VILNIAUS UNIVERSITETAS

KAUNO HUMANITARINIS FAKULTETAS

INFORMATIKOS KATEDRA

Verslo informatikos studijų programa

Kodas 62109P101

INDRĖ JAKIMČIKIENĖ

MAGISTRO BAIGIAMASIS DARBAS

KALBOS SIGNALŲ KOKYBINIŲ CHARAKTERISTIKŲ ANALIZĖS ALGORITMAS

Kaunas 2010

TURINYS

JAKIMČIKIENĖ, Indrė. (2010) *Speech Signals Qualitative Characteristics Analysis System*. MBA*Master Graduation Paper. Kaunas: Vilnius University, Kaunas Faculty of Humanities, Department of Informatics.

SUMMARY

Research for the qualitative characteristics of the speech signal analysis, a description of the potential that exists to accomplish the recognition of sounds, including speech detection criteria to facilitate analysis was performed for qualitative characteristics. The paper used in the analysis of speech corpus in which the accumulated announcer speech recordings. Thus, the study objective - to establish the qualitative characteristics of speech analysis algorithm which can automatically determine whether the word record made the qualifying criteria. To achieve this requires the examination of existing segmentation algorithms, highlighting their advantages and disadvantages of a signal quality determination algorithm, speech preparation of quality research data, to evaluate speech quality determination algorithm, the results of the research findings. We base the scientific study of literature and summarized the methods for the implementation of the experiment - the experimental measurement and statistical analysis. The study we intend to carry out a speech quality analysis algorithm, the choice of speech quality performance criteria.

ĮVADAS

Kalbos signalų atpažinimas – tai reiškinys, su kuriuo mes susiduriame kiekvieną dieną. Girdint žmogaus šneką mes galime visai nesunkiai nusakyti jo lytį, amžių ar tapatybę, jei kalbantįjį pažįstame ar dažnai girdime jo balsą. Vis daugiau naujų šiuolaikinių technologijų taip pat grindžiamos balsų atpažinimu. Sprendžiant kalbos atpažinimo klausimą buvo įdėta daug pastangų ir pasiekta išties nemažai. Tačiau, nepaisant akivaizdžios pažangos, net ir geriausių automatinių atpažinimo sistemų efektyvumas yra prastesnis negu žmogaus (klausytojo) atpažinimo tikslumas, o nepalankioje aplinkoje - dar blogesnis. Taigi būtina ir toliau tirti atpažinimo metodus, siekiant pagerinti kalbos atpažinimo sistemų patikimumą ir dar yra tikrai daug erdvės tokių sistemų tobulinimui.

20 – ojo amžiaus pabaigoje ypatingai suaktyvėjo tyrimai skirti kalbančiojo atpažininimui bei pasiekta didelė pažanga kalbos sintezavimo srityje. Taigi didelė reikšmė šioje srityje skiriama kokybiškam kalbos signalų atskyrimui. Naudojantis šiuolaikinėmis technologijomis, kompiuterine įranga signalų atskyrimo uždavinys žymiai supaprastėja, tačiau vis tiek išlieka ganėtinai sudėtingas. Kalbos signalams analizuoti bei atskirti naudojami įvairūs jų požymiai, atsižvelgiant į žmogaus balso trakto padėtis, tariant įvairius garsus, bandoma susieti tokias padėtis su signalų požymiais. Norint sėkmingai atskirti vienus nuo kitų, svarbu yra išsiaiškinti kokie signalų požymiai geriausiai juos apibūdina.

Balso technologija – viena iš perspektyviausių praktinio taikymo srityse pabandžius suvokti jos galimybes. Labiausiai išsivysčiusiose šalyse kalbos atpažinimo srityje jau sukurtos įvairios komandų sistemos, bandoma kurti rišlios kalbos atpažinimo sistemas. Daug kur jau pasiekta suprantama ir vartotojui priimtina sintezuotos kalbos kokybė, tokią kalbą siekiama priartinti prie natūralios. Jungiant sintezę su atpažinimu, kuriamos dialogo balsu sistemos. Intensyvūs tyrimai atliekami kalbančiojo atpažinimo pagal balsą srityje. Balso kompiuterinio atpažinimo praktika taikoma gan seniai kriminalistikoje, elektroninėje prekyboje, informacijos apsaugos sistemose. Tai yra labai palengvinantis naudingas dalykas. Tačiau Lietuvių kalba yra mažai nagrinėta šiuo aspektu. Pasaulyje sukurtos balso atpažinimo programos yra pritaikytos tik populiariausiom kalbom kaip anglų, vokiečių ir t.t. Norint sėkmingai tirti kalbos atpažinimą – reikalinga turėti pakankamą garsyno duomenų bazę, kurioje būtų sukaupta įvairių tos kalbos signalų. Norint turėti gerą duomenų bazę reikalinga kuo įvairesnė diktorių auditorija, tačiau labai dažnai kalbos signalų surinkimas tampa komplikuotu procesu. Reikalinga programa, kuri be didelių investicijų leistų pasieki įvairaus amžiaus, profesijos, žmones, būtų visiem lengvai prieinama ir nesudarytų vargo vartotojui ja naudotis.

Tiriamajame darbe aptariamas žmogaus kalbos mechanizmas. Sėkmingą kalbos atpažinimo uždavinių sprendimą būtent lemia didelis žinių turėjimas, kaip garsas formuojamas žmogaus balso trakte, koks būna balso trakto elemento išsidėstymas tariant vienus ar kitus garsus, nuo kokių veiksnių priklauso garsų skirtumai. Taip pat svarbu išsiaiškinti kaip garsas sklinda oru, todėl bus aptariamos garso bangos, sklindančios nuo kabančiojo iki kol pagauna žmogaus klausos aparatas bei parodomi tokių bangų užrašymo ir vaizdavimo būdai, kurie ypač svarbūs norint nustatyti kalbos signalų požymius.

Tyrimas skirtas kalbos signalų kokybinių charakteristikų analizei, įvardinant galimybes, kurių pagalba atliekamas garsų atpažinimas, taip pat atskleidžiant kalbos signalų kriterijus, padedančius atliti kokybinės charakteristikos analizę. Darbe analizei naudojamas kalbos signalų garsynas, kuriame sukaupta diktorių šnekos įrašai. Tokiu būdu tyrimo tikslas - sukurti kalbos signalų kokybinių charakteristikų analizės algoritmą, kuris galėtų automatiškai nustatyti ar padarytas žodžio įrašas atitinka nustatytus kriterijus. Šiam tikslui pasiekti reikia išanalizuoti esamus segmentavimo algoritmus, išskiriant jų privalumus ir trūkumus, sukurti signalų kokybės nustatymo algoritmą, pasiruošti kalbos signalų kokybės nustatymo tyrimui duomenis, įvertinti kalbos signalų kokybės nustatymo algoritmo rezultatus, apibendrinti tyrimo rezultatus. Tyrimui atlikti remsimės moksline literatūra bei apibendrintais metodais, eksperimento įgyvendinimui – eksperimentinio matavimo bei statistinės analizės metodais. Tyrimui vykdyti ketiname sudaryti kalbos signalų kokybės analizės algoritmą, parenkant kalbos signalų kokybės charakteristikų kriterijus.

1.1 Tiriamoji problema

Šiame darbe nagrinėjamos kalbos signalų atpažinimo klausimai, jų kokybės charakteristikų kriterijai bei analizė.

1.2 Darbo objektas

Kalbos signalų surinkimo ir kaupimo bei jų pirminių kokybinių charakteristikos.

1.3 Darbo tikslas ir uždaviniai

Tikslas – pasiūlyti sprendimus, kurie leistų pasiekti tikslesnių rezultatų kalbos signalų surinkime, nustatyti informacinės sistemos reikalavimus, specifikaciją, integraciją su signalų analizės įrankiais, sukurti kalbos signalų kokybinių charakteristikų algoritmą, galintį automatiškai nustatyti ar padarytas žodžio įrašas atitinka nustatytus kriterijus. Siekiant savo tikslo buvo sprendžiama šie uždaviniai:

Uždaviniai:

- Aptarti žmogaus kalbos mechanizmą;
- Išsiaiškinti kalbos signalų požymius;
- Apžvelgti kalbos signalų kokybės charateristikų analizės algoritmus;
- Kalbos signalo kokybinių charakteristikų analizės metodo sukūrimas.
- Eksperimentinių duomenų paruošimas;
- Eksperimentinio kalbos signalo kokybinių charakteristikų analizės algoritmo sudarymas ir tyrimas.
- Tyrimo rezultatų apibendrinimas.

2 KALBOS ATPAŽINIMO SISTEMOS

Pagrindinis kalbos atpažinimo sistemų kūrimo ir tobulinimo tikslas yra sukurti mašinas, kurios sugebėtų girdėti, suprasti, kalbėti ir veikti pagal informaciją gautą balsu. Tokių mašinų kūrimas daug lygmenų turintis procesas, kurio pats pirmasis lygmuo yra kalbos atpažinimas, t.y. kompiuteriu žmogaus šnekos pavertimas tekstu.

Kalbos atpažinimas vystomes bene 50 metų, tačiau patys pirmieji dešimtmečiai nepasižymėjo sparčia bei produktyvia technologijų vystymosi sparta. To priežastis – tuometinės skaičiavimo sistemos ribotumas. Tik paskutiniaisiais dešimtmečiais, ištobulėjus skaičiavimo technikai, kalbos technologijos pradėjo vykdyti ypatingai sparčiai – buvo realizuojami ir tobulinami algoritmai, kuriami hibridiniai metodai, atliekami eksperimentai, kaupiami garsynai, kuriamos praktinės kalbos atpažinimo realizacijos realiems uždaviniams spręsti. Tačiau nors ir begalėpastangų bei lėšų tam yra skiriama, atpažinimo sistemos dar vis nėra tobulos. Komercinių atpažinimo sistemų gamintojai yra išgavę 98-99 % atpažinimo tikslumą, tačiau tai tik atskiriems kalbėtojams ir tam tikromis laboratorijos sąlygomis. Praktikoje kalbos atpažinimas labia priklauso nuo paties kalbėtojo – labia svarbu jo fizinės, emocinės, psichologinės būsenos, kalbėjimo būdas, naudojama įranga – visi šie faktoriai sukelia tam tikrus atpažinimo tikslumo svyravimus. Taip pat svarbu kad nėra surandami naujų, optimalių priemonių kalbos atpažinimui, tik tobulinami su atskirais papildomais parametrais, daromi hibridiniai metodai iš jau esamų klasikinių metodų.

Šiame skyriuje aptariama kalbos atpažinimo sistemos, suformuluojami atitinkami apibrėžimai, pateikiama struktūra, sistemos suklasifikuomos pagal atitinkamus parametrus. Nagrinėjamos kalbos atpažinimo problemos bei jų galimi sprendimai. Apžvelgiama atpažinimo metodų darbai atlikti tiek užsienyje, tiek Lietuvoje.

2.1 Kalbos atpažinimo sistema

Kalbos atpažinimo sistema – programinė arba aparatinė įranga, sugebanti pateiktą kalbos signalą sutapatinti su tekstu. Kalbos signalo analizė skirta tam tikrų požymių išskyrimui, kurie leidžia sumažinti nagrinėjamų duomenų kiekį bei padidina galimybę atvaizduoti fonetinius skirtumus. Signalo analizę sudaro keletas etapų:

- ✓ Signalo skaidymas kadrais;
- ✓ Dauginimas iš lango funkcijos;
- ✓ Spektrinė analizė;
- ✓ Reikiamų požymių išskyrimas.

Analizės metu gautieji požymiai panaudojami sistemai apmokyti. Apmokymu metu požymiai ir jų atstovaujamo žodžio fonetionė reikšmė išsaugoma žodyno atmintyje kaip etaloniniai duomenys, kuriais remiantis vėliau ir atliekamas atpažinimas. Vykdant atpažinimo procesą išsiskyrę požymiai klasisifikuojami pagal jų atitikimą etalonams.

Pirmoji kalbos atpažinimo sistema pasirodė 20 amžiuje mažo žaisliuko pavadinto Radio Rex pavidalu[1][2]. Tai pirmoji mašina (komercinis žaisliukas), kuri tam tikru laipsniu atpažino kalbą. Ji buvo pagaminta 1920 m. Tai buvo celiulioidinis šuo, kurio veikimo principas buvo labia paprastas – ištarus šio šuns vardą Rex, jis sureaguodavo ir išlįsdavo iš savo būdos . Pats veikimas buvo pagrįstas akustine energija – pačiame žaisle buvo įtaisyta šuntas, kuris buvo jautrus 500 Hz akustinei energijai, kuri atitiko žodžio „Rex“ balsės energiją. Garso signalo 500 Hz diapazone energija, esanti žodyje Rex atleisdavo grandinę, kai tik buvo ištariamas žodis „Rex”. Tačiau Radio Rex reaguodavo į visus žodžius, garsus, kurių spektre buvo pakankamo lygio 500 Hz dažnio dedamoji. Nors tai buvo labai paprasta sistema, bet pats atpažinimo principas naudojamas ir šiuolaikinių sistemų kūrime.

2.2 Sistemų raida

Nėra duomenų, kada buvo iškelta idėja realizuoti mašininį kalbos atpažinimą ar suformuluoti pradiniai teoriniai teiginiai. Darbų pradžia laikomas šeštasis dešimtmetis. Tuomet prdėtos kurti fonemų, garsų ir žodžių atpažinimo sistemos. Buvo naudojamas pavyzdžių lyginimo principas, kuomet nagrinėjamas signalas yra palyginamas etalonais ir panašiausias pateikiamas kaip atpažinimo rezultatas. Kaip požymiai buvo naudojamos formantės [3], juostinių filtrų požymiai [4, 5], laikinės spektrų savybės [6]. Taip pat buvo naudojami akustiniai - fonetiniai atpažinimo metodai [7], kuriose nagrinėjamas signalas suskaidomas į segmentus su pastoviomis pasirinktomis charakteristikomas, jiems priskiriant atitinkamą tekstą. Fonetiniams vienetams charakterizuoti buvo parenkami požymiai, atspindintys akustines vienetų savybes – nosinumas, friktyvumas, formančių išsidėstymas, garso vokalizuotumas, atitinkamų spektro juostų energija, pagrindinio tono dažnis [5, 7]

Pirmi bandymai sukurti automatinio kalbos atpažinimo sistemą buvo atlikti 1952 m. Belo laboratorijose [15]. Tai buvo izoliuotų skaičių atpažinimo sistema, skirta vienam kalbėtojui ir rėmėsi kiekvieno skaičiaus balsių srityje spektrinių rezonansų matavimu. Sistema matavo dvejose plačiose dažnių juostose spektrinės energijos tam tikrą nesudėtingą funkciją laike, tokiu būdu grubiai

aproksimuodama pirmas dvi formantes. Ji matavo grubų formančių kelią, o ne patį spektrą. Tai yra potencialiai atsparu nereikšmingoms kalbos spektro modifikacijoms. Pvz., paprastas kalbėtojo galvos pasukimas nuo klausytojo dažnai sukelia pastebimus kalbos spektro pasikeitimus (ypač aukštesnių spektro komponenčių amplitudės sumažėjimą). Belo laboratorijos sistemos spektro įvertis buvo gana grubus, žemų ir aukštų dažnių spektro momentų histogramavimas per visą pasisakymą ir tokiu būdu kitimas laike buvo prarastas. Nors idėja gera, to meto technologija buvo per silpna, kad galima būtų šią sistemą stipriai tobulinti. Sistema naudojo analogines elektrines komponentes ir sunkiai buvo modifikuojama. Nepaisant to, išradėjai tvirtino, kad sistema dirba gana gerai, vienam kalbėtojui tariant skaičius, kurie buvo izoliuoti pauzėmis, pasiekia 2% klaidingumą. Kalbos signalas buvo filtruojamas į žemų ir aukštų dažnių komponentes ir kiekviena komponentė stipriai ribojama taip, kad jos amplitudė nepriklausė nuo signalo stiprumo. Šiems signalams buvo skaičiuojami nulio kirtimai ir sistema naudojo nulių kirtimų reikšmes, vertinant kiekvienos dažnių juostos centrinį dažnį. Žemų dažnių juosta buvo kvantuojama į vieną iš šešių 100 Hz subjuostų (tarp 200 ir 800Hz), aukštų dažnių juosta buvo kvantuojama į vieną iš penkių 500 Hz subjuostų. Kartu šios dvi kvantuotos reikšmės atitinka vieną iš 30 galimų dažnių porų. Skaičiai turėjo atskiriamus dažnių porų pasiskirstymus ir taip buvo vienas nuo kito atskiriami.

