VILNIAUS UNIVERSITETAS
KAUNO HUMANITARINIS FAKULTETAS

INFORMATIKOS KATEDRA

Verslo informacijos sistemų studijų programa
Kodas 62603S108

PAULIUS ČIULADIS

MAGISTRO BAIGIAMASIS DARBAS

DIKTORIAUS ATPAŽINIMO TYRIMAS NAUDOJANT STACIONARIĄ FONEMOS DALĮ

Kaunas 2011

VILNIAUS UNIVERSITETAS
KAUNO HUMANITARINIS FAKULTETAS

INFORMATIKOS KATEDRA

PAULIUS ČIULADIS

MAGISTRO BAIGIAMASIS DARBAS

DIKTORIAUS ATPAŽINIMAS NAUDOJANT STACIONARIĄ FONEMOS DALĮ

Leidžiama ginti _____ _____ Magistrantas _____
(parašas)

Darbo vadovas _____
(parašas)

_____
(darbo vadovo mokslo laipsnis, mokslo pedagoginis vardas, vardas ir pavardė)

Darbo įteikimo data _____

Registracijos Nr. _____

Kaunas 2011

TURINYS

SANTRUMPŲ SĄRAŠAS

MDKK	Melo dažnių kepstriniai koeficientai (MFCC – Mel Frequency Cepstral Coefficients)
GFT	Greitoji Forjė transformacija (FFT - Fast Fourier Transform)
PMM	Paslėptieji Markovo modeliai (HMM- Hiden Markov Models)
TP	Tiesinis prognozavimas (LPC –Linear Predictive Coding)
PTP	Percepcinis tiesinis prognozavimas (PLP –Perceptual Linear Predictive)

LENTELIŲ SĄRAŠAS

PAVEIKSLŲ SĄRAŠAS

SANTRAUKA

ČIULADIS, Paulius (2011) Diktoriaus atpažinimo tyrimas naudojant stacionarią fonemos dalį. Magistro baigiamojo darbo ataskaita. Kaunas: Vilniaus universitetas, Kauno humanitarinis fakultetas, Informatikos katedra. 48psl

Magistro baigiamojo darbo tikslas - Nustatyti ar fonemos stacionarioji dalis turi būdingų konkrečiam diktoriui savybių, kurios leidžia jį identifikuoti arba nustatyti, kuriai grupei (vyrų ar moterų) priklauso. Siekiant tikslo darbe sprendžiami uzdaviniai: 1)Išanalizuoti „Diktoriaus atpažinimo naudojant stacionarią fonemos dalį, išskirtą kepsrinių koeficientų pagalba“ mokslinę temą; 2) Sudaryti ir paruošti diktorių atpažinimui reikalingų įrašų grupę; 3) Sukurti skirtingų diktorių fonemos staconarios dalies savybių išskyrimo ir palyginimo algoritmą; 4) Atlikti eksperimentą su paruoštais įrašais naudojant sukurtą algoritmą; 5) Išanalizuoti eksperimentų metu, pasiektus rezultatus.

Išskiriant fonemos stacionarią dalį naudotas segmentavimo metodas. Iš signalo fonemų segmentų atrinktų tolimesnei analizei buvo išskiriami Melų dažnių kepstriniai koeficientai (MDKK).

Atlikus numatytus uždavinius ir realizavus eksperimentą, diktoriaus atpažinimas pagal stacionarius fonemos fragmentus išskiriant kepstrinius koeficientus pirmo bandymo metu laido teisingi atpažinti diktorius vos **26,3 %** visų atvejų, o pagal lytį net **91,8 %** tikslumu. Pagal identiškame kontekste ištartų fonemų stacionarius fragmentus išskiriant kepstrinius koeficientus, antro bandymo metu laido teisingai atpažinti diktorius vos **8,5 %** visų atvejų, o pagal lytį **62,7%** tikslumu. Trečio bandymo rezultatai pagal atskirus kepstrinius požymius tesiekia maksimaliai 1,7% (5,11 karto mažesnis). Lyties atpažinimo tikslumas (antro bandymo metu pasiektas 62,7 %) atliekant palyginimą pagal atskirus požymius. Sumažėja nuo 1,09 karto (naudojant 9 požymį) iki 1,26 karto (naudojant 26 požymį).

ĮVADAS

Informacinių technologijų tobulinimas yra svarbiausias faktorius globalizacijos procese, todėl vystomi įvairūs tyrimai skirtingomis temomis. Viena iš naujausių ir labai svarbi mokslinė tema įvairiuose pasaulio šalyse yra „Kalbos ir kalbančiojo (diktoriaus) atpažinimas“. Kalbos atpažinimo srityje analizuojamas kalbos turinys, žodžių ir frazių prasmė. Šioje srityje yra pasiekti komerciniam naudojimui tinkami rezultatai, o diktoriaus atpažinimo srityje kol kas nėra atliktų tyrimų leidžiančių įvairiomis kintančiomis sąlygomis pasiekti patenkinamą tikslumą, nors bandymų tai padaryti netrūksta. Lietuvoje kalbos technologijų tyrimus atlieka Vilniaus universitetas (VU), Vilniaus Gedimino technikos universitetas (VGTU) Kauno technologijos universitetas (KTU), Vytauto didžiojo universitetas (VDU) bei Matematikos informatikos institutas (MII). Vis dėlto Lietuvių kalbos apdorojimo tyrimai yra kuklesni ir mažiau išvystyti. Pasiekus aukštą tikslumo lygį diktoriaus atpažinimą būtų galima naudoti kriminalistikos laboratorijose, duomenų apsaugai vietoj taip dažnai pamirštamų slaptažodžių identifikuojant vartotoją prieigai prie kompiuterinių sistemų (mobiliojoje bankininkystėje, elektroninėje prekyboje ar tiesiog kompiuterizuotoje darbo vietoje prisijungiant prie savo darbalaukio). Asmens indentifikavimą pagal balsą vartotojams paprasčiau naudoti, nei kitas biometrines atpažinimo technologijas, nes daugelis naudojasi telefonais, delninukais, kurie jau turi reikiamą duomenų priėmimo įrenginį - mikrofoną ir nereiktų nieko atskirai pirkti.

