuojant teksteine forma [7]. Aktyvūs ASR tyrimai vykdomi automobilių pramonės, medicinos, farmacijos, telekomunikacijų, multimedijos ir kitose srityse. „The Telegraph“ duomenimis, pasaulyje yra daugiau nei 70 milijonų automobilių, kuriuose įdiegtos balsu valdomos sistemos (pvz., automatinio stabdymo ar statymo technologijos); sukurti balsu valdomi įrenginiai, matuojantys kraujo spaudimą, gliukozės ir insulino kiekį kraujyje; išmaniuosiuose telefonuose įdiegti virtualūs balso asistentai, galintys pateikti atsakymus į balsu užduotus klausimus bei vykduantys pateiktas užduotis daugiau kaip 40 kalbų (pvz., navigacijos ar tekstinių žinučių rašymo) [7].

Šnekos atpažinimo procesą galima išskirti į tris pagrindinius etapus [8]:

- Duomenų įvedimą;
- Požymių išskyrimą;
- Atpažinimą.

Kitai variant, šnekos atpažinimo procesas prasideda šnekos signalo įvedimu į sistemą ir baigiasi tekstinio atitikmens akustiniam signalui pateikimu. Apibendrintas šnekos atpažinimo kompiuteriu procesas pavaizduotas 1.1 pav.



1.1 pav. Apibendrintas šnekos atpažinimo procesas

Pirmiausiai, mikrofonu įvedamas šnekos signalas ir paverčiamas skaitmeniniu pavidalu. Antrame etape fiksuotais laiko momentais imami tam tikro ilgio šnekos signalo kadrai, kuriuose skaičiuojami tam tikrų dažnio juostų energiją charakterizuojantys parametrai: laikas, dažnis ir amplitudė [8]. Šie parametrai naudojami trečiame (atpažinimo) etape, kuriame nustatoma, kokios fonemos ar žodžiai yra šnekos signale [8]. Fonema – mažiausias kalbos garsinės sistemos vienetas, skiriantis tos kalbos žodžius ar morfemas¹.

Šnekos atpažinimas yra labai sudėtingas uždavinys. Uždavinio sudėtingumą lemia šios priežastys [8]:

1. Keletą kartų išstarto tam tikro garso akustinė realizacija labai skiriasi, net jei jį ištarė tas pats diktorius ir tame pačiame žodyje;
2. Kalbėjimo greitis gali labai kisti, todėl skiriasi kelių to paties žodžių akustinių realizacijų ilgis. Kintant žodžių ilgiui atskirų garsų ilgis kinta netiesiškai;
3. Garso akustinė realizacija priklauso nuo gretimų garsų (tai vadinama koartikuliacija);
4. Kalbėjimo sraute nėra aiškių garsų ar žodžių ribų;
5. Kiekvieno žmogaus tartis yra skirtinga, todėl reikalingas arba apmokymas konkrečiam diktoriui, arba sistema kūrimo metu turi būti apmokyta su kuo didesniu diktorių skaičiumi;
6. Jei kuriama atpažinimo sistema remiasi žodžių atpažinimu, žodžių etalonų skaičius gali būti pernelyg didelis;
7. Kalbėjimo sraute gali būti ir nekalbinių fragmentų (pvz., kosulys), kuriuos reikia atskirti ir pašalinti;
8. Taikant praktikoje papildomų problemų sukelia foninis triukšmas.

1.2.1. Šnekos atpažinimo sistemos, jų veikimo pagrindai

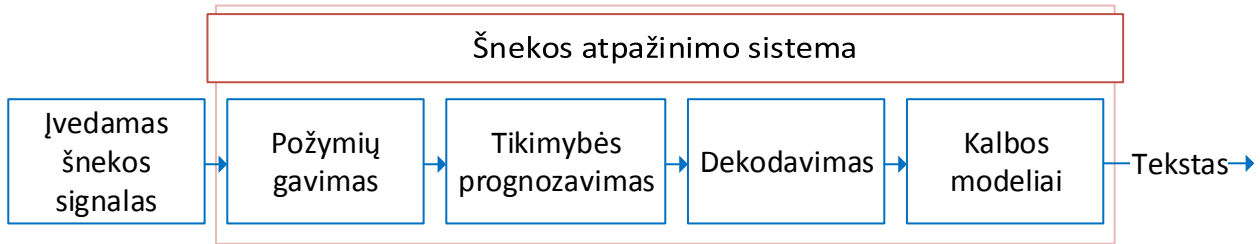
Šnekos atpažinimo sistemos gali būti suskirstytos į kelias skirtingas klases pagal tai, kiek žodžių jos gali atpažinti [1]:

- Pavienių balso komandų (angl. *isolated words*) atpažinimo sistemos. Atpažįstama frazė gali būti sudaryta ir iš daugiau nei vieno žodžių (pvz., sakinys), tačiau kalbant reikia po kiekvieno ištarimo daryti trumpas pauzes (pvz., vienas PAUZĖ du PAUZĖ ir pan.).
- Keleto žodžių sekos (angl. *connected words*) atpažinimo sistemos. Tokios sistemos panašios į pavienių balso komandų atpažinimo sistemas, tačiau tylos tarpai tarp žodžių yra gerokai mažesni (pvz., skaičių sekos atpažinimas).
- Rišlios šnekos (angl. *continuous speech*) atpažinimo sistemos. Tokios sistemos leidžia naudotojui kalbėti beveik natūraliai. Dažniausiai taikoma diktuojamam tekstui atpažinti.
- Spontaniškos šnekos (angl. *spontaneous speech*) atpažinimo sistemos, gebančios apdoroti tokias natūralios kalbos savybes, kaip vienodai tariamus, tačiau skirtingą prasmę (skirtingam kontekste) turinčius žodžius, žodžius, kurie be konteksto neturi prasmės, ir netgi mikčiojimą. Tokią šneką atpažinti labai sunku, todėl dauguma šio tipo šnekos atpažinimo sistemų tebėra prototipo stadijoje.

Bet kokios automatinės šnekos atpažinimo sistemos tikslas – atpažinti išstartus žodžius kiek įmanoma tiksliau, t. y. teoriškai neblogiau, nei tai daro žmogus, nepriklausomai nuo diktorius balso charakteristikų, žodyno dydžio, duomenų perdavimo sąlygų ir t. t. [1] Tačiau dauguma tokio tipo sistemų pasiekia atpažinimo tikslumą didesnę nei 90 % tik prie tam tikrų apribojimų (pvz., keleto skaičių pavadinimų atpažinimo tikslumas per mikrofona (mažas žodynas, aplinka be didesnio pašalinio triukšmo) gali siekti 99 %) [1].

¹ Mažiausia žodžio dalis (šaknis, priešdėlis, priesaga, galūnė), turinti leksinę ar gramatinę reikšmę.

Detalus šnekos atpažinimo procesas pavaizduotas 1.2 pav.



1.2 pav. Detalus šnekos atpažinimo procesas [1]

Pirmasis blokas sudarytas iš akustinės aplinkos bei perdavimo aparatūros (mikrofono, signalo stiprintuvo, filtrų, analoginio keitiklio ir t. t.). Jų kokybė (triukšmas, mikrofono pozicija, tipas ir t. t.) gali smarkiai įtakoti tolesnius procesus [1].

Antrasis blokas (požymių gavimo posistemė) skirtas rasti specifinius šnekos signalo požymius bei akustines reprezentacijas, kad būtų galima kokybiškai atskirti šnekos garsų klases ir efektyviai atmesti nereikalingas variacijas [1].

Trečiasis ir ketvirtasis blokai iliustruoja akustinių požymių lyginimo operacijas [1]. Beveik visose ASR sistemose spektrinės ar keprinės šnekos signalo atskaitos (požymiai) skaičiuojamos tam tikrais intervalais, pavyzdžiui, 100 kartų per sekundę [1]. Norint atpažinti šneką, gauti požymiai lyginami su gautaisiais iš apmokymo duomenų, naudojant tam tikrą panašumo ar atstumo matą [1]. Kiekvienas iš šių palyginimų – lokalusis matas [1]. Globalusis matas – tai geriausios žodžių sekos paieška, kuri nustatoma apjungiant daugelį lokalių matų (pvz., ieškomas visas žodis pagal fonemų rinkinį) [1]. Vietinis atitikmuo paprastai nepateikia vieno geriausio pasirinkimo, bet tam tikrą grupę, atitinkančią galimus garsus [1]. Dar viena dekodavimo bloko funkcija yra laikinių iškraipymų, kurie atsiranda normalioje šnekoje, kompensavimas [1]. Pavyzdžiui, balsiai paprastai sutrumpinami sparčiai kalbant, o priebalsiai lieka tokio paties ilgio [1].

Atpažinimo algoritmas (penktasis blokas) paprastai būna pagrįstas statistiniais modeliais [1]. Plačiai paplitę paslėpti Markovo modeliai (angl. *Hidden Markov Models*, sutr. HMM) [1]. HMM paprastai apibūdinamas kaip baigtinio stochastinio būsenų skaičiaus automatas (angl. *stochastic finite state automaton*) ir priimama, kad jis bus sudarytas iš baigtinio galimų būsenų rinkinio, kur kiekviena būseną atitinka tam tikrą tikimybinį pasiskirstymą (panašumo atveju – tikimybinio tankio funkcija) [1]. Idealiu atveju turėtų būti sudarytas HMM kiekvienam žodžiui, tačiau praktiškai tai sunkiai įgyvendinama [1]. Sakinys modeliuojamas kaip žodžių seka [1]. Kai kurios ASR sistemos veikia žodžio lygyje, tačiau stambesnio žodyno – žodžio dalių lygyje, taip sumažinant reikalingų parametrų (ir tuo pačiu apmokymo duomenų) skaičių [1]. Kiekvienas žodis gali būti suskaidytas į tam tikrą grupę akustinių vienetų, pavyzdžiui, fonemas [1]. Viena ar daugiau HMM būsenų naudojama sumodeliuoti fonemą atitinkančią šnekos segmentą [1]. Žodžių modeliai sudaromi iš fonemų modelių grandinės (apribota pagal žodyną), o sakinių modeliai – iš žodžių modelių grandinės (apribota gramatikos taisyklėmis) [1].

ASR sistemos gali būti skirstomos pagal [2]:

- Šnekos greitį (angl. *speed of speech*). Pertraukiamos šnekos (angl. *discrete speech*) atpažintuvas atpažįsta atskirus žodžius, todėl tarp žodžių/komandų turi būti pauzės. Rišlios šnekos (angl. *continuous speech*) atpažintuvas atpažįsta šneką ir leidžia naudotojui laisvai diktuoti ir sakyti komandas.
- Taikymą (angl. *type of application*). Valdymo ir kontrolės (angl. *command-and-control*) atpažintuvas atpažįsta tik trumpas komandas (pvz., „atidaryk failą“, „uždaryk langą“ ir pan.). Diktavimo (angl. *dictation*) atpažintuvas atpažįsta sakinius, todėl gali būti naudojamas el. laiško turinio diktavimui ir pan.
- Priklausomybę nuo naudotojo (angl. *user-dependency*). Priklausomas nuo kalbėtojo (angl. *speaker dependent*) atpažintuvas gerai veikia tik su konkrečiu naudotoju, su kuriuo buvo apmokomas atpažintuvas. Nepriklausomas nuo kalbėtojo (angl. *speaker independent*) atpažintuvas veikia su bet koku naudotojų skaičiumi be apmokymų.

- Pagrindinį kalbos modelį (angl. *underlying language model*).
- Taikymo sritį (angl. *field of application*). Kompiuteriams skirti (angl. *PC-based*) atpažintuvai skirti dirbti išimtinai su multimedijos programomis personaliniuose kompiuteriuose ir mažu foniniu triukšmu. Telefonams skirti (angl. *telephone-based*) atpažintuvai skirti atpažinti komandas, sakomas telefonu. Dauguma telefonams skirtų atpažintuvų dirba vidutiniškai triukšmingoje aplinkoje. Triukšmui atsparūs (angl. *noise robust*) atpažintuvai skirti veikti ypatingai triukšmingoje aplinkoje (pvz., automobilyje).

1.2.2. Šnekos atpažinimo sistemų charakteristikų įtaka šnekos atpažinimui

Standartinės ASR charakteristikos yra šios: šnekos tipas, žodyno dydis, priklausomumas/nepriklausomumas nuo kalbėtojo, ryšio kanalai ir aplinka [9]. ASR sistemas įmanoma lyginti, jei šie parametrai sutampa [9].

Žodyno dydis. Žodynas apibrėžia ASR sistemos atpažįstamus žodžius. Teigiama [10], kad atpažinimo sudėtingumas didėja logaritmiškai, augant žodžių skaičiui žodyne. ASR sistemas galima klasifikuoti pagal žodyno dydį – mažos, vidutinės ir didelės apimties žodynai [9]. Literatūroje žodyno dydis skaičiais apibrėžiamas labai įvairiai ir sąvokos siejimas su skaitine išraiška nėra nusistovėjęs. Viename literatūros šaltinyje mažos apimties žodynas yra iki 20, o didelės virš 20 000 žodžių [11]. Kitame šaltinyje teigiama, kad mažos apimties žodyne turėtų būti 1–99 žodžiai [10], vidutinės – 100–999, o didelės – virš 1000 žodžių. Daug paprasčiau ASR sistemas priskirti vienai ar kitai kategorijai pagal sistemos taikymo sritį. Mažos apimties žodyno sistemos naudojamos tokių uždavinių, kaip kreditinės kortelės ar telefono numerio skaitmenų atpažinimo užduotims; vidutinio – eksperimentinėms sistemoms laboratorijose, modeliuojant rišlios šnekos atpažinimą; didelės apimties žodyno sistemos naudojamos kuriant komercinius produktus, realizuojant teksto diktavimo ar korespondencijos uždavinius [9].

Priklausomumas/nepriklausomumas nuo kalbėtojo. Viena iš ASR sistemos svarbių charakteristikų – atpažįstamų kalbėtojų skaičius [9]. Sistema, sukurta vieno kalbėtojo poreikiams, vadinama nuo kalbėtojo priklausoma sistema; jei ja gali naudotis daugiau kalbėtojų – tai nuo kalbėtojo nepriklausanti sistema [9]. Pirmosios sistemos didesnis atpažinimo tikslumas paaiškinamas mažesnėmis variacijomis šnekos signaluose [9]. Universalumo prasme antroji sistema yra geresnė, nes nereikalauja pakartotinio akustinių modelių kūrimo naujam kalbėtojui [9]. Norint padidinti atpažinimo tikslumą atliekama kalbėtojo adaptacija esamai sistemai, t. y. sistemos akustiniai modeliai modifikuojami pagal kalbėtojo balso charakteristikas [9]. Ši operacija gali vykti darbo su sistema sesijos metu (angl. *on-line*) arba kalbėtojas prieš pradėdamas naudotis sistema gali būti paprašytas įrašyti garsinę informaciją pagal pateiktą tekstą (angl. *off-line*) [9].

Ryšio kanalai ir aplinka. Vienas iš pagrindinių reikalavimų ASR – atsparumas (robastiškumas) ryšio kanalo iškraipymams ir aplinkos triukšmams [9]. Tiek ryšio kanalai, tiek aplinka šalia šnekos signalo įkomponuoja papildomos informacijos, kuri yra ne tik nenaudinga, bet ir iškraipo patį šnekos signalą [9]. Tokią informaciją įprasta vadinti triukšmu. Triukšmą galima suklasifikuoti į: biuro aplinkos, kanalų, transporto ir pramonės [12]. Galimas ir daug smulkesnis klasifikavimas pagal jo pasiskirstymą laiko ir dažnių srityse: periodiniai, impulsiniai, plačiajuosčiai ir pan. [9] Triukšmui signale išmatuoti naudojamas signalo/triukšmo santykio matas SNR (angl. *signal-to-noise ratio*) ir jo modifikacijos [9].

Teigiama [12], kad ASR sistemų darbingumas nenukenčia SNR mažėjant iki 25 dB, bet triukšmo lygiui didėjant atpažinimas blogėja. Su triukšmu kovojama įvairiais būdais: jis šalinamas, naudojami atsparūs požymiai, akustiniai modeliai adaptuojami triukšmui [9].

Šnekos tipai. Pavienės balso komandos (angl. *isolated words*). Pavienių balso komandų atpažintuvui reikalaujama pateikti ištariamą (tai gali būti atskiras žodis arba frazė), iš abiejų pusių ribojamą kitų signalo charakteristikų (tyla, žemo lygio triukšmas) [9]. Šie atpažintuvai veikia klausymo/neklausymo režimais, realizuojamais signalo galo taškų nustatymo algoritmais (angl. *an endpoint detection algorithm*) [9]. Taip identifikuojami minėtų signalų charakteristikų pasikeitimo momentai [9]. Iš kalbėtojo reikalaujama tarp ištariamų daryti trumpas pauzes, kiekvieną žodį tariant izoliuotai nuo kitų – iš to kilo ir algoritmų pavadinimas [9]. Šis kalbėjimo būdas patogus, jei reikalinga atpažinti atskiras komandas, bet netinkamas kitoms atpažinimo sistemų taikymo sritims [9].

Keleto žodžių seka (angl. *connected words*) yra pereinamasis variantas nuo pavienių balso komandų prie rišlios šnekos [9].

Rišli šneka (angl. *continuous speech*) paprastai būna skaitytinė [9]. Rišlios šnekos atpažintuvams keliamas reikalavimas – atpažinti šneką, kurioje tarp žodžių ar frazių pauzės gali būti arba dažniausiai jų nėra [9]. Kalbėtojas gali kalbėti beveik natūraliai. Panaudojimo sritis – diktavimas kompiuteriui [9].

Spontaniška šneka (angl. *spontaneous speech*) – tai kasdienė šneka. Spontaninį kalbėjimo būdą galima apibūdinti jo savybėmis [9]. Tai nėra sklandi šneka, dažnai joje yra ilgų pauzių, žodžių fragmentų (pradėtų ir nebaigtų žodžių), neteisingų ištarių, mikčiojimų, pasikartojimų, sintaksiškai neteisingų sakinių [9]. Tai paaiškinama tuo, kad kalbėdamas žmogus kartu mąsto [9]. Dažnai tokia šneka lydima papildomų triukšmo šaltinių, kaip juokas, čepsėjimas ir pan. [9] Spontaniška šneka pasižymi ir dideliu skaičiumi žodžių, nesančių sistemos žodyne, dėl ko ją atpažinti tiek akustiškai, tiek gramatiškai yra sudėtinga [9].

Pavienių balso komandų atpažinimas. Pavienių balso komandų atpažinimas palyginti su rišlios ar spontaniškos šnekos atpažinimu laikomas lengvesne užduotimi. Nuo jų prasideda ir šnekos atpažinimo raida: nuo paprastesnių dalykų einama prie sudėtingesnių – nuo pavienių balso komandų (izoliuotų žodžių) atpažinimo prie rišlios šnekos atpažinimo [9]. Standartinis tokios atpažinimo sistemos modelis susidėjo iš požymių išskyrimo, pavyzdžių palyginimo ir sprendimo priėmimo blokų.

Šis šnekos atpažinimo modelis buvo praktiškas dėl kelių priežasčių:

1. invariantiškumo žodynams, naudotojams, pavyzdžių palyginimo algoritmams, sprendimo taisyklėms,
2. lengvo įgyvendinimo ir gero praktinio veikimo [13].

Tokių sistemų etaloniniais ir testiniais pavyzdžiais vadinama požymių pavidalo šnekos signalo spektrinė informacija [9]. Pavyzdžiams palyginti ji taikoma tiesiogiai skaičiuojant atstumą dinaminio laiko skalės kraipymo būdu (angl. *Dynamic Time Warping*, sutr. DTW). Laiko ašių iškraipymai atliekami taikant dinaminį programavimą [9]. Pagal tai, ar atpažinimo sistema priklausoma/nepriklausoma nuo kalbėtojo, skiriasi etalonų formavimo procesas – mokymas ir sprendimo taisyklė [9]. Jei atpažinimo sistema priklausoma nuo kalbėtojo, užtenka vieno ar dviejų kalbėtojo ištarių spektrinės informacijos [9]. Jei atpažinimo sistema nepriklauso nuo kalbėtojo, atliekamas vektorinis kvantavimas, t. y. daugelio kalbėtojų ištarių spektrinės informacijos pavyzdžiai suskirstomi į mažesnių pavyzdžių skaičių [9]. Šio tipo sistema yra veikama tokių veiksnių, kaip: žodyno sudėtingumas, veikiantis atpažinimo tikslumą; perdavimo kanalai, darantys įtaką atpažinimo tikslumui ir šnekos signalo galo taškų nustatymui; pačios sistemos sudėtingumas, turintis įtakos sistemos atsako laikui [9]. Tokio tipo atpažinimo sistemų, besiskiriančių žodyno dydžiu (641–5000 žodžių) ir jo struktūra (žodis neskaidomas ar skaidomas į mažesnius vienetus), priklausomumu/nepriklausomumu nuo kalbėtojo, įvairiais kitais parametrais ir modifikacijomis (kalbos modelio naudojimu) buvo sukonstruota daug [9, 14, 15]. Šių pavienių balso komandų atpažinimo sistemų tikslumas svyravo tarp 94 % ir 96,3 % [9].

Kitas pavienių balso komandų atpažinimo etapas susijęs su netiesioginiu spektrinės informacijos panaudojimo atpažinimui būdu [9]. Didelės apimties garsynų atsiradimas paskatino naudoti statistinius metodus ir naują spektrinės išraiškos būdą. Pavienių balso komandų atpažinimui pradedami naudoti HMM [16], neuroniniai tinklai [17]. Atpažinimo tikslumas yra įvairus 40–99,3 % [9].

Lietuvoje pavienių balso komandų atpažinimas taip pat yra dviejų krypčių: tiesiogiai naudojant spektrinę informaciją šnekos pavyzdžiams palyginti [18] ir naudojant statistinius modelius spektrinės informacijos apibendrinimui [19, 20].

Rišlios šnekos atpažinimas. Rišlios šnekos atpažinimo sistemos lyginant jas su pavienių balso komandų atpažinimo sistemomis yra sudėtingesnės. Pavienių balso komandų atpažinimo sistemose lengviau išskirti žodžio pradžią ir pabaigą, nes paprastai juos žymi pauzės. Rišlios šnekos atpažinimo sistemose šios ribos yra neaiškios, nežinomos ir jų nustatymas yra viena iš papildomų užduočių atpažinimo sistemai [9]. Antras skirtumas yra žodyno dydis, nes rišlios šnekos atpažinimo sistemos neįsivaizduojamos be didelių žodynų (angl. *Large Vocabulary Continuous Speech Recognition*, sutr. LVCSR), todėl šnekos atpažinimo tikslumui didinti naudojami kalbos modeliai [9].

Holmes [21] pateikia 1996–2000 metų šiuolaikinių rišlios ir spontaniškos šnekos atpažinimo sistemų tikslumą, kuris svyruoja tarp 69,6 % ir 98 %. Panašus rišlios šnekos atpažinimo tikslumas

74,5–95,6 % buvo dar 1992–1995 metais [22]. Tai rodo, kad rišlios šnekos atpažinimas netobulėja, išlieka dar vis toje pačioje stadijoje [9].

Rišlios šnekos atpažinimui vien žodžių kalbos vienetai nėra naudojami dėl mažų tokių modelių mokymo imčių. Atskirai sukurti žodžių akustiniai modeliai gali būti integruoti į galutinę akustinių modelių aibę [9].

1.2.3. Šnekos atpažinimo metodai

Šnekos signalo atpažinimo metodai yra skirstomi į tris grupes [23, 5, 24]:

- Akustiniai – fonetiniai metodai;
- Pavyzdžiais grįsti metodai;
- Dirbtinio intelekto atpažinimo metodai.

Akustiniai fonetiniai metodai yra grįsti akustine fonetine teorija ir daro prielaidą, kad šnekamojoje kalboje yra baigtinis fonetinių vienetų skaičius ir kad šie fonetiniai vienetai gali būti aprašomi šnekos signalo laiko arba spektro savybėmis [6]. Pirmajame akustinio fonetinio metodo etape vykdomas segmentavimas, kurio metu šnekos signalas yra suskirstomas į atkarpas, kur signalo akustinės savybės atstovauja vieną (arba kelis) fonetinį vienetą (arba klases) [6]. Tada pagal akustines savybes kiekvienam segmentui yra priskiriama fonetinė klasė [6]. Antrajame etape iš gautų fonetinių klasių sekos nustatomas žodis (ar žodžių seka), kuris atitinka šnekos atpažinimo uždavinio apribojimus, pavyzdžiui, žodžiai turi atitikti žodžius, esančius žodyne, žodžių seka atitinka sintaksines taisykles bei semantinę reikšmę [6].

Pavyzdžiais grįstus metodus sudaro du etapai: apmokymo ir atpažinimo [6]. Apmokymo metu sukuriamas žodynas, kurį sudaro apmokymo pavyzdžių rinkinys [6]. Šie apmokymo pavyzdžiai gali būti šablonai arba modeliai, reprezentuojantys statistines pavyzdžių požymių charakteristikas [6]. Atpažinimo metu nežinomas pavyzdys yra lyginamas su apmokymo pavyzdžiais [6]. Remiantis gautais atstumų rezultatais, nežinomas pavyzdys yra priskiriamas tam tikrai klasei [6]. Pavyzdžiais grįstų metodų konstravime išskiriami du atvejai: pirmuoju atveju naudojamas atstumo matas – nežinomas pavyzdys priskiriamas tai klasei, kuriai gautas atstumas yra mažiausias; antruoju atveju išskiriami statistiniai metodai, kurie remiasi tikimybiniais modeliais [6].

Dirbtinio intelekto atpažinimo metodai grindžiami akustinių fonetinių metodų ir pavyzdžiais grįstų metodų idėjomis [6]. Dirbtinio intelekto atpažinimo metodai remiasi aspektais, pagal kuriuos žmonės, pasitelkę savo intelektą, vizualizuoja, analizuoja, priima sprendimus apie akustinius požymius [6]. Pavyzdžiui, šnekos segmentavimui ir klasių priskyrimui naudojami akustiniai fonetiniai metodai, tuo tarpu dirbtiniai neuroniniai tinklai taikomi apmokymui [6].

Dirbtinio intelekto atpažinimo metodai

Pastaraisiais metais šnekos atpažinimo bendruomenė vėl pradėjo diskutuoti apie neuroninius tinklus, kurie buvo labai populiarūs 1980 – 1990 metais, bet savo savybėmis ženkliai nepralenkė labai sėkmingos HMM ir akustinių modelių, pagrįstų Gauso mišiniais, kombinacijos [25]. Trys pagrindiniai faktoriai, kurie turėjo įtakos neuroninių tinklų, kaip aukštos kokybės akustinių modelių, atsiradimui [25]:

- Darant tinklus gilesniais, jie tampa galingesniais – čia atsiranda sąvoka „gilieji neuroniniai tinklai“ (DNN);
- Protingas svorių inicijavimas [26, 27, 28, 29] ir žymiai greitesnės techninės įrangos naudojimas leidžia efektyviai apmokyti DNN;
- Didelio kiekio nuo konteksto priklausomų įvesties vienetų naudojimas [30, 31, 32, 33, 34] ženkliai pagerina DNN spartą.

Gilusis mokymas (angl. *deep learning*) kaip giliųjų neuroninių tinklų mokymo būdas tapo populiarus 2006 metais [35, 36]. Jis buvo labai sėkmingas su akustiniais modeliais, taikant juos šnekos atpažinime (ypatingai didelės apimties užduotims), ir ši sėkmė labiausiai priklausė nuo „back-propagation“ algoritmo su gana standartiniu „feed-forward“ daugiasluoksnių neuroninių tinklų panaudojimu [25]. Be pagerintų procedūrų naudojimo didelę įtaką DNN sėkmei turėjo tai, kad buvo galimybė naudoti daugiau skaičiavimo resursų, daugiau duomenų mokymui, taip pat buvo naudojama geresnė programinė įranga [25]. Tačiau esminis lūžis akustiniame modeliavime buvo generatyvinių,

sluoksnis po sluoksnio (angl. *layer-by-layer*) apmokymo metodų naudojimas svorių inicijavimui prieš paleidžiant diskriminacines „back-propagation“ mokymo procedūras [25]. Vėlesni tyrimai atskleidė, kad generatyvus apmokymas nebūtinai yra didelis mokymo duomenų kiekis [25].

Pirmieji DNN, kurie turėjo ženklų pranašumą lyginant su Gauso mišinių modeliais (GMM), taikomais akustiniame modeliavime, tarpusavyje mažai skyrėsi, bet jų mokymas buvo toks lėtas, net ir su grafikos apdorojimo blokais (angl. *Graphics processing unit*, sutr. GPU), kad nebuvo įmanoma atlikti išsamių tyrimų ir išsiaiškinti, kurie aspektai DNN šnekos atpažinime daro tokiu sėkmingu [25].

Paskutinius kelerius metus tiriant DNN naudojimą šnekos atpažinime buvo pastebėta, kad [25]:

- DNN gerai veikia su didelio žodyno rišlios šnekos atpažinimu (angl. *large vocabulary continuous speech recognition*, sutr. LVCSR);
- Net kai naudojamas sluoksnis po sluoksnio apmokymas, yra daug alternatyvų panaudoti ribotas Bolcmano mašinas (angl. *Restricted Boltzmann Machines*) kiekvieno sluoksnio apmokymui;
- Nuo kalbėtojo priklausomi metodai turi stebėtinai mažą pranašumą prieš nuo kalbėtojo nepriklausomus DNN;
- DNN gerai atpažįsta šneką su triukšmu;
- Standartinių logistinių neuronų naudojimas yra praktiškas, bet neoptimalus. DNN mokosi žymiai greičiau, jei naudojami rektifikuoti tiesiniai vienetai (angl. *rectified linear units*);
- DNN architektūra gali būti naudojama daugiafunkciniam mokymui skirtingais būdais ir DNN yra gerokai efektyvesni nei GMM keičiant užduoties svarbą, kai norima pagerinti susijusių užduočių vykdymo spartą.

1.2.4. Šnekos atpažinimo vertinimo metodai

Kadangi nėra jokių standartinių metodų, leidžiančių įvertinti ASR kokybę, publikuojami rezultatai negarantuoja deklaruojamos ASR sistemos kokybės bet kokioje aplinkoje, išskyrus tą aplinką, kurioje buvo atlikti bandymai [2]. Norint priimti svarų sprendimą, kuri ASR sistema yra geresnė, būtina jų bandymus atlikti toje pačioje aplinkoje [2].

Atliekant ASR sistemos bandymus galima naudoti gyvus kalbėtojus ir (arba) balso įrašus (įrašytą šneką) [2]. Bandymams naudojant balso įrašus užtikrinama, kad lyginamoms ASR sistemoms naudojama ta pati šnekos įvestis [2]. Tuo tarpu, gyvas kalbėtojas negali užtikrinti, kad tos pačios komandos bus ištartos taip pat daugybę kartų [2].

Kai kurie bandymai yra atliekami tylioje laboratorijos aplinkoje, kur nėra jokio foninio triukšmo [2]. Jei ASR sistemos darbinė aplinka bus triukšminga, tikslinga bandymus atlikti taip pat triukšmingoje aplinkoje, t. y. artimoje darbinei [2]. Foninis triukšmas gali būti generuojamas arba specifiniai garsai įrašyti darbinėje aplinkoje ir vėliau naudojami kaip foninis triukšmas [2]. Yra du būdai panaudoti foninį triukšmą ir įrašytas šnekos komandas kartu: šnekos komandos gali būti įrašytos kartu su foniniu triukšmu arba šnekos komandos ir foninis triukšmas įrašyti atskirai, o vėliau sujungiami [2]. Antrasis variantas, dažniausiai, labiau pageidautinas, nes tai leidžia kontroliuoti signalo ir triukšmo santykį (angl. *signal-to-noise ratio*, sutr. SNR) [2].

Kitas svarbus akcentas atliekant ASR sistemos bandymus yra kalbėtojų skaičius [2]. Kai kurie bandymai yra atliekami su duomenų bazėmis, kuriose saugomi tūkstančiai kalbėtojų; kai kurie bandymai yra atliekami su mažu kiekiu kalbėtojų [2]. Teoriškai yra naudingiau atlikti bandymus su didesne duomenų baze, bet praktiškai yra geriau dirbti su mažesne specifine kalbėtojų duomenų baze [2]. Mažesnio kalbėtojų skaičiaus naudojimas leidžia asmeniui, atliekančiam palyginimą, patvirtinti, kad kalbėtojas yra galutinių naudotojų atstovas [2]. Tai taip pat leidžia šnekos pavyzdžių surinkimo procesui būti kontroliuojamam ir užtikrinti nuoseklumą [2].