Balsių srityje spektrų matavimai, gauti lygiagrečių filtrų pagalba, buvo panaudoti ir bandant atpažinti vieno kalbėtojo dešimt skirtingų skiemenų [16]. Dudley sukūrė klasifikatorių, kuris vertino spektro kitimą laike. Šis būdas buvo plačiai paplitęs, o dabar kalbos atpažinimui dažniausiai naudojama kuri nors iš kintančių laike lokalinių spektro įverčių funkcijų, kurios vaizduoja atpažįstamą kalbą. Fry ir Denes [17] bandė sukurti fonemų atpažinimo mechanizmą, kuris galėtų atpažinti keturias balses ir penkis priebalsius. Atpažinimo tikslumo gerinimui, čia buvo panaudota statistinė informacija apie galimas fonemų sekas ir tam tikrus spektrinių objektų tapatinimo būdus. Tai buvo bandymas šalia akustinės informacijos panaudoti gramatikos tikimybes. Buvo iškelta mintis, kad konkretaus lingvistinio vieneto tikimybė gali būti

priklausoma nuo ankstesnio lingvistinio vieneto, taip, kad žodžio tikimybė nėra priklausoma vien tik nuo akustinio įėjimo. Sekantis to laikotarpio bandymas buvo Forgie balsių atpažinimo aparatas [18]. Šis aparatas galėjo atpažinti 10 balsių, esančių tarp priebalsių /b/ ir /t/ ir atpažinimas buvo nepriklausomas nuo diktoriaus. Spektrinei informacijai gauti buvo naudojamas juostinių filtrų analizatorius. Balso trakto rezonansų radimui buvo naudojamas kintantis laike rezonansų įvertis.

Sakai ir Doshita 1962 metais [19] realizavo fonemų atpažinimo įrenginį: sprendimų priėmimui buvo panaudotas aparatūrinis kalbos segmentavimas ir nulio kirtimų analizė. Tuo metu taip pat aktyviai buvo pradėta spręsti kalbos vienetų laiko skalės nevienodumo problema. Martin (1964 m.) [20] sukūrė eilę nesudėtingų laiko skalės normalizavimo metodų, kurie rėmėsi patikimu kalbos pradžios ir galo momentų nustatymu. Beveik tuo pat metu, T. K. Vinciuk (Ukraina) (1986 m.) [21] laiko skalės vienodinimui pasiūlė naudoti dinaminio programavimo metodus. Pritaikymo idėja buvo panaudoti dinaminį programavimą laiko skalei normuoti, taip sprendžiant iki tol buvusią nevienosdos lyginamų kalbos vienetų trukmės problemą. Šis svarbus mokslinis laimėjimas Vakaruose buvo nežinomas iki aštuntojo dešimtmečio. Šeštajame dešimtmetyje svarbus pasiekimas buvo Reddy tyrimai (1966 m.) [22], kuriuose nepertraukiamos kalbos atpažinimui panaudotas dinaminis fonemų kitimo sekimas.

1964 m. Martin fonemoms atpažinti panaudojo neuroninius tinklus. Widrow 1963 m. panaudojo neuroninius tinklus skaičiams atpažinti.

6- ajame dešimtmetyje buvo sukurti trys spektro įverčių metodai, kurie vėliau tapo labai svarbūs kalbos atpažinimui. Šie metodai pirmiausiai buvo pritaikyti kalbos kodavimui: Greita Furjė transformacija (FFT), kepstrinė (arba homomorfinė) analizė ir tiesinio prognozavimo kodavimas (LPC). Buvo sukurti nauji pavyzdžių tapatinimo metodai:

- deterministinis metodas, vadinamas dinaminiu laiko skalės kraipymu (DTW),
- statistinis metodas, vadinamas paslėptu Markovo modeliu (HMM).

Itakura (1975 m.) parodė, kaip tiesinės prognozės (LPC) idėjos gali būti sėkmingai pritaikytos kalbai atpažinti. Šiuo laikotarpiu prasidėjo labia aktyvus kalbos atpažinimo sistemų kūrimas (1985 m.). Buvo kuriamos tikrai nepriklausančios nuo kalbėtojo kalbos atpažinimo sistemos (1979 m.). Jose buvo panaudoti tobuli klasterizacijos algoritmai daugybės skirtingų požymių, reikalingų įvairių žodžių variacijoms plačioje kalbančiųjų populiacijoje atspindėti, išskyrimui. Vėliau šie metodai buvo dar labiau ištobulinti ir plačiai vartojami.

Septintasis dešimtmetis pasižymėjo ypač svarbiais darbais kalbos technologijoms. 1965 m. Buvo pristatytas greitosios Furje transformacijos (FFT) algoritmas [8], po keleto metų A. V. Oppenheim su kolegomis kalbos signalui apdoroti pritaikė kepstro analizę [9]. Septintojo dešimtmečio pabaigoje – aštunto pradžioje kalbos signalo analizei pasiūlytas tiesinės prognozės modelis (LPC), tuo metu sėkmingai naudotas kalbos signalui koduoti. Šiame dešimtmetyje tyrinėtojai savo atpažinimo sistemose naudojo spektrinius požymių vektorius, LPC ir fonetinius požymius. Buvo sukurti metodai naudojantys DTW, HMM ir neuroninius tinklus ir daugybė atpažinimo sistemų. Buvo stengiamasi sutrumpinti pavyzdžių palyginimo trukmę. Dažnai buvo naudojami dirbtinio intelekto metodai, ypač ARPA programoje. Automatiniam kalbos atpažinimui buvo pritaikyta HMM teorija ir sukurtos sistemos, pagrįstos HMM.

Advanced Research Projects Agency (ARPA) finansavo didelį kalbos suvokimo projektą. Tikslas buvo sukurti 1000 žodžių automatinį kalbos atpažinimą, naudojant kelis kalbėtojus, ištisinę kalbą ir apribotą gramatiką su mažesniu nei 10% semantinių klaidų kiekiu. Tačiau tik HARPY sistema, sukurta CMU doktoranto Bruce Lowerre, tenkino keliamus reikalavimus. Jis naudojo LPC segmentus, gramatikos žinias ir „Baker Dragon system “, bei CMU sistemos Hearsey modifikuotus metodus. Tyrinėtojai savo atpažinimo sistemose naudojo spektrinius požymių vektorius, LPC ir fonetinius požymius. Buvo sukurti metodai, naudojantys DTW, HMM ir neuroninius tinklus. Buvo stengiamasi sutrumpinti pavyzdžių palyginimo trukmę. Dažnai buvo naudojami dirbtinio intelekto metodai, ypač ARPA programoje. Automatiniam kalbos atpažinimui buvo pritaikyta HMM teorija ir buvo sukurtos HMM besiremiančios sistemos.

8-tame dešimtmetyje kalbos atpažinimo tyrėjų dėmesys persikėlė nuo atskirų žodžių atpažinimo prei kalbos, sudarytos iš sujungtų žodžių (connected word), atpažinimo. Tyrimų tikslas buvo sukurti robotinę sistemą, sugebančią atpažinti skalndžiai apsakytą žodžių seką. Tyrimų objektu tapo statistiniai modeliavimo metodai, ypač HMM metodas [26 - 27]. Kita

fundamentalia tyrimų sritimi, kuri pradėjo sparčiai vystytis, tapo neuroninių tinklų panaudojimas kalbos atpažinimo problemų sprendimui [28 - 29]. Taip pat reikėtų paminėti, jog būtent šiame dešimtmetyje buvo sukurta pirmojin komercinė kalbos atpažinimo sistema, skirta komandoms ir duomenim įvestin[10]

1986 m. Buvo pradėta kurti TIMIT balsių bazė, kuri vėliau tapo pirma plačiai naudojama standartine balsų baze. Taip ji pavadinta todėl, kad duomenis rinko Texas Instruments (TI), o anotaciją atliko MIT. Fonetiniam savitumui pavaizduoti buvo parinktas 61 garso alfabetas. Buvo parinkti fonetiškai subalansuoti tekstai, kad mokymo aibėje būtų gerai atstovaujamas kiekvienas garsas. 630 kalbėtojų pasakė po10 sakinių, iš kurių du buvo visų kalbančiųjų tie patys.

Didelio žodyno nepertraukiamos kalbos atpažinimo sistemų kūrimui didelį impulsą suteikė DARPA (Defence Advanced Research Projects Agency) projektai, kuriuos plėtojant buvo labiau orientuojamasi į natūralios kalbos apdorojimą ir taikymus, tokius kaip telefoninių tinklų operatoriaus darbo palengvinimas. 1984 metais ARPA pradėjo finansuoti antrą programą. Prasidėjus šiai programai buvo suformuluotas naujas tikslas - išteklių valdymas (angl. Resource Management - RM) su nauja balsų baze. RM bazės balso įrašai buvo sukurti skaitant tekstą. Sakiniai buvo sukonstruoti iš 1000 žodžių kalbos modelio. Buvo įrašyta 21000 pasakymų, kuriuos ištarė 160 kalbėtojų. Į RM tikslus įėjo nepriklausomas nuo kalbėtojo atpažinimas. Projekte dalyvaujančių sistemų vertinimas vyko kartą ar du per metus. ARPA projektas sudarė labai geras sąlygas tobulinimams. Daug laboratorinių sistemų gali atpažinti naujo kalbėtojo kalbą (be apmokymo konkrečiam kalbėtojui) su 60000 žodžių žodynu realiame laike, gaunant mažesnį nei 10% klaidingumą. Tai įkvėpė kitus tyrinėtojus, kurie nebuvo finansuojami iš šio projekto, tame tarpe ir tyrinėtojus iš Europos.

Vėliau, vykdant RM programą, dėmesio centras persikėlė į Wall Street Journal. Tikslas - atpažinti skaitomą kalbą iš Wall Street Journal.

Lygiagrečiai RM buvo suformuluotas kitas tikslas Air Travel Informatio System (ATIS), kuris rėmėsi spontaniškais užklausimais lėktuvų bilietų rezervavimo srityje. ATIS tikslas yra kalbos supratimas (priešingai negu kalbos atpažinimas). Sistemos ne tik turėjo sukurti žodžių sekas, bet ir bandyti suteikti joms semantinę prasmę, kad galėtų atlikti atitinkamą funkciją.

Devinto dešimtmečio pabaigoje sukurta paslėptuosiu Markovo modelius naudojanti atpažinimo modeliavimo sistema HTK [11], labiausiai paplitusi priemonė statistiniam kalbos atpažinimo modeliuoti ir sėkmingai taikoma iki šiol. Taip pat pradėta plačiai taikyti neuronų tinklai.

Per paskutinį dešimtmetį kalbos atpažinimo srityje nebuvo sukurtas joks bazinis mechanizmas, tačiau buvo daug svarbių pasiekimų: apdorojimo (front-end) pasiekimai (pvz., melų arba barkų skalės kepstrų įverčiai, delta požymiai, kanalo normalizacijos būdai ir balso trakto normalizacija) ir tikimybinis įvertinimas (pvz., maksimalaus tikėtinumo tiesinė regresija, naujų kalbėtojų ar akustinių sistemų adaptacija, arba mokymas maksimizuojant tarpusavio informaciją tarp duomenų ir modelių). Todėl galima teigti, kad dirbančių toje srityje pastangos yra orientuotos link egzistuojančių idėjų efektyvumo pagerinimo nei link naujų idėjų generavimo. Tai turėjo tendenciją konverguoti į geras, panašias sistemas, kiekvienai laboratorijai bandant pasinaudoti patobulinimais, kurie pavykdavo kitiems.

Taip pat šis laikotarpis pasižymi masiniu kalbinių technologijų taikymu įvairiose srityse: telekomunikacijose, informacinėse tarnybose, mokyme, tarptautiniuose komerciniuose ryšiuose. Daugelyje šių taikymų kaip pagrindinė arba kaip sudedamoji sistemos dalis yra kalbos atpažinimas. Išsiplėtus taikymų sričiai, kartu labai pagriežtėjo reikalavimai kalbos atpažinimo tikslumui. Reikia patikimai atpažinti kalbą, esant triukšmingai aplinkai, dideliems ryšio kanalo iškraipymams, ryšio kanalo dažnių juostos apribojimams (pavyzdžiui telefonijoje). Dėl to labai suintensyvėjo tyrimai kalbos atpažinimo srityje. Bandoma surasti požymius, kurie leistų atpažinti kalbą esant sudėtingoms ryšio sąlygoms, triukšmui, panaudoti lingvistines žinias kalbos atpažinimo patikimumo padidinimui.

2.3 Darbai Lietuvoje

Lietuvoje priešingai nei pasaulyje kalbos atpažinimo darbai pradėti gerokai vėlai – aštuntojo dešimtmečio pabaigoje-devinto pradžioje, taigi stipriai atsiliekant nuo pasaulinių darbų.

Pačiose pirmose kalbos atpažinimo sistemose buvo naudojama dinaminio laiko skalės kraipymo metodas, realizuotas naudojant dinaminį programavimą. Kaip požymiai buvo naudojami juostinių filtrų požymiai, vėliau pereita prie tiesinės prognozės modelio, tiesinės prognozės kepstro

koeficientų. Dinaminis laiko skalės kraipymo metodas ypač išpopuliarėjo ir taikomas dar ir šiomis dienomis. Devintajame dečimtmetyje pradedama naudoti paslėptieji Markovo modeliai, nors šis metodas pasaulyje vartojamas kaip efektyviausias ir plačiausiai vartotas, Lietuvoje paplito tik dešimto dešimtmečio pabaigoje. Signalui analizuoti buvo naudojama tiesinės prognozės, kepstrinė analizės, tačiau didžiausias atpažinimo tikslumas pasiektas naudojant melų skalės kepstro požymių sistemą. Jau šio amžiaus pradžioje kalbai atpažinti pritaikyti dirbtiniai neuronų tinklai, kurie deja neprigijo kalbos atpažinime. Bandyta kurti ir hibridines atpažinimo sistemas jungiant vienus metodus su kitais, bet ypatingų rezultatų nepasiekta. Taip pat buvo pasiūlyta keletas orginalių sprendimų: dichotominis klasifikatorius (atliekantis dvinarį dalijimą), dinaminio laiko skalės kraipymo mokidifikacija – projekcijų metodas, žemos ir aukštos eilių nulio kritimo požymiai (nulio kritimo parametras, skaičiuojamas įvairios eilės skirtuminiams signalams), fonemų klasifikacija naudojant suderintą diskriminantinę analizę. Prieš keletą metų pradėta kurti lietuvių kalbos garsynai.

Didžiausi kalbos atpažinimo darbai daromi Matematikos ir informatios institute (Vilnius), Vytauto Didžiojo universitete (Kaunas) bei Kauno technologijos universitete. Didžiausias dėmesys atpažįstant kalbą skiriamas paslėptiems Markovo modeliams (kuriami kalobos medeliai, atliekamo eksperimentai), kalbos duomenų bazėms (kaupiamos pavienių žodžių ir ištisinės kalbos duomenų bazės) bei bazių kaupimo automatizavimui. Taigi Lietuvoje ikalbos atpažinime padaryta didelė pažanga lyginant su pasauliniais pasiekimais, tačiau atsilikimas dar vis jaučiamas.

2.4 Atpažinimo problemos

Kalbos atpažinimui skiriama nemažai laiko ir pastangų, sukurta daug metodų, realizuota daug sistemų, kurios jau yra ir taikomos, tačiau vis dar kyla sunkumų ir neatsakytų klausimų realizuojant triukšmui atsparias, tikslias atpažinimo sistemas. Šiuos sunkumus galima įvardinti kelomis problemomis.

Viena iš problemų – kalbos signalo kintamumas, t.y. neįmanoma realizuoti dviejų visiškai identiškų, to paties lingvistinio vieneto pavyzdžių. Tą patį žodį neįmanoma ištarti vienodai, nors ir galima stengtis iki begalybės, tačiau ištarimai vis tiek skirsis ar tai tempu, ar energijos lygiu, ar tai kažkokiomis kitomis laikinėmis ar spektrinėmis savybėmis. Išskiriami du kalbos kintamumo tipai - vidinis kintamumas ir išorinis. Vidinis kintamumas pasireiškia to paties kalbančiojo kalbos nepastovumu. Viena iš šio nepastovumo priežasčių - kalbėjimo maniera. Kalbantysis savo mintis gali išreikšti pakeltu tonu ar net rėkdamas, šnabždėdamas, bandydamas paslėpti savo akcentą, ir t. t. Be to, įtakos turi ir subjektyvūs veiksniai - kalbančiojo laikysena, nuotaika, sveikatos būklė, amžius, pokalbio tematika. Dėl šių priežasčių netgi to paties asmens ištarti žodžiai tarpusavyje skirsis ir tie skirtumai ilgėjant laikotarpiui tarp ištarimų didės. Prie išvardintųjų vidinio kintamumo priežasčių

prisideda natūra-lios kalbos savybės (koartikuliacija, įsiterpiantys beprasmiai garsai, kalbos tempo varijavimas ir pan.). Ypač akustiniai skirtumai išryškėja tarp skirtingų lyčių, skirtingo amžiaus kalbančiųjų. Kalbos kintamumo problema gali būti sprendžiama dviem būdais. Pirmasis - kalbančiojo adaptacija. Antrasis būdas - kalbančiajam atsparios požymių sistemos naudojimas.