Darbo objektas

Fonetinis diktoriaus atpažinimas

Tyrimo tikslas:

Nustatyti ar fonemos stacionarioji dalis turi būdingų kongrečiam diktoriui savybių, kurios leidžia jį identifikuoti arba nustatyti, kuriai grupei (vyrų ar moterų) priklauso.

Siekiant tikslo sprendžiami uždaviniai:

1. Išanalizuoti „Diktoriaus atpažinimo naudojant stacionarią fonemos dalį, išskirtą kepsrinių koeficientų pagalba“ mokslinę temą;
2. Sudaryti ir paruošti diktorių atpažinimui reikalingų įrašų grupę;
3. Sukurti skirtingų diktorių fonemos staconarios dalies savybių išskyrimo ir palyginimo algoritmą;
4. Atlikti eksperimentą su paruoštais įrašais naudojant sukurtą algoritmą;
5. Išanalizuoti eksperimentų metu, gautus rezultatus.

Glaustas darbo struktūros aprašymas

Darbą sudaro trys pagrindinės dalys: 1) Diktoriaus atpažinimo apžvalga, 2) Teorinis diktoriaus atpažinimo tyrimo modelis, 3) Diktoriaus atpažinimas pagal stacionarius fonemos fragmentus

Pirmajame analitiniame skyriuje aprašoma teorinė temos medžiaga, mokslininkų atlikti tyrimai ir naudoti metodai diktoriaus ir kalbos atpažinimo srityje. Antrame skyriuje remiantis analitinio skyriaus apibendrinimu iškeliama prielaida ir aprašomas tyrimo, skirto prielaidai patikrinti, modelis. Pagal sukurtą modelį atliekamas eksperimentas, kuris aprašomas trečiajame darbo skyriuje.

Darbe naudoti literatūros šaltiniai

Teorinėje darbo dalyje naudotasi užsienio bei Lietuvos autorių moksliniais darbais, empiriniais tyrimais.

Tyrimo metodai

Tyrimo metu numatoma panaudoti eksperimento, skaitmeninių signalų apdorojimo, kalbos signalų analizės, modeliavimo, statistinius ir sintezės metodus.

Darbo struktūra ir apimtis

Magistro baigiamąjį darbą sudaro įvadas, 3 dalys, darbo išvados ir pasiūlymai, literatūros sąrašas. Aiškinamoji medžiaga aprašyta 37 puslapiuose. Literatūros sąraše pateiktos 22 nuorodos.

1. DIKTORIAUS ATPAŽINIMO APŽVALGA

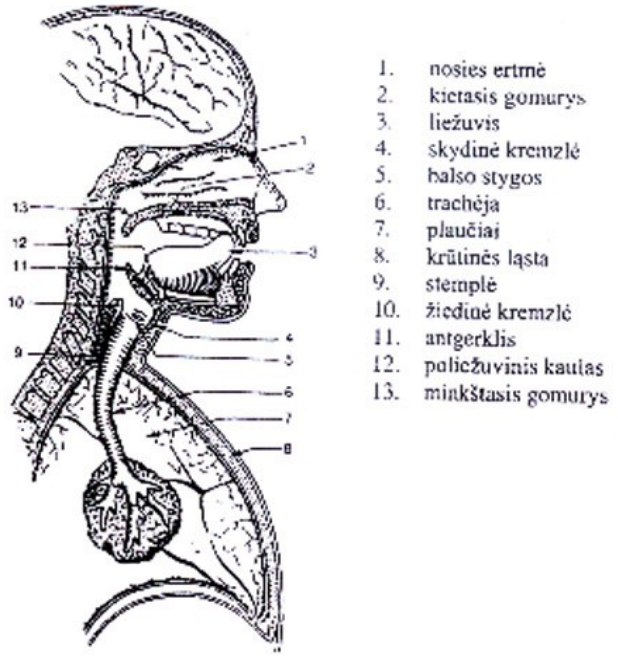
Žmogaus kalbą galima apibrėžti ir interpretuoti įvairiais požiūriais. Kaip savo darbe teigia K. Driaunys (2006), kalbą galima aprašyti informaciniu požiūriu kaip perduodamą informaciją ir kitu aspektu - kaip signalą arba akustinius virpesius, kurie perduoda tiek verbalinę tiek neverbalinę informaciją. Remiantis antruoju požiūriu yra atliekami tyrimai, kuriami algoritmai, kurie leistų kompiuterinėms technologijoms daryti tai kas įprasta žmogui: suvokti kas yra sakoma, identifikuoti kalbantįjį, išskirti ir susikoncentruoti informacijos gavimui į vieną kalbantį asmenį iš minios. Kalbančiojo atpažinimas yra biometrijos mokslo nagrinėjama žmogaus autentifikavimo užduotis.

Biometrija – tai vienareikšmis asmens identifikavimas pagal įvairias fiziologines arba elgesio savybes. Informacinėse technologijose biometrija naudojama indentifikavimo ir prieigos valdymo tikslais suteikiant prisijungusiam vartotojui suteikti atitinkamas teises pagal jo tapatybę. Fiziologinėms savybėms priskiriama žmogaus kūno formų ypatybės. Tai pirštų anspaudai, veido kontūrai, DNR, rankų ir delnų forma ir linijos, rainelės ar akies tinklainės savybės, kvapas. Žmogaus elgesio ypatybėms nagrinėjamoms biometrijoje priskiriama spausdinimo ritmas, eisena bei balsas.