1.2.5. Šnekos atpažintuvai

Be gerai žinomų programinės įrangos kūrėjų, tokių kaip „Google“, „Microsoft“, yra ir kiti, sėkmingai kuriantys ir tobulinantys šnekos atpažintuvus. Komerciniai šnekos atpažintuvai pasižymi tuo, kad visi reikiami skaičiavimai yra atliekami panaudojant debesų kompiuterijos galimybes ir atpažintuvai yra apmokyti panaudojant milžiniškus duomenų kiekius.

Žemiau yra aprašomi komerciniai ir atvirojo kodo šnekos atpažintuvai.

Komerčiniai šnekos atpažintuvai

Vienas iš populiariausių šnekos atpažintuvų yra įmonės „Apple“ sukurtas asmeninis asistentas „Siri“ [37]. Pirmąkart jis buvo išleistas 2011 m. spalio mėn. kaip išmaniojo telefono programėlė, o vėliau tapo pagrindine „iPhone 4S“ ypatybe [37]. Tikriausiai, tai pati sėkmingiausia programinė įranga, apimanti šnekos atpažinimą ir natūralios kalbos supratimą, kontekstą ir samprotavimą [37].

Kaip atsaką „Siri“, „Google“ 2012 m. pristatė savo asmeninį asistentą „Google Now“, veikiančią „Android 4.1“ aplinkoje [37]. Kuriant jį mažiau dėmesio skirta natūralios kalbos supratimui, kontekstui ir samprotavimui, o daugiau akcentuotas paieškos užklauskos nuspėjimas (prognozė) ir nuspėjamų paieškos rezultatų rodymas [37]. Taigi, esminis skirtumas tarp „Google Now“ ir „Siri“ yra tai, kad pirmoji aktyviai teikia informaciją, kai to reikia [37]. Pavyzdžiui, traukinių stotyje „Google Now“ rodo išvykimo ir atvykimo laikus be naudotojo užklauskos [37]. Be „Google Now“, skirtos išmaniesiems telefonams, „Google“ išleido šnekos atpažinimo aplikacijų programavimo sąsają (angl. *application programming interface* (API)), skirtą žiniatinkliui [37]. Ši *Javascript* biblioteka leidžia naudotis šnekos atpažinimo bei sintezės funkcionalumu ir yra galima pradedant naršyklės „Chrome“ 25 versija [37]. Ši „Google“ šnekos atpažinimo aplikacijų programavimo sąsaja įgalina programinės įrangos kūrėjus kurti balsu valdomus naršyklės „Chrome“ įskiepius ir puslapius [37].

Kita plačiai žinoma programinė įranga, pasižyminti labai aukšta natūralios kalbos ir šnekos apdorojimo technologija, yra į klausimus atsakanti IBM sistema „Watson“, išgarsėjusi po to, kai įveikė du pagrindinius viktorinos „Jeopardy“ dalyvius [37]. „Ask Watson“ pasiekama įvairiais kanalais, tokiais kaip pokalbių svetainės, el. paštas, išmaniųjų telefonų programėlės ir SMS [37].

Kiti debesų kompiuterija paremti (angl. *cloud-based*) šnekos atpažintuvai – populiarius šnekos į tekstą vertimo įrankis „Dragon Naturally Speaking“ ir „Siri“ alternatyva „Android“ ir „iOS“ aplinkoms „Vlingo Virtual Assistant“ [37].

Atvirojo kodo šnekos atpažintuvai

Vienas populiariausių atvirojo kodo šnekos atpažintuvų yra „Hidden Markov Model Toolkit“ (HTK) [37]. Be to, kad jis gali būti taikomas bet kokio tipo paslėptojo Markovo modelio (angl. *Hidden Markov Model* (HMM)) apmokymui, HTK, dažniausiai, yra naudojamas šnekos atpažinimo tyrimams [37]. HTK yra sudarytas iš bibliotekų ir įrankių, parašytų C kalba, rinkinio [37]. Įrankiai suteikia galimybę atlikti šnekos analizę, HMM apmokymus, testavimą ir rezultatų analizę [37]. HTK gali būti naudojamas kuriant atpažintuvus, skirtus tūkstančius valandų trunkančios šnekos apdorojimui [37].

Kitas atvirojo kodo šnekos atpažintuvas – SRI „Decipher“ [37]. Jis gali atpažinti natūralią, rišlią šneką be poreikio naudotojui pirmiausiai apmokyti sistemą [37]. „Decipher“ išsiskiria iš kitų šnekos atpažintuvų savo detaliu tarties pokyčių modeliavimu ir atsparumu foniniam triukšmui ir kanalo iškraipymui [37]. Šios galimybės leidžia „Decipher“ pranokti kitus šnekos atpažintuvus atpažįstant skirtingų dialektų spontanišką šneką ir sumažinant atpažinimo kokybės priklausomybę nuo skirtingų mikrofonų ir akustinės aplinkos [37].

Dar vienas atvirojo kodo šnekos atpažintuvas – „Microsoft Research“ sukurtas „Kaldi“ [37]. Tai šnekos atpažinimo sistema, paremta baigtinės būsenos davikliais su detalia dokumentacija ir skriptais, skirtais baigtinių atpažinimo sistemų kūrimui. „Kaldi“ parašytas C++ kalba ir išleistas su neapribojančia licencija „Apache License v2.0“, todėl šis atpažintuvas prieinamas plačiai naudotojų bendruomenei [37].

MIT „WAMI“ suteikia karkasą programinei įrangai kurti, diegti ir internetu prieinamoms jungtinėms sąsajoms, kuriomis naudotojai sąveikauja naudojant šneką, pelę, rašiklį ar lietimą, įvertinti [37]. Šnekos atpažintuvas naudoja šiuolaikines žiniatinklio programavimo technologijas, įgalina kurti naršyklėse veikiančias programas [37]. Taip pat jis suteikia galimybę kaupti, perrašyti ir anotuoti duomenis, gautus iš jungtinių naudotojų sąsajų [37].

„Sphinx“ yra šnekos atpažinimo sistemų, sukurtų „Carnegie Mellon“ universitete, grupės pavadinimas [37]. Tarp šių sistemų yra „Sphinx 3“, šnekos atpažinimo dekoderis, parašytas C kalba, „Sphinx 4“, atpažintuvas, parašytas Java kalba, ir „Pocketsphinx“, šnekos atpažinimo biblioteka, parašyta C kalba [37].

RWTH *Automatic Speech Recognition* (RASR) yra parašytas C++ kalba ir sudarytas iš šnekos atpažintuvo ir įrankių, skirtų akustiniams modeliams kurti [37]. Šnekos atpažinimo sistemos, sukurtos

panaudojant šį karkasą, pasižymi didele sparta keliuose tarptautiniuose mokslinių tyrimų projektuose ir nuolatiniuose vertinimuose (pvz., „GALE“ projekto „DARPA“ apimtyje) [37].

1.3. Tyrimo objekto naudotojų analizė

Sprendimas aktualus programinės įrangos kūrėjams, kurie planuoja savo kuriamus produktus integruoti su „Google“ šnekos atpažintuvu, taip pat įvairių programėlių (maršrutų paieška, parkavimas ir pan.) ir funkcijų, kurios paremtos šnekos įvestimi (paieška telefone, numerio rinkimas ir pan.), naudotojams.

Naudotojai yra suinteresuoti panaudoti kuo kokybiškesnį ir efektyvesnį šnekos atpažintuvą.

1.4. Esamų problemos sprendimo metodų analizė (Lietuvos ir tarptautiniu mastu)

Atpažinimo sistemų kūrimas reikalauja didelių laiko ir žmogiškojo darbo resursų, o tai sąlygoja didelę tokių produktų savikainą [38]. Todėl komerciniai produktai, kuriuose realizuojamas šnekos atpažinimas, kuriami tik didelėms rinkoms (plačiai paplitusioms kalboms), o mažiau paplitusios kalbos lieka be dėmesio [38]. Toks gamintojų atsiribojimas nuo komerciškai nepatrauklių kalbų lemia tų šnekų atpažinimo tyrimų nykimą – tyrimai apsiriboja plačiai paplitusioms šnekoms sukurtų metodų ir technologijų pritaikymu ir modifikavimu [38]. Įvertinus tai, kad šnekos skiriasi savo fonetinėmis savybėmis, gramatika, tokie tyrimai anksčiau ar vėliau gali tapti neperspektyvūs, ir mokslo bei praktinės realizacijos požiūriu beverčiai [38].

Aukščiau pateikti teiginiai pilnai atitinka situaciją lietuvių šnekoje [38]. Komerciniu požiūriu lietuvių šnekos vartojimo sritis yra per maža, kad atkreiptų didžiųjų gamintojų dėmesį [38]. Nepaisant to, technologijų gigantas „Google“ pristatė šnekos atpažintuvą, pritaikytą lietuvių šnekai (jo efektyvumas šiame darbe ir tiriamas).

Šiame poskyryje apžvelgiami Lietuvoje ir užsienyje atliekami darbai, susiję su šnekos atpažinimu.

1.4.1. Lietuvoje atliekami darbai

Lietuvių šnekos tyrimai vyksta įvairiomis kryptimis, todėl kartais sunku susidaryti bendrą vaizdą, koks kalbinių technologijų išsivystymo lygis jau pasiektas [9]. Atskiri darbai nurodo, kiek toli pažengta konkrečioje srityje.

Galima identifikuoti tokias būdingiausias lietuvių šnekos tyrimo sritis [9]:

- skaitmeninių signalų apdorojimas,
- kalbančiojo identifikavimas ir verifikavimas,
- akustinis modeliavimas,
- kalbos modelio tyrimai,
- šnekos sintezavimo tyrimai,
- šnekos tyrimai lingvistiniu aspektu.

Visi šie tyrimai skirti šnekos atpažinimo ir sintezavimo blokams, kuriuos sujungus būtų galima kalbėti apie šnekos atpažinimo-sintezės sistemą, sukurti.

Lietuvoje nuolatos vyksta šnekos atpažinimo tyrimai, šnekos atpažinimo sprendimus bandoma pritaikyti praktiniams uždaviniams spręsti. 2013-2015 metais Vilniaus universitetas (Matematikos ir informatikos institutas ir Filologijos fakultetas) ir partneriai (Lietuvių kalbos institutas, Lietuvos edukologijos universitetas ir Šiaulių universitetas) vykdė projektą „Liepa“, kurio metu buvo sukurtos lietuvių šneka valdomos paslaugos. Detalesnė informacija apie visus šiuos darbus pateikiama žemiau.

Pavienių žodžių atpažinimo ir segmentavimo sistema KAS

Sistemoje KAS (Kalbos Atpažinimas ir Segmentavimas) įdiegti trys darbo režimai: pavienių žodžių atpažinimo, žodžių segmentavimo ir žodžio segmentų (garsų) atpažinimo [38]. Kiekvienam darbo režimui būdingi savi parametrai, valdymo elementų rinkiniai ir rezultatų pateikimas [38].

Pirmasis iš galimų sistemos KAS režimų – pavienių žodžių atpažinimas. Atpažinimas nėra susietas su kalbos pavyzdžiu, todėl atpažinimui gali būti pateiktas tiek garsas, tiek garsų junginys, tiek žodžių junginys [38]. Kaip rezultatas pateikiama artimiausio etalono transkripcija, skaitmeninė atstumo išraiška, įvesto signalo laiko ir energijos diagramos su žodžio ribomis [38]. Jei žodžio riboms nustatyti naudotas energijos slenksčio metodas, energijos diagramoje papildomai atvaizduojamas

slenksčio lygis. Papildomai pateikiama informacija apie sistemos darbo režimą, darbo sesiją, signalo analizės metodą, pasirinkto failo pavadinimą ir atliktų palyginimų skaičių [38].

Antrasis sistemos darbo režimas – žodžių segmentavimas. Persijungus į šį režimą pasikeičia valdymo elementų rinkinys ir parametrų dialogai [38]. Segmentuoti galima žodžius tiek su tylos atkarpomis, tiek be jų [38]. Kaip rezultatai pateikiami kita spalva išskirti naudotojo pasirinkti kadrai pradinėms parametrų reikšmėms vertinti ir garsų ribos signale [38]. Papildomai pateikiamas darbo sesijos pavadinimas, pasirinkto failo pavadinimas ir skaičiavimo iteracijų skaičius [38]. Naudotojas turi galimybę išskirtiems segmentams suteikti fonetinę transkripciją ir visus segmentavimo rezultatus išsaugoti failuose [38].

Trečiasis sistemos darbo režimas – žodžio segmentų atpažinimas. Išskirtieji žodžio segmentai panaudojami kaip akustiniai vienetai atpažinimo sistemoje [38]. Kiekvienam jų sukuriama etalona, o segmentavimo į garsus procesas įdiegiamas kaip pradinis palyginimo proceso etapas [38]. Kaip atpažinimo rezultatas pateikiama žodžio laiko diagrama su segmentų ribomis, atpažintų segmentų fonetinės transkripcijos ir atstumai iki etalonų, laukeliai segmentams transkribuoti [38]. Papildomai pateikiamas darbo sesijos pavadinimas, naudojamas analizės tipas ir eilė, pasirinkto garso failo pavadinimas, nurodomi segmentacijos metu atliktų iteracijų ir palyginimo operacijų skaičiai [38].

Komercinių užsienio kalbos atpažintuvų taikymas Lietuvių šnekai atpažinti

Yra atlikta nemažai tyrimų užsienio kalbų taikymo Lietuvių šnekai atpažinti srityje. Žemiau pateikta keletas pavyzdžių.

„Microsoft Speech“ serveris ir anglų ar ispanų šnekos atpažinimo varikliai buvo panaudoti įvairių prietaisų gyvenamojoje aplinkoje valdymui [39]. Tokia galimybė yra aktuali žmonėms, turintiems tam tikrą fizinę negalią ir neturintiems galimybėms įrenginius valdyti įprastinėmis priemonėmis. Dauguma balso komandų buvo atpažinta daugiau nei 95 % tikslumu tyrimą atliekant su dideliu kiekiu skirtingų kalbėtojų ir skirtingomis akustinėmis sąlygomis [39]. Tyrimo metu pastebėta, kad kai kurios balso komandos negali būti gerai atpažintos ir būtina sukurti tam tinkamą šnekos atpažintuvą [39].

Taip pat buvo atliktas tyrimas anglų ir ispanų šnekos atpažinimo variklius naudojant skaitmenims (nuo 0 iki 9) ir balso komandoms kompiuteriui (interneto naršyklei, „Microsoft Office Word“ ir garso grotuvui) atpažinti [40]. Skaitmenų atveju siekiamas šnekos atpažinimo tikslumas (95 %) nebuvo pasiektas, o balso komandų atveju šnekos atpažinimo tikslumas buvo šiek tiek virš 90 % [40]. Prasčiausiai buvo atpažintos komandos garso grotuvui – atpažinimo tikslumas 85-87 % [40].

Medicinos-farmacijos informacinei sistemai buvo pritaikytas „Microsoft“ ispanų šnekos atpažintuvas ir pačių straipsnio autorių sukurtas paslėptaisiais Markovo modeliais pagrįstas šnekos atpažintuvas su lietuvių šnekos akustiniu modeliu [41]. Eksperimento metu buvo naudojamos frazės iš medicinos srities. Šnekos atpažinimo rezultatai: vidutinis WER 14,5 %, geriausias rezultatas pagal kalbėtojus – 8,53 %, blogiausias – 30,2 % [41]. Parinkus optimalias transkripcijas buvo pasiektas 98,9 % šnekos atpažinimo tikslumas 1000 lietuviškų komandų iš medicinos srities [42].

Projektas „Liepa“

Pasitelkus projekto metu sukurtas lietuvių šneka valdomas paslaugas, moksleiviai, suaugusieji, neįgalieji galės pradėti savo veiklose naudoti lietuvišką šneką suprantančius ir lietuviškai šnekančius kompiuterius [43]. Naudotojams teikiamos šios paslaugos [43]:

- **Tartuvas** – paslauga, leidžianti išgirsti per internetą, kaip tarti lietuvių kalbos naujažodžius;
- **Naršytuvas** – paslauga, leidžianti naršyti internete duodant kompiuteriui nurodymus komandomis balsu;
- **Pažintuvas** – paslauga, leidžianti ugdyti moksleivių ląstelės pažinimo įgūdžius naudojant balso komandas;
- **Valdytuvas** – paslauga, leidžianti valdyti kompiuterį balsu;
- **Ieškotuvus** – paslauga, leidžianti ugdyti moksleivių gebėjimus šneka ieškoti išteklių apie UNESCO kultūros paveldą pasaulyje;
- **Pagalbininkas** – paslauga, padedanti įvairių negalių turintiems žmonėms šnekančiais su kompiuteriu plėtoti socialinius gebėjimus, savitvarką, savitarną, informacinių technologijų naudojimo įgūdžius;

- **Lietuvių šnekos sintezatorius akliesiems** – su SAPI5 standartu suderinamas lietuvių šnekos sintezatorius, skaitantis balsu tai, kas rodoma kompiuterio ekrane.

Galimos šios infrastruktūrinės paslaugos:

- **Elektroninio teksto skaitytuvas** – programinė įranga, skaitanti lietuvišką elektroninį tekstą balsu (lietuvių šnekos sintezatorius).
- **Interneto puslapių atidarymo balsu programinė įranga** – programinė įranga, sudaranti galimybes balso komandomis atverti interneto puslapius.
- **Lietuvių šnekos atpažinimo variklis** – programinė įranga, atpažįstanti lietuvių šneką.
- **Garsynas** – 113 valandų anotuotas lietuvių šnekos garsynas.
- **Balso įrašų laboratorija** – Vilniaus universiteto Filologijos fakultete.

Visi projekto LIEPA metu sukurti produktai nemokami, prieinami visiems norintiems jais naudotis.

1.4.2. Užsienyje atliekami darbai

Norint teisingai įvertinti bet kokios sistemos kokybę labai svarbu pasirinkti teisingas metrikas.

Vertinant „Google Search by Voice“, pritaikytos anglų kalbai, kokybę buvo pasirinktos šios metrikos [44]:

1. **Klaidingų žodžių dažnis** (angl. *Word Error Rate*, WER) – įvertinamas klaidingai atpažintų žodžių dažnis žodžių lygyje; kitaip tariant, palyginamas atpažintuvo grąžintas tekstas su tekstu, kuris buvo kalbėtojo įgarsintas [44]. Apskaičiuojama taip:

$$WER = \frac{\text{pakeitimai} + \text{įterpimai} + \text{panaikinimai}}{\text{Iš viso žodžių}}; \quad (1)$$

čia:

- *pakeitimai* – žodžiai, kurie buvo pakeisti kitais žodžiais;
- *įterpimai* – žodžiai, kurių nebuvo originaliame tekste;
- *panaikinimai* – žodžiai, kurių trūksta lyginant atpažintuvo pateiktus rezultatus su originaliu tekstu.

2. **Semantinė kokybė** (angl. *Semantic Quality*, WebScore) – įvertinama paieškos rezultatų kokybė semantiniu požiūriu, t. y. kaip vykdoma paieška iš užklausos pašalinus ar užklausa papildžius pagalbiniais žodžiais [44]. Apskaičiuojama taip:

$$WebScore = \frac{\text{Teisingi paieškos rezultatai}}{\text{Iš viso paieškos užklausų}}. \quad (2)$$

Kuo aukštesnės kokybės atpažintuvas, tuo aukštesnė *WebScore* metrikos reikšmė [44]. Ši metrika leidžia numanyti, ką naudotojas patiria kai naudojasi paieška balsu [44]. Atliekant tyrimus, dažniausiai, straipsnio autoriai optimizuoja būtent šios metrikos rezultatus, o ne WER [44].

3. **Neapibrėžtumas** (angl. *Perplexity*, PPL) – žodžių rinkinio, kuris gali būti atpažintas remiantis anksčiau atpažintais užklausos žodžiais, dydis [44]. Apskaičiuojama taip:

$$PPL = 2^{\sum_{i=1}^N \frac{1}{N} \log_2 q(x_i)}. \quad (3)$$

Ši metrika leidžia grubiai įvertinti kalbos modelio kokybę – kuo žemesnė šios metrikos reikšmė, tuo kalbos modelis yra geresnis nuspėjant kitą žodį, kurį reikia atpažinti [44].

4. **Žodžių, kurių nėra žodyne, dažnis** (angl. *Out-of-Vocabulary* (OOV) *Rate*) – apskaičiuojamas žodžių, kuriuos ištarė kalbėtojas, bet kurių nėra kalbos modelyje, kiekis procentais [44]. Labai svarbu, kad šis skaičius būtų kaip įmanoma mažesnis, nes tai labai siejasi su atpažinimo klaidų skaičiumi bei gali daryti įtaką gretimų žodžių atpažinimo kokybei [44].
5. **Trukmė** (angl. *latency*) – įvertinamas laikas (sekundėmis), kiek truko apdoroti paieškos balsu užklausa, t. y. įvertinamas laikas nuo to momento, kai kalbėtojas baigė kalbėti, iki to momento, kai paieškos rezultatai buvo atvaizduoti ekrane [44]. Šiai metrikai daro įtaką keletas faktorių [44]:
 - laikas, per kurį sistema įvertina, kad kalbėtojas baigė kalbėti;

- laikas, per kurį buvo apdorota paieškos užklausa;
- laikas, per kurį buvo atlikta žiniatinklio paieška (angl. *web query*);
- laikas, per kurį žiniatinklio paieškos rezultatai tinklu buvo grąžinti klientui;
- laikas, per kurį paieškos rezultatai buvo atvaizduoti naudotojo telefono naršyklėje.

1.5. Darbo tikslas, uždaviniai ir siekiami privalumai

Darbo tikslas – įvertinti „Google“ lietuvių šnekos atpažintuvo efektyvumą.

Darbo uždaviniai:

1. Išanalizuoti šnekos atpažinimo sistemų efektyvumo įvertinimo metodus.
2. Pasirinkti tinkamą šnekos atpažinimo sistemų efektyvumo įvertinimo metodą.
3. Sukurti prototipą pasirinkto metodo efektyvumui įvertinti.
4. Atlikti eksperimentinius tyrimus „Google“ šnekos atpažintuvo efektyvumui įvertinti.
5. Nustatyti, kokie faktoriai (kalbėtojo savybės, kalbėjimo savybės, foninis triukšmas, šnekos tipas) daro įtaką šnekos atpažinimo kokybei.

Atlikus tyrimą bus galima vertinti, kokiomis sąlygomis (kokioje aplinkoje, kaip kalbant, su koku šnekos tipu) „Google“ lietuvių šnekos atpažintuvas veikia tiksliausiai. Taip pat remdamiesi atlikto tyrimo rezultatais suinteresuoti asmenys galės priimti sprendimą, ar šnekos atpažinimo kokybė yra pakankama, kad šis šnekos atpažintuvas galėtų būti integruotas su jų kuriama programine įranga.

1.6. Siekiamo sprendimo apibrėžimas

Norint įvertinti „Google“ lietuvių šnekos atpažintuvo efektyvumą reikia atlikti eksperimentą. Šiuo tikslu turi būti sukurta programinė įranga, gebanti „Google“ šnekos atpažintuvui perduoti balso ar garso (balso įrašas, apjungtas su triukšmu) įrašą, kurį reikia atpažinti, ir gautą atpažinimo rezultatą (tekstą) įrašyti į DB, t. y. veikianti kaip tarpininkas tarp „Google“ šnekos atpažintuvo ir DB, kurioje saugomi balso įrašai ir atpažinimo rezultatai.

Eksperimento dalyvis PĮ pagalba galės įrašyti naujus ir sukelti į DB turimus balso įrašus nurodant balso įrašus aprašančią informaciją. Eksperimento vykdytojas galės atlikti eksperimento dalyvio funkcijas ir redaguoti balso įrašų informaciją, inicijuoti garso įrašų apdorojimą bei peržiūrėti garso įrašų apdorojimo rezultatus. Detalesnė informacija apie sistemos aktorių vykdomas funkcijas pateikta poskyryje „2.1.1 Kompiuterizuojamos sistemos funkcijos“.

Eksperimento vykdytojui inicijavus garso įrašų apdorojimo procesą, įrašų apdorojimas ir gautų šnekos atpažinimo rezultatų saugojimas vyks automatiškai, be eksperimento vykdytojo įsikišimo.

1.7. Analizės išvados

1. Literatūros šaltinių analizė parodė, kad nėra jokių nusistovėjusių metodikų ar standartų, leidžiančių patikimai įvertinti šnekos atpažintuvų efektyvumą.
2. Remiantis literatūros šaltinių analize galima teigti, kad šiuolaikinių šnekos atpažintuvų efektyvumas, lyginant su pirmais sukurtais, pakito labai neženkliai (yra apie 98 %). Tai galima paaiškinti tuo, kad kiekvieno iš jų efektyvumas vertinamas jam palankiomis aplinkybėmis, t. y. su tam tikrais kalbėtojais, tam tikroje aplinkoje ir pan.
3. Literatūros šaltinių analizė parodė, kad didelę įtaką šnekos atpažintuvo efektyvumui turi kalbančiojo fizinė ir psichologinė būseną, kalbėjimo manieras, taip pat foninis triukšmas.
4. Atlikus šnekos atpažintuvų efektyvumo vertinimo metodikų analizę galima teigti, kad dažniausiai šnekos atpažintuvo kokybei vertinti yra naudojama metrika WER (angl. *Word Error Rate*).

2.1.4. Rizikos faktorių analizė

1. „Google“ šnekos atpažintuvo rakto (angl. *API Key*) galiojimas gali būti sustabdytas arba panaikintas. Neturint šio rakto negalimas šnekos atpažinimas siunčiant „Google“ šnekos atpažintuvui garso rinkmenas.
2. „Google“, pastebėję didelį užklausų kiekį iš vieno IP adreso, gali pradėti neteikti lietuvių šnekos atpažinimo paslaugos tam IP adresui. Rizika galėtų būti valdoma garso įrašų apdorojimą vykdant kitu IP adresu (kitas interneto paslaugų teikėjas, tarpiniai (angl. *proxy*) serveriai, mobilusis internetas ir kt.).
3. Gali nepavykti surinkti pakankamo balso įrašų kiekio eksperimentui atlikti (su skirtingais teksta, kalbėjimo greičiu, skirtingas įrašų garsumas; kalbėtojai su akcentu/be akcento, skirtingų lyčių), dėl ko eksperimento rezultatai nebūtų traktuojami kaip svarūs ir patikimi.

Sukūrus du alternatyvius garso įrašų apdorojimo būdus pirma ir antra rizikos galėtų būti valdomos – susidūrus su problemomis apdorojant garso įrašus vienu būdu, būtų galima balso įrašų apdorojimą tęsti kitu būdu.

2.1.5. Rezultato kokybės kriterijai

1. Eksperimentas atliktas su ne mažiau kaip 5000 balso įrašų.
2. Balso įrašai tenkina šiuos kriterijus: bent 50 skirtingų kalbėtojų; skirtingų lyčių kalbėtojai; skirtingi šnekos tipai (pavienės balso komandos, žodžių seka, rišli šneka, spontaniška šneka); skirtingi kalbėjimo garsumo lygiai (garsiai, tyliai); skirtingi kalbėjimo greičiai (greita šneka, lėta šneka).
3. Eksperimente panaudoti bent 4 skirtingi triukšmai ir kiekvienam iš jų bent 3 skirtingi SNR.
4. Bendra balso įrašų trukmė yra ne mažesnė kaip 10 val.

2.2. Tyrimo metodo aprašas

„Google“ šnekos atpažintuvo efektyvumo tyrimas atliekamas trimis etapais:

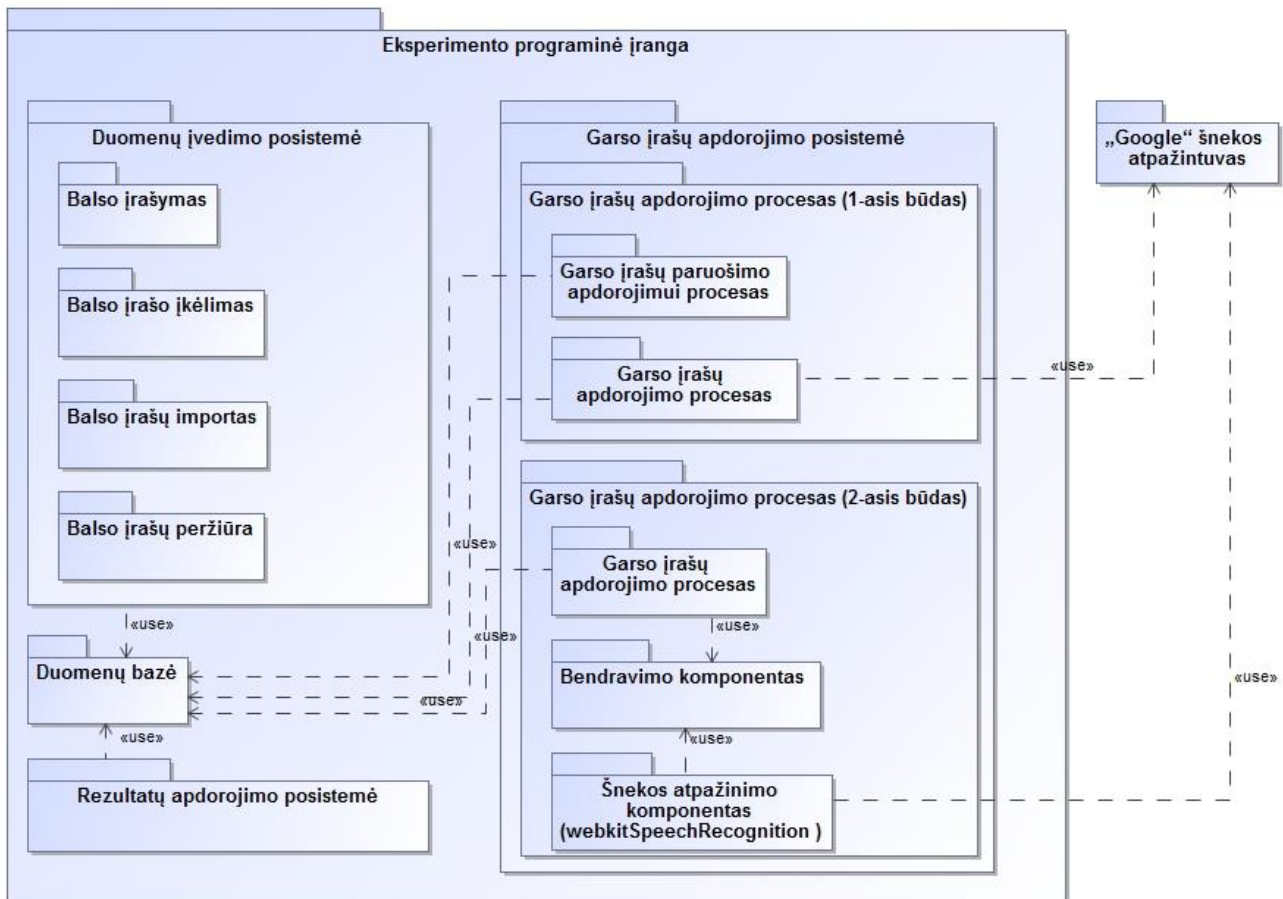
1. Pirmiausiai „Google“ šnekos atpažintuvu bus apdoroti (t. y. gauti šnekos atpažinimo rezultatai) visi balso įrašai be triukšmo.
2. Antrame etape visi balso įrašai apjungiami su visais galimais triukšmais, modifikuotais taip, kad būtų gauti tam tikri SNR.
3. Kadangi „Google“ šnekos atpažintuvas pagrįstas dirbtinio intelekto atpažinimo metodais, vykdydamas šnekos atpažinimą nuolatos apsimoko, tikslinga po mėnesio nuo balso įrašų apdorojimo be triukšmo visus balso įrašus apdoroti be triukšmo pakartotinai. Lyginant pirmame ir trečiame tyrimo etape gautus rezultatus bus galima vertinti, ar šnekos atpažinimo kokybė gerėja.

Žemiau pavaizduotas „Google“ šnekos atpažintuvo efektyvumo tyrimo metodas.



2.2 pav. „Google“ šnekos atpažintuvo efektyvumo tyrimo metodas

2.3. Sistemos loginė architektūra



2.3 pav. Sistemos loginė architektūra

Kuriama eksperimento programinė įranga išskaidyta į posistemas. Kiekviena posistemė realizuota kaip paslauga ir yra nepriklausoma nuo kitų posistemių darbo.

Duomenų įvedimo posistemėje yra viskas, kas susiję su pradinių duomenų įkėlimu ir jų tvarkymu.