Antroji problema - natūralios kalbos savybės. Vienas iš natūralios kalbos reiškinių tai koartikuliacija - gretimų garsų susiliejimas. Susilieję garsai tampa sunkiai atskiriami ar netgi įgyja visiškai kito fonetinio vieneto skambesį (pvz., žodį „čia" girdime kaip „če" ir teisingai mums jį parašyti padeda tik gramatikos žinios). Natūraliai kalbai taip pat būdingi nelingvistiniai garsai (pvz., abejojimo garsas „mmmm", kostelėjimas), kurie gali užpildyti pauzes, įsiterpti į žodį ar netgi nutraukti jį. Žmogaus suvokimo sistema šiuos garsus lengvai išskiria kaip nelingvistinius, tuo tarpu atpažinimo sistema juos gali suprasti kaip žodį ar jo dalį (ypač jei tą rodo pvz., akustinės analizės rezultatai). Kai kuriais atvejais gali būti aktualus ribų tarp žodžių ištisinėje kalboje nebuvimo klausimas. Šios problemos turėtų būti sprendžiamos lingvistiniame lygmenyje - naudojami kalbos modeliai, taikomos papildomos gramatikos, prozodikos, semantikos, pragmatikos žinios, t. y. greta akustinio kalbos signalo apdorojimo atsiranda lingvistinio apdorojimo poreikis. Atpažinimo sistemų žodynai yra dar vienas atpažinimo problemų šaltinis. Dideli žodynai yra painūs - juose yra daug akustiškai panašių pavyzdžių. Kai kurie rėjai teigia, jog kalbos atpažinimo uždavinio sunkumas auga logoritmiškai didėjant žodyno dydžiui [12]. Vienas iš galimų šios problemos sprendimų būdų - kontekstinio (t. y. skirto konkrečiai dalykinei sričiai) žodyno naudojimas.

Dar sunkiau sprendžiama žodyne neesančių žodžių problema. Bet kuri sistema anksčiau ar vėliau susiduria su žodyne neesančiu žodžiu. Tokiu atveju galimi du sprendimai - atmesti žodį kaip neatpažintą arba įtraukti į sistemos žodyną. Antrasis sprendimas sukelia dar aibę sunkiai atsakomų klausimų - kaip garantuoti, kad neatpažintasis pavyzdys yra lingvistiškai prasmingas, kaip sugeneruoti reikalingą transkripciją, kaip atskirti pavyzdį nuo pašalinių triukšmų ir pan. Kol kas nėra pasiūlyta efektyvios procedūros šiems klausimams spręsti, todėl dažnai neatpažintasis pavyzdys tiesiog ignoruojamas.

Ketvirtoji problema - signalo akustinės ir sklidimo aplinkos įtaka signalui. Bet kuris signalo generavimo, sklidimo, priėmimo etape esantis triukšmas gali įtakoti ir būtinai įtakos signalą. Triukšmo šaltiniais gali būti pats kalbantysis (iškvėpimo triukšmas, kalbos padargų mechaniniai triukšmai), sklidimo aplinkos (foninis aplinkos triukšmas, aidas), įvedimo įrenginys (mikrofono elektriniai triukšmai, netiesiniai iškraipymai), perdavimo kanalas (atspindžiai, kanalo netiesiniai iškrai-pymai), priėmimo įrenginys (elektriniai triukšmai, netiesiniai iškraipymai, kvan-tavimo triukšmas). Visų šių poveikių rezultatas - užtriukšmintas signalas, lengvai suprantamas žmogui ir kartais visiškai nepriimtinas techninei sistemai. Be to, kiekvienas įrenginys pasižymi savo individualiomis spektrinėmis charakteristikomis (pvz., ribota pralaidumo juosta), kurios taip pat turi

įtakos apdorojamam signalui, todėl skirtingų techninių priemonių (skirtingos paskirties, įvairių gamintojų) poveikis signalui yra nevienodas. Sistema apmokyta su vieno tipo mikrofonu (ir puikiai su juo veikianti) gali visiškai prarasti savo savybes pakeitus mikrofoną kitu (šiuo atveju sistemos darbingumas yra priklausomas nuo įrangos). Ši problema turėtų būti sprendžiama ieškant triukšmams atsparių požymių sistemų.

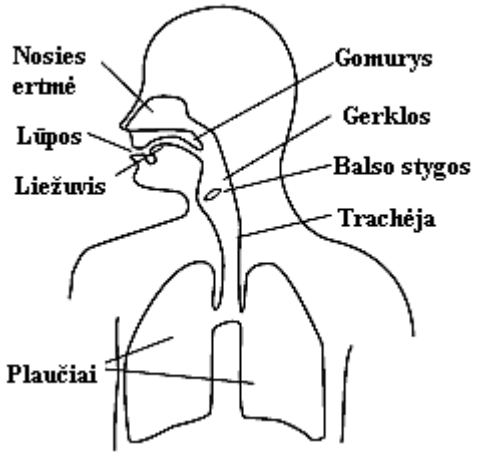
Apibendrinant galima būtų pasakyti, kad ne vien dėl išvardintų problemų automatinės atpažinimo sistemos neprilygsta žmogiškajam kalbos suvokimui. Žmogus bendravime neapsiriboja akustine analize, o panaudoja ir fonetikos, fonologijos, leksikos, sintaksės, prozodikos, semantikos, pragmatikos žinias, pokalbio konteksto duomenis, papildomą informaciją, perduodamą gestais, mimika, laikysena, galbūt net intuiciją ir kitus informacijos šaltinius, kurių technikoje mes negalime realizuoti dėl nežinojimo, sudėtingumo ir savo pačių išankstinių prielaidų.

3 Kalbos signalų analizė

3.1 Garsas.

3.1.1 Garso generavimo mechanizmas

Žmogaus kalbos generavimo mechanizmą sudaro plaučiai, trachėja, gerklos ir nosies traktai. Susikaupęs plaučiuose oras per trachėją patenka į gerklas. Tai vienas iš sudėtingiausių procesų, kadangi jame dalyvauja ir sinchronizuojasi įvairūs raumenys ir fizinis kūnas. Kai žmogus kalba, oras pučiamas iš plaučių aukštyn į balso stygas. Galiausiai, balso stygos vibruoja tam tikru dažniu ir taip susidaro garsas. Oro srautas ir toliau juda burnos link, kur juo manipuliuoja liežuvis, dantys ir lūpos, kad susidarytų garsai, kuriuos mes vadiname žodžiais ir frazėmis. Smegenys vaidina pagrindinį vaidmenį, valdydamos visą procesą ir stebėdamos, kas pasakyta, kad jūsų skleidžiamas garsas atitiktų tai, ką norėjo pasakyti smegenys ir kad ištarti žodžiai būtų tinkamo garsumo, kad juos girdėtų tikslinis klausytojas [35].

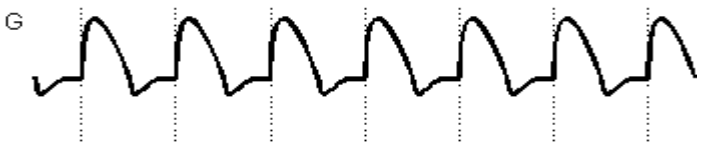


Pav 1 Žmogaus kalbos organai

Plaučiai pro antgerklį stumia orą, balso stygos vibruoja, jos iškreipia oro srovę ir sukelia kvaziperiodines slėgio bangas kitaip vadinamas impulsais. Toliau šie impulsai praleidžiami per gerklės, burnos ir galimai nosies ertmes. Ir priklausomai nuo įvairių artikuliatorių (smakro, liežuvio, minkšto gomurio, lūpų, burnos) padėties yra sukuriami įvairūs garsai.

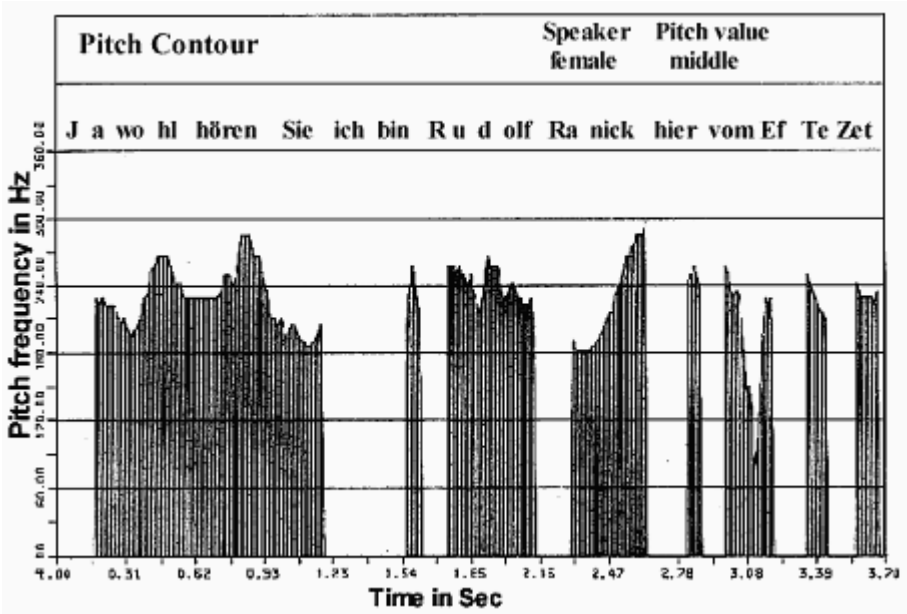
Artikuliatoriai yra svarbūs todėl, kad būtent jų dėka keičiasi burnos trakto forma, o tuo būdu keičiasi ir pats garsas. Artikuliatoriai judėdami ir įvairiai keisdami savo vietą burnos trakte veikia burnos trakto susiaurėjimą, o tuo pačiu ir balsių, priebalsių, bei begarsių garsų, tokių kaip trinamieji-frikatyviniai garsai pokyčius. Pavyzdžiui, burnos trakto susiaurėjimas skatina triukšmo susidarymą oro tėkmėje, tuo būdu skatinant trinamųjų garsų, šnypštimo, ir kitų priebalsių susidarymą. Nevokalizuotiems sprogstamiesiems garsams susidaryti tam tikroje burnos trakto vietoje yra sudaroma aklina pertvara, ir tik po tam tikro laiko tarpo oras staigiai išleidžiamas. Staigus oro išleidimas sukuria trumpalaikį sprogimą. Taigi priklausomai nuo įvairių artikuliatorių padėties yra sukuriami įvairūs garsai. [36]

Slėgio impulsai, gauti vibruojant balso stygoms, paprastai yra vadinami garso impulsais, o slėgio signalo dažnis vadinamas garso dažniu arba pagrindiniu dažniu. 2 pav**.** pavaizduota tipinė impulsų seka (garso slėgio funkcija) sukelta balso stygų tariant vokalizuotą garsą. Bet tai tik dalis garso signalo. Kai mes kalbam pastoviu garso dažniu, kalbos garsai yra monotoniniai, bet paprastai seka nuolatiniai dažnio pasikeitimai. Kaip garso dažnis kinta pavaizduota 3 pav.



Tipinė impulsų seka

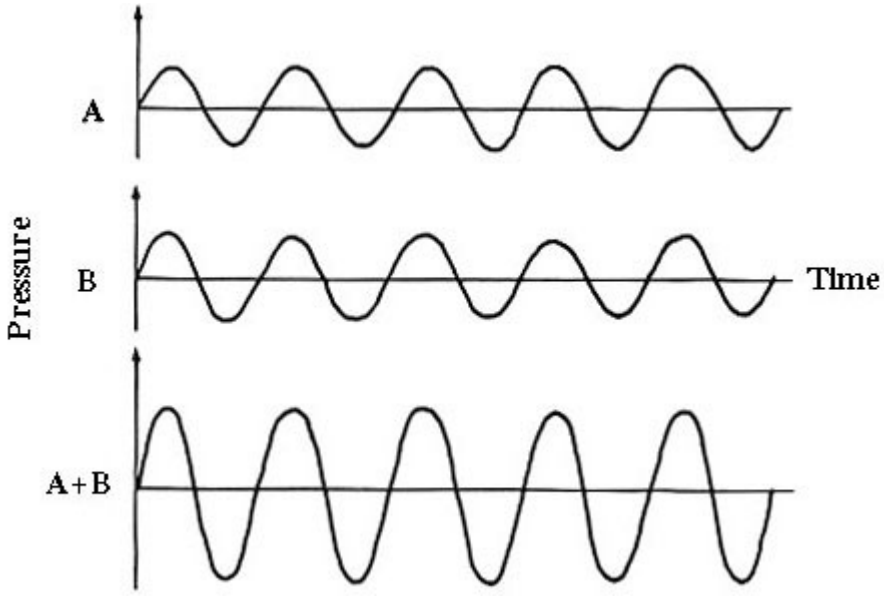
Pav 2



Pav 3 Dažnio svyravimai (vokiškas garsynas)

3.1.2 Garso bangos

Garsas sklinda oru aukšto ir žemo slėgio tarpsniais. Paprasčiausias būdas pavaizduoti garsą diagrama yra nurodyti oro slėgį kiekviename laiko vienete. Taip gaunamos periodinės garso bangos, kaip parodyta 4 paveiksle.



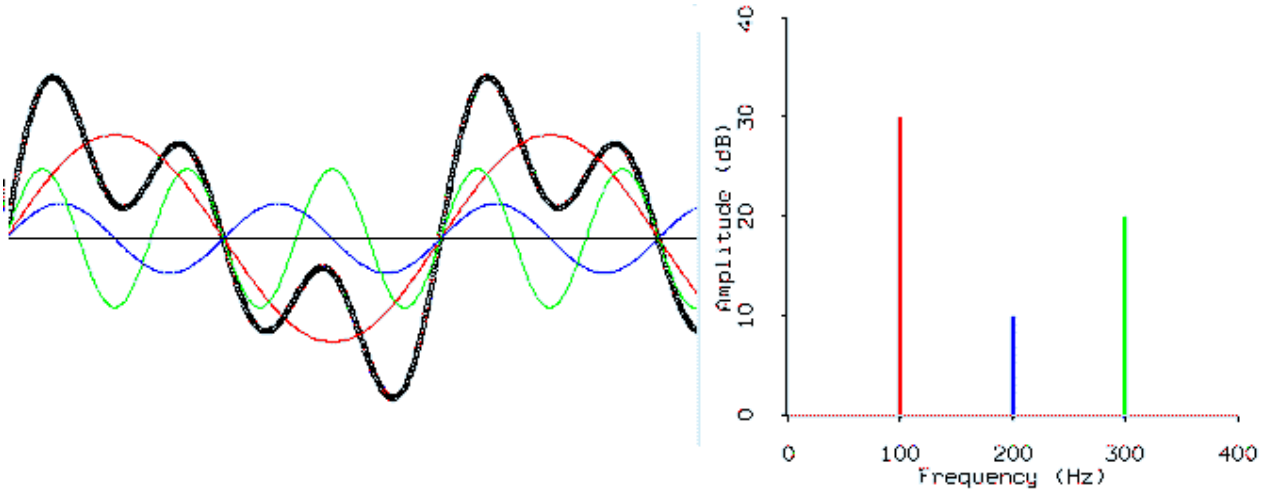
Pav 4 Garso bangos

Galima išskirti garso bangos požymius: amplitudę (slėgių skirtumą arbą garsumą), bangos ilgį bei pasikartojimo dažnį (garso aukštį). Amplitudėm ir dažnis tarpusavyje yra nepriklausomi, tuo tarpu dažnis, priklausomai nuo bangos ilgio, didėja - kai banga trumpėja, ir mažėja – kai banga ilgėja.

3.1.3 Spektras

Iš vieno šaltinio išleistos dvi skirtingos garso bangos susilieja į vieną. Ten, kur abiejų bangų slėgiai yra aukšti, jie papildo vienas kitą, sukurdami dar aukštesnį. Du žemi slėgiai atitinkamai pažemina bendrą slėgį, o aukštas ir žemas slėgiai neutralizuojasi. Sudėtingos garso bangos yra

įvairių dažnių, amplitudžių bei ilgių paprastų bangų kombinacijos. Todėl kiekvieną garsą galima užrašyti kaip tam tikrame dažnių intervale esančių bangų kombinaciją.



Pav 5 Garso bangos ir spektrinė diagrama

5 paveiksle kairėje pavaizduota juoda kreivė yra raudonos, žalios ir mėlynos bangų kombinacija. Raudonos bangos dažnis yra 100 Hz, mėlynos – 200 Hz, o žalios – 300 Hz. Tokias bangų kombinacijas patogu vaizduoti amplitudžių ir dažnių diagramomis, kitaip vadinamomis spektrinėmis diagramomis. 5 paveikslo dešinėje pavaizduota šių bangų spektrinė diagrama. Spektrinės diagramos ypač patogios vaizduojant natūralius garsus, nes šie garsai dažniausiai būna sudaryti iš daugybės paprastesnių garsų [37].