1.1. Balso susidarymas ir charakteristikos

Žmogaus balsas – tai garsas sukuriamas plaučių, balso stygų ir burnos ertmes artikuliavimo mechanizmų, kai kalbama, dainuojama, juokiamasi, verkiama, ar klykiama-rėkiama.

Balso formavimą galimą suskirstyti į tris pagrindinius etapus: 1)plaučių, 2)balso stygų ir gerklų ir 3) artikuliavimo mechanizmo formuojamas balso ypatybes. Plaučiai sudaro reikiamą oro srautą norimam balso stiprumui (garsumui) išgauti. Oro srautas, tekėdamas per gerklas, suvirpina balso stygas. Balso stygos, tai vibruojantys vožtuvai, kurie suskaido oro srautą atitinkamais intervalais. „Jų sluoksnyje yra balso raištis ir balso raumuo. Moters balso klostės yra 18 – 20 mm, vyro 20 – 24 mm ilgio. Tarp balso klosčių yra balso plyšys. Balso klostės virpėjimas sukelia garsus“ (Medicinos enciklopedija 1 dalis, p. 101). Gerklų raumenys reguliuoja balso stygų ilgį ir įtempimą. Taip suformuojama atitinkamo aukščio ir tono garso banga. Toliau suformuota garso banga patenka į viršutinę žmogaus balso trakto dalį (virš gerklų), kurioje yra artikuliatoriai (liežuvis, gomurys, žandai, lūpos, dantys). Artikuliatoriai garsą suskaido, sustiprina arba susilpnina ir taip paverčia artikuliuotais kalbos garsais.



1 pav. Balso trakto struktūra

Šaltinis: DRIAUNYS, K. (2006) Lietuvių šnekamosios kalbos segmentavimo ir fonetinio atpažinimo tyrimas naudojant LTDIGITS garso įrašus, p. 13

Visuose žmogaus balso formavimo etapuose dalyvaujantys organai dėl savo skirtingų ypatybių suformuoja kiekvienam žmogui unikalų balsą, pagal kurį kitas žmogus savo klausos dėka gali identifikuoti pažįstamą kalbantįjį (prisiminti kam priklauso kalbančiojo balsas iš pažįstamų asmenų). Sukūrus galingas, didelius skaičiavimus ir operacijų kiekius galinčias atlikti kompiuterines sistemas, šį žmonėms ir kai kuriems kitiems žinduoliams būdingą gebėjimą siekiama perkelti ir į nająsias kompiuterines sistemas. Dėl šios priežasties vykdomi tyrimai bandant nustatyti balse užkoduotus požymius, pagal kuriuos žmogus geba identifikuoti ar priskirti tam tikrai grupei (vyras ar moteris, jaunas ar senas ir pan.) kalbantįjį.

1.2. Kalbos ir balso apdorojimo istorija

Kalbos garsų analizės pradžia, kaip teigia M.C. McDermott, T. Owen ir F.M. McDermott (1996), galima laikyti prieš beveik šimtmetį Alexander Melville Bell sukurtą tariamų žodžių vizualinį atvaizdavimą, kuris leido aprašyti žymiai daugiau informacijos nei bet kuriame tuometiniame žodyne naudojami kirčiavimo ženklai. Ši mokslininko sukurta sistema buvo pavadinta „matoma kalba“ (angl.: Visible Speech) (McDermott M.C. Owen T. ir McDermott F.M., 1996). JAV mokslininkai B.H. Juang ir Lawrence R. Rabiner (2005), pirmuosiais darbais kalbos apdorojimo srityje laiko 1930 m, ATT&T's Bell Labs pasiūlytą pirmąjį elektroninės kalbos

sintezatorių, kurį kompanija pristatė 1939 m. Pirmieji tyrimai buvo atliekami kariniams tikslams kalbos atpažinimo tematika, kai siekiama atpažinti kas yra sakoma, nesiekiant identifikuoti kalbančiojo (diktoriaus).

1940m ATT&T's Bell Labs mokslininkai sukūrė kalbos signalo atvaizdavimo būdų, kuris ir šiendien plačiai naudojamas – tai spectrogramos. Tai vaizdinis kalbos signalo atvaizdavimo atvaizduojantis tris parametrus: dažnį, intensyvumą ir laiką.

Kalbos atpažinimo srityje, kai 20a. 8-ame dešimtmetyje karinės struktūros išslaptino tyrimus atliktų tyrimų duomenis ir jie tapo prieinami privačių laboratorijų, buvo smarkiai pasistūmėta į priekį. Pasiekti atpažinimo rezultatai leido pradėti komercinių produktų gamybą. Vienas iš žinomiausių programinių produktų yra „Dragon Naturally Speaking" .

Remiantis JAV nacionalinės mokslo ir technologijų tarybos (angl.: National Science and Technology Council – NSTC) pateikiamais duomenimis, kalbančiojo (diktoriaus) atpažinimo srityje pirmieji darbai buvo paskelbti 1960 m. Tais metais, pasak Q. Jin (2007), Bell Labs specialistė Lawrence Kersta atliko vizualų spektrografinį balso signalo indentifikavimą. Šis tyrimas buvo inicijuotas New York (JAV) miesto policijos departamento, siekiant identifikuoti skambinančius anonimus pranešančius apie galimus sprogdinimus įvairiose mieto vietose. Nors vizualinis spektrografų palyginimas ir nebuvo tinkamas fizinių ir lingvistinių kalbos signalo variacijų įvertinimui, vis dėl to tai buvo didžiulis žingsnis, kad kalbančiojo atpažinimas taptų įmanomas ne tik žmogaus bet ir kompiuterinių sistemų pagalba. Kaip rašo savo darbe M.C. McDermott, T. Owen ir F.M. McDermott (1996), Lawrence G. Kersta per dvejus metus trukusius tyrimus sukūrė indentifikavimo metodą, kurio tikslumas tuometiniais skaičiavimais buvo 99,65%.