Garso įrašų apdorojimo posistemė šioje sistemoje yra pati svarbiausia ir labiausiai nepatikima, nes naudoja išorinę sistemą („Google“ šnekos atpažintuvą), kurios veikimas nepriklauso nuo kuriamos sistemos. Šią posistemę sudaro du bendravimo su „Google“ šnekos atpažintuvu procesai. Abu procesai realizuoti skirtingomis technologijomis, kad galėtume valdyti rizikas ir užsitikrinti, jog nebus techninių kliūčių vykdyti eksperimentą. Šios posistemės tikslas – paversti garso bylose esančią informaciją į tekstinę informaciją bendraujant su „Google“ šnekos atpažintuvu.

Rezultatų apdorojimo posistemėje generuojami šnekos atpažinimo rezultatai lentelių formatu.

2.4. Sistemos elgsenos modelis

2.4.1. Garso įrašų apdorojimo procesas

Šiame poskyryje aprašomas garso įrašų apdorojimo procesas.

Garso įrašų apdorojimui „Google“ šnekos atpažintuvu reikalingi du būdai:

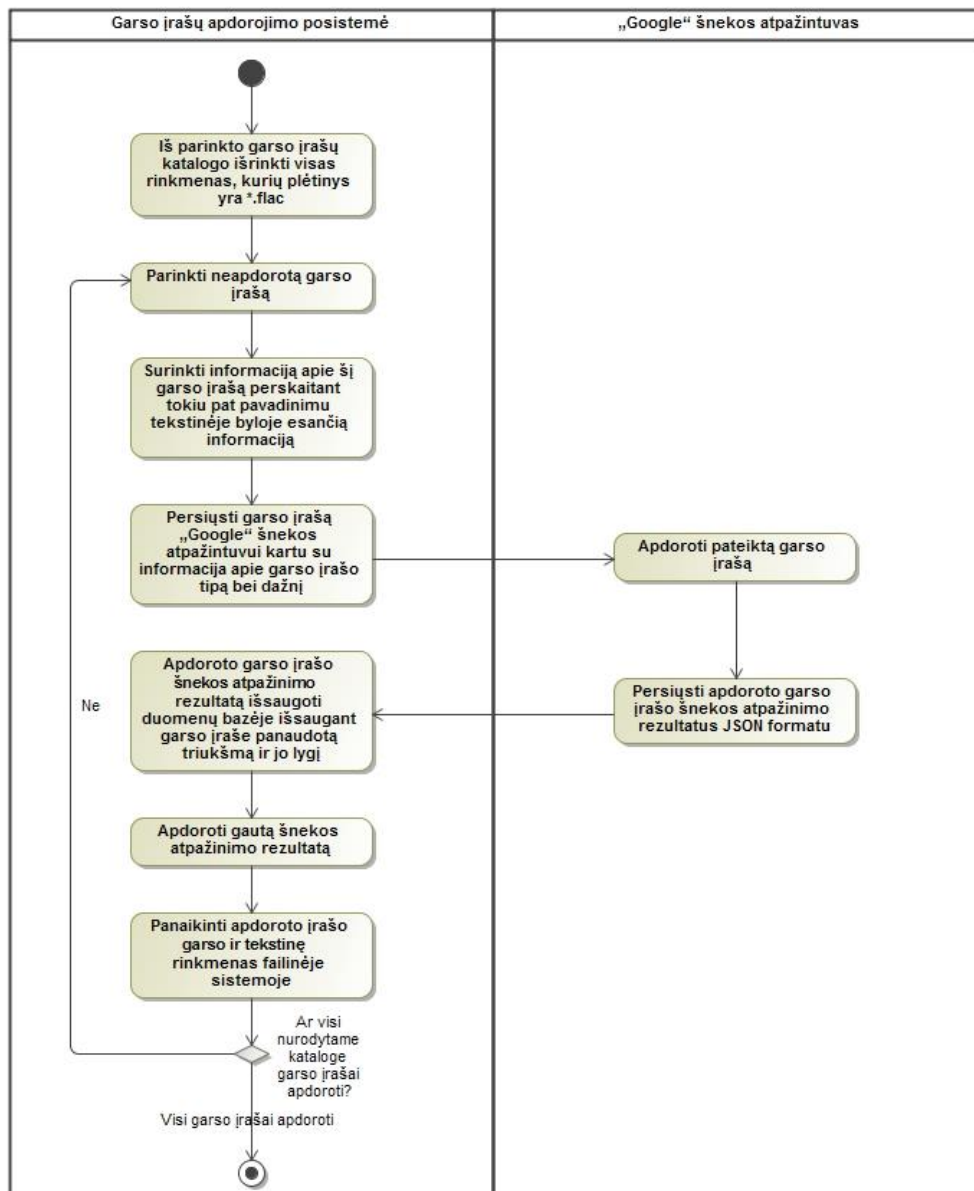
- Pirmasis – vienu metu vienas kompiuteris gali siųsti „Google“ šnekos atpažintuvui apdoroti daug garso įrašų lygiagrečiai. Siunčiami atpažinti garso įrašai būtinai turi būti FLAC formato, todėl prieš naudojant juos šnekos atpažinimo procese pirmiausiai jie turi būti paruošti apdorojimui (garso įrašų paruošimo apdorojimui algoritmas pateiktas 2.4.2 poskyryje). Siunčiamų apdoroti garso įrašų kiekis valdomas gijų skaičiumi (angl. *thread*). Apdorojant garso įrašus pirmuoju būdu garso adapteris nereikalingas. Pasikliauti tik šiuo būdu negalima, nes garso įrašų apdorojimas gali būti bet kada sustabdytas „Google“ panaikinus aktyvų raktą. Anksčiau „Google“ suteikdavo aktyvų raktą bet kuriam asmeniui, bet su apribojimu – galėjo būti atpažinta tik iki 50 balso įrašų per dieną.

Šis garso įrašų apdorojimo procesas pavaizduotas 2.4 pav. Procesas inicijuojamas garso įrašų apdorojimo inicijavimo lange (3.7 pav.).

- Antrasis – vienu metu vienas kompiuteris gali siųsti „Google“ šnekos atpažintuvui apdoroti vieną garso įrašą (įrašo formatas nesvarbus). Garso įrašų apdorojimo procesas gali būti paspartintas sujungiant daugiau kompiuterių. Apdorojant garso įrašus šiuo būdu yra reikalingas garso adapteris (eksperimento PĮ groja įrašą, o „Google“ šnekos atpažintuvas „klaus“). Šis garso įrašų apdorojimo procesas pavaizduotas 2.5 pav. Procesas inicijuojamas garso įrašų apdorojimo inicijavimo lange (3.8 pav.).

Garso įrašų apdorojimo procesas (1-asis būdas)

Šio garso įrašų apdorojimo proceso metu „Google“ šnekos atpažintuvui siunčiamas jau paruoštas apdorojimui garso įrašas (FLAC formatu).



2.4 pav. Garso įrašų apdorojimo procesas (1-asis būdas)

Iš naudotojo parinkto garso įrašų katalogo išrenkami visi garso įrašai, kurių formatas FLAC. Tuomet išrenkamas dar neapdorotas garso įrašas ir iš tekstinės bylos, kurios pavadinimas sutampa su garso įrašo pavadinimu, nuskaityta informacija apie garso įrašą (triukšmas, SNR, dažnis). Garso įrašas kartu su informacija apie jį siunčiama „Google“ šnekos atpažintuvui, kuris gautą garso įrašą apdoroja ir persiunčia apdoroto garso įrašo šnekos atpažinimo rezultatus JSON formatu. Gauti šnekos

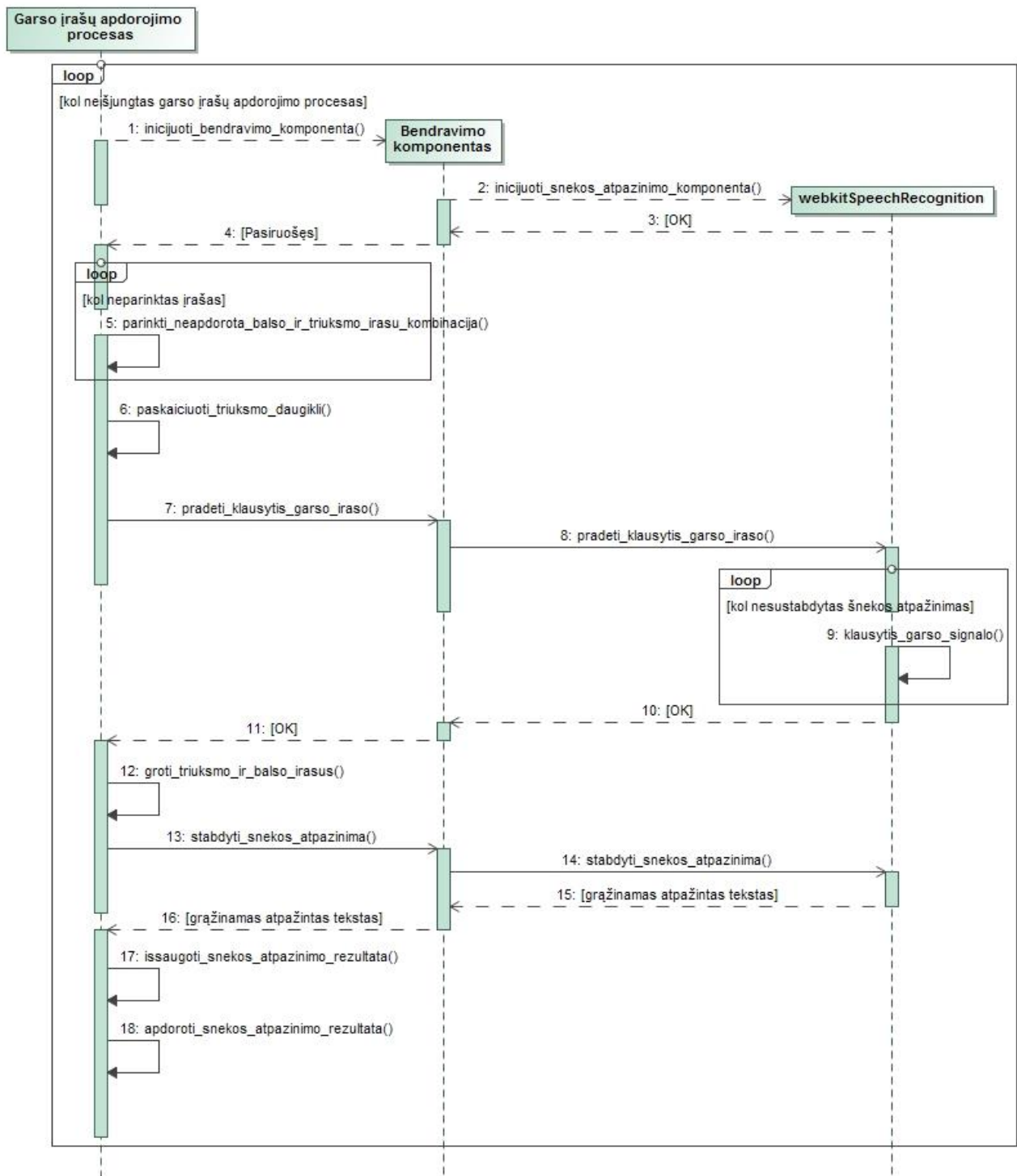
atpažinimo rezultatai ir informacija apie garso įrašė panaudotą triukšmą ir SNR išsaugomi DB. Po to, šnekos atpažinimo rezultatai apdorojami (išskaidomi į pagrindinį ir alternatyvius rezultatus, visi skaitvardžiai paverčiami tekstu, atliekamas WER skaičiavimas visiems išskaidytiems šnekos atpažinimo rezultatams), o apdoroto įrašo garso ir tekstinė rinkmenos pašalinamos iš failinės sistemos. Žemiau pateikiamas detalus kiekvieno proceso veiksmo aprašymas.

2.1 lentelė. Garso įrašų apdorojimo procesas (1-asis būdas)

Eil. Nr.	Veiksmas	Aprašymas
1.	Iš parinkto garso įrašų katalogo išrinkti visas rinkmenas, kurių plėtinys yra *.flac	Iš nurodyto garso įrašų katalogo išrenkami visi garso įrašai, kurių plėtinys yra *.flac.
2.	Parinkti neapdorotą garso įrašą	Išrenkamas pirmas neapdorotas garso įrašas.
3.	Surinkti informaciją apie šį garso įrašą perskaitant tokiu pat pavadinimu tekstinėje byloje esančią informaciją	Iš tokiu pat pavadinimu tekstinės bylos nuskaitoma informacija apie konkretų garso įrašą (triukšmas, SNR, dažnis).
4.	Persiųsti garso įrašą „Google“ šnekos atpažintuvui kartu su informacija apie garso įrašo tipą bei dažnį	Garso įrašas ir informacija apie garso įrašo tipą bei dažnį nusiunčiama „Google“ šnekos atpažintuvui.
5.	Apdoroti pateiktą garso įrašą	Garso įrašas apdorojamas.
6.	Persiųsti apdoroto garso įrašo šnekos atpažinimo rezultatus JSON formatu	Išsiunčiami šnekos atpažinimo rezultatai JSON formatu.
7.	Apdoroto garso įrašo šnekos atpažinimo rezultatus išsaugoti duomenų bazėje išsaugant garso įrašė panaudotą triukšmą ir SNR	Gauti šnekos atpažinimo rezultatai išsaugomi DB. Kartu išsaugoma informacija ir apie garso įrašė panaudotą triukšmą ir SNR.
8.	Apdoroti gautą šnekos atpažinimo rezultata	Atliekamas šnekos atpažinimo rezultato apdorojimas. Detalesnė informacija pateikta 2.4.4 poskyryje.
9.	Panaikinti apdoroto įrašo garso ir tekstinę rinkmenas failinėje sistemoje	Apdoroto įrašo garso ir tekstinė rinkmenos pašalinamos iš failinės sistemos.

Garso įrašų apdorojimo procesas (2-asis būdas)

Šio garso įrašų apdorojimo proceso metu imituojamas gyvas kalbėjimas, t. y. eksperimento PĮ groja įrašą, o „Google“ šnekos atpažintuvas „klausos“.



2.5 pav. Garso įrašų apdorojimo procesas (2-asis būdas)

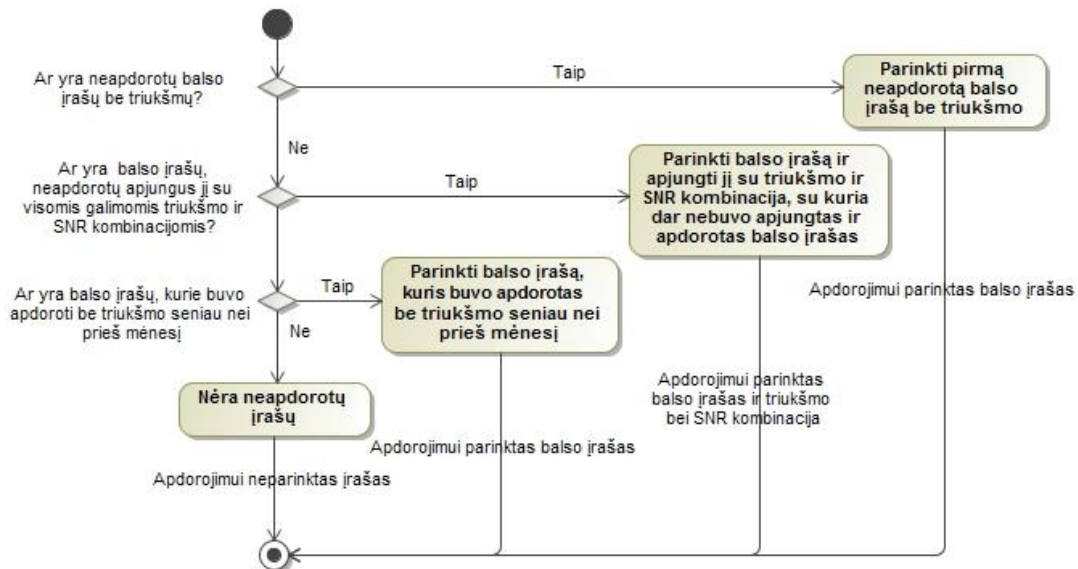
Paleidus garso įrašų apdorojimo procesą automatiškai atveriamas „Google Chrome“ naršyklės puslapis ir inicijuojamas bendravimo komponentas, kuris inicijuoja šnekos atpažinimo komponentą. Tuomet parenkami balso įrašas ir, jeigu reikia, triukšmo įrašas bei balso ir triukšmo santykis (SNR). Jei išrinktas apdoroti balso įrašas su triukšmu, pagal (4) formulę apskaičiuojamas triukšmo daugiklis. Šnekos atpažinimo komponentui (*webkitSpeechRecognition*) perduodama žinutė, kad šis klausytųsi mikrofono ir apdorotų į jį grojamą garsą. Šis klausosi mikrofono, kol bus pranešta apie šnekos pabaigą. Pradedamas groti triukšmas reikalingu garsumu, jeigu jis reikalingas, ir triukšmas grojamas tol, kol nesibaigs balso įrašas. Pradėjus groti triukšmą pradedamas groti balso įrašas. Pasibaigus balso įrašui sustabdomas ir triukšmo grojimas. Pasibaigus balso įrašui perduodama žinutė, kad balso įrašas pasibaigė ir reikia stabdyti šnekos atpažinimo komponentą. Šnekos atpažinimo komponentas sustabdomas ir laukiama šnekos atpažinimo rezultatų. „Google“ šnekos atpažintuvas gautą garso įrašą

apdoruoja ir persiunčia apdoroto garso įrašo šnekos atpažinimo rezultatą JSON formatu. Gautas šnekos atpažinimo rezultatas ir informacija apie grotą triukšmą ir SNR (jei triukšmas buvo panaudotas) išsaugomi DB. Po to, šnekos atpažinimo rezultatai yra apdorojami (išskaidomi į pagrindinį ir alternatyvius rezultatus, visi skaitvardžiai paverčiami tekstu, atliekamas WER skaičiavimas visiems išskaidytiems šnekos atpažinimo rezultatams).

Žemiau pateikiamas detalus kiekvieno proceso veiksmo aprašymas.

2.2 lentelė. Garso įrašų apdorojimo procesas (2-asis būdas)

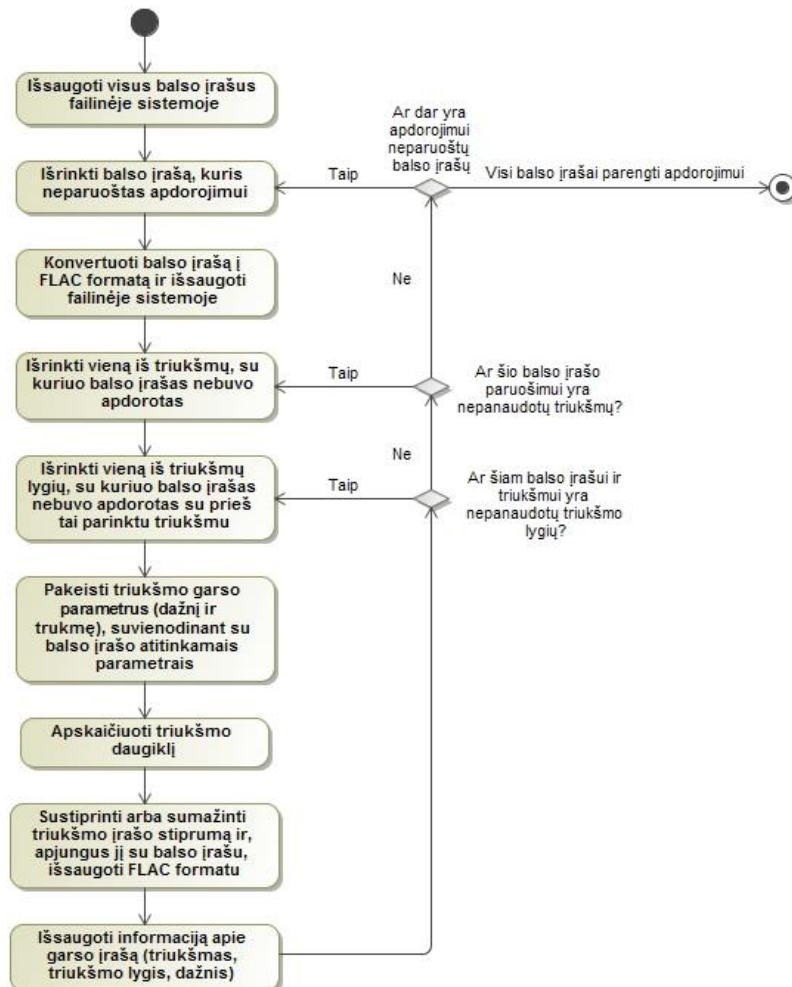
Eil. Nr.	Operacija	Aprašymas
1.	inicijuoti_bendravimo_komponenta()	Atveriamas „Google Chrome“ naršyklės puslapis ir inicijuojamas bendravimo komponentas, kuris prisijungia prie garso įrašų apdorojimo procese esančios programinės įrangos.
2.	inicijuoti_snekos_atpazinimo_komponenta()	Inicijuojamas šnekos atpažinimo komponentas.
5.	parinkti_neapdorota_balso_ir_triuksmo_irasu_kombinacija()	Parenkama balso ir triukšmo įrašų kombinacija, kuri bus perduodama šnekos atpažinimo komponentui. Garso įrašo išrinkimo apdorojimui procesas pavaizduotas 2.6 pav.
6.	paskaiciuoti_triuksmo_daugikli()	Jei išrinktas apdoroti balso įrašas su triukšmu, pagal (4) formulę apskaičiuojamas triukšmo daugiklis (2.4.5 poskyris), kad apjungus balso įrašą su triukšmu būtų gautas parinktas SNR.
7.	pradeti_klausytis_garso_iraso()	Garso įrašų apdorojimo procesas siunčia žinutę bendravimo komponentui dėl mikrofono klausymosi inicijavimo.
8.	pradeti_klausytis_garso_iraso()	Šnekos atpažinimo komponentui (<i>webkitSpeechRecognition</i>) perduodama žinutė, kad šis klausytųsi mikrofono ir apdorotų į jį grojamą garsą.
9.	klausytis_garso_signalo()	Klausomasi mikrofono, kol bus pranešta apie šnekos pabaigą.
12.	groti_triuksmo_ir_balso_irasus()	Pradedamas groti triukšmas, jeigu jis reikalingas, ir triukšmas grojamas tol, kol nesibaigs balso įrašas. Pradėjus groti triukšmą pradedamas groti balso įrašas. Pasibaigus balso įrašui sustabdomas triukšmo grojimas.
13.	stabdyti_snekos_atpazinima()	Perduodama žinutė, kad balso įrašas pasibaigė ir reikia stabdyti šnekos atpažinimo komponentą.
14.	stabdyti_snekos_atpazinima()	Sustabdomas šnekos atpažinimo komponentas ir laukiama šnekos atpažinimo rezultatų.
17.	issaugoti_snekos_atpazinimo_rezultata()	Gautas apdoroto garso įrašo šnekos atpažinimo rezultatas išsaugomas duomenų bazėje. Kartu išsaugoma informacija ir apie garso įrašė panaudotą triukšmą ir SNR (jei triukšmas buvo panaudotas).
18.	apdoroti_snekos_atpazinimo_rezultata()	Atliekamas šnekos atpažinimo rezultato apdorojimas. Detalesnė informacija pateikta 2.4.4 poskyryje.



2.6 pav. Garso įrašo išrinkimo apdorojimui algoritmas

2.4.2. Garso įrašų paruošimo apdorojimui algoritmas

Šiame poskyryje aprašomas garso įrašų paruošimo apdorojimui algoritmas. Paruošti garso įrašus apdorojimui aktualu tik tuomet, kai jie siunčiami apdoroti „Google“ šnekos atpažintuvui pirmuoju būdu (žr. 2.4.1 poskyrį), nes tik tuomet atpažinimui siunčiami garso įrašai, iš anksto apjungti su reikiamais triukšmais, modifikuotais taip, kad būtų gauti tam tikri SNR, taip pat tik pirmuoju būdu siunčiant apdoroti įrašus svarbus garso įrašo formatas (FLAC).



2.7 pav. Garso įrašų paruošimo apdorojimui algoritmas

Naudotojo nurodytame balso įrašų rinkmenų kataloge ir jo visuose vaikiniuose kataloguose surandamos balso įrašų rinkmenos. Išrenkamas balso įrašas, kuris neparuoštas apdorojimui, šis balso įrašas konvertuojamas į FLAC formatą ir išsaugomas failinėje sistemoje kartu su tekstine byla, kurios pavadinimas toks pat, kaip ir balso įrašo. Tekstinėje byloje išsaugomas balso įrašo dažnis (angl. *sample rate*). Parinktasis balso įrašas apjungiamas su visais galimais triukšmais, modifikuotais taip, kad būtų gauti tam tikri SNR (balso įrašų apjungimas su triukšmu aprašytas 2.4.5 poskyryje) ir apjungti garso įrašai taip pat išsaugomi failinėje sistemoje FLAC formatu kartu su tekstinėmis bylomis (bylos pavadinimas toks pat, kaip atitinkamo garso įrašo), kuriose įrašoma informacija apie garso įrašus (triukšmas, SNR, dažnis). Visi šie veiksmai atliekami su visais balso įrašais, kurių yra atrinkti apdorojimui.

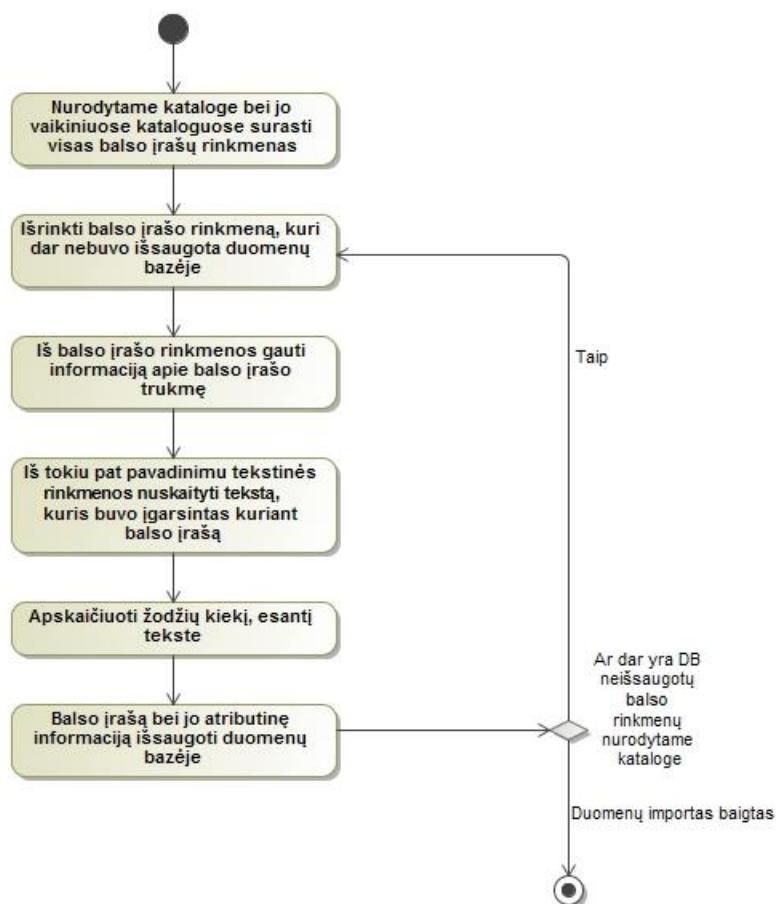
Žemiau pateikiamas detalus kiekvieno proceso veiksmo aprašymas.

2.3 lentelė. Garso įrašų paruošimo apdorojimui algoritmas

Eil. Nr.	Veiksmas	Aprašymas
1.	Išsaugoti visus balso įrašus failinėje sistemoje	Visi balso įrašai išsaugomi failinėje sistemoje.
2.	Išrinkti balso įrašą, kuris neparuoštas apdorojimui	Naudotojo nurodytame balso įrašų rinkmenų kataloge ir jo visuose vaikiniuose kataloguose surandamos balso įrašų rinkmenos. Išrenkamas balso įrašas, kuris dar neparuoštas apdorojimui.
3.	Konvertuoti balso įrašą į FLAC formatą ir išsaugoti failinėje sistemoje	Garso įrašas konvertuojamas į FLAC formatą ir išsaugomas failinėje sistemoje. Be to, sukuriama tekstinė byla tokiu pat pavadinimu, kaip ir balso įrašas, ir joje išsaugomas balso įrašo dažnio parametras.
4.	Išrinkti vieną iš triukšmų, su kuriuo balso įrašas nebuvo apdorotas	Išrenkamas triukšmas, su kuriuo balso įrašas dar nebuvo apdorotas.
5.	Išrinkti vieną iš triukšmų lygių, su kuriuo balso įrašas nebuvo apdorotas su prieš tai parinktu triukšmu	Išrenkamas vienas iš keturių triukšmų lygių (SNR = 10, SNR = 15, SNR = 20, SNR = 25), su kuriuo nebuvo apdorotas balso įrašas su prieš tai parinktu triukšmu.
6.	Pakeisti triukšmo garso parametrus (dažnį ir trukmę), suvienodinant su balso įrašo atitinkamais parametrais	Pakeičiami triukšmo garso parametrai (dažnis ir trukmė), suvienodinant juos su balso įrašo atitinkamais parametrais.
7.	Apskaičiuoti triukšmo daugiklį	Pagal (4) formulę apskaičiuojamas triukšmo daugiklis (2.4.5 poskyris), kad apjungus balso įrašą su triukšmu būtų gautas parinktas SNR.
8.	Sustiprinti arba sumažinti triukšmo įrašo stiprumą ir, apjungus jį su balso įrašu, išsaugoti FLAC formatu	Priklausomai nuo apskaičiuoto triukšmo daugiklio reikšmės sustiprinamas (kai triukšmo daugiklis > 1) arba sumažinamas (kai triukšmo daugiklis < 1) triukšmo įrašo stiprumas ir, apjungus triukšmo įrašą su balso įrašu, gautas garso įrašas išsaugomas FLAC formatu failinėje sistemoje.
9.	Išsaugoti informaciją apie garso įrašą (triukšmas, SNR, dažnis)	Informacija apie garso įrašą (triukšmas, SNR, dažnis) išsaugoma tekstinėje byloje failinėje sistemoje.

2.4.3. Balso įrašų importo procesas

Šiame poskyryje aprašomas turimų balso įrašų importo į eksperimento PĮ procesas. Procesas inicijuojamas iš balso įrašų importo lango (3.6 pav.).



2.8 pav. Balso įrašų importo procesas

Naudotojo parinktame balso įrašų kataloge ir jo visuose vaikiškuose kataloguose surandamos balso įrašų rinkmenos. Tuomet išrenkamas balso įrašas, kuris dar nebuvo išsaugotas duomenų bazėje, ir gaunama informacija apie jo trukmę. Iš tekstinės rinkmenos, kurios pavadinimas yra toks pat, kaip ir balso įrašo, nuskaitytas įgarsintas tekstas, apskaičiuojamas žodžių kiekis įgarsintame tekste. Balso įrašas ir visa jo atributinė informacija (balso įrašo trukmė sekundėmis, tekstas bei žodžių skaičius, esantis įgarsintame tekste) išsaugoma duomenų bazėje. Tai atliekama su visais balso įrašais, esančiais nurodytame kataloge.

Žemiau pateikiamas detalus kiekvieno proceso veiksmo aprašymas.

2.4 lentelė. Balso įrašų importo procesas

Eil. Nr.	Veiksmas	Aprašymas
1.	Nurodytame kataloge bei jo vaikiškuose kataloguose surasti visas balso įrašų rinkmenas	Nurodytame kataloge bei jo vaikiškuose kataloguose surandami visi balso įrašai.
2.	Išrinkti balso įrašo rinkmeną, kuri dar nebuvo išsaugota duomenų bazėje	Išrenkamas balso įrašas, kuris dar nebuvo išsaugotas duomenų bazėje.
3.	Iš balso įrašo rinkmenos gauti informaciją apie balso įrašo trukmę	Iš balso įrašo rinkmenos gaunama informacija apie balso įrašo trukmę.
4.	Iš tokiu pat pavadinimu tekstinės rinkmenos nuskaityti tekstą, kuris buvo įgarsintas kuriant balso įrašą	Iš tokiu pat pavadinimu tekstinės rinkmenos nuskaitytas tekstas, kuris buvo įgarsintas kuriant balso įrašą.
5.	Apskaičiuoti žodžių kiekį, esantį tekste	Apskaičiuojamas žodžių kiekis, esantis įgarsintame tekste.
6.	Balso įrašą bei jo atributinę informaciją išsaugoti duomenų bazėje	Balso įrašas bei jo atributinė informacija (balso įrašo trukmė sekundėmis, tekstas bei žodžių

Eil. Nr.	Veiksmas	Aprašymas
		skaičius, esantis tekste) išsaugoma duomenų bazėje.