3.1.4 Formantės

Tariant skirtingus garsus, balso trakto padėtis ir forma keičiasi. Todėl spekrinės diagramos viršūnės išsidėsto skirtingose vietose. Spektrinės diagramos viršūnės yra vadinamos formantėmis. Jos numeruojamos pradedant nuo pirmosios ir žymimos F_1, F_2, F_3 ir t.t. Formantės parodo intensyviausius taškus dažnių juostoje, jos pakankamai gerai atsispindi spektrogramose, ypač plačiajuostėse.

3.2 Kalbos signalų tikrinimo sistemų analizė

Šiuolaikinėms technologijoms stripriai žengiant į priekį ir didėjant perduodamų žinių kiekiui, vien pagrindinių kompiuterių įvesties priemonių, tokių kaip pelė ir klaviatūra nebeužtenka. Balso apdorojimo technologijos nėra naujas dalykas informacinių technologijų srityje, kuris ypač stipriai tobulinamas. Yra sukurta nemažai balso atpažinimo sistemų, kurių tikslai ir paskirtys yra įvairios, tačiau dauguma yra orientuota į užsienio kalbas, daugiausiai anglų, vokiečių, prancūzų kalbas.

Lietuvoje ši sritis yra nauja sritis. Neturime kalbos signalų garsynų, yra tik keletas mokslo tikslams (tokie kaip LT DIGITS) surinktų garsynų. O komercinių balso atpažinimo sistemų išvis nėra.

3.2.1 Analizės tikslas

Viena iš svarbiausių balso atpažinimo informacinės sistemos dalis tai yra gerai paruoštas kalbos signalų duomenų bazė – garsynas. Norint surinkti tinkamą naudojimui garsyną reikia nemažai išteklių. Didžiausia išteklių problema - diktorių žmogiškieji ištekliai. Kad garsynas būtų skirtas nuo diktorių nepriklausančiom balso atpažinimo sistemom reikia didelio skaičiaus skirtingų diktorių įrašų – skirtingų lyčių, skirtingų regionų, tarmės ir pan. Esami garsynai buvo įrašinėjami profesionalų – tai sudarė dideles išlaidas, be to tokių diktorių profesionalų nėra daug, kas neatitinka vieno iš kriterijaus norint turėti tinkamą naudojimui garsyną

Analizės tikslas būtų ištirti kalbos signalų (garsyno) surinkimo informacinės sistemos reikalavimus, specifikaciją, integraciją su signalų analizės įrankiais.

3.2.2 Objekto charakteristika

Analizuojamas objektas - lietuvių kalbos signalų kaupimo ir pirminės analizės informacinė sistema internete. Šios sistemos paskirtis yra įrašyti tyrimui reikalingus lietuvių kalbos žodžius ir atlikti pirminę kalbos signalo kokybinių charakteristikų analizę.

3.2.3 Pirminis kalbos signalų apdorojimas

Norint atlikti kalbos požymių išskyrimą ir apskaičiavimą, reikia atlikti paruošiamuosius kalbos signalo apdorojimo veiksmus: signalo diskretizavimas, pradinė filtracija, kadrų ir langų funkcijos taikymas. Šie etapai naudojami beveik visose kalbos ir kalbančiojo atpažinimo bei kitose sistemose.

Norint sumažinti kalbos signalo atvaizdavimui reikalingus resursus ir duomenų kiekį, reikia atlikti kalbos signalo diskretizavimą. Tam tikslui kiekvieną sekundė daug kartų yra išmatuojama ir registruojama garso signalo amplitudėje. Pagal iš anksto apsibrėžtus tikslus yra nustatomos maksimaliai leistinos amplitudės reikšmės. Priklausomai nuo įrašymui naudojamų reikšmių skaičiaus, maksimaliai amplitudės reikšmei yra priskiriamas didžiausias įmanomas sveikas skaičius. (Driaunys, 2006)

3.2.3.1 Diskretizavimo dažnis

Diskretizavimo dažniu yra vadinamas dažnis, kuriuo atliekami kalbos signalo amplitudės matavimai. Yra žinoma, kad kuo didesnis diskretizavimo dažnis, tuo labiau skaitmeninis kalbos signalo įrašas atitinka analoginį. Pagal Naikvisto teoremą, diskretizavimo dažnis turi būti nemažiau kaip du kartus didesnis už maksimalų įrašomo signalo dažnį, kad neprarastume signale esančios svarbios informacijos. Turime diskretizuotą, kalbos signalą Y, kurio diskretų skaičius lygus N. (Driaunys, 2006)

$$Y=y(1), y(2),...,y(i),...,y(N). \quad (1)$$

3.2.3.2 Pradinė filtracija

Kalbos signalo spektras daugiausia išsidėstęs žemų dažnių srityje, aukštesniuose dažniuose jo intensyvumas mažesnis ir krenta maždaug 6 dB į oktavą (t. y. du kartus padidėjus dažniui, spektro amplitudė sumažėja maždaug 6 dB). Dėl šios priežasties, pvz. aukštesnės formantės, yra žymiai mažiau išreikštos nei žemesnės, nors jos taip pat turi savyje svarbią informacija apie kalbą bei asmenį. Pradinės filtracijos (Rodman 1999) tikslas – iškelti aukštesniu dažniu spektro komponentes tam, kad būtų galima padidinti jų įtaką bei pagerinti naudojamų požymių kokybę. Tokiu būdu yra nuslopinamos žemesniu dažniu spektro komponentės ir taip spektras „išlyginamas“. Laiko srityje pradinė filtracija atliekama panaudojus žemos eilės skaitmeninį RIR filtrą. Dažniausiai naudojamas pirmos eilės RIR filtras, apibrėžiamas kaip:

$$\tilde{s}(n) = s(n) - \alpha \quad n - \ ) \ (2)$$

čia $\tilde{s}$ (n) yra filtruotas signalas, s(n) yra pradinės diskretizuoto šnekos signalo reikšmės, α – koeficientas, apsprendžiantis šnekos signalo spektro išlyginimo laipsnį, jis yra parenkamas intervale $0,9 \le \alpha \le 1,0$. Šio filtro sistemos funkcija:

$$H(z) = \ -\alpha \quad (3)$$

3.2.3.3 Signalų dalijimas į kadrus

Dalijimas į kadrus yra labai svarbus etapas kuriant kalbos ir kalbančiojo atpažinimo sistemas. Spektriniai kalbos signalų požymiai yra išskiriami iš trumpų kalbos signalo intervalų – kadrų, kurių trukmė apie 20–25 ms. Taip elgiamasi todėl, kadangi remiamasi daroma prielaida, kad per trumpą laiko intervalą žmogaus balso trakto parametrai nespėja pasikeisti (Hui-Ling 2002), t. y. trumpame laiko intervale žmogaus balso traktą galima aprašyti pastoviais parametrais.

Dažniausiai šie kadrai persikloja, t. y. sekantis kadras prasideda nuo prieš tai buvusio kadro tam tikros dalies, tuomet tarp kadru atsiranda tam tikra koreliacija. Atskiras kadras gali buti išreikštas:

$$s(j,n) = \text{(}j*(N-\text{)})+\text{)} \quad (4)$$

čia s(n) – originalus signalas, s(j,n) – j-tasis kadras, N – kadro ilgis atskaitomis, O – gretimų kadrų persiklojimo ilgis atskaitomis.

3.2.3.4 Lango funkcijos taikymas

Prieš tolimesnį apdorojimą, kalbos signalo kadrai yra dauginami iš tam tikros lango funkcijos w(n) . Signalą po lango funkcijos galime išreikšti:

$$s_w(n) = \text{(}n)*w(n) \quad (5)$$

Tiesiog paėmus signalo kadrą tai yra tolygu jį padauginti iš stačiakampio lango funkcijos:

$$w(n) = \begin{cases} 1, kai 1 \le n \le N; \\ 0, kitur. \end{cases} \quad (6)$$

Kadangi stačiakampį frontą suformuoti reikia begalinio spektro harmonikų skaičiaus, todėl stačiakampis langas turi labai prastas spektrines charakteristikas, t. y. toks langas labai iškraipo kalbos signalo spektrą. Dėl to parenkamos langų funkcijos, kurios ties lango pradžia ir pabaiga artėja prie nulio. Naudojami įvairūs langai: Hemingo, Haningo, Gauso, Barleto (Bartlett) ir t. t.

Hemingo lango funkcija:

$$w(n) = \begin{cases} 0.54 - 0.46*\cos(2\pi \frac{n}{N}), \ 1 \le \text{n} \le \text{N}; \\ 0, \text{kitur.} \end{cases} \quad (7)$$

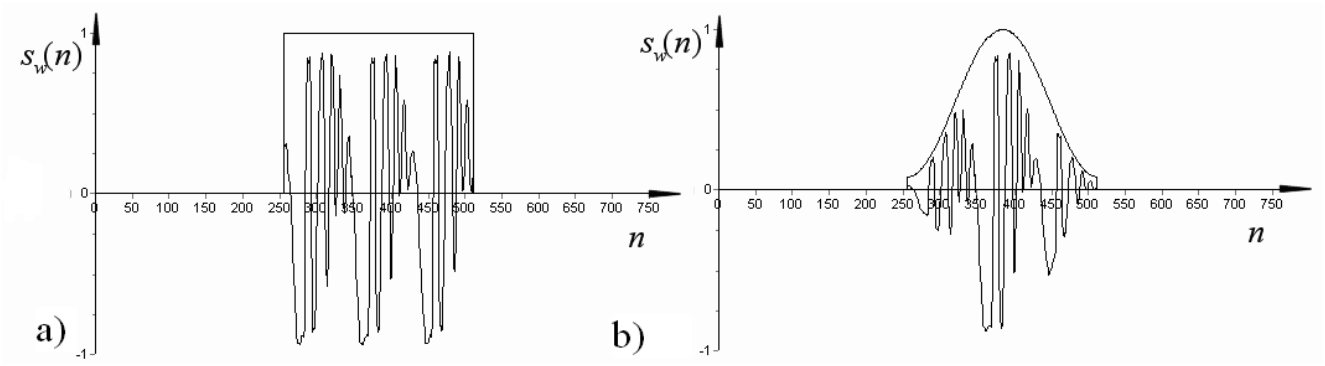
Haningo (Hanning) lango funkcija:

$$w(n) = \begin{cases} 0.5 - 0.5*\cos(2\pi \frac{n}{N}), 1 \le \text{n} \le \text{N}; \\ 0, \text{kitur.} \end{cases} \quad (8)$$

Gauso lango funkcija:

$$w(n) = \begin{cases} e^{-\frac{1}{2}(\alpha \frac{n}{N/2})^2}, 1 \le \text{n} \le \text{N}; \\ 0, \text{kitur.} \end{cases} \quad (9)$$

Kalbos signalo kadro dauginimas iš stačiakampio bei Hemingo lango pavaizduotas 6 paveiksle.



Pav 6 Signalo kadras, padaugintas iš: a) stačiakampio lango funkcijos; b) Hemingo lango funkcijos

3.2.1 Kalbos signalų duomenų bazės

Kalbos atpažinimo sistemas reikia apmokyti. Tam kuriami garsynai – specialiai surinktų kalbos signalų duomenų bazės (angl. speech corpora). Sukurti universalų garsyną yra sudėtingas uždavinys, nes toks garsynas turėtų gerai aprašyti kalboje sutinkamų fonetinių vienetų įvairovę, įvertinti kontekstinius efektus, diktorių ir kalbėjimo stilių įvairovę ir pan. Todėl garsynai sudaromi orientuojantis į tam tikros uždavinių klasės sprendimą. Šiuo metu pasaulyje sukurta nemažai įvairių garsynų, kuriuose surinkta skirtingų kalbų medžiaga. Tačiau keli garsynai yra tapę savotiškais standartais, į kuriuos orientuojamasi sudarant naujas kalbos signalų duomenų bazes. Visų pirma tai TIMIT garsynas ir jo variantai (NTIMIT, CTIMIT).

Lietuviškų kalbos signalų technologijų vystymas neįmanomas be lietuvių kalbos signalų duomenų bazių. Tai būtų instrumentai mokslinių ir praktinių uždavinių vystymui, metrologijos priemonės. Pirmas sisteminis žingsnis – tai LTDIGITS garsynas, sudarytas KTU ir VU. Tačiau LTDIGITS medžiaga - tik nedidelė dalis lietuviškų fonetinių vienetų bei žodžių įvairovės, todėl būtina ir toliau kaupti lietuviškus garsynus, surinktus papildyti nauja medžiaga bei apdoroti, t.y. atlikti leksinį bei fonetinį segmentavimą.

Pagal paskirtį garsynus būtų galima išskirti į tokias grupes: garsynai kurie buvo kurti balso atpažinimo nuo diktoriaus nepriklausančias sistemas ir yra sistemos, kurios balsą atpažįsta, tačiau turi būt pritaikomos (apmokamos) prie diktoriaus.

3.2.1.1 TIMIT garsynas [34]

Bazės pavadinimas sudarytas iš pagrindinių autorių (Texas Instruments ir Massachussets Institute of Technology) pavadinimų abreviatūrų. Tai viena iš pirmųjų sistemingai surinktų kalbos signalų duomenų bazių.

TIMIT duomenis vieno įrašų seanso metu perskaitė 630 diktorių (po 10 sakinių kiekvienas), atstovaujantys 8 JAV dialektinius regionus. Saugantis pašalinių triukšmų, įrašai padaryti akustinėje kabinoje, naudojantis plačiajuosčiu mikrofonu.

Kiekvienas diktorius perskaitė po du tuos pačius sakinius, kuriuose buvo siekiama atspindėti dialekto ypatybes (*dialect senteces - SA*).

Fonetinių porų įvairovė vaizduojama taip vadinamais reprezentatyviais - kompaktiniais sakiniais (*phonetically-compact senteces SX*). Kiekvienas diktorius perskaitė po 5 šio tipo sakinius, o kiekvieną parinktą tekstą perskaitė 7 diktoriai.

Maksimizuojant tekstuose aptiktų alofonų įvairovę, kiekvienam diktoriui buvo pasiūlyta perskaityti po 3 tik jam skirtus sakinius (*phonetically-diverse senteces SI*).

Ši bazė unikali tuo, kad turi labai kruopščiai sužymėtas fonemų ribas. Daugelyje kitų bazių pateikiamos tik žodžių ribos. Būtent todėl ji tapo plačiai naudojama ir palaipsniui, adaptavus įvairiems ryšio kanalams, buvo transformuota į kitas (NTIMIT , CTIMIT , FFMTIMIT, HTIMIT).

Prie šios duomenų bazės priėjimo per internetą nėra, bazė platinama CD ROM diskelyje.

3.2.1.2 Diktoriaus nustatymo duomenų bazės [34]

Diktoriaus nustatymo uždaviniams skirtų garsynų kategorijoje populiariausiomis tapo YOHO ir SWITCHBOARD duomenų bazės.

YOHO duomenų bazė buvo sukurta užsakius JAV vyriausybei, siekiant suformuoti kalbos signalais paremtas diktoriaus identifikavimo/verifikavimo priemones. Čia naudotas telefoninis aukštos kokybės mikrofonas ofiso tipo akustinėje aplinkoje. Duomenis perskaitė 138 diktoriai, pakartotinai perskaitydami tekstus kelių mėnesių bėgyje.

SWITCHBOARD bazė skirta finansinių operacijų, naudojantis kredito kortelėmis ir perduodant balso komandas telefono kanalu, vykdymui.

Eilė pažangių principų realizuota europinėje daugiakalbėje POLYCOST duomenų bazėje.

3.2.1.3 Fonetinės duomenų bazės [34]

Šioje kategorijoje paminėtini ISOLET ir OGI garsynai.

ISOLET bazėje yra sukaupti 150 diktorių perskaityti anglų kalbos raidžių pavadinimų įrašai, naudojant aukštos kokybės mikrofoną (diskretizavimo dažnis 16 kHz, spektras iki 7.6 kHz). Fonetinį segmentavimą reikia atlikti vartotojui.

3650 diktorių per telefoną sudiktavo OGI duomenis. Čia yra diktorių vardai, pavardės, angliškai ištarti žodžiai taip/ ne, diktoriaus gyvenamoji vieta, jo gimtinė, anglų kalbos raidžių pavadinimai bei kita. Svarbu, jog tam tikra duomenų dalis yra fonetiškai segmentuota.

Kuriant LTDIGITS buvo panaudoti kai kurios aukščiau aprašytų garsynų savybės:

- analogiška TIMIT bazei duomenų išdėstymo sistemą;

- POLYCOST pavyzdžiu paįvairinti medžiagą: t.y., šalia skaičių sekų padiktuoti kai kurių valdymo komandų pavadinimus;

- papildant ISOLET spragas, sudiktuoti porą sekų, kuriose yra akustiškai artimi skiemenis, kad būtų galima spręsti fonemų skyrimo uždavinius.

Galima daryti prielaidą, kad ir kuriant kitus lietuviškus garsynus reikia naudoti šių duomenų bazių bei LTDIGITS savybėmis. Tai padėtų išspręsti ir visą eilę suderinamumo problemų.