Studijuojant su informacinėmis technologijomis susijusios šaltinius daugiausia randama aprašomų diktoriaus atpažinimo tyrimų anglų, rusų, vokiečių, kinų kalboms. Mažiau paplitusioms kalboms, kurios nėra taip plačiai ir universaliai vartojamos bendraujant oficialiuose tarptautiniuose renginiuose diktoriaus atpažinimo tyrimų yra kur kas mažiau. Taip yra, nes: „1) pažangių šnekos technologijų produktų, skirtų mažiau paplitusioms kalboms, rinka dažniausiai yra nedidelė; 2) daugelis mažiau paplitusių kalbų turi daug sudėtingesnę struktūrą nei kitos, labiau paplitusios kalbos, kurios laikui bėgant buvo supaprastintos dėl jų plataus paplitimo pasaulyje ir naudojimo." (Filipovič M., 2005, p. 10)

Pirmus kalbos signalų tyrimus Lietuvoje inicijavo P. Kemėšis bei L. Telksnys 20a. septintojo dešimtmečio viduryje. Palaipsniui susiformavo kelios grupės atliekančios mokslinius tiriamuosius darbus kalbos technologijų srityje: Kauno Technologijos universitetas (KTU) ir Vilniaus universitetas (VU), Vytauto Didžiojo universitetas (VDU), bei Matematikos informatikos institutas (MII). Vykdant Lietuvių kalbos informacinėje visuomenėje 2000-2006 metų programą. 2000 metais pradėti sistemingi lietuvių kalbos automatinio atpažinimo tyrimai. Šiuose tyrimuose

dalyvavo MII, Vytauto Didžiojo universiteto ir Lietuvos teismo ekspertizės centro mokslininkai. Tyrimų metu buvo dirbama ne tik kalbos, bet ir diktoriaus atpažinimo srityse. Šiuo metu atlikti diktoriaus atpažinimo tyrimai lietuvių kalba orientuoti į kalbančiojo išskyrimą naudojant Povilo Treigio ir Antano Lipeikos (2006) aprašomą vidutinio atstumo tarp tiriamojo ir lyginamųjų kalbėtojų nustatymą ir balso trakto charakteristikas. Vis dėlto Lietuvių kalbos apdorojimo tyrimai yra kuklesni ir mažiau išvystyti nei anglų, ar kitų plačiai tarptautiniame lygmenyje naudojamų kalbų.

1.3. Kalbos atpažinimui naudojami požymiai ir jų išskyrimas

Siekient identifikuoti kas yra sakoma iš kalbos signalo turi būti išskiriami požymiai, būdingi tik analizuojamam signalo fragmento turiniui ir kuriuos matematinių bei statistinių algoritmų pagalba galima analizuoti informacinėse sistemose.

Vienas pirmųjų skaitmeniniai kalbos analizei pradėtų naudoti metodų buvo tiesinis prognozavimas, apie kurį savo darbe 1971m. rašė mokslininkai B.S.Atal ir S.L. Hanauer. Tai prognozės metodas leižiantis tiesinės kombinacijos pagalba pagal prieš tai buvusius kalbos signalus numatyti busimąjį. Šį metodą savo tyrimo metu 1993 m. naudojo L.R. Rabiner ir B.H.Juang mokslininkai nustatė, kad metodas geriausiai modeliuoja skardžiuosius kalbos signalus – fonemas.

A. Rudžionis, K Ratkevičius, T. Dumbliauskas, V. Rudžionis 2008m. sukurė sistemą „Balsas“, kuri skirta valdyti kompiuterį lietuviškomis komandomis. Komandas sudarė žodžiai arba žodžių junginiai. Mokslininkų sukurta sistema žožius skaidė į segmentus, kurie apytikriai lygūs fonemoms. Iš šių segmentų sistemoje „Balsas“ išskiriami požymių vektoriai, kurie naudojami palyginimuose atpažįstant diktoriaus pasakytas kompiuterinės sistemos valdymo komandas.

M.C. McDermott, T. Owen, F.M. McDermott 1996m savo darbe „Voice indentifikation:Aural/Spectrographic Method“ aprašė mišrų spectrogramų ir akustinės analizės panaudojimą nepriklausomam nuo teksto kalbos atpažinimui. Darbe nėra koncentrujamasi į atpažinimą pagal tam skalbos signalo dalis, tačiau aptariami faktai, kurie ištisinėje kalboje leidžia pagerinti arba neigiamai paveikia atpažinimą. Autoriai pažymi, kad labai svarbu kaip žodis ar jo dalis yra tariamas. Kaip pavyzdį, jie pateikia angliško priedėlio „The“ tarimą. Pasak autorių vieni paskutinį balsį taria - [e], o kiti - [i:]. Priklausomai nuo tarimo būtų gauti skirtingi tiek spektirnės, tiek akustinės analizės rezultatai todėl atpažinimas būtų netikslus. Taigi, itin svarbu sitemoje suvesti visus galimus tarimo variantus, arba ją kas kart apmokyti pagal kiekvieno diktoriaus tarimo manierą. Didelę įtaka atpažinimui, pasak mokslininkų, turi ir tai kaip tariami balsiai, tarpų tarp žodžių darymas (ar jie ryškūs ar žodžiai susilieja).