2.4.4. Šnekos atpažinimo rezultatų apdorojimo procesas

Šiame poskyryje aprašomas šnekos atpažinimo rezultatų apdorojimo procesas, kuris yra inicijuojamas vykdant garso įrašų apdorojimo procesą (2.4.1 poskyris).



2.9 pav. Šnekos atpažinimo rezultatų apdorojimo procesas

Iš „Google“ šnekos atpažintuvo gavus šnekos atpažinimo rezultatus jie pirmiausiai išskaidomi į pagrindinį ir alternatyvius (informacija apie „Google“ šnekos atpažintuvo gražinamus šnekos atpažinimo rezultatus pateikta 3.1.2 poskyryje). Norint užtikrinti kuo tikslesnius šnekos atpažintuvo efektyvumo tyrimo rezultatus, prieš atliekant WER skaičiavimus šnekos atpažinimo rezultatuose visi skaitvardžiai, kurie rezultatuose pateikti skaičių forma ar kaip skaičiai su simboliais, paverčiami atitinkamu tekstu. Kai visi skaitvardžiai apdoroti, atliekamas WER skaičiavimas visiems išskaidytiems šnekos atpažinimo rezultatams. Galiausiai, pagrindinis ir alternatyvūs rezultatai su WER reikšmėmis išsaugomi duomenų bazėje.

Žemiau pateikiamas detalus kiekvieno proceso veiksmo aprašymas.

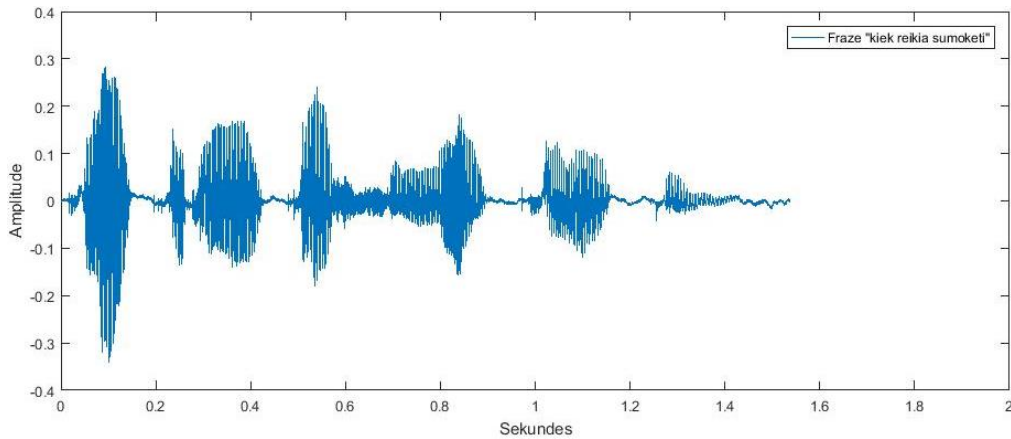
2.5 lentelė. Šnekos atpažinimo rezultatų apdorojimo procesas

Eil. Nr.	Veiksmas	Aprašymas
1.	Paimti šnekos atpažinimo rezultatą ir jį išskaidyti į pagrindinį bei alternatyvius rezultatus	Gautas šnekos atpažinimo rezultatas išskaidomas į pagrindinį bei alternatyvius rezultatus. Gauto šnekos atpažinimo rezultato pavyzdys pateiktas 3.1.2 poskyryje.
2.	Pagrindiniame ir alternatyviuose šnekos atpažinimo rezultatuose apdoroti skaitvardžius	Pagrindiniame ir alternatyviuose šnekos atpažinimo rezultatuose apdorojami skaitvardžiai. Skaitvardžių apdorojimas aprašytas 2.4.6 poskyryje.
3.	Apskaičiuoti WER visiems išskaidytiems šnekos atpažinimo rezultatams	Visiems išskaidytiems šnekos atpažinimo rezultatams paskaičiuojamos WER reikšmės. Skaičiavimas aprašytas 2.4.7 poskyryje.
4.	Išsaugoti pagrindinį ir alternatyvius rezultatus ir WER	Išsaugomi pagrindinis, alternatyvūs rezultatai ir WER reikšmės.

2.4.5. Triukšmo daugiklio skaičiavimas

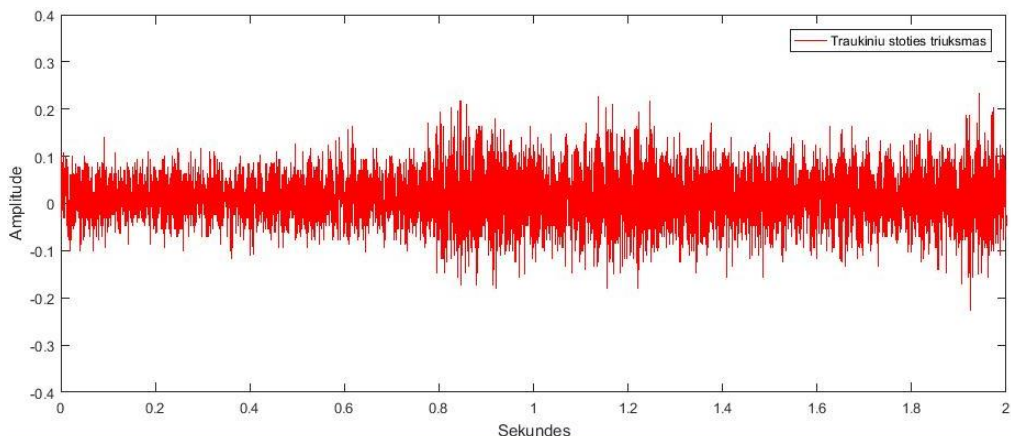
Balso įrašą apjungiant su triukšmu yra naudojamas SNR – balso ir triukšmo signalų santykis. Jei SNR reikšmė yra teigiama, tai balso signalas yra stipresnis už triukšmo signalą, jei neigiama – balso signalas yra silpnesnis už triukšmo signalą. Žemiau pateikti garso signalų pavyzdžiai, gauti „Matlab“ programinės įrangos pagalba.

2.10 pav. pavaizduotas frazės „kiek reikia sumokėti“ garso signalas. Vertikalioje ašyje vaizduojama garso signalo galia momentu, kuris vaizduojamas horizontalioje ašyje.



2.10 pav. Frazės „kiek reikia sumokėti“ garso signalas

2.11 pav. pavaizduotas triukšmo „Traukinių stotis“ signalo pirmosios 2 sekundės. Vertikalioje ašyje vaizduojama garso signalo galia momentu, kuris vaizduojamas horizontalioje ašyje.



2.11 pav. Triukšmo „Traukinių stotis“ signalas

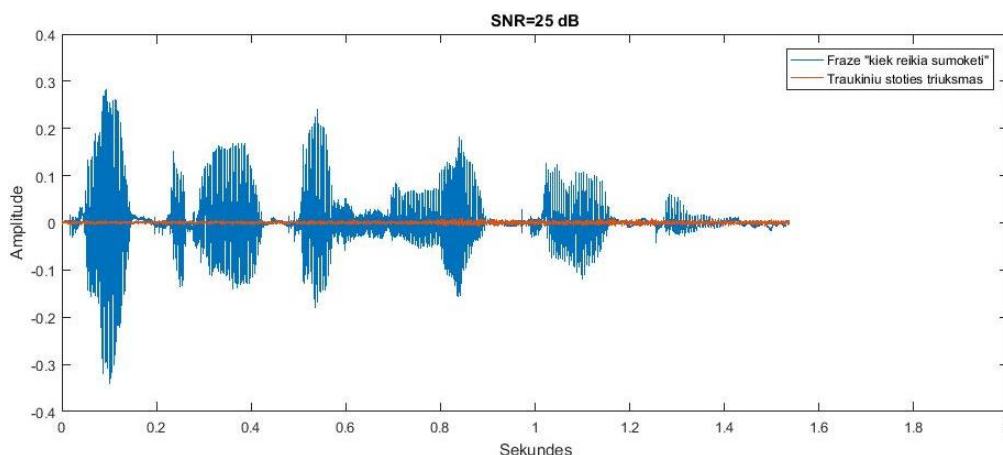
2.12 pav. pavaizduotas nekeistas frazės „kiek reikia sumokėti“ garso signalas (2.10 pav.), apjungtas su modifikuotu triukšmo „Traukinių stotis“ signalu. Mėlyna spalva vaizduojamas balso įrašas (nekeistas signalas), o raudona – triukšmo signalas (modifikuotas signalas).

Norint gauti šį rezultatą, reikia:

- Nuskaityti balso ir triukšmo įrašus;
- Triukšmo signalo dažnį (angl. *sample rate*) suvienodinti su balso įrašo dažniu;
- Triukšmo signalo ilgį (trukmę) suvienodinti su balso įrašo signalu (šiuo atveju, balso įrašas trunka 1,53 sekundes, todėl triukšmo signalas nuo 1,53 sekundės buvo pašalintas);
- Pagal (4) formulę apskaičiuoti triukšmo daugiklį;
- Triukšmo signalą padauginti iš triukšmo daugiklio.

Šiuo atveju, triukšmo signalas sumažintas tiek, kad balso įrašo ir triukšmo signalų santykis (SNR) būtų lygus 25 dB.

Analogiški veiksmai buvo atlikti, kai balso įrašo ir triukšmo signalų santykis (SNR) lygus 20 dB (2.13 pav.), 15 dB (2.14 pav.), 10 dB (2.15 pav.), 5 dB (2.16 pav.), 0 dB (2.17 pav.) ir -5 dB (2.18 pav.).

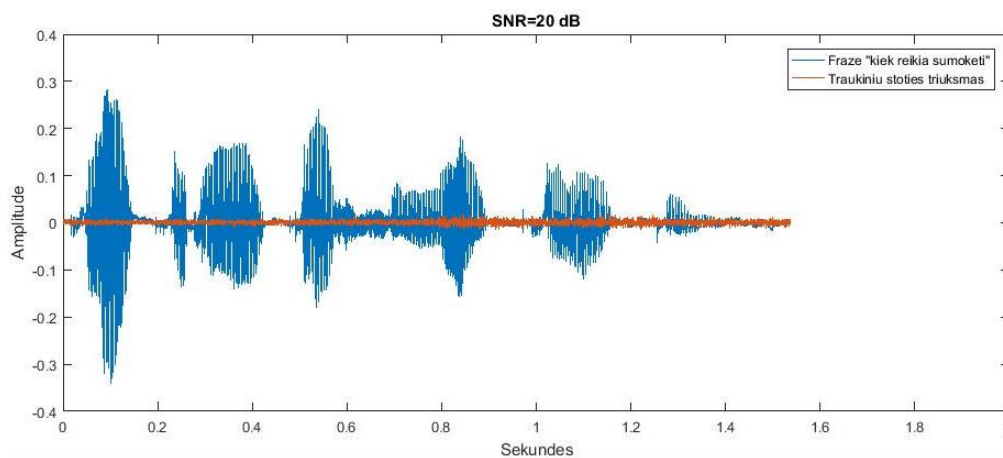


2.12 pav. Apjungtas frazės „kiek reikia sumokėti“ garso signalas su triukšmo „Traukinių stotis“ signalu (SNR=25 dB)

2.13 pav. pavaizduotas nekeistas frazės „kiek reikia sumokėti“ garso signalas (2.10 pav.), apjungtas su modifikuotu triukšmo „Traukinių stotis“ signalu. Mėlyna spalva vaizduojamas balso įrašas (nekeistas signalas), o raudona – triukšmo signalas (modifikuotas signalas).

Šiuo atveju, triukšmo signalas sumažintas tiek, kad balso įrašo ir triukšmo signalų santykis (SNR) būtų lygus 20 dB.

Vertikalojoje ašyje vaizduojama garso signalo galia momentu, kuris vaizduojamas horizontaliojoje ašyje.

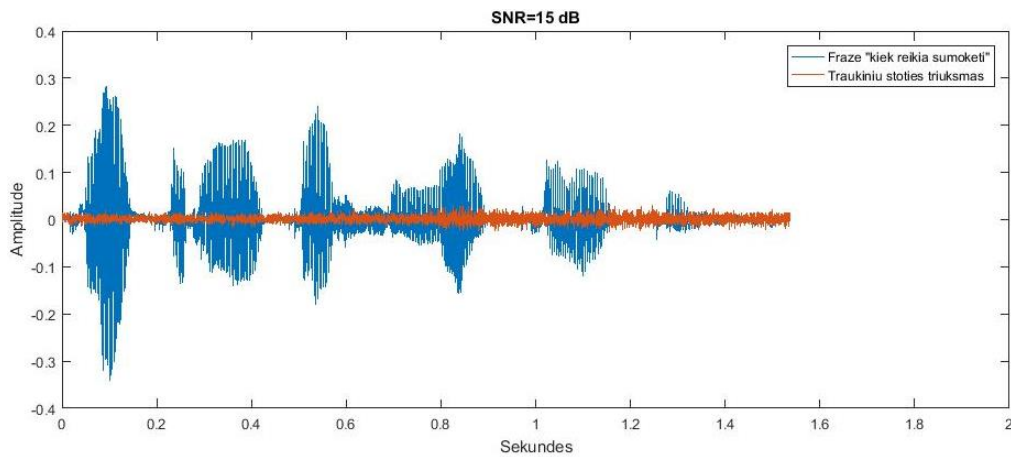


2.13 pav. Apjungtas frazės „kiek reikia sumokėti“ garso signalas su triukšmo „Traukinių stotis“ signalu (SNR=20 dB)

2.14 pav. pavaizduotas nekeistas frazės „kiek reikia sumokėti“ garso signalas (2.10 pav.), apjungtas su modifikuotu triukšmo „Traukinių stotis“ signalu. Mėlyna spalva vaizduojamas balso įrašas (nekeistas signalas), o raudona – triukšmo signalas (modifikuotas signalas).

Šiuo atveju, triukšmo signalas sumažintas tiek, kad balso įrašo ir triukšmo signalų santykis (SNR) būtų lygus 15 dB.

Vertikalojoje ašyje vaizduojama garso signalo galia momentu, kuris vaizduojamas horizontaliojoje ašyje.

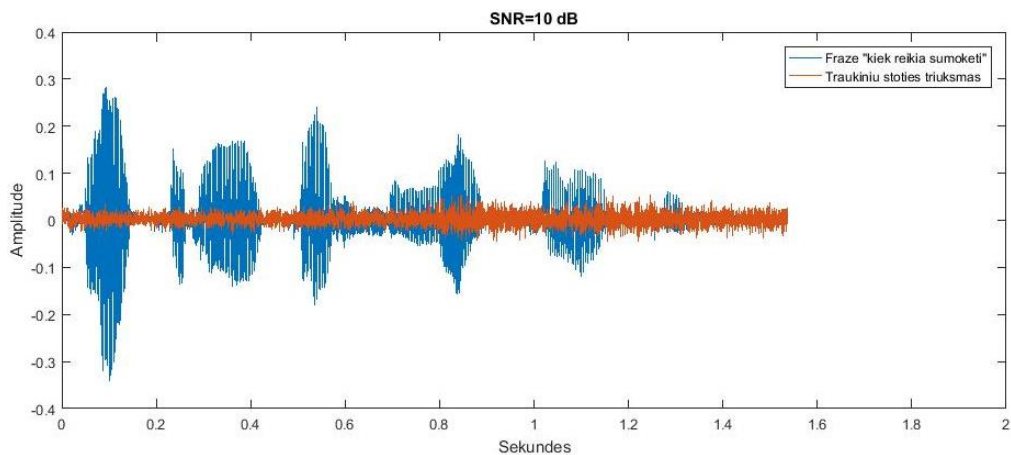


2.14 pav. Apjungtas frazės „kiek reikia sumokėti“ garso signalas su triukšmo „Traukinių stotis“ signalu (SNR=15 dB)

2.15 pav. pavaizduotas nekeistas frazės „kiek reikia sumokėti“ garso signalas (2.10 pav.), apjungtas su modifikuotu triukšmo „Traukinių stotis“ signalu. Mėlyna spalva vaizduojamas balso įrašas (nekeistas signalas), o raudona – triukšmo signalas (modifikuotas signalas).

Šiuo atveju, triukšmo signalas sumažintas tiek, kad balso įrašo ir triukšmo signalų santykis (SNR) būtų lygus 10 dB.

Vertikaloje ašyje vaizduojama garso signalo galia momentu, kuris vaizduojamas horizontalioje ašyje.

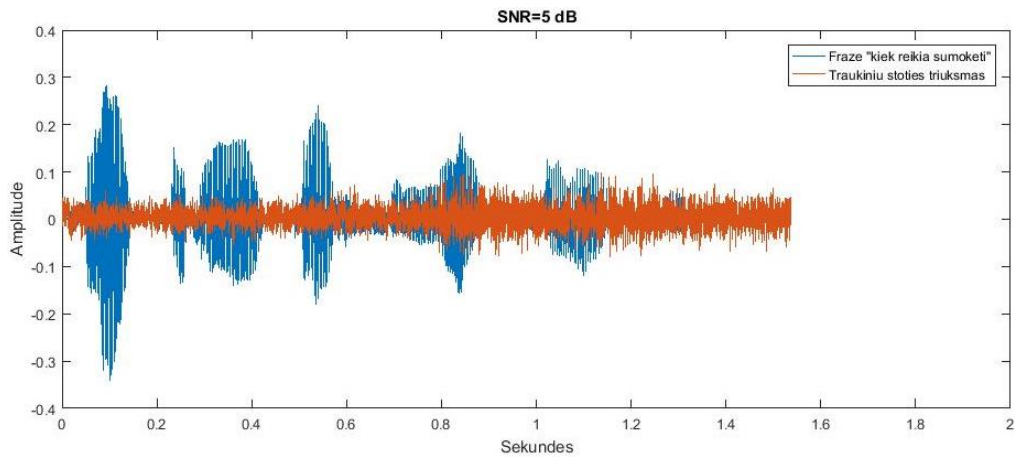


2.15 pav. Apjungtas frazės „kiek reikia sumokėti“ garso signalas su triukšmo „Traukinių stotis“ signalu (SNR=10 dB)

2.16 pav. pavaizduotas nekeistas frazės „kiek reikia sumokėti“ garso signalas (2.10 pav.), apjungtas su modifikuotu triukšmo „Traukinių stotis“ signalu. Mėlyna spalva vaizduojamas balso įrašas (nekeistas signalas), o raudona – triukšmo signalas (modifikuotas signalas).

Šiuo atveju, triukšmo signalas sumažintas tiek, kad balso įrašo ir triukšmo signalų santykis (SNR) būtų lygus 5 dB.

Vertikaloje ašyje vaizduojama garso signalo galia momentu, kuris vaizduojamas horizontalioje ašyje.

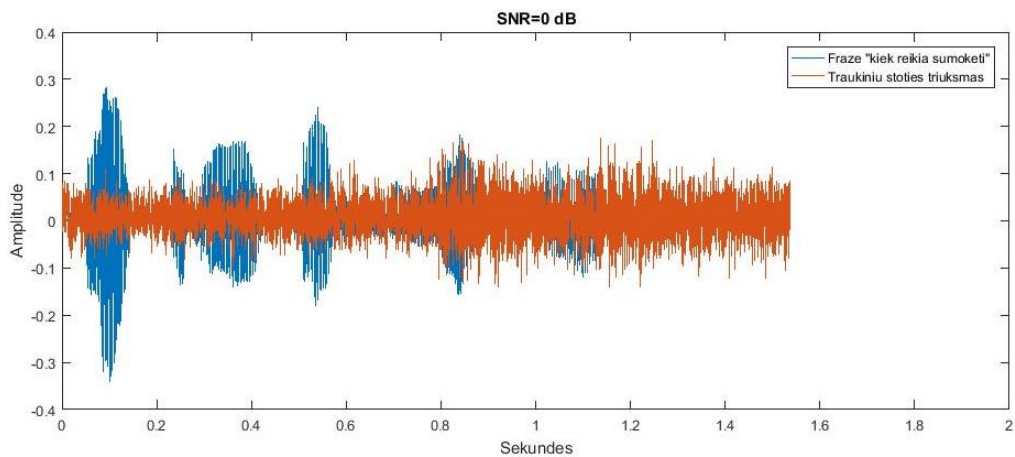


2.16 pav. Apjungtas frazės „kiek reikia sumokėti“ garso signalas su triukšmo „Traukinių stotis“ signalu (SNR=5 dB)

2.17 pav. pavaizduotas nekeistas frazės „kiek reikia sumokėti“ garso signalas (2.10 pav.), apjungtas su modifikuotu triukšmo „Traukinių stotis“ signalu. Mėlyna spalva vaizduojamas balso įrašas (nekeistas signalas), o raudona – triukšmo signalas (modifikuotas signalas).

Šiuo atveju, triukšmo signalas sumažintas tiek, kad balso įrašo ir triukšmo signalų santykis (SNR) būtų lygus 0 dB.

Vertikaliajoje ašyje vaizduojama garso signalo galia momentu, kuris vaizduojamas horizontalioje ašyje.

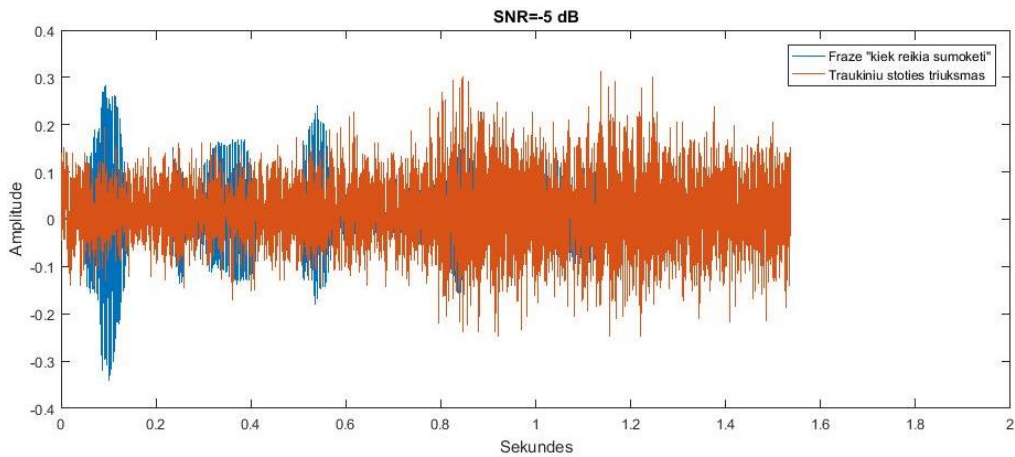


2.17 pav. Apjungtas frazės „kiek reikia sumokėti“ garso signalas su triukšmo „Traukinių stotis“ signalu (SNR=0 dB)

2.18 pav. pavaizduotas nekeistas frazės „kiek reikia sumokėti“ garso signalas (2.10 pav.), apjungtas su modifikuotu triukšmo „Traukinių stotis“ signalu. Mėlyna spalva vaizduojamas balso įrašas (nekeistas signalas), o raudona – triukšmo signalas (modifikuotas signalas).

Šiuo atveju, triukšmo signalas sumažintas tiek, kad balso įrašo ir triukšmo signalų santykis (SNR) būtų lygus -5 dB.

Vertikaliajoje ašyje vaizduojama garso signalo galia momentu, kuris vaizduojamas horizontalioje ašyje.



2.18 pav. Apjungtas frazės „kiek reikia sumokėti“ garso signalas su triukšmo „Traukinių stoties“ signalu (SNR=-5 dB)

Norint gauti nurodytą SNR, reikia apskaičiuoti triukšmo daugiklį. Triukšmo daugiklio skaičiavimas atliekamas „Matlab“ programinėje įrangoje. Apskaičiuojama taip:

$$\mathbf{Triukšmo\ daugiklis} = \frac{\mathbf{rms(balso\ įrašo)}}{\mathbf{rms(triukšmo) * \sqrt{db2pow(SNR)}}}; \quad (4)$$

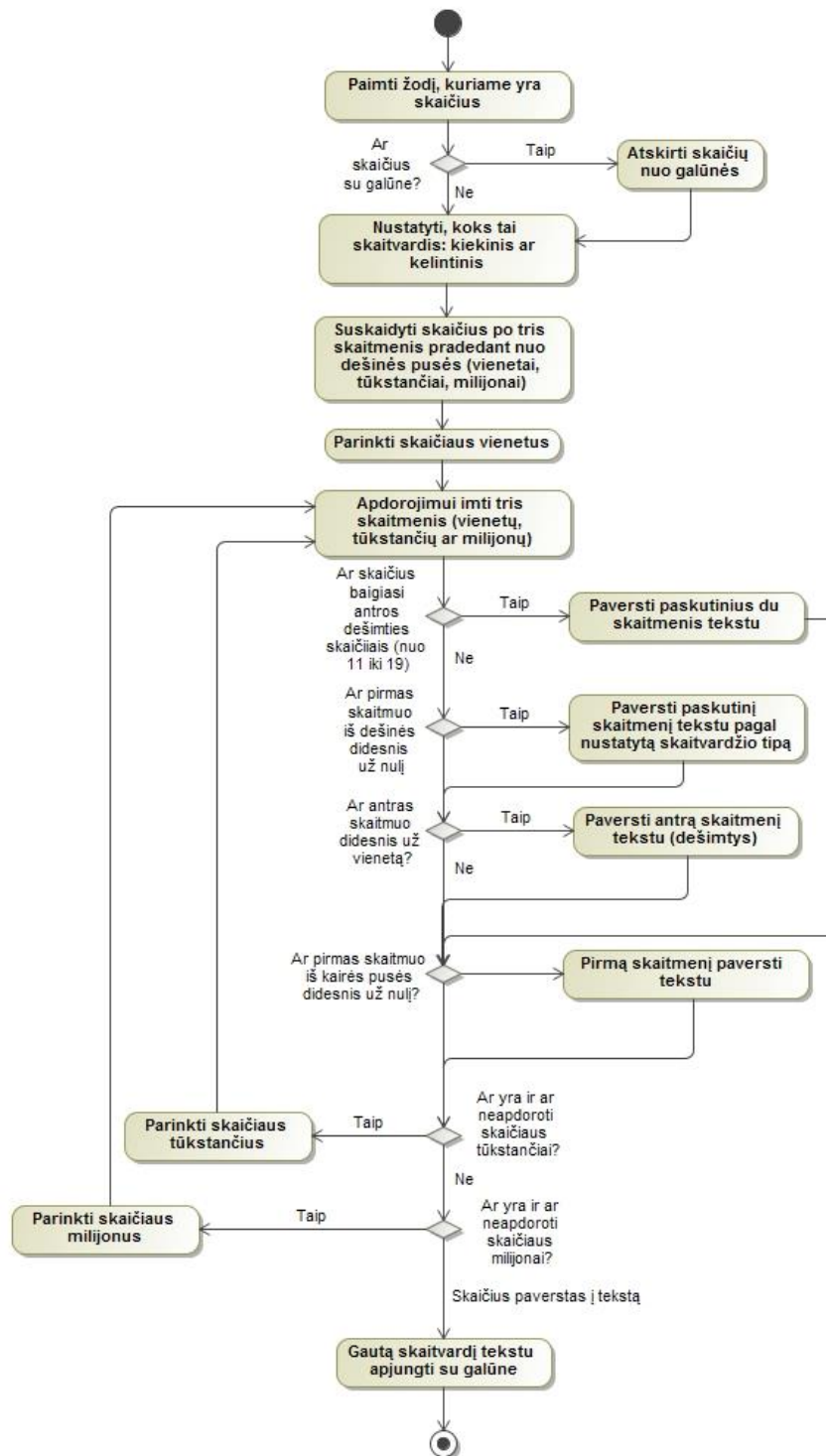
čia:

- *rms()* – funkcija, skirta apskaičiuoti reikšmių kvadrato vidurkio šaknį (angl. *root-mean-square value of a signal*);
- *db2pow()* – funkcija, paverčianti decibelus į galią;
- *SNR* – balso įrašo ir triukšmo signalų santykis, kurį norime pasiekti apjungę balso ir triukšmo signalus. Išreiškiamas decibelais (dB).

2.4.6. Skaitvardžių apdorojimas

Įgarsinant tekstą priklausomai nuo konteksto skaitvardžiai buvo atitinkamai linksniuojami. „Google“ šnekos atpažintuvo pateiktuose šnekos atpažinimo rezultatuose skaitvardžiai yra pateikiami ir skaičiais, ir tekstu (net ir siunčiant atpažinti tą patį balso įrašą skaitvardžiai šnekos atpažinimo rezultate gali būti pateikti skirtinga forma). Pvz., frazė „tūkstantis devyni šimtai keturiasdešimtaisiais metais“ šnekos atpažinimo rezultatuose gali būti pateikta kaip „1940 metais“ ir „1940-aisiais metais“. Lyginti šių šnekos atpažinimo rezultatų su įgarsintu tekstu negalima, nes tai darytų įtaką WER skaičiavimams. Dėl šios priežasties prieš skaičiuojant WER šnekos atpažinimo rezultatuose atitinkamai apdorojami skaitvardžiai, kurie rezultatuose pateikiami kaip skaičiai (pvz., „1940“) ir kaip skaičiai su simboliais (pvz., „1940-aisiais“, „1940i“), t. y. atpažinimo rezultatuose skaičiai paverčiami atitinkamu tekstu, kaip ir yra įgarsintame tekste.

Žemiau pateiktas skaitvardžių apdorojimo procesas.



2.19 pav. Skaitvardžių apdorojimo procesas

„Google“ šnekos atpažintuvas dažnai, bet ne visada, atpažinimo rezultatuose naudoja skaičius. Norint išgauti kuo teisingesnį tekstų palyginimo rezultatą būtina skaičius paversti tekstu. Atpažintas tekstas suskaidomas į žodžius, tada surandamas pirmas žodis, kuriame yra skaičius. Dažnu atveju po skaičiaus būna nurodyta galūnė, taigi skaičius yra atskiriamas nuo teksto (galūnės) ir suskaidomas po tris skaitmenis, skaičius apdorojamas paverčiant jį tekstu. Jeigu šnekos atpažintuvo pateiktame tekste buvo nurodyta skaičiaus galūnė, tai paskutinis žodis šiame iš skaičiaus paverstame tekste nustatomas be galūnės, o pati galūnė, kuri pradžioje buvo atskirta nuo skaičiaus, yra pridama paversto skaičiaus į tekstą pabaigoje.

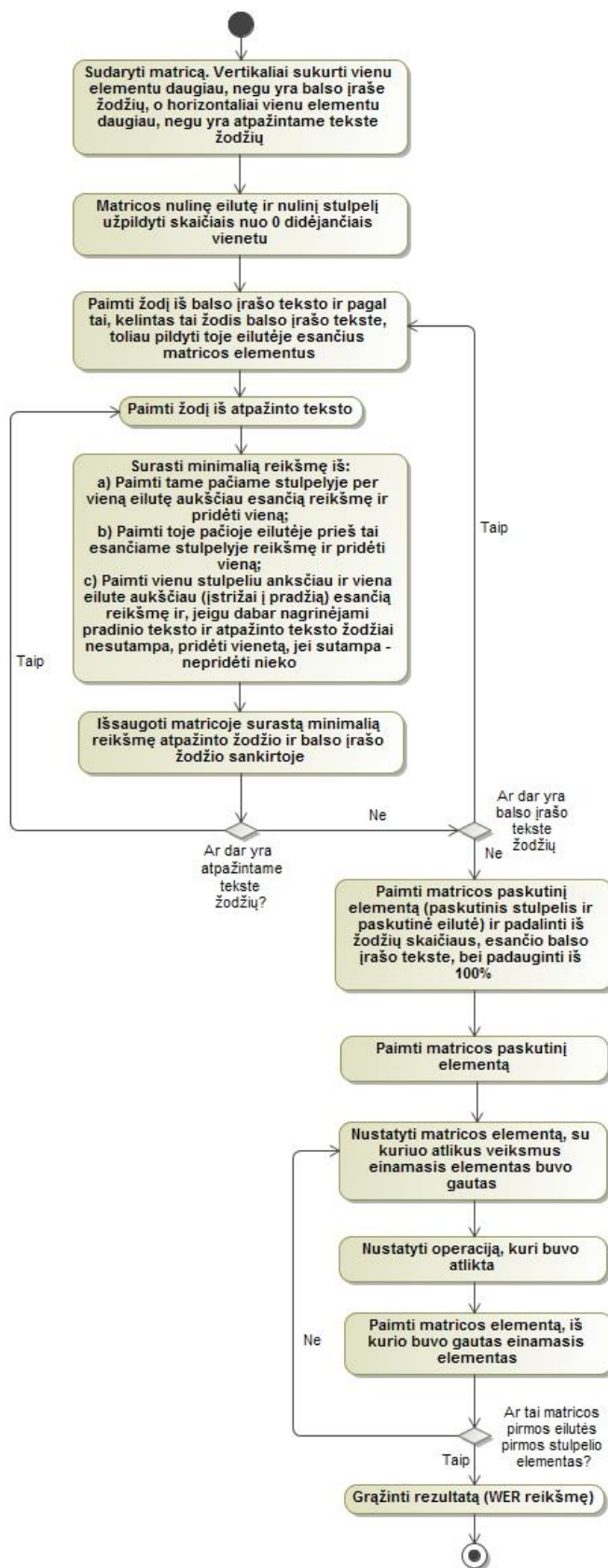
Žemiau pateikiamas detalus kiekvieno proceso veiksmo aprašymas.