3.2.1.4 Lietuviški garsynai

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

Garsyną kaip produktą apsprendžia jį ruošianti žmonių grupė. Svarbu, kad į grupę patektų kuo įvairesnės specializacijos žmonių – lingvistų, programuotojų, vartotojų. Tada galima tikėtis, kad garsynas atspindės vartotojų poreikius, atitiks galiojančias kalbos normas ir bus aprūpintas programine įranga, leidžiančia
lanksčiai dirbti su garsyne esančiais duomenimis. Išskirtiniu požiūriu į lietuvių šnekos garsinės sistemos ypatumus pasižymi VDU kuriami garsynai. Jie anotuojami keliais lygiais, pagristai parenkant fonetinių vienetų

sistemą (Raškinis et al. 2003a). Matematikos ir Informatikos institutas labiau specializuojasi kurti šnekos atpažinimo technologijas ir jas tirti, todėl čia ruošiami garsynai tenkina tik pagrindinius programinių įrankių reikalavimus.

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

Atliekant kalbos sistemų tyrimus, populiarumo atžvilgiu pirmauja nedidelės apimties, konkrečiam uždaviniui skirtos kalbos signalų duomenų bazės. 1999 metais buvo pradėtas rinkti pirmasis sisteminis lietuviškas garsynas – LTDIGITS [44].

Tai pat galime rasti dar kelis lietuviškus garsynus:

• 2002 metais VDU sukurtas bendrinės lietuvių kalbos pavieniui tariamų žodžių universalus anotuotas garsynas (4 diktoriai, 731 skirtingas žodis, 1 valandos garso įrašų trukmė) (Raškinis ir kt., 2004b).

• Matematikos ir informatikos instituto (MII) ir VDU surinktas lietuvių kalba transliuojamų žinių garsynas LRN0 (23 diktoriai, daugiau nei 18000 žodžių). (Šilingas, 2005).

3.2.1.4.1 LT DIGITS garsynas

Garsyne įgarsino 294 diktoriai. Iš jų 120 moterų, 129 vyrai. Garsyne yra daugiau kaip 350 įrašų, kur kiekvienas diktorius perskaitė 10 žodžių sekų (frazių). Vienos sekos ilgis yra 6 žodžiai arba skiemenys. Visos sekos toliau bus pateiktos 1 lentelėje:

LT DIGITS ŽODŽIŲ SEKOS **LENTELĖ 1**

Sekos nr.	Sekos turinys					
1 - 5	trys	penki	keturi	vienas	devyni	nulis
6	pradėti	baigti	sustoti	pauzė	laukti	tęsti
7	pirmyn	atgal	į pradžią	į pabaigą	sekantis	perduoti
8	taip	ne	pagalba	saugoti	start	stop
9	ma	na	mu	nu	mi	ni
10	Mikas	mato	nišoje	mūsų	namo	numerį

Šaltinis: LT DIGITS garsyno projektas: Rudžionis A., Rudžionis V., Žvinys P. Lietuvių kalbos signalų duomenų bazės LTDIGITS akustinės-fonetinės charakteristikos/ Materialy XXVIII mežvuzovskoj naučno–metodičeskoj konferencii prepodavatelej i aspirantov/ Sekcija baltistiki.- Sankt Peterburgskij Gosudarstvennyj universitet, 2-4 marta, s.29-30.

Iš lentelės matyti, kad penkios pirmosios sekos (1 – 5) yra rišliai ištartų lietuviškų skaitmenų pavadinimai nuo 0 iki 9, parinkti atsitiktine tvarka (PIN kodo, asmens kodo ar telefono numerio rinkimo modeliavimas). Sekančios trys sekos (6 – 8) yra izoliuotai tariamų lietuviškų valdymo komandų pavadinimai. Devintoji seka yra šeši skiemenys, kur du nosiniai priebalsiai (m, n) tariami

prieš tris kontrastingiausius balsius (a, i, u). Dešimtoje sekoje visi 6 devintos sekos skiemenys kirčiuotose pozicijose yra sujungti į rišlią frazę.

Diktorių pasiskirstymas pagal amžių nėra įvairus. Daugiausiai frazes perskaitė asmenys nuo 18-25 metų (71 %).

Didelė dalis diktorių buvo studentai ir jie buvo suskirstyti į dvi grupes: filologai ir ekonomistai. Ekonomistų grupei priskirti taip pat informatikai, inžinieriai ir kitų profesijų darbuotojai. Bazės apraše yra paliktas originalus profesijos pavadinimas, ir paaiškėjus, kad tarsenai įtakos turi profesija, ją bus galima atsekti.

Fiziškai „LT DIGITS“ duomenų bazė susideda iš dviejų dalių, įrašytų skirtinguose disko kataloguose LTDIGITS\MALE**Mx** (vyrų balsai) ir LTDIGITS\FEMALE**Fx** (moterų balsai), kuriuose yra audio ir tekstiniai failai. Simbolis x reiškia diktoriaus arba įrašo numerį. Pavyzdžiui, 5 diktoriaus vyro duomenys būtų kataloge LTDIGITS\MALE\M005. Kiekvieno diktoriaus kataloge yra po tris failus kiekvienai frazei. Failuose įrašytos signalo diskretos, frazės ortografija ir frazės transkripcija.

Failų specifikacijos pavyzdys pateiktas 2 lentelėje.

LT DIGITS garsyno failų specifikacija

Lentelė 2

Garsyno failų specifikacijos pavyzdys	Turinys
…\LTDIGITS\MALE\M017\T05.WAV	Signalo diskretos
…\LTDIGITS\MALE\M017\T05.TXT	Ortografija
…\LTDIGITS\MALE\M017\T05.TRN	Transkripcija
…\LTDIGITS\MALE\M017\T05.WRD	Leksinės žymės
…\LTDIGITS\MALE\M017\T05.PHN	Fonetinės žymės

Šaltinis: LT DIGITS garsyno projektas: Rudžionis A., Rudžionis V., Žvinys P. Lietuvių kalbos signalų duomenų bazės LTDIGITS akustinės-fonetinės charakteristikos/ Materialy XXVIII mežvuzovskoj naučno– metodičeskoj konferencii prepodavatelej i aspirantov/ Sekcija baltistiki.- Sankt Peterburgskij Gosudarstvennyj universitet, 2-4 marta, s.29-30.

Viena iš didesnių problemų kurią pateikė garsyno autoriai ta, kad garsyno įrašams buvo pasirinkti neprofesionalus, t.y. netreniruoti diktoriai, kurie atsidūrę prieš mikrofoną dažnai pasimeta, pradeda kikenti, o besistengdami nesuklysti pradeda daryti rimtas tarties klaidas. Šios problemos sprendimo būdai gali būt 2: arba parinkti profesionalius diktorius, tačiau tai labai padidina kaštus, arba priartinti garsų įrašymo aplinką prie diktoriui artimos aplinkos [43].

3.2.2 Garso signalų pirminės analizės sistemos

Kokybiškai sudaryti garsynai - tai vienas iš pagrindinių aspektų gerai balso atpažinimo sistemai. Anksčiau sudarytų garsynų kūrimo patirtis parodė, kad išgauti kokybiškus garsus neužtenka parinkti tik gerą įrašinėjimo aplinką bei technologijos. Didelę įtaką sudaro žmogiškasis faktorius. Pavyzdžiui, kuriant LT DIGITS garsyną problema buvo, kad pasirinkti neprofesionalus, t.y.

netreniruoti diktoriai, kurie atsidūrę prieš mikrofoną dažnai pasimeta, pradeda kikenti, o besistengdami nesuklysti pradeda daryti rimtas tarties klaidas.

Kuriama informacinė sistema veiks internete, tad diktoriai bus tiesiogiai neprieinami sistemos administratoriams, todėl reikalinga įrašytų signalų pirminės analizės sistema.

Kalbos signalų analizavimo būdų yra sukurta daug ir įvairių (pvz. Skaitmeniniai filtrai, žodžių ribų išskyrimas, laiko skaičiavimas, diskretų pokyčių skaičiavimas ir t.t.). Tačiau šiai sistemai reikalinga tik greita ir pirminė analizė kuri leistų atmesti nekokybiškus signalus.

3.2.2.1 Numatytos ir realios įrašo trukmės lyginimas

Toks skaitmeninio įrašo analizės metodas remiasi iškart žinomu analizuojamų žodžių ilgiu sekundėmis: imamas teorinis įrašo laikas ir lyginamas su realiu įrašo ilgiu sekundėmis. Jei skirtumas yra didesnis už galimą paklaidą, laikoma, kad įrašas netinkamas.

- Privalumai: labai greit apskaičiuojama ir rezultatai visada būna tikslūs.
- Trūkumai: sunku parinkti teorinį įrašo ilgį sekundėmis. Tai reikalauja daug atitinkamų statistinių žinių.

3.2.2.2 Įrašų diskretų didžiausių ir mažiausių reikšmių lyginimas

Šis metodas yra skirtas nustatyti garso bangos amplitudės pokyčių dydį. Tai leidžia surasti pokyčius tarp vidutinių didžiausių ir vidutinių mažiausių diskretų reikšmių. Palyginus procentalias reikšmes galima spręsti, kad įraše yra atskiri žodžiai. Jei pokyčiai yra mažesni už nustatytą statistinę reikšmę, galima spręsti, kad įraše yra tik vientisas garso signalas ir yra netinkamas.

- Privalumai: galima atskirti tuščius (be žodžių) ar nekokybiškus balso įrašus.
- Trūkumai: užtrunka apskaičiavimas. Ilgiems įrašam netinka dėl amplitudės vidurkio vienodėjimo. Sunku surasti tinkamas statistines ribas.

3.2.2.3 Įrašų diskretų (energijos) maksimalios reikšmės analizė

Šiuo būdu galima nustatyti ar įrašas nėra įrašytas per garsiai šnekant į mikrofoną. Kai žmogaus balso garso bangos viršija mikrofono maksimalias dažnio ribas, įraše signalas būna nekokybiškas. Tai galima apskaičiuoti maksimalių diskretų reikšmių sumai. Ji yra palyginama su visų diskretų skaičiumi, jei skirtumas viršija nustatytą reikšmę – laikoma, kad įrašas netinkamas

- Privalumai: gan tiksliai galima atpažinti įrašus, kurie viršija mikrofono technines galimybes.
- Trūkumai: užtrunka apskaičiavimas. Sunku nustatyti statistines ribas, kurios nurodo įrašo netinkamumą.

3.2.3 Siūlomas algoritmas

Apibendrinus esamas atpažinimo sistemas bei kalbos signalų garsynus, juose esančias klaidas, nustatyta, kad tokios sistemos, kuri atliktų tik kalbos signalų kokybinės charakteristikų tikrinimą nėra. Daugiausia naudojama kompleksinės sistemos, kuriose vykdomas kalbos atpažinimas ir naudojami jau esami garsynai. Kadangi esamuose garsynuose yra klaidų, taip pat jie nėra dideli apimtimi siūloma algoritmas, kuris atliktų pirminę kalbos signalų kokybės charakteristikų analizę, kas pagerintų garsyno kokybę bei padidintų apimtis. Siūlomas algoritmas leidžia atlikti įrašą diktoriui patogiu metu, nepaliekant jam įprastos aplinkos, taip pat būtų galima realizuoti kokybinių įrašo charakteristikų patikrinimą realiu laiku, ir pastebėjus klaidas prašyti diktoriaus pakartoti reikiamą frazęAlgoritmo realizacija:

- Kalbos signalo įvedimas;
- Kalbos signalo kokybinių charakteristikų tikrinimas:
    - skaičiuojama ir tikrinama įrašo trukmė,
    - tikrinama įrašo energija,
    - ieškoma didžiausia/mažiausia amplitudės reikšmė,
    - skaičiuojama signalo-triukšmo santykis.
- Įrašai neatitikę kokybinių charakteristikų gražinami vartotojui su klaidos apibūdinimu (pagal tai kurios charakteristikos netenkina) pataisymui;
- Neradus klaidų įrašas priimamas, įrašomas į duomenų bazę.

. Šio algoritmo pagalba taupyta laikas bei žmogiškieji ištekliai surenkant garsyno įrašus. Taip pat didėja garsyno apimtys, kas lemia įrašų įvairumą. Tai būtų didelis privalumas tolimesniuose kalbos technologijų etapuose, tokiuose kaip rišlios kalbos atpažinimas, segmentavimas, žodžio ribų ieškojimas, atskirų kalbos garsų tyrimas ir pan.

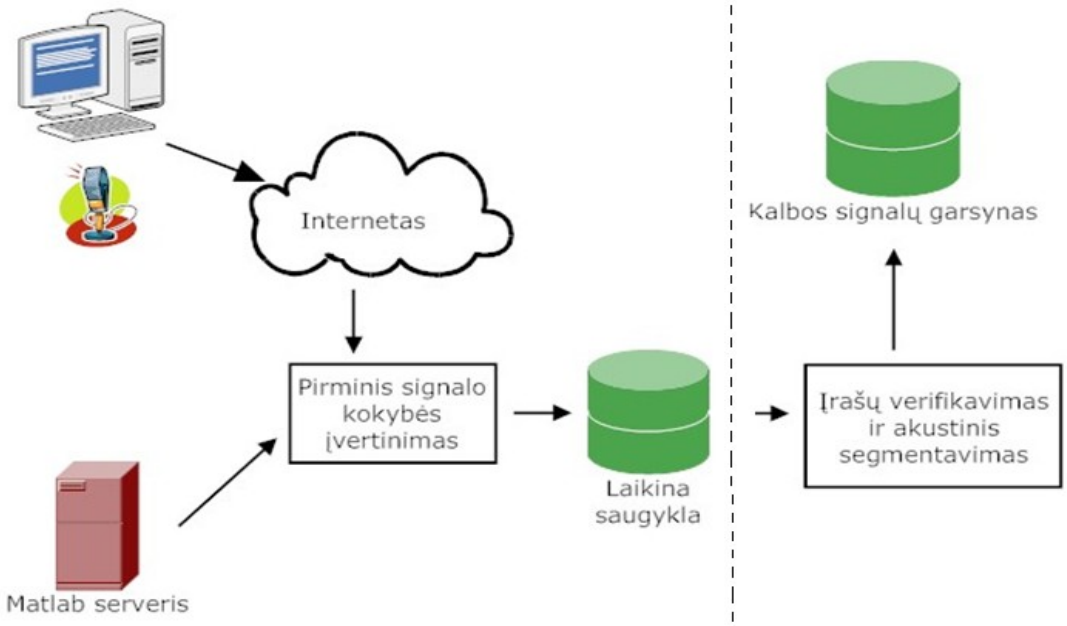
4 KALBOS SIGNALŲ KOKYBINIŲ CHARAKTERISTIKŲ ANALIZĖS SISTEMA

Bandant bent jau iš dalies eliminuoti pastebėtus LTDIGITS trūkumus ir norint tinkamai valdyti ir analizuoti kalbos signalus ir su jais susijusius meta duomenis, reikalinga automatizuota programinė aplinka, kuri suteiktų tokias galimybes:

- Įrašyti, patikrinti kokybines charakteristikas ir išsaugoti kalbos signalą bei su juo susijusius meta duomenis duomenų bazėje;
- Įrašyti ir peržiūrėti kalbos signalą per Internetą;
- Atrinkti pagal dominančius kriterijus, peržiūrėti ir paruošti duomenis analitiniam apdorojimui;
- Sisteminti gautus rezultatus.