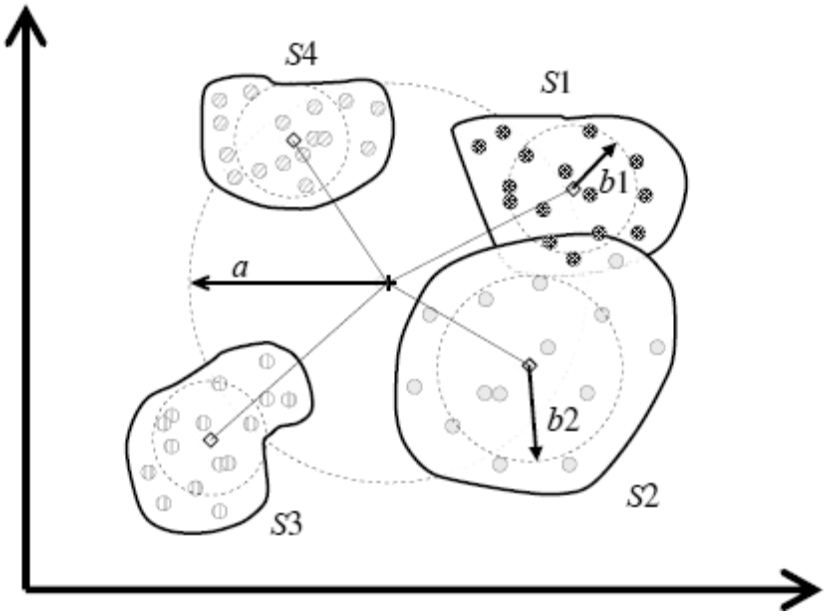
Eksperimente, kurį Kęstutis Driaunys aprašė savo disertacijoje „Lietuvių šnekamosios kalbos sekmentavimo ir fonetinio atpažinimo tyrimas naudojant LTDIGITS garsyno įrašus“, iš kalbos signalo išskiriami požymiai buvo melų dažnių kepstrinius koeficientui. Aprašomame eksperimente buvo atliktas žodžių sekmentavimą į fonemas. Siekient išskirti tik pasirinktoms fonemoms būdingus požymius, kurių neįtakoja gretimos žodžio fonemos, buvo iškirptos iš kalbos signalo stacionarios fonemų dalys.

Taigi, kalbos atpažinimo tyrimuose mokslininkai naudojasi Melų dažnių kepstriniais koeficientais, geriausiai leidžiainčiais išskirti kalbos signalo turinio savybes požymiais. Negrinėtuose tyrimuose reikšminiams požymiams išskirti orientuojamasi neretai ne į visą žodį, o į jo dalį, fonemas.

1.4. Diktoriaus atpažinimas

Diktorių atpažinimo srityje pasiekti kur kas kuklesni tikslumo rezultatai nei atpažįstant kalbos turinį. Naudojantis paieškos sistemomis pagal raktinius žodžius (kalbančiojo atpažinimas, diktoriaus atpažinimas; angl. Speaker recognition, voice recognition) pavyksta rasti tik bandomąsias tyrimų metu sukurtas nedideles, nemokamai prieinamas sistemas. Taip yra todėl, anot J. Kamarausko (2009), kad kol kas nėra vieningos teorijos kaip žmogus sugeba savo jutimais atskirti vieno kalbančiojo balsą nuo kito akustiniame lygmenyje. Taip pat nėra sukurti metodai, kaip įvertinti skirtingus balsus, kai pasakomos skirtingos frazės, esant skirtingoms aplinkoms, kuriose skiriasi triukšmo lygis, bei kaip įvertinti žmogaus balso pakitimus jam užkimus ar pasikeitus jo nuotaikai, nuovargiui. Žmogaus balsas yra greičiausiai kintanti biometrinė savybė lyginant su kitomis (pirštų anspaudais, delno linijomis, akies rainele, DNR). Kalbos atpažinime tai neturi lemiamos įtakos, nes analizuojami skirtingų žodžių ir frazių tarimo ypatumai. Diktoriaus atpažinime tai yra itin svarbus faktorius, kuris apsunkina tyrimus siekiančius žmogaus balsą pritaikyti efektyviam asmens indentifikavimui.

Kaip pavyzdys balso mutacijų pateikiamas H. Melin (2006) keturių žmonių balso energijos skirtumai priklausomai nuo kalbėjimo laiko (2 pav.). Vidutinė kalbančiojo (S1, S2, S3 arba S4) energija žymi energijų apskritimų centrus. Kaip matome iš paveikslėlio kalbančiųjų S1 ir S2 balso energijos tam tikru laiku persidengia. Visų keturių kalbančių energijų vidurkį žymi vektoriaus α pradžia.



2 pav. Keturių žmonių balso energijos skirtumai

Šaltinis: MELIN H (2006) Automatic speaker verification on site and by telephone: methods, applications and assessment, p.11

Dėl tokių požymių kaip diktoriaus balso energijos pasikeitimas, kurie konkrečiu laiku ir konkrečiomis sąlygomis gali identifikuoti ir leisti atpažinti diktorių, o pasikeitus aplinkybėms nebegali to padaryti, kuriami, tobulinami sujungiami įvairūs modeliai, algoritmai, išskiriami skirtingi požymiai. Šių tyrimų skeptikai teigia, kad mažesni požymių skirtumai yra tarp skirtingų asmenų balso, nei to paties žmogaus skirtingu laiku. Hagen M, Kamarauskas J. Jin Q. Savo darbuose aprašo kalbančiojo atpažinimui naudojamus Gausinių mišinių, vektorinio kvantavimo, atraminių vektorių, dirbtinių neuroninių tinklų modelius ir kt. Detaliau apie šiuos modelius aprašoma 1.6 skyriuje.(IŠSAMIAU)