2.6 lentelė. Skaitvardžių apdorojimo procesas

Eil. Nr.	Veiksmas	Aprašymas
1.	Paimti žodį, kuriame yra skaičius	Frazė suskaidoma į žodžius ir išrenkamas žodis, kuriame yra skaitmenų.
2.	Atskirti skaičių nuo galūnės	Iš žodžio, kuriame yra skaitmenų, paimamas tik skaičius, kurį reikia paversti tekstu.
3.	Nustatyti, koks tai skaitvardis: kiekinis ar kelintinis	Jeigu po žodyje esančio skaičiaus buvo dar teksto, tai šis tekstas analizuojamas ir nustatoma, koks tai skaitvardis. Jei galūnės nebuvo automatiškai nustatoma, jog tai kiekinis skaitvardis ir jo galūnė turi būti vyriškos giminės vardininko linksnyje.
4.	Suskaidyti skaičius po tris skaitmenis pradedant nuo dešinės pusės (vienetai, tūkstančiai, milijonai)	Gautas skaičius suskaidomas į vienetus, tūkstančius ir milijonus – tris atskirus skaičius.
5.	Parinkti skaičiaus vienetus	Perduodama, jog reikia apdoroti vienetus.
6.	Apdorojimui imti tris skaitmenis (vienetų, tūkstančių ar milijonų)	Paimami nurodyti skaičiaus skaitmenys.
7.	Paversti paskutinius du skaitmenis tekstu	Jeigu skaičiaus dešimtys ir vienetai sudaro skaičių nuo 11 iki 19, skaičius paverčiamas tekstu. Jeigu tai yra pirmas žodis ir galūnė buvo žodyje su skaičiumi, tai šis skaitvardis pateikiamas be galūnės.
8.	Paversti paskutinį skaitmenį tekstu pagal nustatytą skaitvardžio tipą	Paskutinis skaitmuo paverčiamas tekstu.
9.	Paversti antrą skaitmenį tekstu (dešimtys)	Skaičiaus dešimtys paverčiami tekstu.
10.	Pirmą skaitmenį paversti tekstu	Skaičiaus šimtai paverčiami tekstu.
11.	Parinkti skaičiaus tūkstančius	Nustatoma, kad turi būti apdoroti skaičiaus tūkstančių dalis, taip pat nusprendžiama, kokią galūnę turės žodis „tūkstančiai“.
12.	Parinkti skaičiaus milijonus	Nustatoma, kad turi būti apdoroti skaičiaus milijonų dalis, taip pat nusprendžiama, kokią galūnę turės žodis „milijonai“.
13.	Gautą skaitvardį tekstu apjungti su galūne	Jeigu galūnė buvo aptikta ir atskirta antrame proceso žingsnyje, gautas tekstas apjungiamas su galūne.

2.4.7. WER skaičiavimas

WER skaičiavimui naudojamas Levenšteino atstumo skaičiavimo algoritmas (Levenshtein, 1965), pritaikytas žodžiams. WER skaičiavimo procesas inicijuojamas vykdant šnekos atpažinimo rezultatų apdorojimo procesą (2.4.4 poskyris).



2.20 pav. WER skaičiavimo algoritmas

Klaidingų žodžių santykio (WER) paieškos algoritmas lygina dvi frazes – pradinę ir atpažintą. Įterpimų, pakeitimų ir panaikinimų reikšmės parodo, kiek ir kokių pakeitimų šnekos atpažintuvas padarė palyginus su įgarsinta fraze. Norint apskaičiuoti šias reikšmes sudaroma matrica, į kurią surašomos apskaičiuotos reikšmės. Šios reikšmės parodo trumpiausią kelią pakeisti vieną frazę į kitą.

Žemiau pateikiamas detalus kiekvieno proceso veiksmo aprašymas.

2.7 lentelė. WER skaičiavimo algoritmas

Eil. Nr.	Veiksmas	Aprašymas
1.	Sudaryti matricą. Vertikaliai sukurti vienu elementu daugiau, negu yra balso įrašė žodžių, o horizontaliai vienu elementu daugiau, negu yra atpažintame tekste žodžių	Sudaroma matrica (pavyzdys pateiktas 2.21 pav.). Vertikaliai sukuriama vienu elementu daugiau, negu yra balso įrašė žodžių, o horizontaliai vienu elementu daugiau, negu yra atpažintame tekste žodžių.
2.	Matricos nulinę eilutę ir nulinių stulpelių užpildyti skaičiais nuo 0 didėjančiais vienetu	Matricos nulinė eilutė ir nulinis stulpelis užpildomas skaičiais nuo 0, didėjančiais vienetu.
3.	Paimti žodį iš balso įrašo teksto ir pagal tai, kelintas tai žodis balso įrašo tekste, toliau pildyti toje eilutėje esančius matricos elementus	Paimamas žodis iš balso įrašo teksto ir pagal tai, kelintas tai žodis balso įrašo tekste, toliau užpildomi toje eilutėje esantys matricos elementai.
4.	Paimti žodį iš atpažinto teksto	Paimamas žodis iš atpažinto teksto.
5.	Surasti minimalią reikšmę iš: a) Paimti tame pačiame stulpelyje per vieną eilutę aukščiau esančią reikšmę ir pridėti vieną; b) Paimti toje pačioje eilutėje prieš tai esančiame stulpelyje reikšmę ir pridėti vieną; c) Paimti vienu stulpeliu anksčiau ir viena eilute aukščiau (įstrižai į pradžią) esančią reikšmę ir, jeigu dabar nagrinėjami pradinio teksto ir atpažinto teksto žodžiai nesutampa, pridėti vieneta, jei sutampa - nepridėti nieko	Surandama minimali reikšmė iš: a) Paimama tame pačiame stulpelyje per vieną eilutę aukščiau esanti reikšmė ir pridamas vienetas; b) Paimama toje pačioje eilutėje prieš tai esančiame stulpelyje esanti reikšmė ir pridamas vienetas; c) Paimamas vienu stulpeliu anksčiau ir viena eilute aukščiau (įstrižai į pradžią) esanti reikšmė ir, jeigu dabar nagrinėjami pradinio teksto ir atpažinto teksto žodžiai nesutampa, pridamas vienetas, jei sutampa – nepridedama.
6.	Išsaugoti matricoje surastą minimalią reikšmę atpažinto žodžio ir balso įrašo žodžio sankirtoje	Atpažinto žodžio ir balso įrašo žodžio sankirtoje išsaugoma 5 žingsnyje surasta minimali reikšmė.
7.	Paimti matricos paskutinį elementą (paskutinis stulpelis ir paskutinė eilutė) ir padalinti iš žodžių skaičiaus, esančio balso įrašo tekste, bei padauginti iš 100%	Reikšmė, esanti paskutiniame matricos stulpelyje ir paskutinėje eilutėje, padalinama iš žodžių skaičiaus, esančio balso įrašo tekste, ir padauginama iš 100 % – taip gaunama WER reikšmė.
8.	Paimti matricos paskutinį elementą	Norint paskaičiuoti, kiek buvo atlikta įterpimų, pakeitimų ir panaikinimų, pradedama nagrinėti matrica nuo paskutinio elemento.
9.	Nustatyti matricos elementą, su kuriuo atlikus veiksmus einamasis elementas buvo gautas	Patikrinama, ar esamos matricos celėje susikertantys atpažinto ir balso įrašo tekstų atitinkami žodžiai sutampa. Šis elementas buvo gautas iš: a) Eilutėje aukščiau ir stulpeliu anksčiau esančio matricos elemento, jeigu eilutėje aukščiau ir stulpeliu anksčiau esantis matricos elementas yra lygus esamam elementui, kai žodžiai sutampa, arba yra mažesnis vienetu nei einamasis elementas ir žodžiai nesutampa; b) Stulpeliu anksčiau esančio elemento toje pačioje eilutėje, jeigu stulpeliu

Eil. Nr.	Veiksmas	Aprašymas
		<p>anksčiau esantis elementas yra vienetu mažesnis nei einamasis elementas;</p> <p>c) Eilute anksčiau tame pačiame stulpelyje esančio elemento, jeigu šio elemento reikšmė yra vienetu mažesnė.</p> <p>Pastaba: lyginama su aukščiau esančios eilutės reikšmėmis tik tokiu atveju, kai egzistuoja aukščiau esanti eilutė, ir tikrinama anksčiau esančiame stulpelyje reikšmė tik jeigu egzistuoja anksčiau esantis stulpelis.</p>
10.	Nustatyti operaciją, kuri buvo atlikta	<p>Jeigu einamasis elementas buvo gautas iš kairėje pusėje nuo jo esančio elemento, tai buvo atlikta įterpimo operacija.</p> <p>Jei einamasis elementas buvo gautas iš elemento, esančio per vieną aukščiau nei einamasis, tai buvo atlikta panaikinimo operacija.</p> <p>Jei einamasis elementas buvo gautas iš įstrižai į kairę ir į viršų esančio elemento ir jų reikšmės nesutampa, tai buvo atlikta pakeitimo operacija.</p> <p>Jei žodžiai sutampa, tai nebuvo atlikta jokia operacija ir žodžiai yra lygūs.</p>
11.	Paimti matricos elementą, iš kurio buvo gautas einamasis elementas	Parenkamas matricos elementas, kuris buvo nustatytas 9-ame žingsnyje.
12.	Grąžinti rezultatą (WER reikšmę)	Pateikiama WER reikšmė.

Žemiau pateiktas matricos pavyzdys. Buvo ištarta frazė „aš esu aš“, o gautas šnekos atpažinimo rezultatas – „esu tu“. Paskutiniame matricos elemente esantis skaičius parodo klaidingų žodžių skaičių atpažintame tekste.

	<i>esu</i>	<i>tu</i>
	0	2
<i>aš</i>	1	2
<i>esu</i>	2	1
<i>aš</i>	3	2

2.21 pav. Matricos pavyzdys

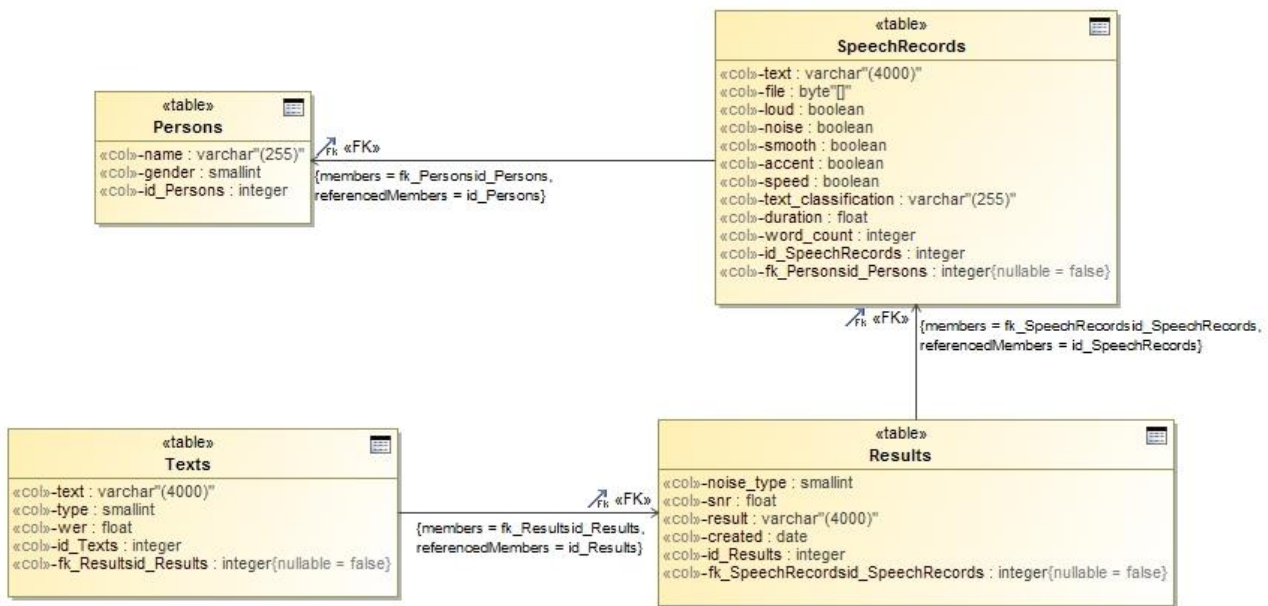
Žemiau pateiktame įterpimų, pakeitimų ir panaikinimų skaičiavimo pavyzdyje parodoma, kaip iš 2.21 pav. sudarytos matricos, geltonai pažymėtame „trumpiausio atstumo“ kelyje, eidami iš paskutinio matricos elemento galime nustatyti, kokios operacijos buvo atliktos norint iš ištartos frazės teksto gauti šnekos atpažinimo rezultatą.

	<i>esu</i>	<i>tu</i>
<i>aš</i>	Ištrinta	
<i>esu</i>		Lygu
<i>aš</i>		Pakeista

2.22 pav. Įterpimų, pakeitimų, panaikinimų skaičiavimo pavyzdys

2.5. Duomenų modelis

Žemiau pateikta duomenų bazės schema.



2.23 pav. Duomenų bazės schema

Lentelėje „Persons“ saugoma informacija apie kalbėtojus (t. y. apie asmenis, įgarsinusius balso įrašus). Detalesnė informacija pateikta 2.8 lentelėje.

2.8 lentelė. Lentelės „Persons“ aprašas

Atributas	Duomenų tipas	Aprašymas
name	varchar"(255)	Kalbėtojo vardas, pavardė
gender	smallint	Kalbėtojo lytis (0 – moteris; 1 – vyras)
id_Persons	integer	Lentelės „Persons“ identifikatorius, pirminis raktas (angl. <i>primary key</i>)

Lentelėje „SpeechRecords“ saugomi balso įrašai bei informacija apie juos. Detalesnė informacija pateikta 2.9 lentelėje.

2.9 lentelė. Lentelės „SpeechRecords“ aprašas

Atributas	Duomenų tipas	Aprašymas
text	varchar"(4000)	Tekstas, kuris yra įgarsintas balso įrašė
file	byte"[]"	Balso įrašas
loud	boolean	Požymis, nurodantis, kad kalbėtojas kalba garsiai
noise	boolean	Požymis, nurodantis, kad balso įrašas yra su foniniu triukšmu
smooth	boolean	Požymis, nurodantis, kad kalbėtojas kalba sklandžiai
accent	boolean	Požymis, nurodantis, kad kalbėtojas kalba su akcentu
speed	boolean	Požymis, nurodantis, kad kalbėtojas kalba labai greitai
text_classification	varchar"(255)	Šnekos tipas (keleto žodžių seka; pavienės balso komandos; rišli šneka; spontaniška šneka)
duration	float	Balso įrašo trukmė sekundėmis
word_count	integer	Žodžių skaičius įgarsintame tekste
id_SpeechRecords	integer	Lentelės „SpeechRecord“ identifikatorius, pirminis raktas (angl. <i>primary key</i>)

Atributas	Duomenų tipas	Aprašymas
fk_Personsid_Persons	integer	Nuoroda į lentelę „Persons“, išorinis raktas (angl. <i>foreign key</i>)

Lentelėje „Results“ saugomi šnekos atpažinimo rezultatai. Detalesnė informacija pateikta 2.10 lentelėje.

2.10 lentelė. Lentelės „Results“ aprašas

Atributas	Duomenų tipas	Aprašymas
noise_type	smallint	Triukšmo tipas (1 – baltasis triukšmas; 2 – traukinių stotis; 3 – važiavimas automobiliu; 4 – automobilių spūstis)
snr	float	Balso įrašo ir triukšmo signalų santykis decibelais (dB)
result	varchar"(4000)	JSON formatu grąžintas šnekos atpažinimo rezultatas
created	date	Šnekos atpažinimo rezultato gavimo data ir laikas
id_Results	integer	Lentelės „Results“ identifikatorius, pirminis raktas (angl. <i>primary key</i>)
fk_SpeechRecordsid_SpeechRecords	integer	Nuoroda į lentelę „SpeechRecords“, išorinis raktas (angl. <i>foreign key</i>)

Lentelėje „Texts“ saugomi išskaidyti šnekos atpažinimo rezultatai (pagrindinis ir alternatyvūs). Detalesnė informacija pateikta 2.11 lentelėje.

2.11 lentelė. Lentelės „Texts“ aprašas

Atributas	Duomenų tipas	Aprašymas
text	varchar"(4000)	Šnekos atpažinimo rezultatas (tekstas)
type	smallint	Šnekos atpažinimo rezultato tipas (1 – pagrindinis; 2 – alternatyvus)
wer	float	Apskaičiuota WER (klaidingų žodžių dažnis) reikšmė
id_Texts	integer	Lentelės „Texts“ identifikatorius, pirminis raktas (angl. <i>primary key</i>)
fk_Resultsid_Results	integer	Nuoroda į lentelę „Results“, išorinis raktas (angl. <i>foreign key</i>)

3. SPRENDIMO REALIZACIJA IR TESTAVIMAS

Šiame skyriuje pateikiamas eksperimento programinės įrangos aprašymas: jos sandara, realizacija, pateikiama keletas PĮ langų ir jų aprašymas. Taip pat pateikiama informacija apie eksperimento PĮ testavimą.

3.1. Sprendimo realizacijos ir veikimo aprašas

Sukurta eksperimento programinė įranga susideda iš:

- **Duomenų įvedimo posistemė** – skirta turimiems balso įrašams į DB sukelti, įrašyti naujus balso įrašus, nurodyti balso įrašus aprašančią informaciją. Šia posisteme naudojami eksperimento dalyvis ir eksperimento vykdytojas. Vykdomi šie panaudos atvejai (2.1 pav.): „PA 1 Peržiūrėti balso įrašų sąrašą“; „PA 1.1 Įrašyti balso įrašą“; „PA 1.2 Įkelti balso įrašus“; „PA 1.3 Redaguoti balso įrašo informaciją“. Realizuota ASP.NET technologijų pagrindu eksperimento dalyvio vykdomoms funkcijoms ir „Windows Forms“ technologijų pagrindu eksperimento vykdytojo atliekamoms funkcijoms.
- **Garso įrašų apdorojimo posistemė** – atsakinga už balso įrašų apdorojimą. Šia posisteme naudojami tik eksperimento vykdytojas. Vykdomas panaudos atvejis „PA 2 Inicijuoti garso įrašų apdorojimą“ (2.1 pav.). Detalesnė informacija apie šią posistemę pateikta žemiau.
- **Rezultatų apdorojimo posistemė** – apdoroja gautus šnekos atpažinimo rezultatus ir pateikia agreguotą informaciją. Šia posisteme naudojami tik eksperimento vykdytojas. Vykdomas panaudos atvejis „PA 3 Peržiūrėti garso įrašų apdorojimo rezultatus“ (2.1 pav.). Realizuota „Windows Forms“ technologijų pagrindu.



3.1 pav. Sprendimo konteksto diagrama

Garso įrašų apdorojimo posistemė susideda iš:

I. Garso įrašų apdorojimo procesas (1-asis būdas)

- **Garso įrašų paruošimo apdorojimui procesas** – šis procesas visus reikiamus balso įrašus konvertuoja į FLAC formatą, taip pat balso įrašus apjungia su visais galimais triukšmais, modifikuotais taip, kad būtų gauti tam tikri SNR. Daugiau informacijos pateikta 2.4.2 poskyryje.
- **Garso įrašų apdorojimo procesas** – tai procesas, kuris siunčia „Google“ šnekos atpažintuvui apdoroti daug garso įrašų (kartu su informacija apie juos: garso įrašo tipas bei dažnis) lygiagrečiai.

II. Garso įrašų apdorojimo procesas (2-asis būdas)

- **Garso įrašų apdorojimo procesas** – tai pagrindinis procesas, kuris paima balso įrašą iš DB, paleidžia „Google Chrome“ naršyklę automatinio mikrofono panaudojimo būsenoje ir atveria žiniatinklio svetainės puslapį, kuriame yra bendravimo komponentas bei šnekos atpažinimo (*webkit.SpeechRecognition*) komponentas. Per bendravimo komponentą bendrauja su šnekos atpažinimo komponentu ir išsaugo šnekos atpažinimo rezultatus. Realizuota „Windows Forms“ technologijų pagrindu.

- **Bendravimo komponentas** – šis komponentas atlieka tarpininko vaidmenį tarp garso įrašų apdorojimo proceso ir šnekos atpažinimo komponento. Realizuota ASP.NET *SignalR* technologijų pagrindu.
- **Šnekos atpažinimo komponentas** (*webkitSpeechRecognition*) – tai „Google Chrome“ naršyklėje veikiantis šnekos atpažinimo komponentas, apdorojantis garso įrašą ir grąžinantis šnekos atpažinimo rezultataž tekstu.

Šių komponentų sąveika pavaizduota 2.5 pav.

3.1.1. Eksperimento programinė įranga

Šiame skyriuje pateikti eksperimento programinės įrangos langai.

Duomenų įvedimo posistemė

Balso įrašo įkėlimo langas (3.2 pav.) matomas eksperimento dalyviui. Lango paskirtis – įkelti turimus balso įrašus (po vieną) į DB, nurodant juos aprašančią informaciją. Apie balso įrašą nurodoma ši informacija:

- **Asmuo** – kalbėtojo vardas ir pavardė;
- **Lytis** – kalbėtojo lytis;
- **Įrašas su triukšmu** – požymis, nurodantis, kad balso įrašas yra su foniniu triukšmu;
- **Sklandi šneka** – požymis, nurodantis, kad kalbėtojas kalba sklandžiai arba ne (pvz., ne iš pirmo karto teisingai pasako žodį);
- **Šneka su akcentu** – požymis, nurodantis, kad kalbėtojas kalba su akcentu;
- **Greita šneka** – požymis, nurodantis, kad kalbėtojas kalba labai greitai;
- **Garsi šneka** – požymis, nurodantis, kad kalbėtojas kalba garsiai;
- **Šnekos tipas** – šnekos tipas, kuriam priskiriamas balso įrašas;
- **Tekstas** – įgarsintas tekstas.

3.2 pav. Balso įrašo įkėlimo langas

Šnekos įrašymo langas (3.3 pav.) matomas eksperimento dalyviui. Lango paskirtis – sukurti naują balso įrašą ir išsaugoti jį DB, nurodant šią informaciją apie balso įrašą:

- **Asmuo** – kalbėtojo vardas ir pavardė;
- **Lytis** – kalbėtojo lytis;
- **Tekstas** – įgarsintas tekstas.

Šnekos įrašymas

Asmuo

Lytis Nenurodyta Vyras Moteris

Tekstas

3.3 pav. Šnekos įrašymo langas

Balso įrašų sąrašo langas (3.4 pav.) matomas eksperimento vykdytojui. Lango paskirtis – peržiūrėti DB esančius balso įrašus, iškviešti informacijos apie balso įrašą redagavimo langą (3.5 pav.).

id	name	word	second	lytis	su_tni	sklandi	su_ak	greita	garsiai	bal	text
194142	Vardenė Pavardenė	9	5,80	moteris	True	False	False	True	True	True	lietuviškieji kryžiai liaudies mažosios architekt...
194152	Vardenė Pavardenė	10	5,64	moteris	True	False	False	True	True	True	ije yra sudaryti ir iš architektūrinių dekoratyvin
194162	Vardenė Pavardenė	7	3,25	moteris	True	False	False	True	True	True	tai savotiška amato meno bei tikėjimo sintezė
194172	Vardenė Pavardenė	7	3,93	moteris	True	False	False	True	True	True	joje jungiasi skulptūros kalvystė bei tapybos e
194182	Vardenė Pavardenė	7	3,23	moteris	True	False	False	True	True	True	tradicinių kryžių darymas lietuvių liaudies men
194192	Vardenė Pavardenė	10	6,43	moteris	True	False	False	True	True	True	įrašyta į unesco žmonijos nematerialaus ir žoc
194202	Vardenė Pavardenė	5	4,29	moteris	True	False	False	True	True	True	koplytstulpiai koplytelės pasižymi formų įvairo
194212	Vardenė Pavardenė	8	5,32	moteris	True	False	False	True	True	True	lotyniškojo kryžiaus skersinė sija yra trumpesr
194222	Vardenė Pavardenė	11	5,33	moteris	True	False	False	True	True	True	graikiškojo kryžiaus abi sijos yra vienodo ilgio
194232	Vardenė Pavardenė	6	4,86	moteris	True	False	False	True	True	True	kryžiai geležinėmis viršūnėmis ornamentuoti k
194242	Vardenė Pavardenė	7	4,21	moteris	True	False	False	True	True	True	įvairių siužetų medinės skulptūros laikoma ve
194252	Vardenė Pavardenė	10	7,16	moteris	True	False	False	True	True	True	kryždirbystę tyrinėję krašto muziejaus speciali
194262	Vardenė Pavardenė	6	3,12	moteris	True	False	False	True	True	True	džamo minaretas stovi kalnų apsuptame slėn
194272	Vardenė Pavardenė	10	5,53	moteris	True	False	False	True	True	True	ankoras didžiulis šventyklų rūmų vandens tel
194282	Vardenė Pavardenė	6	3,96	moteris	True	False	False	True	True	True	susidomėjimo stounhendžui duoda nežinoma

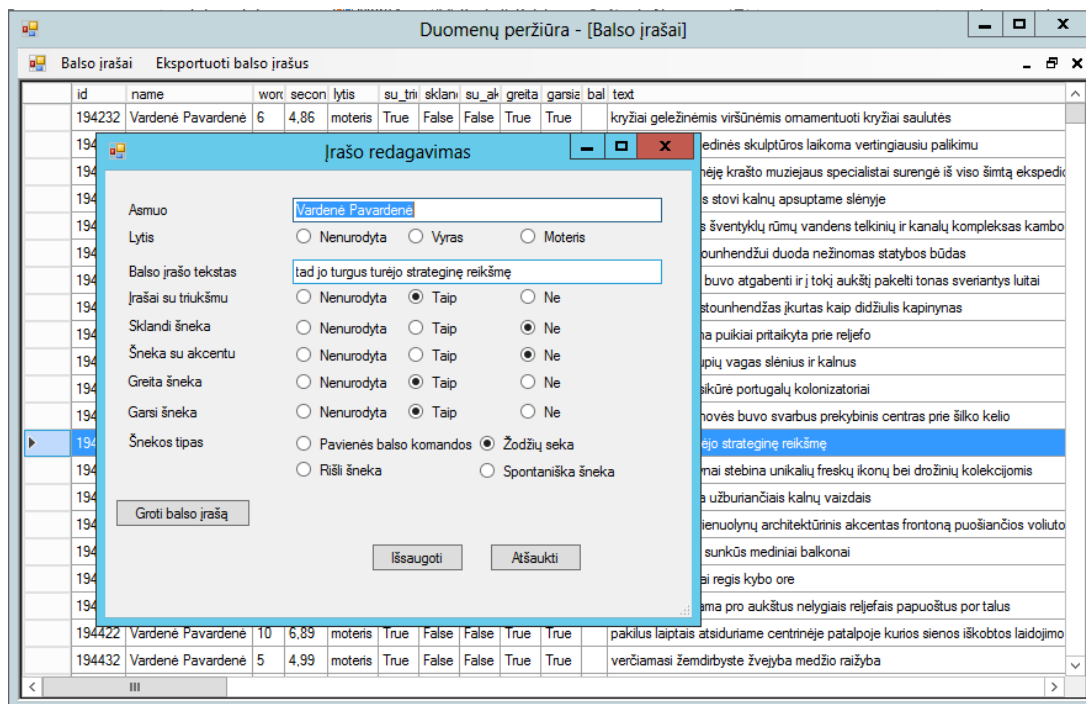
3.4 pav. Balso įrašų sąrašo langas

Balso įrašo informacijos redagavimo langas (3.5 pav.) matomas eksperimento vykdytojui. Lango paskirtis – peržiūrėti ir, jei reikia, atnaujinti DB esančio balso įrašo informaciją.

Balso įrašo informacijos redagavimo lange gali būti keičiama ši informacija:

- **Asmuo** – kalbėtojo vardas ir pavardė;
- **Lytis** – kalbėtojo lytis;
- **Tekstas** – įgarsintas tekstas;
- **Įrašas su triukšmu** – požymis, nurodantis, kad balso įrašas yra su foniniu triukšmu;
- **Sklandi šneka** – požymis, nurodantis, kad kalbėtojas kalba sklandžiai arba ne (pvz., ne iš pirmo karto teisingai pasako žodį);
- **Šneka su akcentu** – požymis, nurodantis, kad kalbėtojas kalba su akcentu;
- **Greita šneka** – požymis, nurodantis, kad kalbėtojas kalba labai greitai;
- **Garsi šneka** – požymis, nurodantis, kad kalbėtojas kalba garsiai;
- **Šnekos tipas** – nurodomas šnekos tipas, kuriam priskiriamas balso įrašas.

Paspaudus mygtuką [Groti balso įrašą] yra galimybė perklausyti balso įrašą.

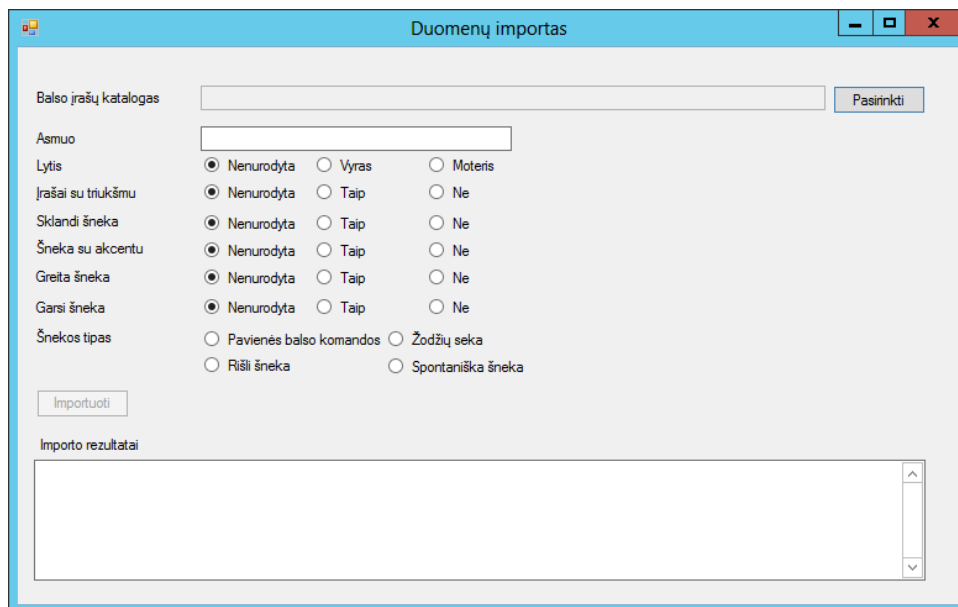


3.5 pav. Balso įrašo informacijos redagavimo langas

Balso įrašų importo langas (3.6 pav.) matomas eksperimento vykdytojui. Lango paskirtis – įkelti turimus balsų įrašus (pasirenkant atitinkamą balsų įrašų katalogą) į DB, nurodant juos aprašančią informaciją.

Balso įrašų importo metu nurodoma ši informacija:

- **Asmuo** – kalbėtojo vardas ir pavardė;
- **Lytis** – kalbėtojo lytis;
- **Įrašas su triukšmu** – požymis, nurodantis, kad balsų įrašas yra su foniniu triukšmu;
- **Sklandi šneka** – požymis, nurodantis, kad kalbėtojas kalba sklandžiai arba ne (pvz., ne iš pirmo karto teisingai pasako žodį);
- **Šneka su akcentu** – požymis, nurodantis, kad kalbėtojas kalba su akcentu;
- **Greita šneka** – požymis, nurodantis, kad kalbėtojas kalba labai greitai;
- **Garsi šneka** – požymis, nurodantis, kad kalbėtojas kalba garsiai;
- **Šnekos tipas** – nurodomas šnekos tipas, kuriam priskiriamas balsų įrašas.