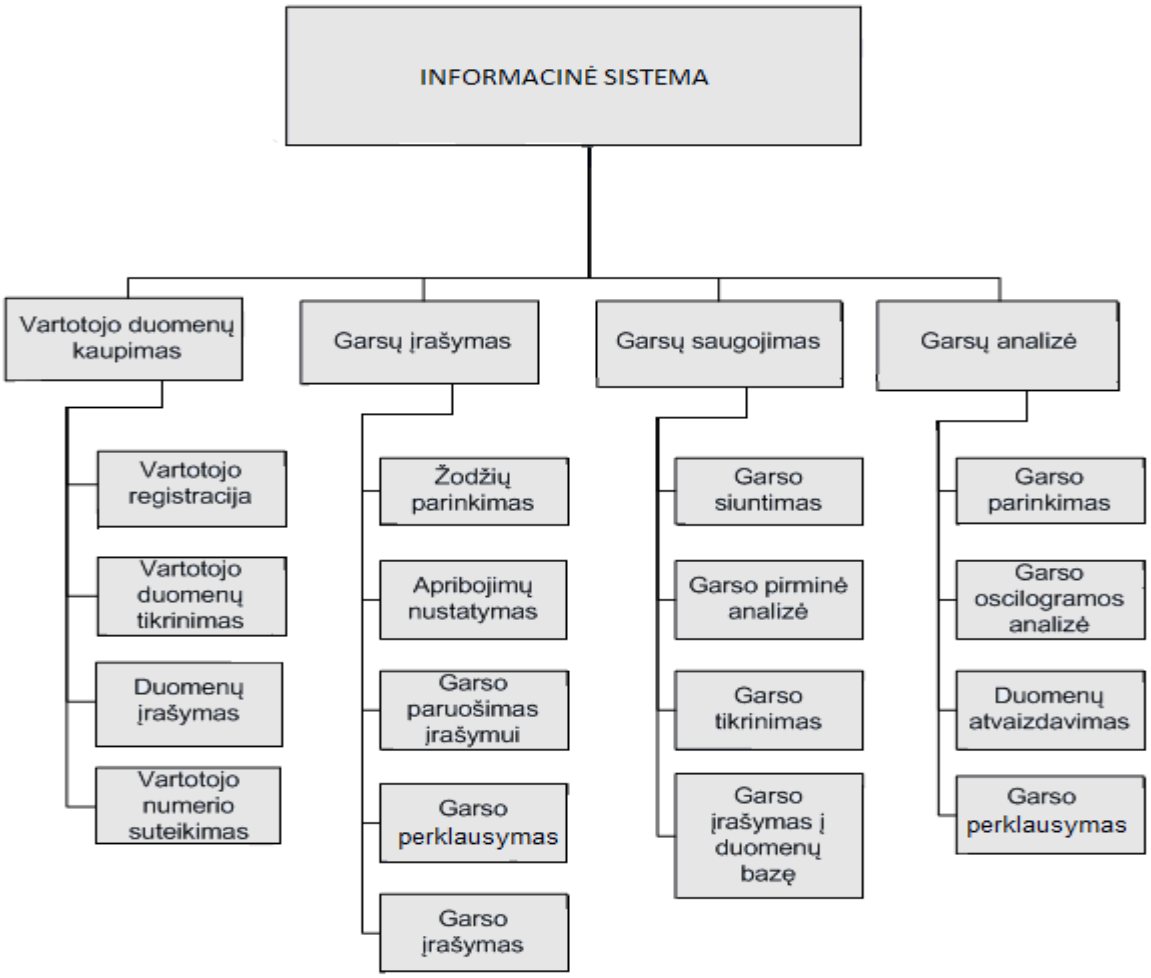
4.1 Tyrime naudojamo garsyno garsų įrašymo sistemos kūrimas

Sistemai sukurti pasinaudota PHP ir Java programavimo kalbos, bei Matlab įrankis. Sekančiame (7) paveikslėlyje atvaizduota sistemos architektūra. Kadangi tyrimas skirtas tik pirminei analizei (kalbos garsų signalų kokybinėms charakteristikoms nustatyti), realizuojama tai kas paveikslėlyje matoma iki brūkšneliais atskirtos vietos.



Pav 7 Sistemos architektūra

Taip pat pateikiamas Garsų įrašymo bei analizės IS funkcijų hierarchijos diagrama (8 pav.).

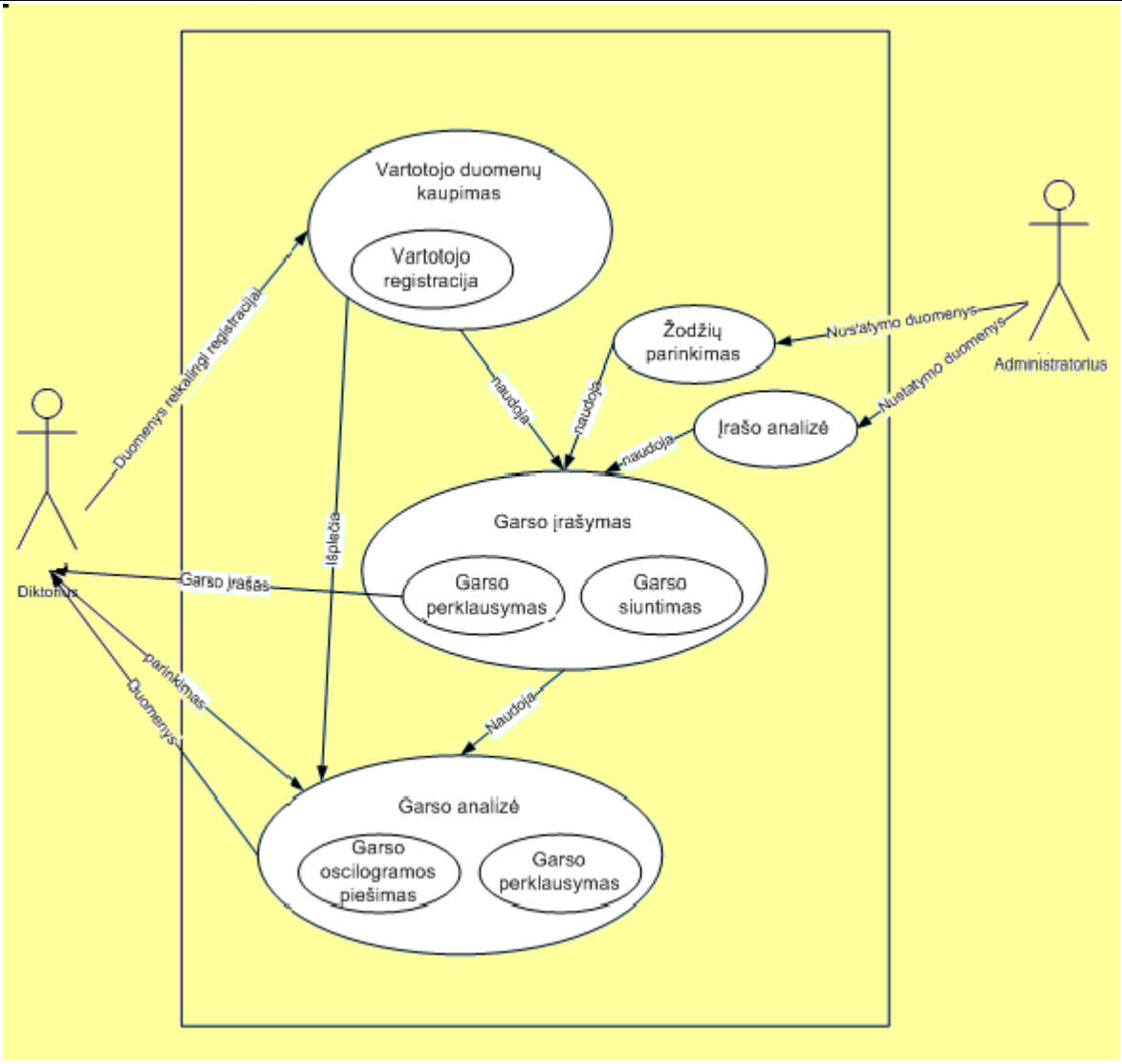


Pav 8 . „Garsų įrašymo bei analizės informacinės sistemos funkcijų hierarchijos diagrama“.
Šaltinis: sukurta autorės

Iš paveikslėlio matyti, kad yra sudarytos pagrindinės keturios funkcijos.

- Kaupiami vartotojų duomenys: skirta kaupti duomenis apie naują vartotoją;
- Garsų įrašymas: skirta naujo garso įrašymui bei perklausai;
- Garsų saugojimas: skirta garso nusiuntimui ir pirminės jo analizės apdorojimui bei įrašymui į duomenų bazę;
- Garsų analizė: skirta atvaizduoti visą informaciją apie garsą vartotojui.

Toliau (9 pav.) yra pateikta analizuojamos IS procesai bei vartotojai, kurie juos naudoja.



Pav 9 „Informacinės sistemos vartotojų poreikių specifikacija“.

Vartotojų poreikių specifikacijoje matome du pagrindinius vartotojus: tai diktorius, ir sistemos administratorius. Nors vartotojai yra du, tačiau visa sistema yra orientuota į diktorių.

Diktoriaus pirminis žingsnis tai yra vartotojo duomenų registracija. Kadangi garsyne reikalinga kaupti duomenis ne tik apie garsus bet ir specifinius diktoriaus duomenis.

Kitas diktoriaus žingsnis yra garso įrašymas. Garso įrašymas yra sudėtinis procesas, kuris reikalauja ir administratoriaus įsikišimo. Administratorius nurodo garso analizavimo kriterijus pagal kuriuos naujas įrašas bus tikrinamas ir analizuojamas. Garso įrašymas yra ciklinis procesas ir neatitikus tinkamo įrašo kriterijų jis nebus įtraukiamas į pagrindinę duomenų bazę.

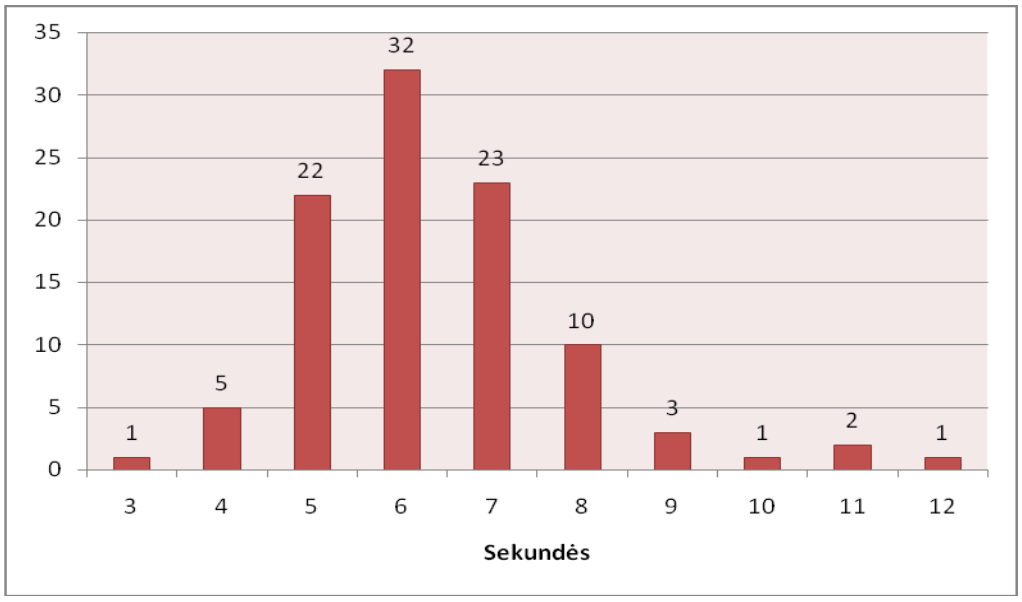
Kai naujas garsas sėkmingai įrašomas į duomenų bazę, tada būna paskutinis diktoriaus žingsnis – garso analizė. Jos metu diktorius gauna pirminius duomenis apie įrašytą garsą: garso oscilograma, garso perklausymas, kiti techniniai įrašo parametrai.

5 EKSPERIMENTINĖ DALIS

Signalų kokybės tikrinimui paskaičiuota kiekvieno įrašo vidutinė trukmė bei vidutinė energijos reikšmė, tiriamas kiekvienos frazės amplitudinio apribojimo santykis ir STS santykis. Kokybės analizė atlikta 10 diktorių – 5 vyrų ir 5 moterų. Kiekvienas iš jų perskaitė po 10 frazių, kurios sudarytos iš skaičių nuo 0 iki 9. Frazėje skaičiai sugeneruojami atsitiktine tvarka.

Pirmoji kalbos signalų kokybės charakteristika, kuri tiriama tai įrašo trukmė. Įrašytų kalbos signalų trukmė vyravo nuo 3,46s iki 12,25s.

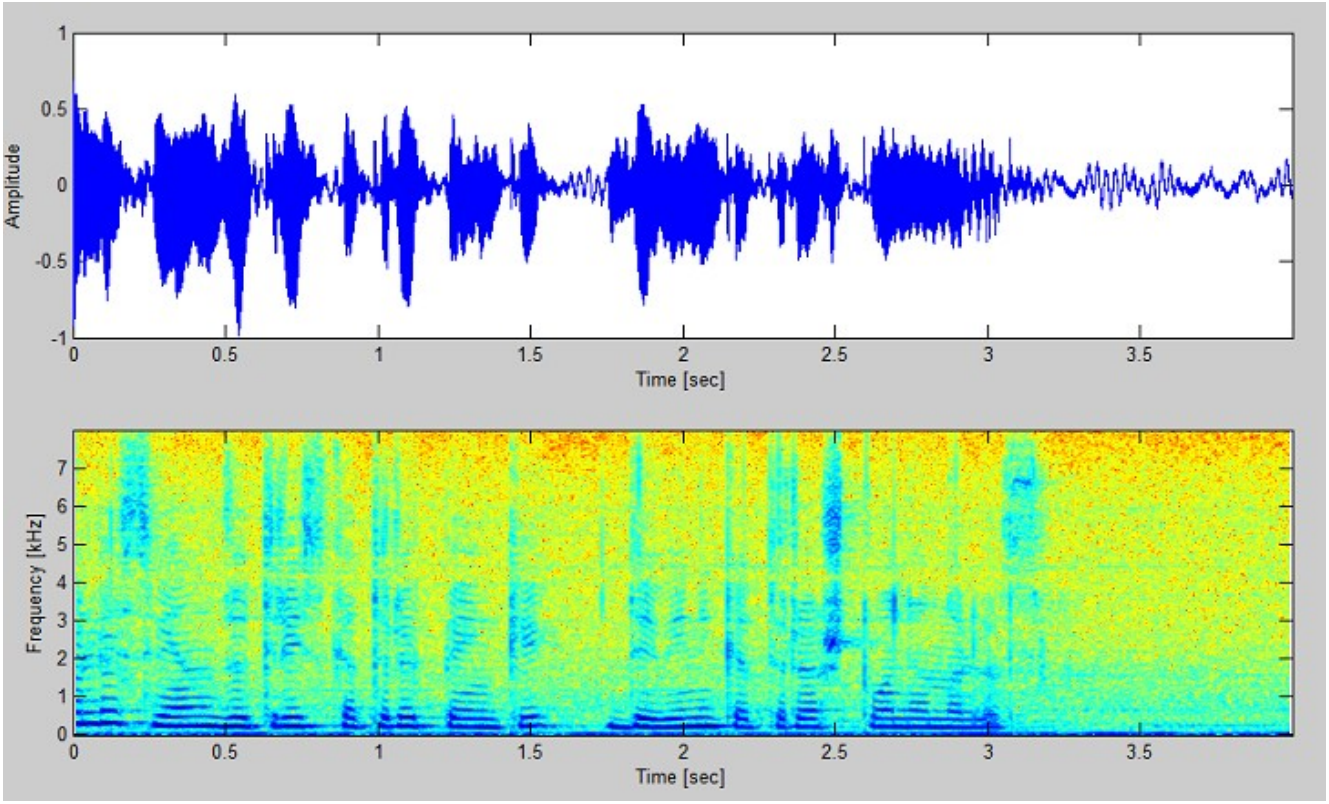
Apskaičiuotas kiekvieno įrašo ilgis (sekundėmis), siekiant nustatyti, ar nėra kritiškai ilgų arba trumpų įrašų. Gautas trukmės skirstinys pateikiamas 10 pav.



Pav 10 Garsyno frazių trukmių skirstinys

Patį trumpiausią įrašą įrašė 1 diktorius- tai buvo moteris (Pav 11). Įrašo trukmė 3,46s. Įrašų nesiekiančių 4 sek daugiau nebuvo. Kalbėjimas yra greitas, labai trumpos pauzės tarp žodžių, nors triukšmų nejaučiama, gana aiškiai kalbama, tačiau tokią frazę turime atmesti. Įrašo trukmė yra kritiškai trumpa, per mažos pauzės tarp žodžių, dėl šios priežasties bus sunku ar net neįmanoma atlikti sekančių žingsnių – segmentavimas, žodžių ribų nustatymas. Taigi, šį įrašą garantuotai atmetame, vartotojui parodoma klaida, kad jis per greitai kalba, kad sulėtintų tempą, padarytų pauzių tarp žodžių.

Atmetame ir įrašus kurių trukmė siekia 4s, kadangi tai taip pat kritiškai trumpa įrašo frazė, vartotojui parodoma klaida, kad jis per greitai kalba, kad sulėtintų tempą, padarytų pauzių tarp žodžių.



Pav 11 Trumpiausias įrašas

Įrašų, kurių trukmė buvo tarp 4 ir 5 sekundžių buvo 5. Juos įrašė 3 moterys ir 2 vyrai. Kalbos tempas suprantamas, tačiau nepakankamai kad tiktų tolimesnei analizei, pauzės tarp žodžių trumpos, kai kurių žodžių pabaiga su sekančio pradžia susilieja. Taigi skaičiuojant pauzių skaičių nesutampa su skaičiumi realiai turinčių būti pauzių. Įrašai taipogi atmetami, traktuojama kritiškai trumpo įrašo klaida.