1.5. Diktoriaus atpažinimo tyrimų uždavinių paskirtys

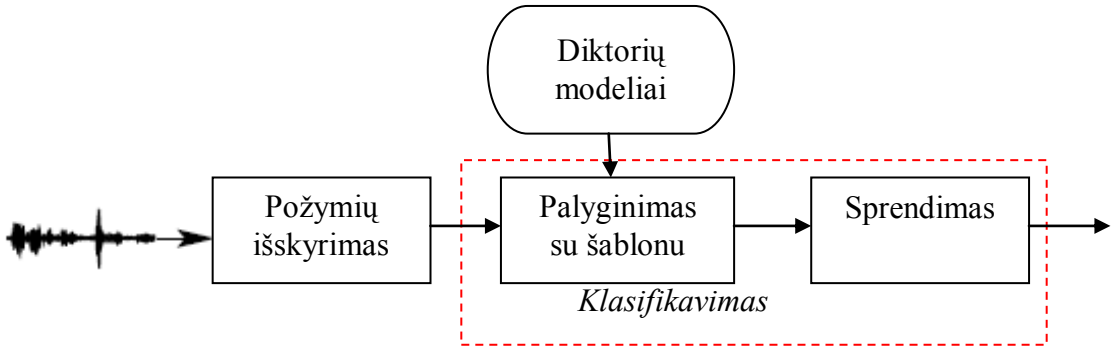
Nepaisant to, kad mokslininkų atlikti tyrimai diktoriaus atpažinimo srityje dar neleidžia pasiekti norimo tikslumo, atliktų darbų šioje srityje yra nemažai. Apžvelgdamas diktoriaus atpažinimo srityje atliktus tyrimus Q. Jin (2007) teigia, kad juos galima suskirstyti į tris grupes: kalbančiojo indentifikavimas, kalbančiojo verifikavimas ir kalbančiojo išskyrimas.

Kalbančiojo identifikavimas yra skirtas nustatyti ar testuojamas asmuo priklauso tam tikrai žinomai asmenų grupei, o jei jos yra kelios, tai kuriai.

Kalbančiojo verifikavimo (patvirtinimo) tikslas yra nustatyti ar testuojamas asmuo yra tas, kuriuo jis dedasi. Testuojamo asmens balsas palyginamas su sistemoje saugomu balso pavyzdžiu ir pagal tam tikrus požymius sulyginamas.

Kalbančiojo išskyrimas yra naudojamas tam, kad būtų išskirtas vienas kalbantysis iš kelių tuo metu kalbančių žmonių arba foninio triukšmo. H. Melin (2006) savo darbe tai įvardija kaip kalbančiojo sekimą (angl.: Speaker traking). Ši technologija glaudžiai siejasi ir su kalbos atpažinimu tik naudojami skirtingi algoritmai ir požymiai. Kalbos atpažinime tai yra reikšminio signalo išskyrimas iš foninio triukšmo.

Kaip ir daugelyje biometrinėmis savybėmis pagrįstuose žmogaus ar jo savybių (nuotaikų, sveikatos sutrikimų) indentifikavimo tyrimuose, diktoriaus atpažinime, kaip teigia Q. Jin (2007) galima išskirti du pagrindinius etapus: 1) požymių išskyrimą ir 2) klasifikavimą (3 pav.).



3 pav. Apibendrinta kalbančiojo atpažinimo sistema

Šaltinis: sukurta autoriaus pagal JIN, Q (2007) Robust Speaker Recognition., p. 5.

Klasifikavimas atliekamas pirmiausia išskirtus požymius palyginant su kalbančiojo modeliu palyginimo su šablonu modulyje. Kalbančiojo modelis sudaromas dažniausiai apmokymų metu, įrašant ne vieną atskiro kalbančiojo įrašą ir pagal pasitelktus algoritmus išskiriant šabloninius požymius. Gautas palyginimo rezultatas siunčiamas į sprendimo priėmimo modulį, kuriame statistinių ar deterministinių metodų pagalba priimamas sprendimas apie kalbančiojo tapatybę, arba ar jis priklauso tam tikrai grupei.

Toliau šiame skyriuje bus apžvelgiami atliktuose kalbančiojo atpažinimo tyrimuose ir eksperimentuose mokslininkų naudojami ir tobulinami algoritmai ir metodai.

1.6. Diktoriaus atpažinimui naudojami modeliai ir algoritmai

Kalbančiojo identifikavimas sudėtingas ir daugiafunkcinis procesas. Žmogus kalbantį asmenį atpažįsta priėmęs kalbos signalą klausos organų pagalba ir apdorojęs jį smegenyse vykstančių procesų metu. Mokslininkai aiškinasi, kokie procesai vyksta žmonių ar gyvūnų gebančių, atpažinti skleidžiančius garsus objektus, smegenyse. Pagal šiuos procesus siekia sukurti

jiems analogišką rezultatą duodančius matematinius ir statistinius algoritmus, kurie leistų išanalizuoti ir pagal analizės rezultatus atpažinti kalbantįjį kompiuterinėms sistemoms.