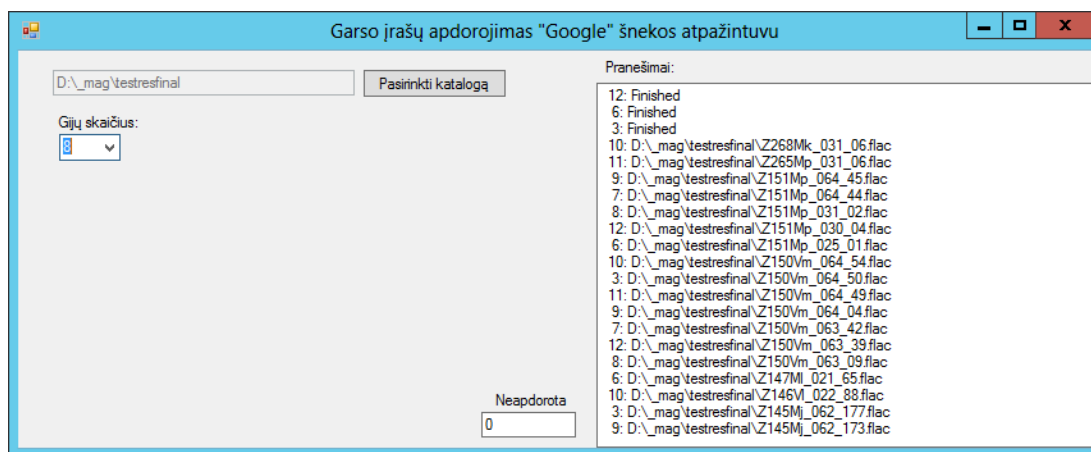


3.6 pav. Balso įrašų importo langas

Garso įrašų apdorojimo posistemė

Garso įrašų apdorojimas pirmuoju būdu (kai „Google“ šnekos atpažintuvui siunčiama apdoroti daug garso įrašų vienu metu) pavaizduotas 3.7 pav.

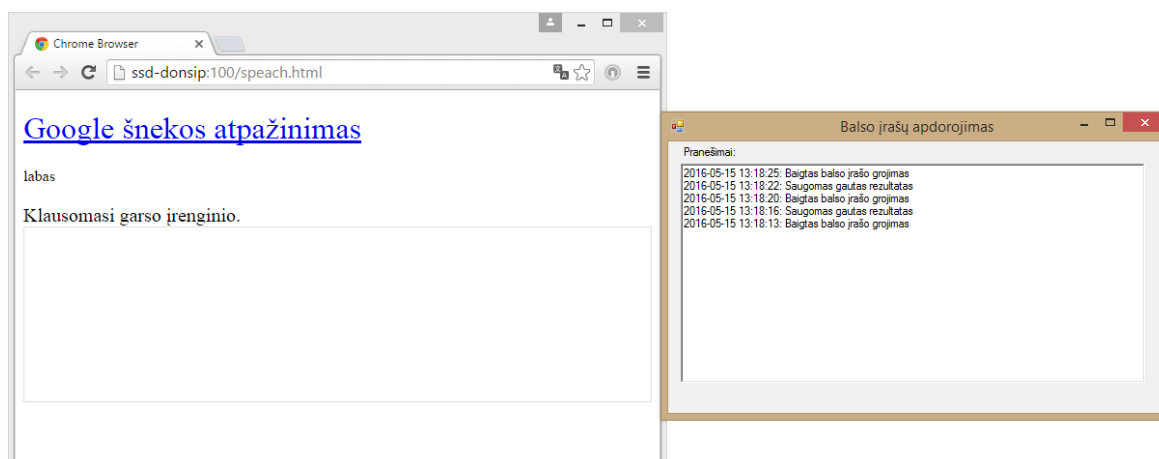
Lauke „Gijų skaičius“ nurodoma, kiek gijų (angl. *thread*) turėtų dalyvauti garso įrašų apdorojimo procese (maksimali reikšmė – 50). Nuo šiamo lauke nurodytos reikšmės priklauso, kiek garso įrašų bus apdorojama vienu metu. Lauke „Neapdorota“ rodoma, kiek dar liko neapdorotų garso įrašų. Lauke „Pranešimai“ matomi informaciniai pranešimai apie vykstantį garso įrašų apdorojimo procesą: pranešimo priekyje rodomas gijos ID, o po dvitaškio – jo veiklos rezultatas.



3.7 pav. Garso įrašų apdorojimo langas (1 būdas)

Garso įrašų apdorojimas antruoju būdu (kai imituojamas gyvas kalbėjimas ir apdorojamas tik vienas įrašas vienu metu) pavaizduotas 3.8 pav.

Dešinėje pusėje yra garso įrašų apdorojimo procesas, kuris parenka balso įrašą, triukšmą bei SNR, su kuriais bus atliekamas garso įrašo apdorojimas. Lange matomi informaciniai pranešimai apie vykstantį garso įrašų apdorojimo procesą. Kairėje pusėje, „Google Chrome“ naršyklėje, yra šnekos atpažinimo (*webkitSpeechRecognition*) komponentas. Naršyklės lange matomas informacinis pranešimas bei šnekos atpažintuvo pagrindinis rezultatas.



3.8 pav. Garso įrašų apdorojimo langas (2 būdas)

Rezultatų apdorojimo posistemė

Šnekos atpažinimo rezultatai pateikiami skirtingų vaizdų (angl. *view*) forma „Microsoft SQL Server 2014“. Šnekos atpažinimo rezultatai matomi tik eksperimento vykdytojui. Žemiau pateiktas suagreguotų šnekos atpažinimo rezultatų pavyzdys.

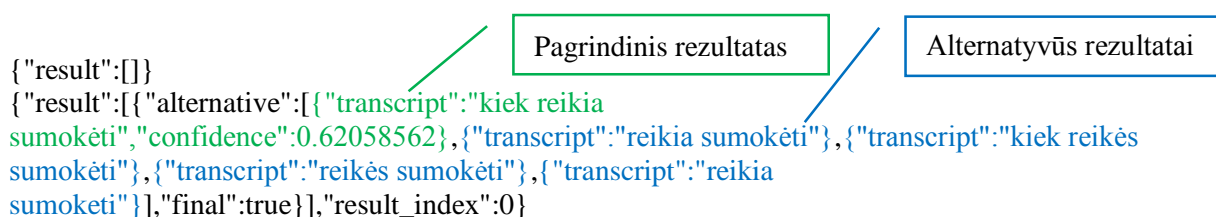
	SNR	Baltasis	Traukinių stotis	Važavimas automobiliu	Automobilių spūstis
1	25	49430	47918	44500	48927
2	15	44702	43264	43934	37339
3	20	48594	46080	46971	45357
4	10	33142	28911	37984	37572

3.9 pav. Šnekos atpažinimo rezultatų lango pavyzdys

3.1.2. Šnekos atpažinimo rezultatai

Šnekos atpažinimo rezultatas pateikiamas JSON formatu. Rezultatas sudarytas iš 1 pagrindinio rezultato, kuriam pateikiama įsitikinimo atpažinimo teisingumu reikšmė (angl. *confidence*), ir 3-4 alternatyvių rezultatų.

Žemiau pateikiamas vieno iš rezultatų pavyzdys.



3.2. Testavimo modelis, duomenys, rezultatai

Duomenų įvedimo posistemės testavimo tikslas – patikrinti, ar balso įrašai ir juos aprašanti informacija į DB įrašoma teisingai. Funkcinio testavimo metu buvo įrašyti nauji balso įrašai, ikelti balso įrašai po vieną nurodant skirtingą juos aprašančią informaciją, ikelti balso įrašai pasirinkus visą katalogą ir nurodžius skirtingą juos aprašančią informaciją, pakeista balso įrašus aprašanti informacija. Testavimas buvo sėkmingas.

Garso įrašų apdorojimo posistemės testavimo tikslas – patikrinti, ar sėkmingai vyksta automatinis balso įrašų apdorojimo procesas. Funkcinio testavimo metu buvo vykdyti testavimo atvejai, aprašyti prieduose, 7.3 poskyryje. Testavimo metu šnekos atpažinimui buvo panaudota 50 balso įrašų. Šie balso įrašai buvo apdoroti su dviem triukšmais (automobilių spūstis, baltasis triukšmas), dviem SNR: 25 dB ir 20 dB. Testavimas buvo sėkmingas. Vykdam šį testavimą buvo stebėta ir greitaveika: 50 balso įrašų, kurių bendra trukmė 401,9 sek., su dviem triukšmais ir dviem SNR „Google“ šnekos atpažintuvus apdorojo:

- Garso įrašų apdorojimo procesas (1-asis būdas, 50 gijų) – per 394 sek.;
- Garso įrašų apdorojimo procesas (2-asis būdas) – per 2209 sek.

Garso įrašų paruošimas apdorojimui pirmuoju būdu truko 39 sek., taigi tų pačių 50-ies balso įrašų apdorojimas abiem būdais truko atitinkamai 433 sek. (garso įrašų paruošimas apdorojimui ir garso įrašų apdorojimas pirmuoju būdu) ir 2209 sek. (garso įrašų apdorojimas antruoju būdu), t. y. pirmasis garso įrašų apdorojimo būdas yra apie penkiskart greitesnis.

Rezultatų apdorojimo posistemės testavimo tikslas – patikrinti, ar teisingai agreguojami gauti šnekos atpažinimo rezultatai. Funkcinio testavimo metu buvo patikrinta, ar rezultatų apdorojimo posistemės pateikti agreguoti atpažinimo rezultatai atitinka laukiamus rezultatus. Testavimas buvo sėkmingas.

4. EKSPERIMENTINIS SPRENDIMO TYRIMAS

Šiame skyriuje pateikiama informacija apie eksperimentinį sprendimo tyrimą: eksperimento planas, eksperimento duomenys, detalūs eksperimento rezultatai. Taip pat atlikta sprendimo veikimo ir savybių analizė, atliktas kokybės kriterijų įvertinimas bei pateiktos sprendimo taikymo rekomendacijos.

4.1. Eksperimento planas

1. Surinkti reikiamus balso įrašus.
2. Sukelti balso įrašus į DB, nurodyti juos aprašančią informaciją (t. y. suklasifikuoti).
3. Paruošti (konvertuoti visus balso įrašus į FLAC formatą, apjungti visus balso įrašus su reikiamais triukšmais ir gautus garso įrašus konvertuoti į FLAC formatą) balso įrašus apdorojimui.
4. Apdoroti visus balso įrašus be triukšmo.
5. Apdoroti visus balso įrašus, apjungtus su skirtingais triukšmais, modifikuotais taip, kad būtų gauti tam tikri SNR.
6. Po mėnesio nuo balso įrašų apdoravimo be triukšmo (4 punktas) pakartotinai apdoroti visus balso įrašus be triukšmo (tikimasi, kad „Google“ šnekos atpažintuvas yra nuolatos tobulinamas, „apsimoko“ su atpažįstamomis frazėmis ir bus geresni šnekos atpažinimo rezultatai).
7. Agreguoti gautus šnekos atpažinimo rezultatus.

4.2. Eksperimento duomenys

Eksperimentas buvo atliktas su 63299 balso įrašais, kurių bendra trukmė – 86,06 val., vidutinė trukmė – 4,89 sek. Detali informacija apie balso įrašus pateikta 4.1 – 4.3 lentelėse.

4.1 lentelė. Informacija apie kalbėtojus

Skirtingi kalbėtojai:	359			
Iš jų:	Vyrai		Moterys	
	Vnt.	%	Vnt.	%
	111	30,92	248	69,08

4.2 lentelė. Informacija apie balso įrašus

Balso įrašo požymis	Taip		Ne	
	Vnt.	%	Vnt.	%
Įrašas su triukšmu	18788	29,68	44511	70,32
Sklandi šneka	61930	97,84	1369	2,16
Šneka su akcentu	1640	2,59	61659	97,41
Greita šneka	23773	37,56	39526	62,44
Garsi šneka	62869	99,32	430	0,68

4.3 lentelė. Balso įrašų pasiskirstymas pagal šnekos tipą

Pavienės balso komandos		Keleto žodžių sekos		Rišli šneka		Spontaniška šneka	
Vnt.	%	Vnt.	%	Vnt.	%	Vnt.	%
39349	62,16	35	0,06	23887	37,74	28	0,04

Visi balso įrašai buvo apdoroti be triukšmo ir su 4 skirtingais triukšmais: traukinių stotis, automobilių spūstis, važiavimas automobiliu (angl. *car driving*), baltasis triukšmas (angl. *white noise*). Kiekvienas triukšmas buvo panaudotas 4 skirtingais SNR: 25 dB, 20 dB, 15 dB ir 10 dB.

4.3. Eksperimento rezultatai

„Google“ šnekos atpažintuvui vertinti buvo pasirinkta metrika WER – klaidingų žodžių dažnis. Ši metrika apskaičiuojama pagal (1) formulę. Eksperimento rezultatuose naudojama procentinė WER išraiška.

Šiame skyriuje aprašoma, kokie buvo šnekos atpažinimo rezultatai apdorojus visus turimus balso įrašus be triukšmo (4.3.1 Šnekos atpažinimo rezultatai (be triukšmo)), kokie buvo šnekos atpažinimo

rezultatai apdorojus visus turimus balso įrašus su triukšmais (4.3.2 Šnekos atpažinimo rezultatai (su triukšmu)) ir kokie buvo šnekos atpažinimo rezultatai apdorojus visus turimus balso įrašus be triukšmo po 1 mėnesio nuo eksperimento pradžios (4.3.3 „Google“ šnekos atpažintuvo vystymo vertinimas).

4.3.1. Šnekos atpažinimo rezultatai (be triukšmo)

Apdorojus visus 63299 balso įrašus pirmą kartą (be triukšmo) 20784 balso įrašams nebuvo gautas šnekos atpažinimo rezultatas. Šiame poskyryje pateikiama statistika tik tiems 42515 balso įrašų, kuriems „Google“ šnekos atpažintuvas pateikė atpažinimo rezultatus. Vidutinė WER reikšmė visiems 42515 balso įrašų yra 40,74 % (žr. 4.4 lentelę). Ši reikšmė gauta apskaičiavus visų balso įrašų, kuriems „Google“ šnekos atpažintuvas pateikė atpažinimo rezultatus, WER reikšmių vidurkį. WER standartinis nuokrypis yra 37,70. Tai parodo vidutinę duomenų sklaidą apie vidurkį, o $\pm 0,30$ reikšmė yra paskaičiuotas pasikliautinas intervalas su 90 % tikimybe, jog reikšmė yra šiame intervale.

4.4 lentelė. Apibendrinti šnekos atpažinimo rezultatai

Balso įrašų kiekis	Atpažinimo rezultatai									
	Frazių kiekis	Žodžių kiekis frazėse	WER						Vidutinis WER, %	WER standartinis nuokrypis
			Įterpimai		Pakeitimai		Panaikinimai			
			Vnt.	%	Vnt.	%	Vnt.	%		
63299	42515	238885	1295	0,54	50495	21,14	65618	27,47	40,74	37,70 \pm 0,30

4.5 lentelėje pateikiami šnekos atpažinimo rezultatai pagal kalbėtojo lytį. Kaip matome, WER reikšmės skiriasi labai neženkliai, todėl galima daryti išvadą, kad kalbėtojo lytis šnekos atpažinimui įtakos neturi.

4.5 lentelė. Šnekos atpažinimo rezultatai pagal kalbėtojo lytį

Kalbėtojo lytis	Atpažinimo rezultatai									
	Frazių kiekis	Žodžių kiekis frazėse	WER						Vidutinis WER, %	WER standartinis nuokrypis
			Įterpimai		Pakeitimai		Panaikinimai			
			Vnt.	%	Vnt.	%	Vnt.	%		
Moteris	30963	160531	892	0,56	34801	21,68	43135	26,87	39,91	38,54 \pm 0,36
Vyras	11552	78354	403	0,51	15694	20,03	22483	28,69	42,96	35,27 \pm 0,54

4.6 lentelėje pateikti atpažinimo rezultatai pagal kalbėtojus.

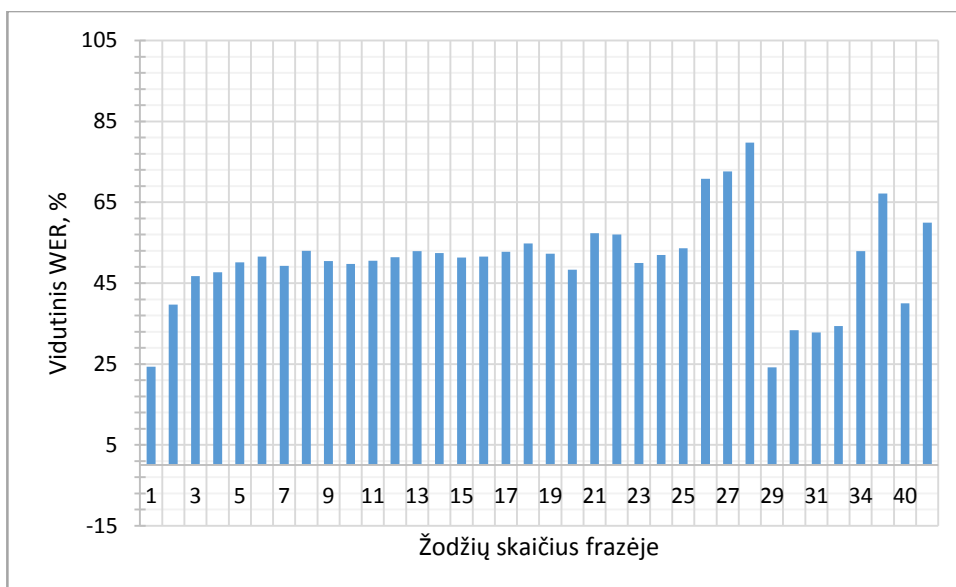
Pastaba: į blogiausius rezultatus netraukti tie kalbėtojai, kurių balso įrašams „Google“ šnekos atpažintuvas nepateikė nei vieno rezultato (tokių kalbėtojų yra 7), t. y. atpažintų frazių kiekis yra 0.

Detalūs šnekos atpažinimo rezultatai pateikti prieduose, 7.4.1 poskyryje (7.16 lent.).

4.6 lentelė. Šnekos atpažinimo rezultatai pagal kalbėtojus

	Rezultatai pagal kalbėtojus				
	Vidurkis	Geriausias	3 geriausių vidurkis	Blogiausias	3 blogiausių vidurkis
WER, %	40,74	10,00	14,74	100,00	96,39

Žemiau (4.1 pav.) pavaizduota, kokie šnekos atpažinimo rezultatai pagal žodžių kiekį frazėje. Grafike nerodomi rezultatai, kur atpažintų frazių kiekis yra 1 ar neatpažinta nei viena frazė. Detali informacija apie šnekos atpažinimo rezultatus pagal žodžių kiekį frazėje pateikta prieduose, 7.4.2 poskyryje (7.17 lent.).



4.1 pav. Šnekos atpažinimo rezultatai pagal žodžių kiekį frazėje

Kaip matome, aiškios tendencijos nėra, todėl negalima daryti išvadų dėl žodžių kiekio frazėje daromos įtakos šnekos atpažinimui.

Žemiau pateikiama informacija, kokie šnekos atpažinimo rezultatai atsižvelgiant į balso įrašo požymius.

4.7 lentelė. Šnekos atpažinimo rezultatai pagal balso įrašo požymį „Įrašas su triukšmu“

Balso įrašo požymis	Atpažinimo rezultatai									
	Frazių kiekis	Žodžių kiekis frazėse	WER						Vidutinis WER, %	WER standartinis nuokrypis
			Įterpimai		Pakeitimai		Panaikinimai			
			Vnt.	%	Vnt.	%	Vnt.	%		
Be triukšmo	27556	145662	418	0,29	23985	16,47	53726	36,88	41,26	38,77±0,38
Su triukšmu	14959	93223	877	0,94	26510	28,44	11892	12,76	39,77	35,64±0,48

Kaip matome, nežymus foninis triukšmas šnekos atpažinimo kokybei įtakos beveik neturi.

4.8 lentelė. Šnekos atpažinimo rezultatai pagal balso įrašo požymį „Sklandi šneka“

Balso įrašo požymis	Atpažinimo rezultatai									
	Frazių kiekis	Žodžių kiekis frazėse	WER						Vidutinis WER, %	WER standartinis nuokrypis
			Įterpimai		Pakeitimai		Panaikinimai			
			Vnt.	%	Vnt.	%	Vnt.	%		
Nesklaidi šneka	1075	6393	82	1,28	1884	29,47	713	11,15	38,71	34,91±1,75
Sklandi šneka	41440	232492	1213	0,52	48611	20,91	64905	27,92	40,79	37,77±0,31

Kaip matome, kalbėjimo sklandumas šnekos atpažinimo kokybei įtakos beveik neturi.

4.9 lentelė. Šnekos atpažinimo rezultatai pagal balso įrašo požymį „Šneka su akcentu“

Balso įrašo požymis	Atpažinimo rezultatai									
	Frazių kiekis	Žodžių kiekis frazėse	WER						Vidutinis WER, %	WER standartinis nuokrypis
			Įterpimai		Pakeitimai		Panaikinimai			
			Vnt.	%	Vnt.	%	Vnt.	%		
Be akcento	41549	231978	1255	0,54	48774	21,03	63658	27,44	40,56	37,76±0,30
Su akcentu	966	6907	40	0,58	1721	24,92	1960	28,38	48,39	34,25±1,81

Kaip matome, šneka su akcentu daro įtaką šnekos atpažinimo kokybei.

4.10 lentelė. Šnekos atpažinimo rezultatai pagal balso įrašo požymį „Greita šneka“

Balso įrašo požymis	Atpažinimo rezultatai									
	Frazių kiekis	Žodžių kiekis frazėse	WER						Vidutinis WER, %	WER standartinis nuokrypis
			Įterpimai		Pakeitimai		Panaikinimai			
			Vnt.	%	Vnt.	%	Vnt.	%		
Normali šneka	27913	160592	1018	0,63	36042	22,44	37162	23,14	39,90	36,95±0,36
Greita šneka	14602	78293	277	0,35	14453	18,46	28456	36,35	42,33	39,06±0,53

Kaip matome, kalbėjimo greitis šnekos atpažinimo kokybei įtakos beveik neturi.

4.11 lentelė. Šnekos atpažinimo rezultatai pagal balso įrašo požymį „Garsi šneka“

Balso įrašo požymis	Atpažinimo rezultatai									
	Frazių kiekis	Žodžių kiekis frazėse	WER						Vidutinis WER, %	WER standartinis nuokrypis
			Įterpimai		Pakeitimai		Panaikinimai			
			Vnt.	%	Vnt.	%	Vnt.	%		
Tyli šneka	174	1703	18	1,06	460	27,01	289	16,97	43,59	27,96±3,49
Garsi šneka	42341	237182	1277	0,54	50035	21,10	65329	27,54	40,73	37,74±0,30

Kaip matome, kalbėjimo garsumas šnekos atpažinimo kokybei įtakos beveik neturi.

4.12 lentelė. Šnekos atpažinimo rezultatai pagal šnekos tipą

Šnekos tipas	Atpažinimo rezultatai									
	Frazių kiekis	Žodžių kiekis frazėse	WER						Vidutinis WER, %	WER standartinis nuokrypis
			Įterpimai		Pakeitimai		Panaikinimai			
			Vnt.	%	Vnt.	%	Vnt.	%		
Pavienės balso komandos	22206	40555	277	0,68	11586	28,57	3810	9,39	31,55	43,29±0,48
Keleto žodžių seka	12	246	11	4,47	105	42,68	26	10,57	64,55	30,98±14,71
Rišli šneka	20297	198084	1007	0,51	38804	19,59	61782	31,19	50,77	27,09±0,31
Spontaniška šneka	0	-	-	-	-	-	-	-	-	-

Kaip matome, šnekos tipas turi ženkliai įtaką šnekos atpažinimo kokybei: pavienės balso komandos buvo atpažintos geriausiai, keleto žodžių seka – su aukščiausia vidutine WER reikšme (buvo siųstos atpažinti 35 frazės), tuo tarpu spontaniškos šnekos nebuvo atpažinta nei viena frazė iš 28 siųstų atpažinti frazių.

Yra 853 frazės, kurių vidutinis WER yra 0,00 %. Lentelėje pateikiami šių frazių pavyzdžiai.

4.13 lentelė. Geriausiai atpažintos frazės

Frazė	Siūsta atpažinti kartų	Atpažinta kartų
pridėti priedą	32	32
valdymo skydelis	32	32
didelis	29	29
devyniolika	29	29
parduoda	28	28
...		
berniukas	57	56
dokumentai	32	31

Frazė	Siūsta atpažinti kartų	Atpažinta kartų
įrašyti kitu vardu	32	31
atsakyti	32	31
mano kompiuteris	32	31
...		
lietuvoje labai reta rūšis įrašyta į lietuvos raudonąją knygą	15	14
...		
suvalkija lietuvos etnokultūrinis regionas kairiajame nemuno krante	14	11
suskaičiuojama iki dviejų šimtų šešiasdešimt paukščių rūšių	6	3
...		
miltai cukrus medus vaisiai ir uogos pienas ir grietinėle riebalai	9	3
ledo sluoksnyje susidariusiame prieš keturis tūkstančius metų rastas maždaug trijų centimetrų storio dulkių sluoksnis	8	2
jos suklestėjo septynioliktojo amžiaus antroje pusėje	7	1
...		

Žemiau pateikiami blogiausiai atpažintų frazių pavyzdžiai.

4.14 lentelė. Blogiausiai atpažintos frazės

Frazė	Šnekos atpažinimo rezultatas	Įterpimai	Pakeitimai	Panaikinimai	WER
penktas	tanki test	1	1	0	2,00
išsijunk	iš trijų	1	1	0	2,00
sviestas	speed test	1	1	0	2,00
lizosoma	visos stoma	1	1	0	2,00
taikyk	tai kiek	1	1	0	2,00
dešinėn	dešimt min	1	1	0	2,00
padidink	lady zippy	1	1	0	2,00
šunkelis	šunų kelis	1	1	0	2,00
išsijunk	iš jų	1	1	0	2,00
bjaurus	į eurus	1	1	0	2,00
rnr	prie neries	1	1	0	2,00
aštuntas	pašto kodas	1	1	0	2,00
...					
išjunk dešininę lygiuotę	iš jų dešimt yra likviduota	2	3	0	1,67
...					
išjunk programą	iš jų programa	1	2	0	1,50
atverk įrašus	perku ir aš	1	2	0	1,50
pristabdyk grojimą	kristen wiig rajone	1	2	0	1,50
parodyk darbalaukį	rodyti arba lauke	1	2	0	1,50
užverk dokumentą	už darbą dokumentai	1	2	0	1,50
atverk įrašus	verkiu ir aš	1	2	0	1,50
...					
išjunk pusjuodį šriftą	iš širdies į širdį	1	3	0	1,33
įjunk centrinę lygiuotę	tyrimų centras smėlio juosta	1	3	0	1,33
įjunk teksto pabraukimą	young sex to paburkimą	1	3	0	1,33
išjunk pusjuodį šriftą	iš visų pusių apsupta	1	3	0	1,33

Frazė	Šnekos atpažinimo rezultatas	Įterpimai	Pakeitimai	Panaikinimai	WER
kurie gali būti vienbalsiškumas homofoninis daugiabalsumas heterofonija ir polifonija	gali būti vienas iš kumas ir kuma foninis daugiau balsu masketuri of one ir polifonija	7	4	1	1,33

...

kaip veikia ši observatorija	sveiki aš esu vyras nori	1	4	0	1,25
ji skrodžia įspūdingus tarpeklis	iš kraujo spaudimas ir pulsas	1	4	0	1,25

...

vienuolynai yra pastatyti ant imponantiškų natūralių smiltainio uolų stulpų	imponantiškas mūsų	0	2	7	1,00
marienburgio pilį įkūrė kryžiuočiai užkariautoje prūsų žemėje	žemė	0	1	6	1,00
atonas į poseidoną metė uolą kuri įkrito į egėjo jūrą ir tapo atono kalnu	vietoje gėlių jūra	0	3	11	1,00

...

4.3.2. Šnekos atpažinimo rezultatai (su triukšmu)

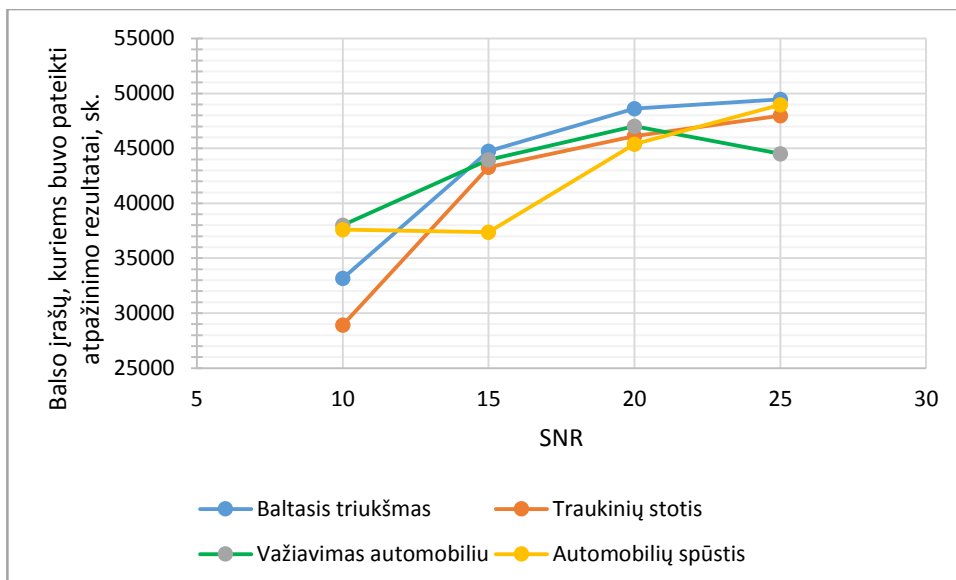
Šiame poskyryje pateikti šnekos atpažinimo rezultatai, kai visi 63299 balso įrašai buvo apdoroti su triukšmu (4 skirtingi triukšmai, 4 SNR).

4.15 lentelėje pateikti šnekos atpažinimo rezultatai, kurie parodo, kaip kinta balso įrašų, kuriems buvo pateikti atpažinimo rezultatai, skaičius priklausomai nuo triukšmo tipo ir SNR.

4.15 lentelė. Balso įrašų, kuriems buvo pateikti atpažinimo rezultatai, skaičiaus priklausomybė nuo triukšmo tipo ir SNR

SNR	Balso įrašų, kuriems buvo pateikti atpažinimo rezultatai, skaičius			
	Baltasis triukšmas	Traukinių stotis	Važiavimas automobiliu	Automobilių spūstis
10	33183	28934	38014	37601
15	44738	43302	43964	37369
20	48625	46111	47008	45391
25	49468	47966	44526	48964

4.2 pav. šie rezultatai pateikti grafiškai. Kaip matome, daugiausiai šnekos atpažinimo rezultatų (nebūtinai teisingų) gražinta, kai SNR reikšmė 25 dB (išimtis – triukšmas „Važiavimas automobiliu“, su kuriuo daugiausiai šnekos atpažinimo rezultatų gauta, kai SNR=20 dB).