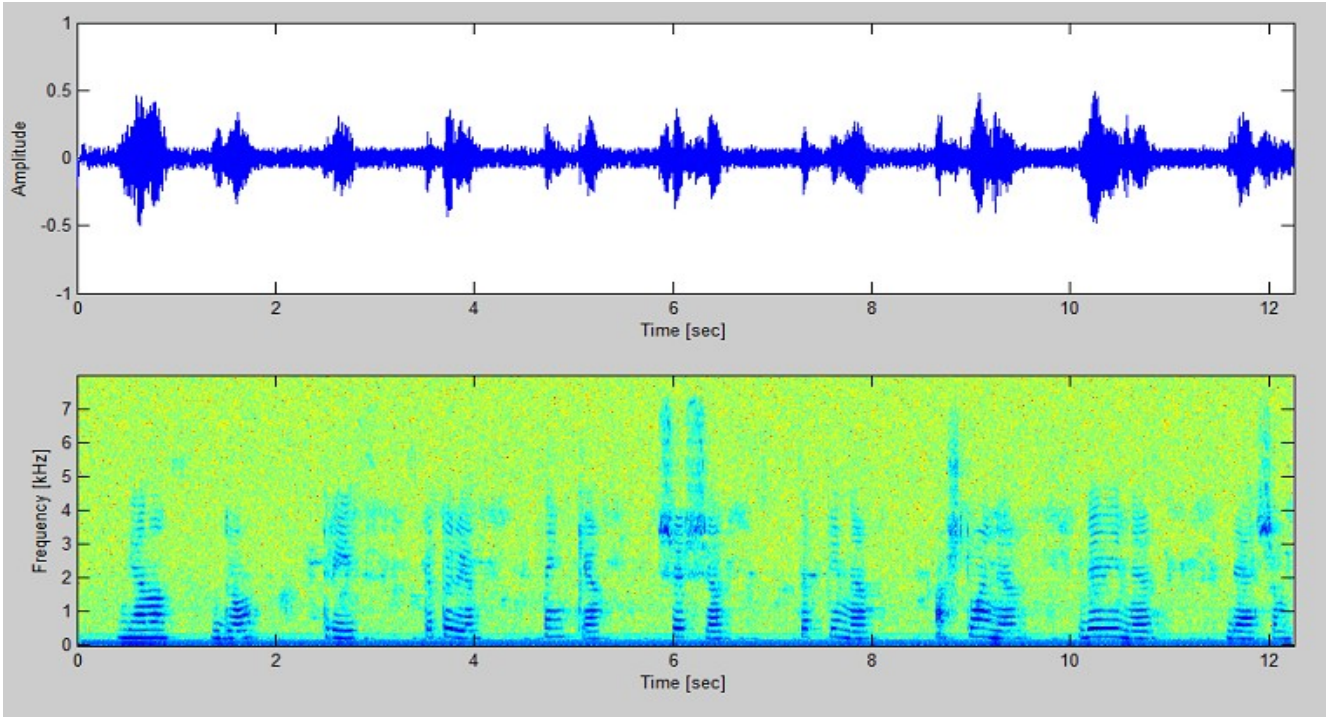
Įrašų tarp 5 ir 6 sekundžių buvo 22, kalbėjimo tempas nėra per greitas, lengviau išsiskiria pauzės tarp žodžių. Didžiausią dalį tarp įrašų sudarė tie, kurių vidutinė trukmė buvo tarp 6 ir 7 sekundžių, jų yra 32. Kalbos tempas vidutiniškas, apskaičiuotas pauzių skaičius tarp žodžių atitinka realų, bei yra pakankamos tolimesnei analizei.

Įrašų tarp 7 ir 8 sekundžių yra 23, įrašų kokybė analogiška kaip ir esančių tarp 5 ir 6. Duomenys analizei tinka, nėra kritiškai trumpų įrašo frazių.

Tarp 8 ir 9 sekundžių yra 10 garso įrašų, kalbos tempas sulėtėjęs, pauzės tarp žodžių pailgėję, tačiau dar nėra kritinių frazių.

Atmetame įrašus kurių trukmė siekia 10s ir 11s, kadangi tai jau kritiškai ilga įrašo frazė, vartotojui parodoma klaida, kad jis per lėtai kalba, kad pagreitintų tempą, padarytų mažesnes pauzes tarp žodžių.

Patį ilgiausią įrašą įrašė moteris (Pav 12).Įrašo trukmė – 12,25s. Kalbėjimas yra lėtas, ilgos pauzės tarp žodžių, nelabai aiškiai kalbama, tokią frazę turime atmesti. Įrašo trukmė yra kritiškai ilga, per ilgos pauzės tarp žodžių, dėl šios priežasties bus sunku ar net neįmanoma atlikti sekančių žingsnių – segmentavimas, žodžių ribų nustatymas. Taigi, šį įrašą garantuotai atmetame, vartotojui parodoma klaida, kad jis per lėtai kalba, kad pagreitintų tempą, padarytų mažesnes pauzes tarp žodžių.



Pav 12 Ilgiausias įrašas

Analizuojant kritiškos trukmės (tiek ilgas tiek trumpas) frazes pastebėta, kad įrašai yra geri tačiau su tam tikrais neatitikimais:

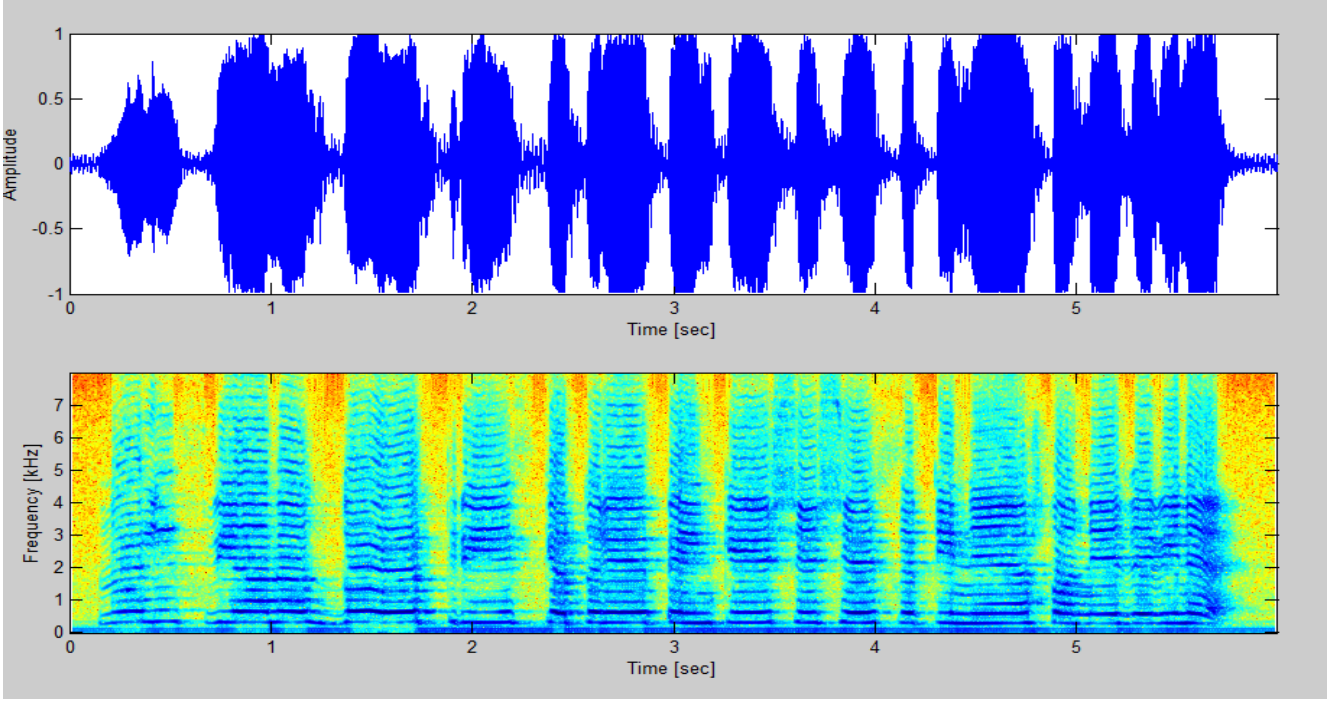
- daugumoje trumpų frazių diktorių kalbėjimo tempas yra didelis, nepadaroma minimali pauzė tarp žodžių;
- ilgose frazėse priešingai, kalbama labai lėtai, daromos didelės pauzės tarp žodžių;

Dėl tokių netikslumų frazės atmetamos, kaip netinkamos. Tokias klaidas galima paaiškinti diktorių neprofesionalumu, galbūt tai lemia susijaudinimas, pasimetimas, dėl to pasitaikė kad įrašant garsus dar prieš ištariant pirmą žodį būna padaryta ilgos pauzės.

Kad išvengt klaidų, įrašai su šiais neatitikimais nepatvirtinami.

Antroji kalbos signalų kokybės charakteristika, kuri tiriama tai amplitudinis apribojimas. Paprastai amplitudinio apribojimo santykis yra išreiškiamas maksimaliai/minimaliai galimai reikšmei lygių reikšmių skaičių padalijant iš failo diskretų skaičiaus. Buvo paskaičiuota kiekvienos frazės amplitudinis apribojimas ir gauta, kad 4 frazėse buvo užfiksuotas amplitudinis ribojimas. Trijose frazėse signalas viršijo amplitudės ribas ir viename įraše signalas buvo labai silpnas, tik viena kartą palietė amplitudės ribas ir visą laiką laikėsi arti nulio.

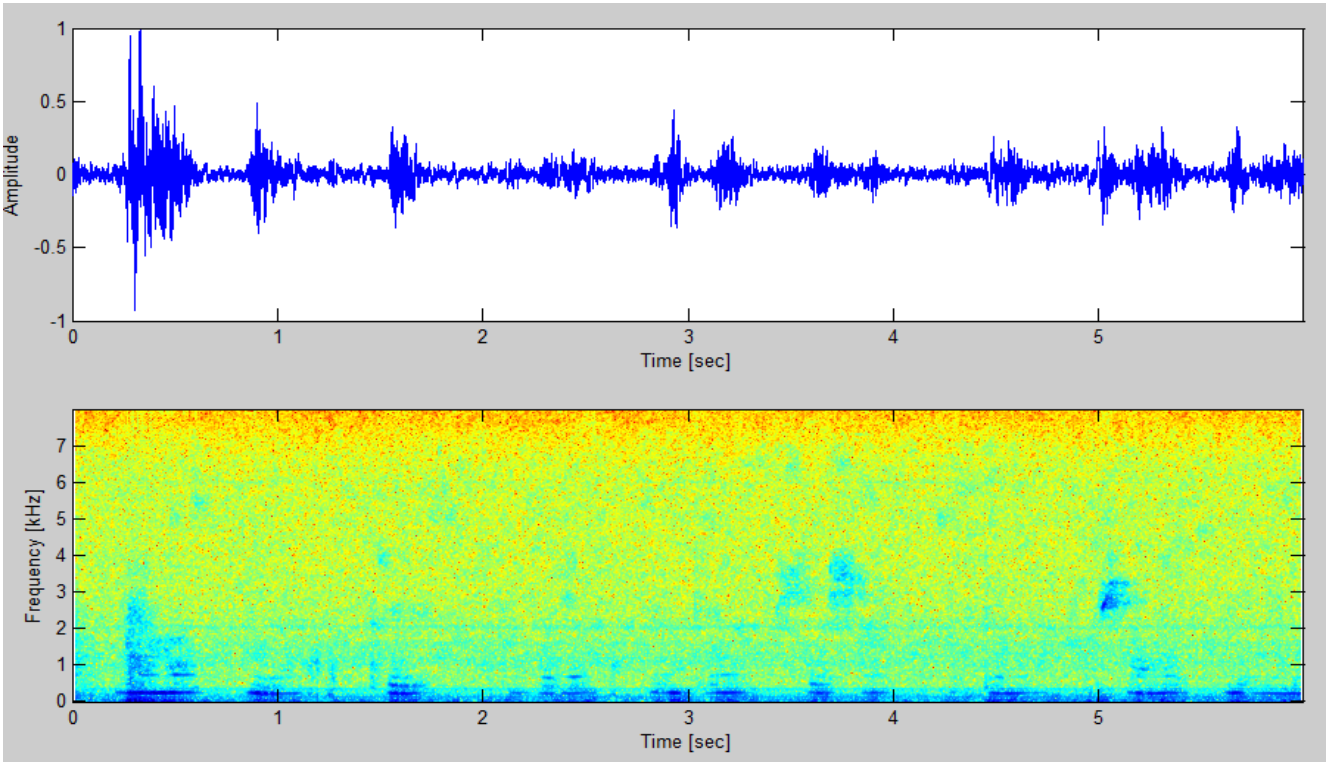
Pav. 13 aiškiai matosi, kad yra užfiksuota amplitudinis ribojimas, garso bangos viršija amplitudės ribas. Šiuo atveju tai reikštų, jog diktorius kalbėjo per garsiai, arba laikė labai arti savęs mikrofoną. Paskaičiuota amplitudinė reikšmė yra -1 ir 1, tačiau kaip grafike matosi signalo bangos išeina už ribų. Taigi, šį įrašą garantuotai atmetame, vartotojui parodoma klaida, kad jis per garsiai kalba, kad kalbėtų tyliau, arba patrauktų mikrofoną toliau nuo savęs.



Pav 13 Amplitudinis ribojimas – viršija normas

Pav. 14 taip pat yra užfiksuota amplitudinis ribojimas, šiuo atveju garso bangos arti nulio, vietomis susitapatina su triukšmu. Šiuo atveju tai reikštų, jog diktorius kalbėjo per tyliai, arba laikė toli nuo savęs mikrofoną. Taigi, šį

įrašą garantuotai atmetame, vartotojui parodoma klaida, kad jis per tyliai kalba, kad kalbėtų garsiau, arba patrauktų mikrofoną arčiau savęs.



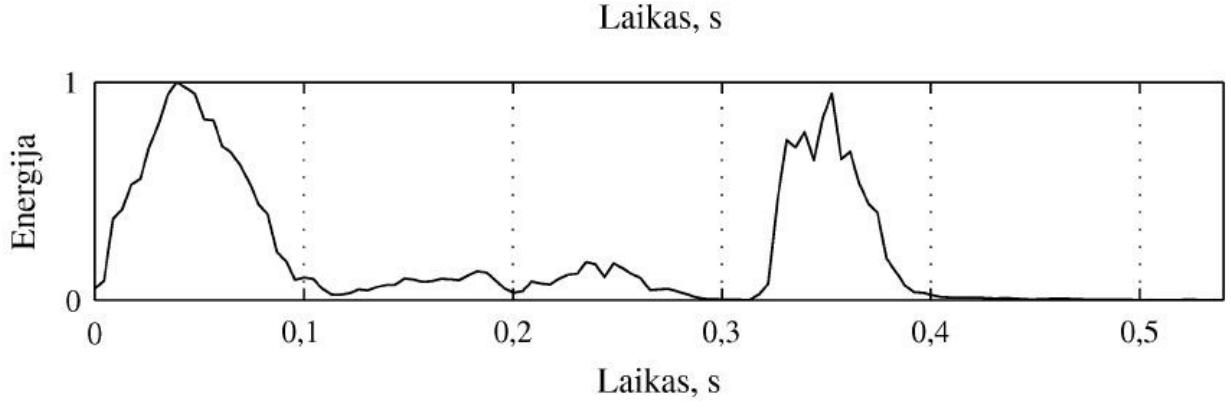
Pav 14 Amplitudinis ribojimas

Signalo energija dažniausiai naudojama kalbos signalų segmentavimui, kai reikia atskirti kalbos signalą nuo triukšmo, kadangi kalbos signalo energija yra didesnė už triukšmo energija. Fiksuoto ilgio diskretinio laiko signalo energija gali buti išreikšta:

$$E = \sum_{n=1}^{N_{sum}} s^2(n) \quad , \qquad (11)$$

čia N_{sum} - signalo ilgis atskaitomis.

Skaičiuojant frazės energiją, gaunama tokia kreivė:



Pav 15 Frazės energija

Tačiau siekiant norint padaryti sprendimą, skaičiuojama vidutinė visos frazės energijos reikšmė.