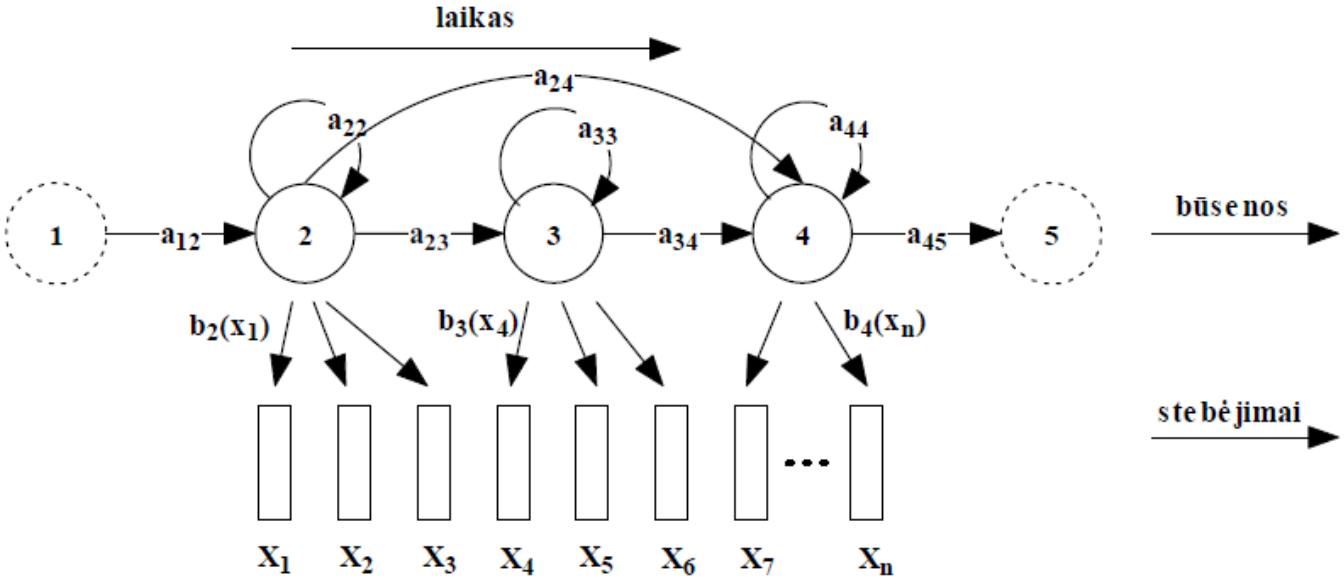
1.6.1. Paslėpti Markovo modeliai

Paslėptieji Markovo modeliai (PMM) (angl.: Hidden Markov Models –HMM), pirmiausia pradėti naudoti kalbos atpažinime. PMM pagrindas rusų matematiko Andrey Markov sukurtas stochastinis procesas kuris vėliau tapo žinomas kaip „Markovo grandinė. Modelio principinė schema yra pateikta 4 pav.

PMM pagrindas – Markovo grandinė. Atsitiktinių dydžių seka $X_1,X_2,...,X_n,...$ vadiname Markovo grandine, jeigu:

$$P(X_{n+} = z \mid X_n = z_n, X_{n-} = z_{n-}, ..., X_1 = z_1) = P(X_{n+} = z \mid X_n = z_n) \quad (1)$$

Apskritai Markovo procesas – tai būsenų SS =ss0,ss1,..,ssT seka. 4 pav. pateiktas 5 būsenų PMM modelis. Kiekviena būsena atvaizduota apskritimu, kurio viduje įrašytas indekso numeris. Visi perėjimai tarp būsenų (tikimybės a_{ij}) bei pasilikimo toje pačioje būsenoje tikimybės (a_{ii}) pavaizduotos rodyklėmis. Stebėjimų priklausomybės vienai iš būsenų žymimos tiesia linija (išėjimo tikimybės $b_j(x_t)$).



4 pav. Paslėptų Markovo modelių struktūra

Šaltinis: Driaunys, Kęstutis (2000) *Lietuvių šnekamosios kalbos segmentavimo ir fonetinio atpažinimo tyrimas naudojant LTDIGITS garso įrašus.* p. 39.

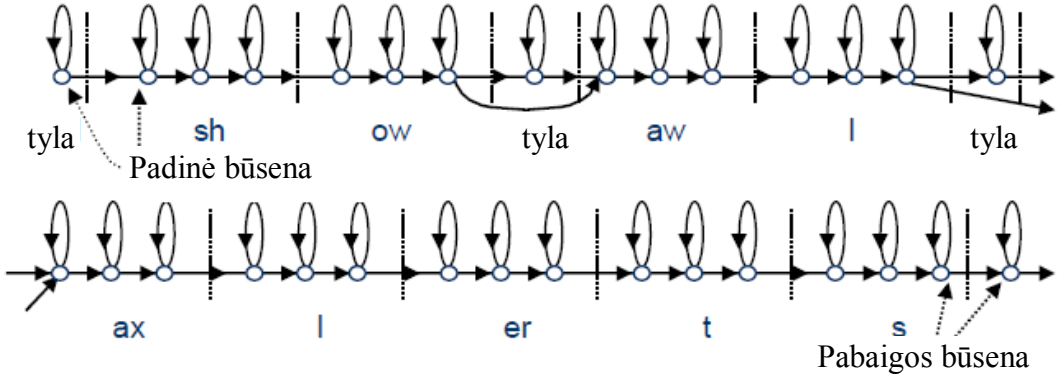
Vienas iš paslėptus Markovo modelius apibūdinančių parametrų yra grandinėlės būsenų skaičius. Egzistuoja nemažai PMM topologijų, tačiau dažniausiai yra naudojama taip vadinamoji „iš kairės į dešinę trijų būsenų“ PMM topologija, kaip parodyta 5 pav. Trys būsenos reiškia, kad modelis, esantis būsenoje 2, gali pereiti į tą pačią būseną 2 arba į būsenas 3, arba 4 būseną.

Kitas parametras, apibūdinantis PMM vadinamosios perėjimų į kitą būseną tikimybės, kurios apibrėžiamos taip:

$$a_{ij} = P(SS_t = i \mid SS_{t-} =), (i, j = ,2,..., n). \quad (2)$$

Ši išraiška taip pat gali būti užrašyta kaip matrica A ={aij}. Taigi konkrečiu laiko momentu *t* egzistuoja tam tikra tikimybė pereiti į bet kurią iš sekančių dviejų būsenų.