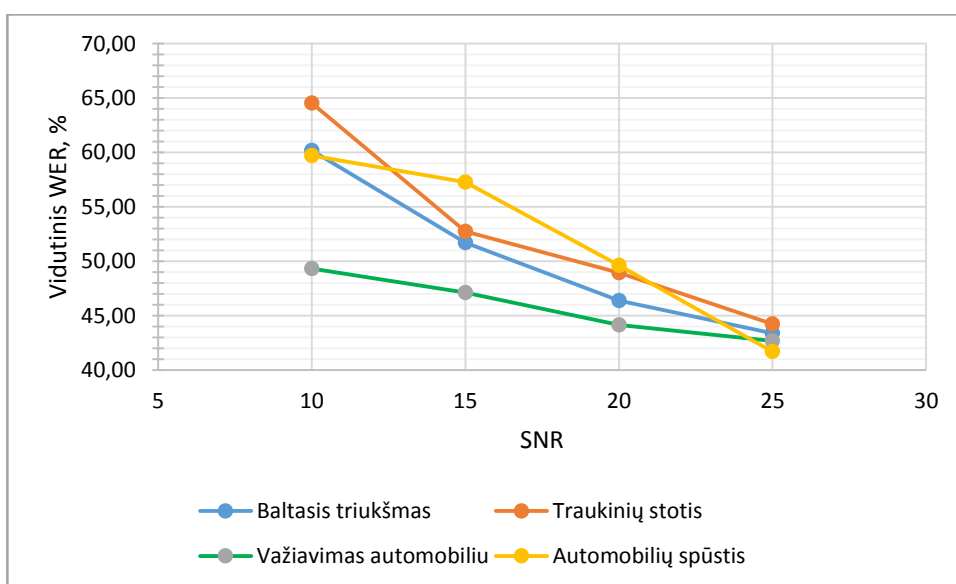
4.2 pav. Balso įrašų, kuriems buvo pateikti atpažinimo rezultatai, skaičiaus priklausomybė nuo triukšmo tipo ir SNR

4.16 lentelėje pateikti šnekos atpažinimo rezultatai, kurie parodo, kaip kinta šnekos atpažinimo kokybė priklausomai nuo triukšmo tipo ir SNR.

4.16 lentelė. Šnekos atpažinimo kokybės priklausomybė nuo triukšmo tipo ir SNR

SNR	Vidutinis WER, %			
	Baltasis triukšmas	Traukinių stotis	Važiavimas automobiliu	Automobilių spūstis
10	60,20	64,55	49,31	59,71
15	51,72	52,73	47,13	57,27
20	46,39	48,93	44,15	49,62
25	43,39	44,23	42,67	41,70

4.3 pav. šie rezultatai pateikti grafiškai. Kaip matome, geriausias šnekos atpažinimas yra kai SNR=25 dB (t. y. su silpniausiu triukšmu).



4.3 pav. Šnekos atpažinimo kokybės priklausomybė nuo triukšmo tipo ir SNR

4.3.3. „Google“ šnekos atpažintuvo vystymo vertinimas

Apdorojus visus 63299 balso įrašus (be triukšmo) pakartotinai po 1 mėnesio 20868 balso įrašams nebuvo gautas šnekos atpažinimo rezultatas (84 įrašais daugiau, nei prieš mėnesį). Vidutinė WER reikšmė visiems 42431 balso įrašų yra 40,82 % (žr. 4.17 lentelę). Ši reikšmė gauta apskaičiuavus visų balso įrašų, kuriems „Google“ šnekos atpažintuvas pateikė atpažinimo rezultatus, WER reikšmių vidurkį.

4.17 lentelė. Apibendrinti šnekos atpažinimo rezultatai

Balso įrašų kiekis	Pirminiai atpažinimo rezultatai			Atpažinimo rezultatai po mėn.			Vidutinio WER pokytis, %
	Frazių kiekis	Žodžių kiekis frazėse	Vidutinis WER, %	Frazių kiekis	Žodžių kiekis frazėse	Vidutinis WER, %	
63299	42515	238885	40,74	42431	237056	40,82	-0,08

Kaip matome, šnekos atpažinimo rezultatai (vidutinis WER) yra 0,08 % prastesni, nei pirminio eksperimento metu.

4.4. Sprendimo veikimo ir savybių analizė, kokybės kriterijų įvertinimas

Realizuoti du balso įrašų apdorojimo būdai. Apdorojant balso įrašus pirmuoju būdu garso adapteris nereikalingas ir „Google“ šnekos atpažintuvui gali būti siunčiama daug balso įrašų (įrašai būtinai turi būti FLAC formato) lygiagrečiai (tai valdoma gijų skaičiumi (angl. *thread*)), tačiau balso įrašų apdorojimas gali būti bet kada sustabdytas „Google“ panaikinus aktyvų raktą (angl. *API Key*), todėl vien šiuo būdu neturėtų būti pasitikima. Balso įrašų apdorojimo spartą riboja šie faktoriai: kliento pusėje – kietojo disko sparta ir tinklo pralaidumas, serverio pusėje – „Google“ skiriami resursai šnekos atpažinimui konkrečiam raktui. Pastebėta, kad apdorojant balso įrašus pirmuoju būdu kartais įvyksta klaidos „Google“ pusėje (galimai, resursų trūkumas) ir negaunami šnekos atpažinimo rezultatai.

Apdorojant balso įrašus antruoju būdu (2.5 pav.) reikalingas garso adapteris (eksperimento PĮ groja įrašą (įrašo formatas nesvarbus), o „Google“ šnekos atpažintuvas „klausos“) ir vienu metu galima apdoroti tik vieną balso įrašą. Balso įrašų apdorojimo procesą galima paspartinti sujungiant daugiau kompiuterių.

Žemiau pateiktas kokybės kriterijų įvertinimas.

4.18 lentelė. Kokybės kriterijų įvertinimas

Eil. Nr.	Kokybės kriterijus	Įvertinimas	Pastaba
1.	Eksperimentas atliktas su ne mažiau kaip 5000 balso įrašų	+	Eksperimentas buvo atliktas su 63299 balso įrašais.
2.	Balso įrašai tenkina šiuos kriterijus: bent 50 skirtingų kalbėtojų; skirtingų lyčių kalbėtojai; skirtingi šnekos tipai (pavienės balso komandos, žodžių seka, rišli šneka, spontaniška šneka); skirtingi kalbėjimo garsumo lygiai (garsiai, tyliai); skirtingi kalbėjimo greičiai (greita šneka, lėta šneka).	+	359 skirtingi kalbėtojai; visų šnekos tipų balso įrašai; skirtingos balso įrašų charakteristikos (įrašas su triukšmu, šneka su akcentu, greita šneka, garsi šneka).
3.	Eksperimente panaudoti bent 4 skirtingi triukšmai ir kiekvienam iš jų bent 3 skirtingi SNR.	+	Eksperimente panaudoti 4 skirtingi triukšmai ir 4 skirtingi SNR.
4.	Bendra balso įrašų trukmė yra ne mažesnė kaip 10 val.	+	Bendra balso įrašų trukmė yra 86,06 val.

4.5. Sprendimo taikymo rekomendacijos

Atlikus eksperimentą su vienodu balso įrašų kiekiu kiekvienoje klasifikuojamoje įrašų grupėje (vienodos frazės visiems kalbėtojams, vienodas balso įrašų skaičius kiekvienam šnekos tipui ir pan.) šnekos atpažintuvo efektyvumo vertinimas galimai būtų tikslesnis.

Vertinant šnekos atpažinimo rezultatus būtų tikslinga atlikti ir semantinę gauto šnekos atpažinimo rezultato analizę, nes WER neparodo, ar gauto rezultato semantinė reikšmė yra tokia pati, kaip įgarsinto teksto.

Eksperimentą reiktų atlikti periodiškai su tais pačiais balsu įrašais ir vertinti, kaip kinta „Google“ šnekos atpažintuvo efektyvumas.

5. IŠVADOS

1. Šiame darbe iškeltas tikslas – įvertinti „Google“ lietuvių šnekos atpažintuvo efektyvumą – buvo pasiektas sukūrus eksperimento PĮ ir atlikus eksperimentą su įvairiais aspektais skirtingais 63299 balso įrašais.
2. Literatūros šaltinių analizė parodė, kad nėra jokių nusistovėjusių metodikų ar standartų, leidžiančių patikimai įvertinti šnekos atpažintuvų efektyvumą, tačiau dažniausiai šnekos atpažintuvo kokybei vertinti yra naudojama metrika WER (angl. *Word Error Rate*).
3. Eksperimento PĮ testavimo metu pastebėta, kad šnekos atpažinimas siunčiant „Google“ šnekos atpažintuvui garso rinkmenas yra apie penkis kartus greitesnis (kai garso įrašų apdorojimo procese dalyvauja 50 gijų), nei imituojant gyvą kalbėjimą, kai apdorojamas tik vienas įrašas vienu metu.
4. Atlikus eksperimento PĮ testavimą pastebėta, kad skaitvardžiai šnekos atpažinimo rezultate gali būti pateikiami kaip skaičiai, tekstas ir kaip skaičiai su simboliais (t. y. „1940“, „tūkstantis devyni šimtai keturiasdešimtieji“, „1940-ieji“, „1940i“) net ir keliskart „Google“ šnekos atpažintuvui apdorojus tą pačią frazę. Tai kelia keblumų norint palyginti įgarsintą tekstą ir šnekos atpažinimo rezultatą (reikia papildomai apdoroti visus šnekos atpažinimo rezultatus, visus skaitvardžius šnekos atpažinimo rezultatuose pakeičiant tekstu), tačiau semantine prasme šnekos atpažinimo rezultatai yra teisingi.
5. Eksperimento pirmo etapo metu pastebėta, kad iš 63299 „Google“ šnekos atpažintuvui pateiktų apdoroti (be triukšmo) balso įrašų 20784 balso įrašams išvis nebuvo pateikti šnekos atpažinimo rezultatai, o tai sudaro net 32,83 % visų balso įrašų. Trečiojo eksperimento etapo metu, kai tikėtasi, kad šnekos atpažintuvas apsimokė ir šnekos atpažinimo rezultatai bus geresni, balso įrašų, kuriems nebuvo gauti šnekos atpažinimo rezultatai, skaičius 84 balso įrašais didesnis.
6. Vidutinė WER reikšmė visiems 42515 balso įrašų, kuriems „Google“ šnekos atpažintuvas pateikė atpažinimo rezultatus, yra 40,74 %. WER standartinis nuokrypis – 37,70 % su 0,30 pasikliautinumo intervalu (paskaičiuota su 90 % tikimybe, jog reikšmė yra šiame intervale).
7. Remiantis atlikto eksperimento rezultatų analize galima teigti, kad šnekos atpažinimo kokybei mažą įtaką daro kalbėtojo lytis – vidutinio WER pokytis tarp skirtingų lyčių tik 3,05 %. Taip pat šnekos atpažinimo kokybei mažą įtaką turi neįžymūs foninis triukšmas, kalbėjimo sklandumas, kalbėjimo greitis, kalbėjimo garsumas (vidutinio WER pokytis tarp lyginamų grupių 1,49-2,86 %), tuo tarpu šnekos su akcentu atpažinimo rezultatas yra 7,83 % prastesnis. Didelę įtaką šnekos atpažinimui turi ir šnekos tipas: geriausiai atpažįstamos pavienės balso komandos, prasčiausiai – spontaniška šneka.
8. Atlikus eksperimentą su 4 skirtingais balso įrašo ir triukšmo signalų santykiais (SNR) pastebėta, kad geriausias šnekos atpažinimas yra kai SNR=25 dB (t. y. su silpniausiu triukšmu). Daugiausiai šnekos atpažinimo rezultatų (nebūtinai teisingų) gražinta, kai SNR reikšmė taip pat 25 dB (išimtis – triukšmas „Važiavimas automobiliu“, su kuriuo daugiausiai šnekos atpažinimo rezultatų gauta, kai SNR=20 dB).
9. Sprendimo eksperimentinio tyrimo rezultatai parodė, kad neretu atveju vienas iš alternatyvių rezultatų yra tikslesnis už pagrindinį rezultatą. Tai rodo, kad šnekos atpažinimo kokybė galėtų būti geresnė, jei „Google“ tiksliau vertintų, kuris rezultatas yra tinkamiausias.
10. Pakartotinai atlikto eksperimento po 1 mėn. rezultatai parodė, kad „Google“ šnekos atpažintuvo šnekos atpažinimo kokybė nepagerėjo, nes šnekos atpažinimo rezultatai (vidutinis WER) yra 0,08 % prastesni, nei pirminio eksperimento metu. Šiuo metu „Google“ šnekos atpažinimo paslauga galima naudotis nemokamai, todėl galima daryti prielaidą, kad „Google“ taip pat mato šnekos atpažintuvo trūkumus ir jis dar bus vystomas, tobulinamas.
11. Norint padidinti eksperimento rezultatų patikimumą eksperimentą reiktų atlikti su vienodu balso įrašų kiekiu kiekvienoje klasifikuojamoje įrašų grupėje.

6. LITERATŪRA

- [1] Maskeliūnas, R. Lietuviškų balso komandų atpažinimas daugybinių transkripcijų pagrindu. Daktaro disertacija (2009)
- [2] Vinciguerra, B. A comparison of commercial speech recognition components for use with the Project 54 system. Thesis (2002)
- [3] Google says its speech recognition technology now has only an 8% word error rate [interaktyvus]. 2015 [žiūrėta 2016-04-25]. Prieiga per: <http://venturebeat.com/2015/05/28/google-says-its-speech-recognition-technology-now-has-only-an-8-word-error-rate/>
- [4] Juang, B. H. , Rabiner, L. R. Automatic Speech Recognition - A Brief History of the Technology. Elsevier Encyclopedia of Language and Linguistics, Second Edition. 2005
- [5] Anusuya, M. A., Katti, S. K. Speech Recognition by Machine: A Review. (IJCSIS) International Journal of Computer Science and Information Security 2(4): 910–954. 2009
- [6] Lileikytė, R. Šnekos atpažinimo požymių kokybės vertinimas. Daktaro disertacija (2012)
- [7] Bivainis, R. Balso atpažinimo programų lietuvinimo galimybių tyrimas. Magistro baigiamasis darbas (2013)
- [8] Kasparaitis, P. Kompiuterinė lingvistika. Kalbos atpažinimas.[žiūrėta 2015-01-24], prieiga per internetą:
<http://www.mif.vu.lt/~pijus/CL/Atpaz.pdf>
- [9] Laurinčiukaitė, S. Lietuvių šnekos atpažinimo akustinis modeliavimas. Daktaro disertacija (2008)
- [10] Deller, J. H., Hansen, J. H. L., Proakis, J. G. Discrete-time processing of speech signals. New York: Macmillan., 988 p. 1993
- [11] Cole, R., Mariani, J., Uszkoreit, H., Zaenen, A., Zue, V., Varile, G. B., Zampolli, A. Survey of the state-of-the-art in human language technology. Cambridge: Cambridge University Press, 533 p. 1998
- [12] Junqua, J. C., Haton, J. P. Robustness in automatic speech recognition: fundamentals and applications. USA: Kluwer academic publishers, 437 p. 1996
- [13] Rabiner, L. R., Levinson, S. E. Isolated and connected word recognition – theory and selected applications. IEEE Trans. on Communications 29(5): 621–659. 1981
- [14] Sugamura, N., Shikano, K., Furui, S. Isolated word recognition using phoneme-like templates. Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP'83 8: 723–726. 1983
- [15] Jelinek, F. A real-time, isolated-word, speech recognition system for dictation transcription. Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP'85 10: 858–861. 1985
- [16] Murviet, H., Weintraub, M. 1000-word speaker-independent continuous-speech recognition using hidden Markov models. Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP'88 1: 1115–1118. 1988
- [17] Sakoe, H., Isotani, K., Yoshida, K. I., Watanabe, T. Speaker independent word recognition using dynamic programming neural networks. Readings in Speech Recognition, ed. A. Waibel, K.-F. Lee, 439–442. 1989
- [18] Lipeika, A., Lipeikienė, J., Telksnys, L. Development of isolated word speech recognition system. Informatica 13(1): 37–46. 2002
- [19] Filipovič, M., Lipeika, A. Development of HMM/Neural Network-based medium-vocabulary isolated-word Lithuanian speech recognition system. Informatica 15(4): 465–474. 2004
- [20] Raškinis, G., Raškinienė, D. Building medium-vocabulary isolated-word Lithuanian HMM speech recognition system. Informatica 14(1): 75–84. 2003
- [21] Holmes, J., Holmes, W. Speech synthesis and recognition. Second Edition. New York: Taylor & Francis, 298 p. 2001
- [22] Gauvain, J. L., Lamel, L. Large vocabulary continuous speech recognition: from laboratory systems toward real-world applications. Trans. of Inst. of Electronics, Information and Communication Engineers 2: 2005–2021. 1996
- [23] Rabiner, L. R., Juang, B. H. Fundamentals of speech recognition. Prentice-Hall, New Jersey. 507 p. 1993

- [24] Anusuya, M. A., Katti, S. K. Classification Techniques used in Speech Recognition Applications: A review. (IJCSIS) International Journal of Computer Science and Information Security 6(3): 181–205. 2011
- [25] Deng, L., Hinton, G., Kingsbury, B. New types of deep neural network learning for speech recognition and related applications: an overview. IEEE International Conference on Acoustics, Speech and Signal Processing, ISSN 1520-6149, 2013
- [26] Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., Kingsbury, B. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine, Vol. 29 (6), p. 82-97, 2012
- [27] Mohamed, A., Dahl, G., Hinton, G. Acoustic modeling using deep belief networks. IEEE Trans. on Audio, Speech, and Language Processing, vol. 20, no. 1, pp. 14–22, 2012
- [28] Deng, L., Yu, D., Platt, J. Scalable stacking and learning for building deep architectures. ICASSP, 2012
- [29] Yu, D., Deng, L., Li, G., Seide, F. Discriminative pretraining of deep neural networks. U.S. Patent Filing, Nov. 2011
- [30] Dahl, G., Yu, D., Deng, L., Acero, A. Large vocabulary continuous speech recognition with context-dependent DBN-HMMs,” ICASSP, 2011
- [31] Dahl, G., Yu, D., Deng, L., Acero, A. Context-dependent pre-trained deep neural networks for large vocabulary speech recognition. IEEE Trans. Speech and Audio Proc., vol. 20, no. 1, pp. 30 – 42, 2012
- [32] Sainath, T., Mohamed, A., Kingsbury, B., Ramabhadran, B. Convolutional neural networks for LVCSR. ICASSP, 2013
- [33] Seide, F., Li, G., Yu, D. Conversational speech transcription using context-dependent deep neural networks. Interspeech, 2011, pp. 437-440, 2011
- [34] Yu, D., Deng, L., Dahl, G. Roles of pretraining and fine-tuning in context-dependent DNN-HMMs for real-world speech recognition. NIPS Workshop on Deep Learning and Unsupervised Feature Learning, 2010
- [35] Hinton, G., Osindero, S., Teh, Y. A fast learning algorithm for deep belief nets. Neural Computation, Vol. 18, 1527-1554, 2006
- [36] Hinton, G., Salakhutdinov, R. Reducing the dimensionality of data with neural networks. Science, Vol. 313. no. 5786, pp. 504 - 507, 2006
- [37] Lange, P., Suendermann-Oeft, D. Tuning Sphinx To Outperform Google’s Speech Recognition API.
- [38] Tamulevičius, G. Pavienių žodžių atpažinimo sistemų kūrimas. Daktaro disertacija (2008)
- [39] Rudzionis, V., Maskeliunas, R., Driaunys, K. Voice controlled environment for the assistive tools and living space control. In: Proceedings of the Federated Conference on Computer Science and Information Systems, p. 1075–1080. PROCEEDINGS OF THE FEDCSIS. WROCLAW (2012). ISBN 978-83-60810-51-4
- [40] Maskeliunas, R., Ratkevicius, K., Rudzionis, V. Some Aspects of Voice User Interfaces Development for Internet and Computer Control Applications. ELEKTRONIKA IR ELEKTROTECHNIKA, ISSN 1392-1215, VOL. 19, NO. 2, 2013
- [41] Rudzionis, V., Raskinis, G., Maskeliunas, R., Rudzionis, A., Ratkevicius, K. Comparative Analysis of Adapted Foreign Language and Native Lithuanian Speech Recognizers for Voice User Interface. ELEKTRONIKA IR ELEKTROTECHNIKA, ISSN 1392-1215, VOL. 19, NO. 7, 2013
- [42] Rudzionis, V., Raskinis, G., Maskeliunas, R., Rudzionis, A., Ratkevicius, K., Bartisiute, G. Web Services Based Hybrid Recognizer of Lithuanian Voice Commands. ELEKTRONIKA IR ELEKTROTECHNIKA, ISSN 1392-1215, VOL. 20, NO. 9, 2014
- [43] Apie projektą LIEPA.[žiūrėta 2016-05-16], prieiga per internetą:
<https://www.raštija.lt/liepa>
- [44] Schalkwyk, J., Beeferman, D., Beaufays, F., Byrne, B., Chelba, C., Cohen, M., Garret, M., Strophe, B. Google Search by Voice: A case study.

7. PRIEDAI

7.1. Panaudos atvejų aprašas

Panaudos atvejų diagrama pateikta 2.1 pav.

7.1 lentelė. Panaudos atvejo „PA 1 Peržiūrėti balso įrašų sąrašą“ specifikacija

PA 1 Peržiūrėti balso įrašų sąrašą		
Tikslas. Peržiūrėti balso įrašų sąrašą		
Aprašymas. Šis PA aprašo balso įrašų sąrašo išskvietimą		
Prieš sąlyga	Duomenų bazėje yra saugomas bent vienas balso įrašas	
Aktorius	Eksperimento vykdytojas	
Sužadinimo sąlyga	Naudotojas nori peržiūrėti balso įrašų sąrašą	
Susiję panaudos atvejai	Išplečia PA	Nėra.
	Apima PA	Nėra.
	Specializuoja PA	Nėra.
Pagrindinis įvykių srautas		
Sistemos reakcija ir sprendimai		
1. Naudotojas pasirenka balso įrašų peržiūros funkciją.	1.1. Atveriamas balso įrašų sąrašo langas.	
2. Naudotojas baigia PA.		
Po sąlyga:		

7.2 lentelė. Panaudos atvejo „PA 1.1 Įrašyti balso įrašą“ specifikacija

PA 1.1 Įrašyti balso įrašą		
Tikslas. Įrašyti naują balso įrašą ir išsaugoti DB		
Aprašymas. Šis PA aprašo naujo balso įrašo išsaugojimą DB		
Prieš sąlyga	Prie kompiuterio prijungtas mikrofonas	
Aktorius	Eksperimento dalyvis; Eksperimento vykdytojas	
Sužadinimo sąlyga	Naudotojas nori įrašyti naują balso įrašą	
Susiję panaudos atvejai	Išplečia PA	PA 1 Peržiūrėti balso įrašų sąrašą
	Apima PA	Nėra.
	Specializuoja PA	Nėra.
Pagrindinis įvykių srautas		
Sistemos reakcija ir sprendimai		
1. Naudotojas pasirenka įrašymo funkciją.	1.1. Sistema atveria įrašymo langą.	
2. Naudotojas užpildo informacinius laukus.	2.1. Įrašymo lange rodoma užpildyta informacija.	
3. Naudotojas spaudžia mygtuką [Pradėti įrašymą].	3.1. Rodomas pranešimas, kad vyksta įrašymas. 3.2. Mygtukas [Pradėti įrašymą] tampa neaktyvus. 3.3. Aktyvuojamas mygtukas [Stabdyti ir išsaugoti].	
4. Perskaitęs reikiamą tekstą naudotojas spaudžia [Stabdyti ir išsaugoti].	4.1. Įrašymas baigiamas. 4.2. Mygtukas [Stabdyti ir išsaugoti] tampa neaktyvus.	
5. Naudotojas baigia PA.		
Po sąlyga:		
Balso įrašas išsaugotas DB kartu su jį aprašančia informacija.		
Alternatyvūs scenarijai		
4a. Naudotojas spaudžia mygtuką [Atšaukti].	Informacija neišsaugoma. Atveriamas balso įrašų sąrašo langas.	

7.3 lentelė. Panaudos atvejo „PA 1.2 Įkelti balso įrašus“ specifikacija

PA 1.2 Įkelti balso įrašus		
Tikslas. Įkelti turimus balso įrašus į DB tolesniam apdorojimui		
Aprašymas. Šis PA aprašo balso įrašų išsaugojimą DB		
Prieš sąlyga	Turimi norimi įkelti balso įrašai	
Aktorius	Eksperimento dalyvis; Eksperimento vykdytojas	
Sužadinimo sąlyga	Naudotojas nori įkelti balso įrašus	
Susiję panaudos atvejai	Išplečia PA	PA 1 Peržiūrėti balso įrašų sąrašą
	Apima PA	Nėra.
	Specializuoja PA	Nėra.
Pagrindinis įvykių srautas		
Sistemos reakcija ir sprendimai		
1. Naudotojas pasirenka balso įrašų įkėlimo funkciją.	1.1. Sistema atveria balso įrašų įkėlimo langą.	
2. Naudotojas pasirenka norimą balso įrašą arba balso įrašų katalogą.	2.1. Balso įrašų įkėlimo lange rodomas pasirinkto failo ar katalogo pavadinimas.	
3. Naudotojas užpildo informacinius laukus.	3.1. Balso įrašų įkėlimo lange rodoma užpildyta informacija.	
4. Naudotojas spaudžia mygtuką [Išsaugoti].	4.1. Informacija išsaugoma. 4.2. Atveriamas balso įrašų sąrašo langas.	
5. Naudotojas baigia PA.		
Po sąlyga:	Balso įrašai išsaugoti DB kartu su juos aprašančia informacija.	
Alternatyvūs scenarijai		
4a. Naudotojas spaudžia mygtuką [Atšaukti].	Informacija neišsaugoma. Atveriamas balso įrašų sąrašo langas.	

7.4 lentelė. Panaudos atvejo „PA 1.3 Redaguoti balso įrašo informaciją“ specifikacija

PA 1.3 Redaguoti balso įrašo informaciją		
Tikslas. Patikslinti informaciją apie balso įrašus		
Aprašymas. Šis PA aprašo informacijos apie balso įrašus patikslinimą		
Prieš sąlyga	DB yra saugomas bent vienas balso įrašas	
Aktorius	Eksperimento vykdytojas	
Sužadinimo sąlyga	Naudotojas nori patikslinti informaciją apie balso įrašą	
Susiję panaudos atvejai	Išplečia PA	PA 1 Peržiūrėti balso įrašų sąrašą
	Apima PA	Nėra.
	Specializuoja PA	Nėra.
Pagrindinis įvykių srautas		
Sistemos reakcija ir sprendimai		
1. Naudotojas pasirenka balso įrašo informacijos redagavimo funkciją.	1.1. Sistema atveria balso įrašo informacijos langą.	
2. Naudotojas užpildo informacinius laukus.	2.1. Lange rodoma užpildyta informacija.	
3. Naudotojas spaudžia mygtuką [Išsaugoti].	3.1. Informacija išsaugoma. 3.2. Atveriamas balso įrašų sąrašo langas.	
4. Naudotojas baigia PA.		
Po sąlyga:	Balso įrašai išsaugoti DB kartu su juos aprašančia informacija.	
Alternatyvūs scenarijai		
3a. Naudotojas spaudžia mygtuką [Atšaukti].	Informacija neišsaugoma. Atveriamas balso įrašų sąrašo langas.	

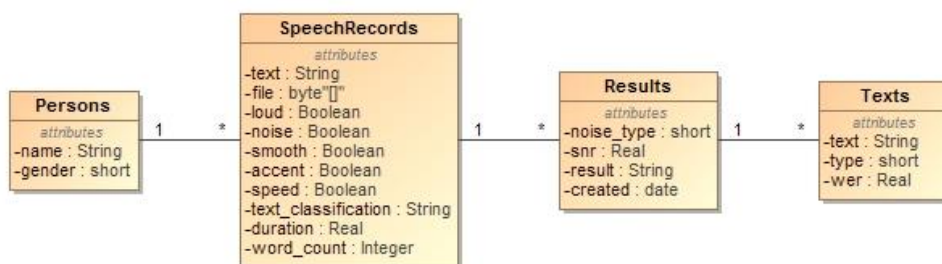
7.5 lentelė. Panaudos atvejo „PA 2 Inicijuoti garso įrašų apdorojimą“ specifikacija

PA 2 Inicijuoti garso įrašų apdorojimą²		
Tikslas. Inicijuoti DB esančių balso įrašų apdorojimą		
Aprašymas. Šis PA aprašo DB saugomų balso įrašų apdorojimo inicijavimą		
Prieš sąlyga		DB yra saugomas bent vienas balso įrašas
Aktorius		Eksperto vykdytojas
Sužadinimo sąlyga		Naudotojas nori inicijuoti DB saugomų balso įrašų apdorojimą
Susiję panaudos atvejai	Išplečia PA	Nėra.
	Apima PA	Nėra.
	Specializuoja PA	Nėra.
Pagrindinis įvykių šrautas		Sistemos reakcija ir sprendimai
1. Naudotojas pasirenka DB saugomų balso įrašų apdorojimo inicijavimo funkciją.		1.1. Sistema vykdo garso įrašų apdorojimo procesą (2.4.1 poskyris). 1.2. Naudotojui rodoma progreso juosta. 1.3. Naudotojui parodomas sėkmės arba klaidos pranešimas.
2. Naudotojas spaudžia mygtuką [OK].		2.1. Sėkmės arba klaidos pranešimas uždaromas.
3. Naudotojas baigia PA.		
Po sąlyga:		Visi DB saugomi balso įrašai yra apdoroti ir parodytas sėkmės pranešimas ARBA parodytas klaidos pranešimas.

7.6 lentelė. Panaudos atvejo „PA 3 Peržiūrėti garso įrašų apdorojimo rezultatus“ specifikacija

PA 3 Peržiūrėti garso įrašų apdorojimo rezultatus		
Tikslas. Peržiūrėti garso įrašų apdorojimo rezultatus		
Aprašymas. Šis PA aprašo garso įrašų apdorojimo rezultatų peržiūrą		
Prieš sąlyga		Įvykdytas „PA 2 Inicijuoti garso įrašų apdorojimą“
Aktorius		Eksperto vykdytojas
Sužadinimo sąlyga		Naudotojas nori peržiūrėti garso įrašų apdorojimo rezultatus
Susiję panaudos atvejai	Išplečia PA	Nėra.
	Apima PA	Nėra.
	Specializuoja PA	Nėra.
Pagrindinis įvykių šrautas		Sistemos reakcija ir sprendimai
1. Naudotojas pasirenka garso įrašų apdorojimo rezultatų peržiūros funkciją.		1.1. Atveriamas garso įrašų apdorojimo rezultatų langas.
2. Naudotojas baigia PA.		
Po sąlyga:		Peržiūrėti garso įrašų apdorojimo rezultatai.

7.2. Dalykinės srities esybių klasių modelis



7.1 pav. Dalykinės srities esybių klasių modelis

² Apdoroti garso įrašą – gauti šnekos atpažinimo rezultatus tekstu.

7.3. Testavimo atvejai

7.7 lentelė. Atpažinti skaitomą tekstą (komandos)

Testavimo atvejis TST-0001: Atpažinti skaitomą tekstą (komandos)		
Pradinės sąlygos:		
<ul style="list-style-type: none"> Nėra. 		
<u>Žingsnis:</u>	<u>Veiksmas:</u>	<u>Laukiami rezultatai:</u>
1.	„Google Chrome“ naršyklėje paleidžiama šnekos atpažintuvo demonstracinė versija.	Šnekos atpažintuvo demonstracinė versija paleista.
2.	Šnekos atpažintuve parenkama lietuvių kalba bei spaudžiamas įrašymo mygtukas.	Šnekos atpažintuvas klausosi mikrofono.
3.	Paleidžiama standartinė garso įrašymo programinė įranga.	Standartinė garso įrašymo programinė įranga paleista.
4.	Spaudžiamas įrašymo paleidimo mygtukas ir skaitomas tekstas, kurį norima įgarsinti.	Vyksta balso įrašymas.
5.	Spaudžiamas įrašymo stabdymo mygtukas.	Įrašymas sustabdomas.
6.	Sustabdomas „Google Chrome“ naršyklėje paleistas šnekos atpažintuvas.	Šnekos atpažintuvas sustabdytas.
7.	Užfiksuojamas atpažintas tekstas.	Atpažintas tekstas užfiksuotas.

7.8 lentelė. Atpažinti skaitomą tekstą (žodžių seka)

Testavimo atvejis TST-0002: Atpažinti skaitomą tekstą (žodžių seka)		
Pradinės sąlygos:		
<ul style="list-style-type: none"> Nėra. 		
<u>Žingsnis:</u>	<u>Veiksmas:</u>	<u>Laukiami rezultatai:</u>
1.	„Google Chrome“ naršyklėje paleidžiama šnekos atpažintuvo demonstracinė versija.	Šnekos atpažintuvo demonstracinė versija paleista.
2.	Šnekos atpažintuve parenkama lietuvių kalba bei spaudžiamas įrašymo mygtukas.	Šnekos atpažintuvas klausosi mikrofono.
3.	Paleidžiama standartinė garso įrašymo programinė įranga.	Standartinė garso įrašymo programinė įranga paleista.
4.	Spaudžiamas įrašymo paleidimo mygtukas ir skaitomas tekstas, kurį norima įgarsinti.	Vyksta balso įrašymas.
5.	Spaudžiamas įrašymo stabdymo mygtukas.	Įrašymas sustabdomas.
6.	Sustabdomas „Google Chrome“ naršyklėje paleistas šnekos atpažintuvas.	Šnekos atpažintuvas sustabdytas.
7.	Užfiksuojamas atpažintas tekstas.	Atpažintas tekstas užfiksuotas.