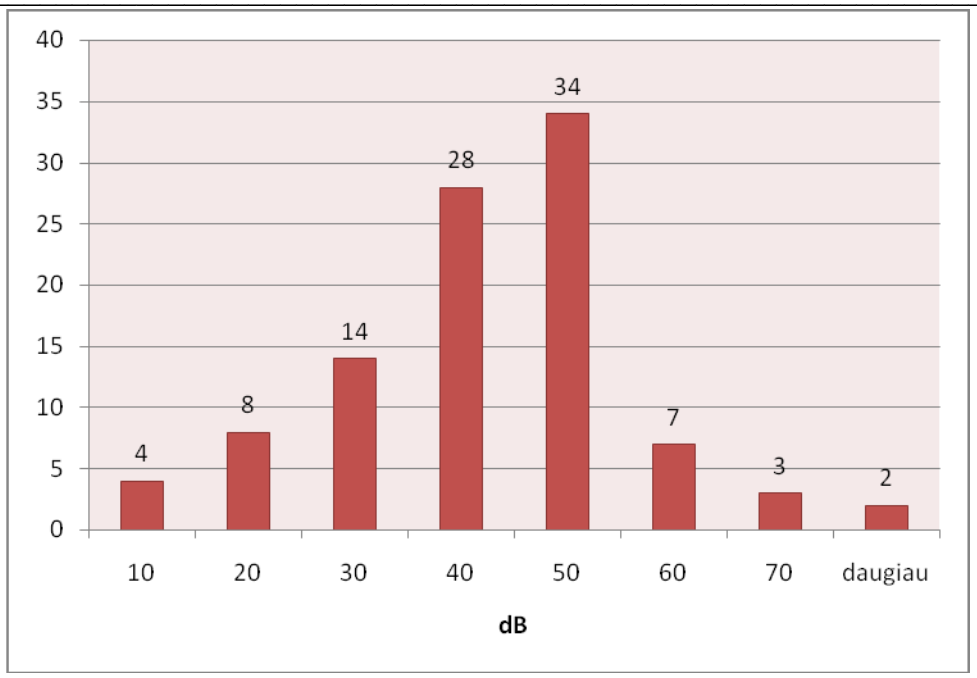
Paskaičiavus kiekvienos frazės vidutinę energijos reikšmę gautas toks skirstinys: 32 frazės, kurių vidutinė energija tarp 1 ir 0 ir 68 frazės, kurių vidutinė energija tarp 0 ir 1. Tokiu būdu galima daryti išvadą, kad nei viename faile nėra kritiškų. Visos frazės atitinka akustinę kokybę.

SNR (Signal-to-noise ratio) - signalas-triukšmas santykis - tai apibrėžiama kaip santykis signalo intensyvumo su triukšmo intensyvumu, išreikštu decibelais. Kaip triukšmas buvo traktuojamas įvestas naujas įrašas, kuriame įrašyta aplinkos fonas. Analizei naudojamas 50 ms su 25ms persidengimu langas (naudojama Hamming langas). Kiekvieno lango SNR skaičiuojama pagal sekančią formulę:

$$SNR_{(k)} = 10 * \log_{10}\left(\frac{P_{(k)signalo}}{P_{(k)triukšmo}}\right) \qquad (10)$$

kur $P_{(k)signalo}$ - k-tojo signalo energija, $P_{(k)triukšmo}$ - k-tojo lango triukšmo energija.

Frazės SNR lygis nustatomas apskaičiuojant visų frazę sudarančių langų SNR medianą. Gautas frazių SNR skirstinys pateikiamas 15 pav.



Pav 16 Frazių SNR skirstinys

Kalbos atpažinime 10 dB ir mažiau yra laikoma kritišku SNR lygiu, kadangi nuo šios ribos stipriai krinta automatinio kalbos atpažinimo tikslumas (Lippmann, 1997).

Įrašų, kurių SNR buvo iki 10dB rasta 4. Juos įrašė 3 moterys ir 1 vyras. Mažiausia reikšmė gauta - 0,448, kitos trys atitinkamai – 1,255, 7,751 ir 9,365. Frazės diktorių ištartos labai silpnai ir akustinis signalas yra nepakankamo lygio kad jį būtų galima priimti. Šios frazės buvo pažymėtos kaip nekokybiškos ir kad išvengti tolimesnių klaidų testuojant atmetami, traktuojama kritiškai silpno akustinio įrašo klaida.

Didžiausią dalį garso įrašų sudarė , kurių SNR lygis buvo tarp 30dB ir 60dB ir tik du diktoriai įrašė frazes, kurių SNR lygis viršino 70dB.

Eksperimentinės dalies rezultatai:

Į garsyno duomenų bazę buvo surinkta 100 garso įrašų, kuriuos įrašė 10 diktorių – 5 moterys ir 5 vyrai. Frazes sudarė po 10 žodžių – skaičiai nuo 0 iki 9 surikiuoti atsitiktine tvarka. Pritaikius algoritmą atlikta garsyno kokybinių charakteristikų analizė, pirmiausiai tikrinta įrašų trukmės, ko pasekoje atmesta 10 frazių, t,y, net 10% viso garsyno įrašo. Vien tiriant šią savybę pasiekta kokybiškesnio garsyno. Atlikus trukmės analizę, skaičiuota amplitudinis ribojimas, kurio metu rasta kritinių klaidų ir atmesta 4 frazės. Tai dar padidina garsyno kokybę. Sekančiais žingsniais skaičiuota frazių įrašų vidutinė energija. Atlikus šiuos skaičiavimus klaidų neaptikta, visi įrašai

atitiko keliamus reikalavimus. Galiausiai buvo ieškoma signalo-triukšmo santykis, kuriuo buvo siekiama atrasti frazes, kuriose fiksuojamas didelis triukšmas, kas iškraipo frazėje esančius duomenis. Šiuo atveju apskaičiavus SNR atmestos 4 frazės. Jose aptiktas per didelis triukšmas frazėje, taip pat paskaičiavus decibelais frazės nesiekė 10 dB normos, kad būtų toliau galima frazes naudoti.

Atlikto eksperimento metu surinktame garsyne patikrinus kokybines charakteristikas rasta 18 klaidų – tiek frazių neatitiko keliamų reikalavimų. Šis skaičius tikrai nemažas net 18% visų įrašo pasirodė nekokybiški ir nepriimti. Diktoriams šios klaidos buvo parodomos, įvardininamos priežastys kodėl įrašas gali būti nekokybiškas ir leidžiama diktoriui pasitaisyti. Gautas eksperimento rezultatas naudojant pasiūlytą algoritmą pasiteisino nekokybiškų signalų charakteristikų aptikime.

6 IŠVADOS, PASIŪLYMAI

Išanalizavus Lietuvoje bei pasaulyje atliktus tyrimus, galime drąsiai teigti, jog kuriant naujus kalbos garsynus įrašinėjant tiek frazės, tiek atskirai ištartus žodžius, būtinai reikia atlikti kokybinių kalbos signalų charakteristikų patikrinimą. Tikslus įrašų atrinkimas tiesiogiai susijęs su tiksliu žodžių atpažinimu. Daugeliu atvejų atliekant fonetinį nagrinėjimą, kalbos signalų kokybinių charakteristikų nustatymo algoritmas yra neatskiriamas komponentas atpažinimo sistemų, tačiau publikacijų susijusiu su šita užduotimi yra gana nedaug.

Lietuviškame LTDIGITS garsyne pastebėta kad dalis įrašų – nekokybiški – signalo fone juntami fono triukšmai. Tokie garsyne esantys neatitikimai gali sąlygoti klaidingus arba iškreiptus kalbos atpažinimo tyrimų rezultatus, todėl būtina taisyti esamas klaidas. Šis procesas atima daug laiko atsižvelgiant į tai, kad garsynai yra nuolat plečiami. Todėl buvo nuspręsta pasiūlyti šį algoritmą, kuris galėtų realiu laiku, tik įrašius garsus aptitikti neatitikimus, tokiu būdu būtų išvengiama nekokybiškų įrašų garsynuose.

Atlikus analizę pasiūlyta kalbos signalų kokybinių charakteristikų nustatymui skaičiuoti ir tirti reikalingai pirmine analizei geriausiai tiktų metodai kurie reikalautu minimaliausių laiko resursų ir būtų tiksliausi , tai: vidutinė įrašo trukmė, amplitudinis ribojimas, vidutinės energijos paskaičiavims ir signalo-triukšmo santykio skaičiavimas.

Šis darbas yra naujas dėl sukurtos realaus laiko informacinės sistemos leidžiančios surinkti ir plėsti garsyną bei tuo pat metu atrinkti kokybiškus kalbos įrašus.

Atlikus bandymus pasiekti tokie rezultatai:

Atlikto tyrimo rezultatus sunku objektyviai įvertinti naudojant informacine sistema surinktą garsyną, nes tokio pobūdžio tyrimas yra pirmasis. Kaip bebūtų garsyne rasta 18% nekokybiškų įrašų.

Šis tyrimas ir sukurtas algoritmas žymiai pagerina garsyno įrašų kokybę, kas labai svarbu sekančiuose kalbos garsų tyrimo etapuose.

Pasiūlymai:

Būtų tikslinga vykdyti garsyno plėtrą renkant ne tik ištisas frazes, bet ir atskirus žodžius pritaikant šį algoritmą, taip pat įvesti papildomus požymius, kurie dar patikslintų kokybę. Tokie garsynai labai palengvintų kalbos atpažinimo darbus.

7 Literatūra

1. Davis, E. E.; Selfride, O. G. Eyes and ears for computers. Prof. of IRE, 1962, t. 50, nr. 5, p. 1093-1101.
2. Rabiner, L.;Juang, B. H. Fundamentals of speeech recognition. PrenticeHall PTR, 1993. 507 p.
3. Davis, K.H.; Biddulph, R.; Balashek, S. Automatic recognition of spoken digits. The Journal o the Acoustical Society of America, 1952, t. 24, nr. 6, p. 637 – 642.
4. Olson, H. R; Belar, H. Phonetic typewriter. *IRE Transactions onAudio,* 1957, t. 5, nr. 4, p. 90-95
5. Hughes, G. W. The recognition of speech by machine. Technical Report 395, MIT Research Laboratory of Electronics, 1961
6. G.Tamulevičius, A.Lipeika Žodžio pradžios ir galo nustatymas atpažįstant atskirai sakomus žodžius (2004) [Interaktyvus]. Prieiga per internetą http://internet.ktu.lt/lt/mokslas/zurnalai/elektr/z58/Tamulevicius.pdf [žiūrėta 2008 11 21].
7. Rabiner, L. R.; Juang, B.-H. Fundamentals of speech recognition. Prentice-Hall, New Jersey, 1993. ISBN 0-13-285826-6
8. Cooley, J. W.; Tukey, J. W. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation,* 1965, t. 19, nr. 90, p. 297-301
9. Oppenheim, A. V; Schafer, R. W.; Stockham, T. G. Nonlinear filtering of multiplied and convolved signals. *Proceedings ofthe IEEE,* 1968, t. 56, nr. 8, p. 1264-1291
10. Hercher, M. B.; Cox, R. B. Source data entry using voice input. Iš *IEEE international conference on acoustics, speech, and signal processing, ICASSP'76,* 1976, t. 1, p. 190-193
11. Woodland, R; Evermann, G. HTK history.,
12. Deller, J. R.; Hansen, J. H. L.; Proakis, J. G. Discrete-time proces-sing of speech signals. IEEE Press, Piscataway, 2000. ISBN 0-7803-5386-
13. Rudžionis A., Rudžionis V., Žvinys P. *Lietuvių šnekamosios kalbos garsynas LTDIGITS: rezultatai ir problemos.* Informacinės technologijos 2000: konferencijos pranešimų medžiaga. Kaunas. Technologija. 2000. p. 162-166.
14. Balvočius B., Telksnys L. *Garsynų duomenų modeliai ir programinės įrangos architektūros.* Informacinės technologijos 2003. Konferencijos pranešimų medžiaga. Kaunas. Technologija. 2003. IX-1-8.
15. G. Tamulevičius, A. Lipeika. Žodžio pradžios ir galo nustatymas atpažįstant atskirai sakomus žodžius // Elektronika ir elektrotechnika. Kaunas: Technologija, 2005. – Nr. 2(58). – P. 61–64**.**
16. K.H. Davis; R. Biddulph; S. Balashek, *"Automtic Recognition of Spoken Digits"*, J.Acoust. Soc. Am., 24(6), 637-642, 1952.
17. H.F. Olson; H. Belar, *"Phonetic Typewriter"*, J. Acoust. Soc. Am., 28(6), 1072- 1081, 1956.
18. D.B. Fry, *"Theoretical Aspects of Mechanical Speech Recognition"*, and P. Denes, *"The Design and Operation of the Mechanical Speech Recognizer at University College London"*, J. British Inst. Radio Engr., 19(4), 211-229, 1959.
19. J.W. Forgie; C.D. Forgie, *"Results Obtained From a Vowel Recognition Computer Program"*, J. Acoust. Soc. Am., 31(11), 1480-1489, 1959.
20. T. Sakai; S. Doshita, *"The Phonetic Typewriter, Information Processing 1962"*, Proc. IFIP Congress, Munich, 1962.
21. T.B. Martin; A.L. Nelson; H.J. Zadell, *"Speech recognition by Feature Abstraction Techniques"*, Tech. Report AL-TDR-64-176, Air Force Avionics Lab, 1964.
22. T.K. Vintsyuk, *"Speech Discrimination by Dynamic Programming"*, Kibernetika, 4(2), 81-88, Jan.-Feb., 1968.

23. D.R. Reddy, *"An Approach to Computer Speech Recognition by Direct Analysis of the Speech Wave"*, Tech. Report No. C549, Computer Science Dept., Stanford Univ., September 1966.
24. H. Sakoe, *"Two Level DP Matching - A Dynamic Programming Based Pattern Matching Algorithm for Connected Word Recognition"*, IEEE Trans. Acoustics, Speech, Signal Proc., SSP-27, 588-595, December 1979.
25. C.S. Myers; L.R. Rabiner, *"A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition"*, IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-29, 284-297, April 1981.
26. C.H. Lee; L.R. Rabiner, *"A Frame Synchronous Network Search Algorithm for Connected Word Recognition"*, IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-37 (11), 1649-1658, November, 1989.
27. J. Ferguson, Ed., *"Hidden Markov Models for Speech "*, IDA, Princeton, NJ, 1980.
28. L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proc. IEEE, 77(2), 257-286, February 1989.
29. R.P. Lippmann, *"An Introduction to Computing with Neural Nets"*, IEEE ASSP Magazine, 4(2), 4-22, April 1987. 52
30. A. Weibel; T. Hanazawa; G. Hinton; K. Shikano; K. Lang, *"Phoneme Recognition Using Time-Delay Neural Networks"*, IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-37, 393-404, 1989.
31. Gražulevičius A., Gražulevičius G. (2005) Garsų atpažinimo galimybių tyrimas [Interaktyvus]. Prieiga per internetą http://www.ee.ktu.lt/journal/2005/8/Grazulevicius.pdf [žiūrėta 2008 11 19].
32. Mark Filipovič (2003) Atskirai pasakytų žodžių atpažinimo, naudojant neuroninius tinklus, tyrimas [Interaktyvus]. Prieiga per internetą http://internet.ktu.lt/lt/mokslas/konf05/konf_02/IT2003/Sekcija09.pdf [žiūrėta 2008 11 21].
33. B.Balvočius, L.Telksnys (2003) Garsynų duomenų modeliai programinės įrangos architekrūros [Interaktyvus]. Prieiga per internetą http://internet.ktu.lt/lt/mokslas/konf05/konf_02/IT2003/Sekcija09.pdf .
34. "Balso technologijų pasiekimai pasaulyje" [paskutinį kartą žiūrėta: 2010-01-20], prieiga per internetą http://www.likit.lt/all/balso_tech/03_pasiekimai.htm
35. Balso analizės technologijos [Interaktyvus] Prieiga per internetą: http://www.mikolsoft.com/lt/technologijos/
36. Kalbos technologijų mokymo Internete tyrimai. [Interaktyvus] Prieiga per internetą: http://kalba.mch.mii.lt/technologijos.htm
37. Bendroji fonetika. [Interaktyvus] Prieiga per internetą: http://www.umanitoba.ca/faculties/arts/linguistics/russell/138/course.htm.]
38. Rudžionis A., Rudžionis V., Žvinys P. (1999a) *Lietuvių kalbos signalų duomenų bazės LTDIGITS akustinės-fonetinės charakteristikos*. Baltų kalbų fonetikos ir akcentologijos problemos. St Peterburgas.
39. Rudžionis A., Rudžionis V., Žvinys P. (1999b) *Lietuvių kalbos garsynas.* Kompiuterininkų dienos- 99, I dalis. Devintosios mokslinės-praktinės kompiuterininkų konferencijos ir ketvirtosios mokyklinės informatikos mokslo darbai. Vilnius. Žara. p. 86-96.
40. Rudžionis A., Rudžionis V., Žvinys P. (2001) *Lietuvių kalbos garsynas.* Informacijos mokslai. ISSN 1392-0561. Nr. 17. p. 77-84.
41. Raškinis A., Raškinis G., Kazlauskienė A. (2004) *VDU bendrinės lietuvių šnekos*

universalus anotuotas garsynas. Informacinė technologijos -2004. Kaunas.Technologija.

42. Šilingas D. (2005) *Akustinių lietuvių šnekos atpažinimo modelių parinkimas, naudojant paslėptus Markovo modelius.* Daktaro disertacija. Kaunas.
43. Draiunys K. (2006) *Lietuvių šnekamosios kalbos segmentavimo ir fonetinio atpažinimo tyrimas naudojant LTDIGITS garsyno įrašus. Daktaro desertacija.Kaunas.*
44. „*Lietuvių šnekamosios kalbos duomenų bazių sudarymas*“ (KTU) [paskutinį kartą žiūrėta: 2010-04-25], prieiga per internetą http://www.ktu.lt/lt/apie_struktura/fakultetai/elektr/vald_tech_kat/mokslas.asp.