7.9 lentelė. Atpažinti skaitomą tekstą (rišli šneka)

Testavimo atvejis TST-0003: Atpažinti skaitomą tekstą (rišli šneka)		
Pradinės sąlygos:		
<ul style="list-style-type: none"> Nėra. 		
<u>Žingsnis:</u>	<u>Veiksmas:</u>	<u>Laukiami rezultatai:</u>
1.	„Google Chrome“ naršyklėje paleidžiama šnekos atpažintuvo demonstracinė versija.	Šnekos atpažintuvo demonstracinė versija paleista.
2.	Šnekos atpažintuve parenkama lietuvių kalba bei spaudžiamas įrašymo mygtukas.	Šnekos atpažintuvas klausosi mikrofono.
3.	Paleidžiama standartinė garso įrašymo programinė įranga.	Standartinė garso įrašymo programinė įranga paleista.

4.	Spaudžiamas įrašymo paleidimo mygtukas ir skaitomas tekstas, kurį norima įgarsinti.	Vyksta balso įrašymas.
5.	Spaudžiamas įrašymo stabdymo mygtukas.	Įrašymas sustabdomas.
6.	Sustabdomas „Google Chrome“ naršyklėje paleistas šnekos atpažintuvas.	Šnekos atpažintuvas sustabdytas.
7.	Užfiksuojamas atpažintas tekstas.	Atpažintas tekstas užfiksuotas.

7.10 lentelė. Atpažinti skaitomą tekstą (spontaniška šneka)

Testavimo atvejis TST-0004: Atpažinti skaitomą tekstą (spontaniška šneka)		
Pradinės sąlygos:		
<ul style="list-style-type: none"> Nėra. 		
<u>Žingsnis:</u>	<u>Veiksmas:</u>	<u>Laukiami rezultatai:</u>
1.	„Google Chrome“ naršyklėje paleidžiama šnekos atpažintuvo demonstracinė versija.	Šnekos atpažintuvo demonstracinė versija paleista.
2.	Šnekos atpažintuve parenkama lietuvių kalba bei spaudžiamas įrašymo mygtukas.	Šnekos atpažintuvas klausosi mikrofono.
3.	Paleidžiama standartinė garso įrašymo programinė įranga.	Standartinė garso įrašymo programinė įranga paleista.
4.	Spaudžiamas įrašymo paleidimo mygtukas ir skaitomas tekstas, kurį norima įgarsinti.	Vyksta balso įrašymas.
5.	Spaudžiamas įrašymo stabdymo mygtukas.	Įrašymas sustabdomas.
6.	Sustabdomas „Google Chrome“ naršyklėje paleistas šnekos atpažintuvas.	Šnekos atpažintuvas sustabdytas.
7.	Užfiksuojamas atpažintas tekstas.	Atpažintas tekstas užfiksuotas.

7.11 lentelė. Atpažinti skaitomą tekstą (komandos; su triukšmu)

Testavimo atvejis TST-0005: Atpažinti skaitomą tekstą (komandos; su triukšmu)		
Pradinės sąlygos:		
<ul style="list-style-type: none"> Nėra. 		
<u>Žingsnis:</u>	<u>Veiksmas:</u>	<u>Laukiami rezultatai:</u>
1.	„Google Chrome“ naršyklėje paleidžiama šnekos atpažintuvo demonstracinė versija.	Šnekos atpažintuvo demonstracinė versija paleista.
2.	Šnekos atpažintuve parenkama lietuvių kalba bei spaudžiamas įrašymo mygtukas.	Šnekos atpažintuvas klausosi mikrofono.
3.	Paleidžiama standartinė garso įrašymo programinė įranga.	Standartinė garso įrašymo programinė įranga paleista.
4.	Spaudžiamas įrašymo paleidimo mygtukas ir skaitomas tekstas, kurį norima įgarsinti.	Vyksta balso įrašymas.
5.	Spaudžiamas įrašymo stabdymo mygtukas.	Įrašymas sustabdomas.
6.	Sustabdomas „Google Chrome“ naršyklėje paleistas šnekos atpažintuvas.	Šnekos atpažintuvas sustabdytas.
7.	Užfiksuojamas atpažintas tekstas.	Atpažintas tekstas užfiksuotas.

7.12 lentelė. Atpažinti skaitomą tekstą (žodžių seka; su triukšmu)

Testavimo atvejis TST-0006: Atpažinti skaitomą tekstą (žodžių seka; su triukšmu)		
Pradinės sąlygos:		
<ul style="list-style-type: none"> Nėra. 		
<u>Žingsnis:</u>	<u>Veiksmas:</u>	<u>Laukiami rezultatai:</u>

1.	„Google Chrome“ naršyklėje paleidžiama šnekos atpažintuvo demonstracinė versija.	Šnekos atpažintuvo demonstracinė versija paleista.
2.	Šnekos atpažintuve parenkama lietuvių kalba bei spaudžiamas įrašymo mygtukas.	Šnekos atpažintuvas klausosi mikrofono.
3.	Paleidžiama standartinė garso įrašymo programinė įranga.	Standartinė garso įrašymo programinė įranga paleista.
4.	Spaudžiamas įrašymo paleidimo mygtukas ir skaitomas tekstas, kurį norima įgarsinti.	Vyksta balso įrašymas.
5.	Spaudžiamas įrašymo stabdymo mygtukas.	Įrašymas sustabdomas.
6.	Sustabdomas „Google Chrome“ naršyklėje paleistas šnekos atpažintuvas.	Šnekos atpažintuvas sustabdytas.
7.	Užfiksuojamas atpažintas tekstas.	Atpažintas tekstas užfiksuotas.

7.13 lentelė. Atpažinti skaitomą tekstą (rišli šneka; su triukšmu)

Testavimo atvejis TST-0007: Atpažinti skaitomą tekstą (rišli šneka; su triukšmu)		
Pradinės sąlygos:		
<ul style="list-style-type: none"> Nėra. 		
<u>Žingsnis:</u>	<u>Veiksmas:</u>	<u>Laukiami rezultatai:</u>
1.	„Google Chrome“ naršyklėje paleidžiama šnekos atpažintuvo demonstracinė versija.	Šnekos atpažintuvo demonstracinė versija paleista.
2.	Šnekos atpažintuve parenkama lietuvių kalba bei spaudžiamas įrašymo mygtukas.	Šnekos atpažintuvas klausosi mikrofono.
3.	Paleidžiama standartinė garso įrašymo programinė įranga.	Standartinė garso įrašymo programinė įranga paleista.
4.	Spaudžiamas įrašymo paleidimo mygtukas ir skaitomas tekstas, kurį norima įgarsinti.	Vyksta balso įrašymas.
5.	Spaudžiamas įrašymo stabdymo mygtukas.	Įrašymas sustabdomas.
6.	Sustabdomas „Google Chrome“ naršyklėje paleistas šnekos atpažintuvas.	Šnekos atpažintuvas sustabdytas.
7.	Užfiksuojamas atpažintas tekstas.	Atpažintas tekstas užfiksuotas.

7.14 lentelė. Atpažinti skaitomą tekstą (spontaniška šneka; su triukšmu)

Testavimo atvejis TST-0008: Atpažinti skaitomą tekstą (spontaniška šneka; su triukšmu)		
Pradinės sąlygos:		
<ul style="list-style-type: none"> Nėra. 		
<u>Žingsnis:</u>	<u>Veiksmas:</u>	<u>Laukiami rezultatai:</u>
1.	„Google Chrome“ naršyklėje paleidžiama šnekos atpažintuvo demonstracinė versija.	Šnekos atpažintuvo demonstracinė versija paleista.
2.	Šnekos atpažintuve parenkama lietuvių kalba bei spaudžiamas įrašymo mygtukas.	Šnekos atpažintuvas klausosi mikrofono.
3.	Paleidžiama standartinė garso įrašymo programinė įranga.	Standartinė garso įrašymo programinė įranga paleista.
4.	Spaudžiamas įrašymo paleidimo mygtukas ir skaitomas tekstas, kurį norima įgarsinti.	Vyksta balso įrašymas.
5.	Spaudžiamas įrašymo stabdymo mygtukas.	Įrašymas sustabdomas.
6.	Sustabdomas „Google Chrome“ naršyklėje paleistas šnekos atpažintuvas.	Šnekos atpažintuvas sustabdytas.
7.	Užfiksuojamas atpažintas tekstas.	Atpažintas tekstas užfiksuotas.

7.15 lentelė. Automatinis įrašų apdorojimas

Testavimo atvejis TST-0009: Automatinis įrašų apdorojimas		
Pradinės sąlygos:		
<ul style="list-style-type: none"> • Įvykdyti testavimo atvejai TST-0001 – TST-0008. 		
<u>Žingsnis:</u>	<u>Veiksmas:</u>	<u>Laukiami rezultatai:</u>
1.	Balso įrašai importuojami į PL.	Balso įrašai suimportuoti.
2.	Paleidžiama garso įrašų apdorojimo posistemė.	Vyksta automatinis garso įrašų apdorojimas.
3.	Gauti automatinio garso įrašų apdorojimo rezultatai palyginami su gautais rezultatais per naršyklę.	Gauti šnekos atpažinimo rezultatai sutampa.

7.4. Detalūs šnekos atpažinimo rezultatai

7.4.1. Šnekos atpažinimo rezultatai pagal kalbėtoją

7.16 lentelė. Šnekos atpažinimo rezultatai pagal kalbėtoją (be triukšmo)

Kalbėtojo ID	Siūstų atpažinti frazių kiekis	Atpažinimo rezultatai					
		Frazių kiekis	Žodžių kiekis frazėse	Įterpimai	Pakeitimai	Panaikinimai	Vidutinis WER, %
2	111	83	445	13	131	23	31,93
3	222	207	749	7	164	33	30,59
4	209	172	699	6	148	16	28,09
5	218	195	737	7	211	34	32,79
6	220	192	730	9	192	33	28,99
7	216	165	697	9	243	82	40,87
8	218	203	745	9	223	47	31,61
9	216	200	742	7	157	29	25,87
10	221	171	710	6	158	74	29,32
11	220	149	668	8	177	66	38,16
12	222	165	694	9	164	52	33,18
13	215	149	670	4	185	154	42,47
14	221	2	8	0	2	6	100,00
16	221	84	495	1	45	256	46,45
17	264	19	156	1	17	109	62,37
18	205	100	719	0	83	317	46,41
21	209	3	17	0	1	12	76,19
23	212	118	788	3	111	244	36,91
25	207	102	691	1	108	341	48,21
26	207	101	674	2	121	306	51,37
27	218	187	922	8	225	55	29,88
28	219	172	909	6	205	157	29,29
29	219	107	775	2	122	319	53,70
30	221	126	850	0	116	210	33,76
31	221	93	683	1	60	373	49,33
32	221	118	777	0	124	313	48,07
35	206	115	694	3	143	216	43,65
36	232	85	577	0	75	299	51,05
37	257	106	716	4	112	206	38,36
41	199	94	575	1	85	244	50,72
42	184	58	371	0	53	230	64,48

Kalbėtojo ID	Siųstų atpažinti frazių kiekis	Atpažinimo rezultatai					
		Frazių kiekis	Žodžių kiekis frazėse	Įterpimai	Pakeitimai	Panaikinimai	Vidutinis WER, %
43	757	541	1514	2	288	406	31,42
44	216	136	707	5	143	183	38,54
45	216	124	698	0	94	162	30,21
46	222	128	802	3	140	231	39,27
47	216	85	517	0	75	289	61,28
48	222	133	822	4	123	180	33,84
49	212	29	124	0	12	71	37,53
50	211	49	337	0	29	231	54,13
51	275	88	444	2	57	224	53,68
52	257	93	489	0	70	249	53,43
53	209	67	463	0	41	317	55,52
54	261	152	856	10	125	149	26,27
55	208	92	648	0	89	329	52,49
56	261	67	370	2	40	218	47,50
57	206	37	262	0	16	171	50,97
58	205	83	626	0	70	291	47,88
59	217	57	469	0	54	324	69,02
60	219	82	477	0	42	303	54,76
61	221	120	690	1	82	364	43,26
62	221	161	847	1	111	390	42,28
63	208	157	753	2	108	290	34,16
64	208	99	465	0	51	252	39,26
65	207	152	748	0	111	309	38,14
66	207	102	452	0	44	249	43,43
67	213	106	499	1	72	278	48,49
68	212	45	238	0	27	170	61,44
69	209	57	279	0	25	166	47,03
70	207	52	321	0	30	214	54,66
71	213	107	670	0	126	247	49,71
73	217	107	730	5	83	312	45,17
74	194	59	481	0	52	287	62,80
75	220	58	299	1	33	180	48,16
76	213	17	116	0	13	84	70,19
77	211	62	288	0	23	181	42,51
78	209	96	396	0	48	198	35,13
79	265	135	673	0	98	314	45,10
80	212	123	547	0	59	295	41,78
81	207	107	534	0	51	275	39,01
82	207	108	398	0	46	205	36,86
83	219	154	779	0	82	385	38,96
84	219	111	523	0	62	272	41,84
85	219	189	923	1	171	299	39,58
86	221	158	855	1	125	281	34,95
87	221	178	911	2	148	247	36,38
88	221	180	905	0	134	263	32,90
89	201	156	719	0	133	223	37,36

Kalbėtojo ID	Siųstų atpažinti frazių kiekis	Atpažinimo rezultatai					
		Frazių kiekis	Žodžių kiekis frazėse	Įterpimai	Pakeitimai	Panaikinimai	Vidutinis WER, %
90	203	165	770	3	117	264	36,80
91	200	123	552	1	83	269	45,62
92	208	138	641	0	77	291	37,21
93	207	157	792	1	173	156	33,59
94	208	159	770	0	129	263	35,25
95	212	155	834	1	104	324	35,49
96	213	134	629	0	59	315	35,27
97	213	170	860	4	163	184	36,68
98	210	167	711	2	140	138	31,72
99	209	158	668	0	104	252	35,47
100	210	99	390	0	57	214	48,63
101	213	160	752	2	116	190	32,06
102	213	166	733	0	116	268	38,27
103	213	180	773	11	196	75	34,15
104	219	180	839	4	122	120	30,41
105	219	170	731	2	104	304	40,75
106	219	180	821	4	169	214	38,82
107	222	175	714	1	135	196	35,83
108	222	177	696	1	135	155	36,50
109	222	157	628	1	89	265	42,80
110	211	137	740	2	121	318	43,34
111	548	381	772	3	177	181	30,42
112	547	339	764	2	167	149	32,35
113	547	420	853	8	208	153	33,08
114	558	485	1006	1	201	163	24,35
115	558	479	1028	2	284	130	32,55
116	555	475	1013	3	206	157	26,27
117	557	498	1039	3	276	152	31,95
118	555	361	810	4	177	187	32,50
119	558	448	954	1	233	152	32,12
121	557	224	461	1	93	135	33,72
122	558	460	1051	6	243	101	28,37
123	553	534	1078	5	299	120	30,86
124	551	340	804	2	144	208	29,40
125	553	368	828	2	171	214	34,36
126	552	376	773	3	176	166	33,25
127	551	352	749	1	164	163	31,99
128	549	430	894	4	211	152	28,64
129	165	122	443	1	93	118	37,04
130	44	28	308	3	61	116	59,48
131	42	19	193	0	39	88	63,08
132	56	40	381	3	72	168	64,20
133	53	29	282	1	37	156	69,42
134	52	41	379	0	56	137	51,09
135	566	174	574	0	112	215	44,50
136	554	250	718	2	154	223	44,93

Kalbėtojo ID	Siūstų atpažinti frazių kiekis	Atpažinimo rezultatai					
		Frazių kiekis	Žodžių kiekis frazėse	Įterpimai	Pakeitimai	Panaikinimai	Vidutinis WER, %
137	549	417	967	2	232	177	35,86
138	551	443	1005	4	202	129	26,31
139	553	396	934	3	231	193	35,27
140	540	390	900	1	230	219	37,37
141	61	41	382	1	53	227	70,50
142	61	42	382	0	59	186	61,71
143	58	43	374	0	71	166	64,84
144	552	369	840	0	170	238	30,82
145	546	270	677	3	138	204	38,19
146	549	340	634	1	165	158	34,71
147	564	437	1031	1	283	307	39,60
148	567	390	925	5	189	279	38,09
149	55	45	461	1	80	151	52,24
150	179	88	562	3	121	351	87,87
151	157	101	1037	10	313	148	62,20
152	44	9	188	0	26	141	88,43
153	41	34	388	0	47	189	59,01
154	40	34	400	1	61	189	64,12
155	77	54	588	2	71	376	74,85
156	72	66	704	2	92	285	51,54
157	77	42	571	0	50	435	81,65
158	89	66	709	0	76	420	70,39
159	88	60	663	0	53	416	69,76
160	88	70	782	2	110	371	62,47
161	96	70	722	0	69	416	66,25
162	93	77	794	1	116	348	59,18
163	95	85	876	5	125	276	46,24
164	106	90	858	0	115	443	63,00
165	105	91	883	1	96	407	54,68
166	106	72	707	3	91	409	70,23
167	103	85	910	0	129	478	66,58
168	107	79	868	0	106	498	67,88
169	107	78	831	0	93	433	64,28
170	127	116	1041	2	131	316	42,88
171	126	115	1024	2	132	353	46,76
172	126	96	905	0	110	471	61,98
173	127	73	797	0	108	468	70,12
174	127	119	1179	8	266	242	45,17
175	127	45	481	0	46	325	74,82
176	109	66	739	0	56	474	71,28
177	109	56	660	0	45	481	79,00
178	110	47	554	0	52	394	79,08
179	127	113	1170	1	98	502	52,65
180	127	119	1225	0	195	371	46,10
182	94	88	909	0	124	270	42,25
183	91	86	896	1	101	255	38,35

Kalbėtojo ID	Siūstų atpažinti frazių kiekis	Atpažinimo rezultatai					
		Frazių kiekis	Žodžių kiekis frazėse	Įterpimai	Pakeitimai	Panaikinimai	Vidutinis WER, %
184	92	82	858	4	117	343	53,62
185	106	99	936	4	181	437	65,15
186	108	94	897	4	139	377	55,91
187	107	87	851	2	133	414	62,46
188	104	101	1048	3	183	289	46,16
189	107	33	366	0	26	276	82,03
190	105	15	165	0	12	138	89,18
191	129	83	818	3	95	452	65,83
192	130	92	833	3	78	474	63,98
193	129	118	1038	2	195	350	51,39
194	128	103	1066	4	185	424	57,30
195	125	122	1181	2	195	322	44,62
196	127	109	1091	1	165	496	61,26
197	113	53	582	0	56	408	77,73
198	113	93	1025	1	137	467	59,46
199	113	88	984	1	126	557	68,99
200	113	69	818	0	84	576	78,64
201	127	104	1086	1	136	660	71,88
202	127	20	270	0	9	228	82,95
203	128	77	829	1	70	535	73,07
204	92	60	628	2	59	340	63,57
205	93	85	881	1	116	337	51,57
206	94	85	885	2	104	403	56,05
207	107	55	536	1	44	355	74,14
208	108	58	584	0	66	393	77,01
209	107	69	753	0	64	498	73,63
210	107	31	338	0	22	258	80,45
211	107	95	1005	4	160	330	50,24
212	107	71	748	1	101	416	69,51
213	126	116	1033	2	163	263	41,27
214	132	111	1010	3	198	379	56,43
216	126	101	991	2	160	448	60,21
228	108	101	957	3	175	261	45,93
230	104	104	965	14	152	64	24,37
234	127	125	1098	4	197	243	40,88
235	130	104	950	1	129	455	59,71
238	127	123	1215	3	217	195	34,38
240	109	107	1152	1	171	375	48,41
241	109	96	1071	1	164	523	65,00
242	108	106	1146	2	181	367	50,72
243	126	124	1287	7	317	193	41,47
244	125	125	1332	4	207	263	37,25
245	123	122	1249	6	209	94	25,25
247	92	90	920	7	128	227	38,02
248	67	6	126	5	30	9	34,30
250	53	41	794	18	329	21	42,56

Kalbėtojo ID	Siųstų atpažinti frazių kiekis	Atpažinimo rezultatai					
		Frazių kiekis	Žodžių kiekis frazėse	Įterpimai	Pakeitimai	Panaikinimai	Vidutinis WER, %
251	53	49	975	15	449	35	47,06
262	53	2	17	0	5	0	33,57
263	53	4	35	2	9	0	30,19
264	53	3	25	1	12	0	48,05
265	51	0	-	-	-	-	-
266	53	0	-	-	-	-	-
267	53	3	28	1	6	2	32,46
268	51	1	7	0	2	0	28,57
269	52	1	10	1	2	0	30,00
270	53	3	28	0	9	1	37,23
271	53	3	40	1	16	17	77,14
272	52	3	25	2	9	1	43,29
273	53	2	14	3	5	0	57,14
274	53	2	21	2	7	1	45,91
275	53	3	25	0	11	2	51,51
276	53	5	55	1	20	0	44,44
277	53	2	21	1	11	1	60,00
278	53	4	35	10	15	1	70,84
279	53	2	17	0	5	0	33,57
280	51	1	7	0	3	0	42,85
281	53	1	11	3	8	0	100,00
282	53	1	11	1	7	0	72,72
283	33	0	-	-	-	-	-
284	53	0	-	-	-	-	-
285	53	2	17	0	2	0	14,29
286	53	0	-	-	-	-	-
287	53	3	28	1	10	0	38,83
288	53	0	-	-	-	-	-
289	53	0	-	-	-	-	-
290	52	1	10	0	1	0	10,00
291	50	2	34	0	28	3	81,25
292	52	2	17	0	4	0	26,43
301	74	5	95	2	52	12	70,59
302	74	5	60	1	19	1	33,33
303	74	1	7	0	4	0	57,14
304	74	1	7	0	4	0	57,14
305	73	4	48	2	9	1	25,73
307	74	1	20	1	10	0	55,00
308	74	11	186	1	118	15	66,74
309	74	3	37	5	11	2	48,52
500	169	118	413	1	89	26	32,82
502	185	185	1852	23	426	106	30,09
503	207	206	2045	23	630	247	45,34
504	41	40	293	3	77	33	39,14
505	221	221	2234	14	487	155	30,51
506	123	123	1233	9	201	36	19,92

Kalbėtojo ID	Siųstų atpažinti frazių kiekis	Atpažinimo rezultatai					
		Frazių kiekis	Žodžių kiekis frazėse	Įterpimai	Pakeitimai	Panaikinimai	Vidutinis WER, %
507	227	227	2171	27	758	233	47,59
508	205	205	1911	29	411	61	27,65
509	215	215	2081	18	504	273	38,93
510	213	213	2124	21	474	169	31,53
511	215	163	882	5	269	64	34,40
512	196	144	609	5	180	50	34,28
513	147	81	175	2	40	11	22,06
514	205	145	761	7	180	52	25,45
515	221	149	868	2	244	154	39,11
516	102	92	460	3	114	58	33,74
517	190	149	791	9	204	78	34,90
518	137	68	467	9	134	60	44,34
519	187	135	782	7	237	109	43,00
520	167	106	706	5	206	92	38,35
522	213	172	754	12	252	68	42,60
523	222	184	813	14	231	50	35,15
524	219	175	876	8	298	117	40,85
525	141	122	168	2	59	3	37,43
526	199	168	710	7	251	95	43,29
527	176	147	517	4	132	22	28,70
528	142	107	152	4	53	4	41,67
529	218	176	801	7	250	108	42,22
530	159	156	748	4	225	77	34,20
531	157	156	828	8	271	63	33,14
532	219	186	818	11	219	61	32,34
533	209	179	835	6	238	78	35,76
534	141	115	161	0	46	2	32,03
535	208	172	832	7	300	119	44,41
536	218	189	928	5	337	198	51,93
537	218	201	933	8	275	87	37,77
538	207	181	776	4	261	139	39,95
539	154	146	806	7	266	207	51,83
540	116	101	706	4	270	148	59,73
541	103	29	43	1	28	1	70,69
543	142	71	116	0	29	9	33,45
544	214	146	858	5	248	80	42,34
545	97	68	454	2	177	147	67,66
546	200	138	761	8	244	93	39,21
547	216	127	713	8	322	138	63,44
548	205	149	846	3	180	40	26,37
549	213	178	876	6	222	62	32,01
550	213	169	761	7	247	76	40,02
551	215	160	819	5	188	73	30,02
552	153	147	784	10	275	86	40,26
554	204	177	766	5	192	47	30,46
555	169	139	912	1	220	298	55,56

Kalbėtojo ID	Siųstų atpažinti frazių kiekis	Atpažinimo rezultatai					
		Frazių kiekis	Žodžių kiekis frazėse	Įterpimai	Pakeitimai	Panaikinimai	Vidutinis WER, %
556	207	152	774	11	291	109	49,94
557	214	164	899	5	265	83	41,31
558	171	162	839	11	281	46	36,40
561	7	5	94	0	34	33	73,99
562	29	23	206	2	54	32	40,72
563	139	134	1366	21	309	90	31,35
564	73	68	658	10	182	56	38,33
565	74	72	787	3	240	111	43,38
566	62	61	595	3	173	114	49,79
567	67	63	782	4	220	55	36,13
568	65	60	607	7	160	31	31,48
569	42	42	401	3	95	39	33,55
570	76	76	838	9	302	107	50,76
571	88	84	761	6	175	71	34,78
572	78	72	688	6	187	80	38,92
573	75	72	721	4	150	32	26,02
574	97	95	835	8	248	140	50,09
575	85	80	882	5	209	195	47,65
576	74	71	695	3	183	108	40,91
577	74	69	721	7	274	146	61,17
578	61	58	604	3	166	84	42,08
579	65	64	671	5	163	64	34,92
580	57	54	561	2	180	98	50,16
581	83	83	830	1	157	118	31,73
582	68	65	638	3	193	88	46,07
583	82	78	636	6	125	65	29,95
584	45	45	461	4	92	36	30,14
585	70	67	655	3	150	57	32,71
586	81	79	830	13	256	139	49,91
587	98	94	902	9	325	119	49,61
588	97	91	847	13	234	73	39,92
589	72	72	854	2	211	54	29,09
590	57	54	537	10	191	74	47,75
593	79	76	810	11	181	91	35,56
595	96	91	854	15	252	119	46,87
596	86	84	852	14	260	72	41,61
597	41	37	379	1	135	140	73,14
598	80	75	731	11	234	100	48,21
599	78	70	663	4	260	153	63,33
600	284	266	1092	16	431	106	55,76
601	202	199	838	6	322	105	50,71
602	98	94	728	1	168	287	47,60
603	256	220	928	4	351	249	62,41
604	227	212	929	6	300	121	44,99
605	209	208	801	19	304	76	48,64
606	34	31	350	5	153	57	64,34

Kalbėtojo ID	Siųstų atpažinti frazių kiekis	Atpažinimo rezultatai					
		Frazių kiekis	Žodžių kiekis frazėse	Įterpimai	Pakeitimai	Panaikinimai	Vidutinis WER, %
607	79	76	718	11	191	66	36,82
608	86	79	703	1	190	236	58,18
609	72	68	733	7	324	138	66,28
610	57	53	528	8	128	39	33,08
611	80	78	757	5	173	89	35,70
907	65	31	539	19	217	58	53,04

7.4.2. Šnekos atpažinimo rezultatai pagal žodžių kiekį frazėje

7.17 lentelė. Šnekos atpažinimo rezultatai pagal žodžių kiekį frazėje (be triukšmo)

Žodžių skaičius frazėje	Atpažinimo rezultatai								Vidutinis WER, %
	Frazių kiekis	Žodžių kiekis frazėse	WER						
			Įterpimai		Pakeitimai		Panaikinimai		
			Vnt.	%	Vnt.	%	Vnt.	%	
1	13741	13741	61	0,44	3286	23,91	0	0,00	24,36
2	5174	10348	40	0,39	2938	28,39	1132	10,94	39,72
3	2763	8289	25	0,30	2498	30,14	1351	16,30	46,73
4	994	3976	13	0,33	895	22,51	988	24,85	47,69
5	1464	7320	35	0,48	1613	22,04	2020	27,60	50,11
6	1782	10692	93	0,87	2468	23,08	2952	27,61	51,56
7	2120	14840	56	0,38	3131	21,10	4122	27,78	49,25
8	1929	15432	74	0,48	3563	23,09	4545	29,45	53,02
9	2284	20556	99	0,48	4164	20,26	6114	29,74	50,48
10	2308	23080	150	0,65	4365	18,91	6958	30,15	49,71
11	1485	16335	153	0,94	3120	19,10	4988	30,54	50,57
12	1594	19128	90	0,47	3643	19,05	6098	31,88	51,39
13	1351	17563	118	0,67	3543	20,17	5627	32,04	52,88
14	910	12740	52	0,41	2461	19,32	4171	32,74	52,46
15	818	12270	65	0,53	2274	18,53	3961	32,28	51,34
16	623	9968	47	0,47	1700	17,05	3391	34,02	51,54
17	390	6630	26	0,39	1124	16,95	2348	35,41	52,76
18	313	5634	13	0,23	1051	18,65	2024	35,92	54,81
19	178	3382	8	0,24	535	15,82	1226	36,25	52,3
20	67	1340	20	1,49	375	27,99	253	18,88	48,36
21	84	1764	2	0,11	345	19,56	664	37,64	57,31
22	9	198	4	2,02	86	43,43	23	11,62	57,07
23	24	552	12	2,17	222	40,22	42	7,61	50
24	32	768	5	0,65	201	26,17	193	25,13	51,95
25	20	500	5	1,00	242	48,40	21	4,20	53,6
26	10	260	3	1,15	147	56,54	34	13,08	70,77
27	10	270	1	0,37	140	51,85	55	20,37	72,59
28	6	168	3	1,79	107	63,69	24	14,29	79,76
29	2	58	0	0,00	10	17,24	4	6,90	24,14
30	3	90	0	0,00	18	20,00	12	13,33	33,33
31	11	341	20	5,87	80	23,46	12	3,52	32,84
32	2	64	1	1,56	17	26,56	4	6,25	34,38

Žodžių skaičius frazėje	Atpažinimo rezultatai								Vidutinis WER, %
	Frazių kiekis	Žodžių kiekis frazėse	WER						
			Įterpimai		Pakeitimai		Panaikinimai		
			Vnt.	%	Vnt.	%	Vnt.	%	
33	1	33	0	0,00	18	54,55	7	21,21	75,75
34	2	68	1	1,47	14	20,59	21	30,88	52,94
35	2	70	0	0,00	21	30,00	26	37,14	67,14
36	0	-	-	-	-	-	-	-	-
37	0	-	-	-	-	-	-	-	-
38	0	-	-	-	-	-	-	-	-
39	0	-	-	-	-	-	-	-	-
40	2	80	0	0,00	28	35,00	4	5,00	40
41	0	-	-	-	-	-	-	-	-
42	0	-	-	-	-	-	-	-	-
43	1	43	0	0,00	1	2,33	42	97,67	100
44	1	44	0	0,00	9	20,45	34	77,27	97,72
45	2	90	0	0,00	26	28,89	28	31,11	60
46	0	-	-	-	-	-	-	-	-
47	1	47	0	0,00	8	17,02	36	76,60	93,61
48	0	-	-	-	-	-	-	-	-
49	0	-	-	-	-	-	-	-	-
50	0	-	-	-	-	-	-	-	-
51	0	-	-	-	-	-	-	-	-
52	0	-	-	-	-	-	-	-	-
53	1	53	0	0,00	7	13,21	4	7,55	20,75
54	0	-	-	-	-	-	-	-	-
55	0	-	-	-	-	-	-	-	-
56	0	-	-	-	-	-	-	-	-
57	0	-	-	-	-	-	-	-	-
58	0	-	-	-	-	-	-	-	-
59	0	-	-	-	-	-	-	-	-
60	1	60	0	0,00	1	1,67	59	98,33	100
61	0	-	-	-	-	-	-	-	-
62	0	-	-	-	-	-	-	-	-
63	0	-	-	-	-	-	-	-	-
64	0	-	-	-	-	-	-	-	-
65	0	-	-	-	-	-	-	-	-
66	0	-	-	-	-	-	-	-	-
67	0	-	-	-	-	-	-	-	-
68	0	-	-	-	-	-	-	-	-
69	0	-	-	-	-	-	-	-	-
70	0	-	-	-	-	-	-	-	-
71	0	-	-	-	-	-	-	-	-
76	0	-	-	-	-	-	-	-	-
109	0	-	-	-	-	-	-	-	-