Kalbos atpažinime Markovo grandinės būsenas atitinka fonemos, kurios yra kalbos nedalomi vienetai.



5 pav. Baigtinių būsenų Markovo grandinė išreiškianti frazę „Show all alerts“

JUANG B.H. ir RABINER L. R. (2005) *Automatic Speech Recognition – A Brief History of the Technology Development.* 14 p

Kalbančiojo atpažinime Markovo modeliai, kurie buvo sukurti ir panaudoti kalbos atpažinime, naudojami nuo teksto priklausančiame atpažinime. Tokio atpažinimo metu diktoriai turi ištarti standartinę numatytą frazę. Kalbančiojo atpažinime PMM buvo modifikuoti ir sujungti autoregresiniais metodais pritaikant juos nuo teksto nepriklausančiam diktorių atpažinimui.

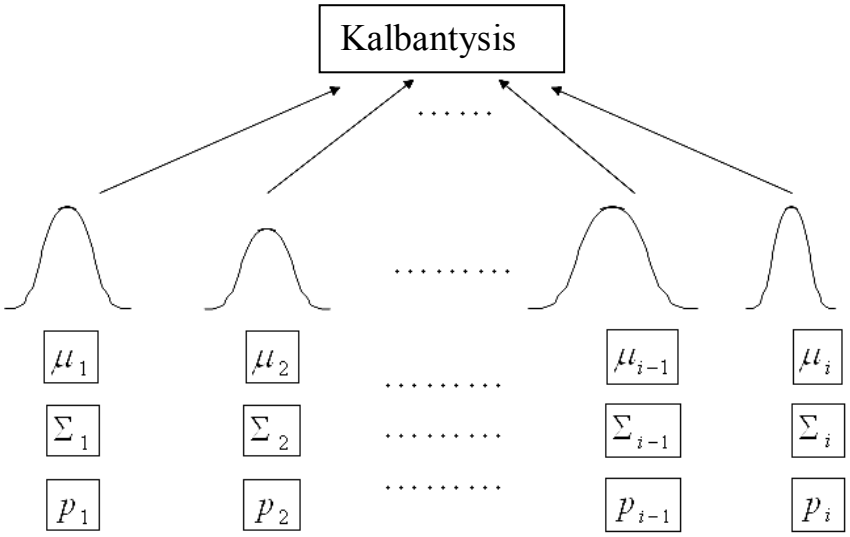
1.6.2. Gauso mišinių modelis.

Šiai dienai, Gauso mišinių modelis (angl.: Gaussian Mixture Model - GMM) ya vienas iš naujausių modelių naudojamų diktoriaus atpažinimui. Jį aprašė ir pristatė savo darbuose 1995 m R. Rose.

Šis metodas remiasi statistiniu pavaizdavimu kaip kalbėtojas taria garsus. Kiekvieno kalbančiojo kalbos signalas gali būti išreikštas Gausiniu parametrų mišiniu:

$$\lambda = \{p_i, \mu_i, \Sigma_i\} \quad i = 1,2,...n; M \quad (3)$$

Gausinių mišinių principinė schema pateikiama 6 paveiksle.



6 pav. Gausinio mišinių modelio principinė schema

Šaltinis: sukurta autoriaus pagal LIU, R. Speech signal processing. Lecture 16, sk. 25.

Dažniausiai Gausinių parametrų mišinys išreiškiamas Gauso pasiskirstymo tankio funkcija:

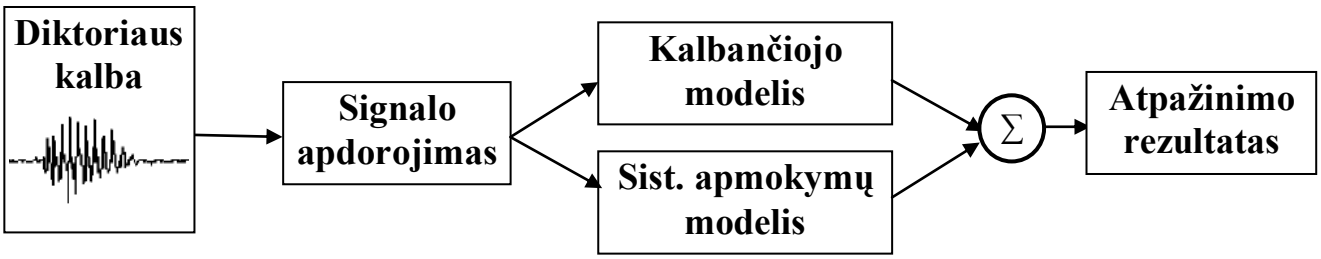
$$\lambda = \frac{1}{(\sqrt{2}\pi)^d \sqrt{|\Sigma_i|}} e^{-\frac{(x-\mu)' \Sigma^{-} (x-\mu)}{2}} \quad \textbf{(4)}$$

kur i=1, 2,..., M, i=1, 2,...,p_i –statistinė kalbėtojo tariamo garso požymių tikimybė (p_i >0, $\sum_{i=}^{n} p_i =$) M – mišinių svorių koeficientai, μ – požymių vidurkio vektorius, $\sum_i$ – kovariacinė matrica.

Galimi du Gausinių mišinių panaudojimo variantai. Pirmasis, kai individualūs komponenčių tankiai gali būti išreiškiami tam tikru akustiniu rinkiniu. Pavyzdžiui tokie rinkiniai gali būti kalbėtojo tariami balsiai, nosiniai garsai, pučiamieji priebalsiai. Šie akustiniai rinkiniai leidžia spręsti apie individualias diktoriaus balso trakto savybes.

Antrasis Gausinių mišinių panaudojimo atvejis, kai Gauso funkcijų tiesinė kombinacija gali atvaizduoti didelę pavyzdžių pasiskirstymų aibę.

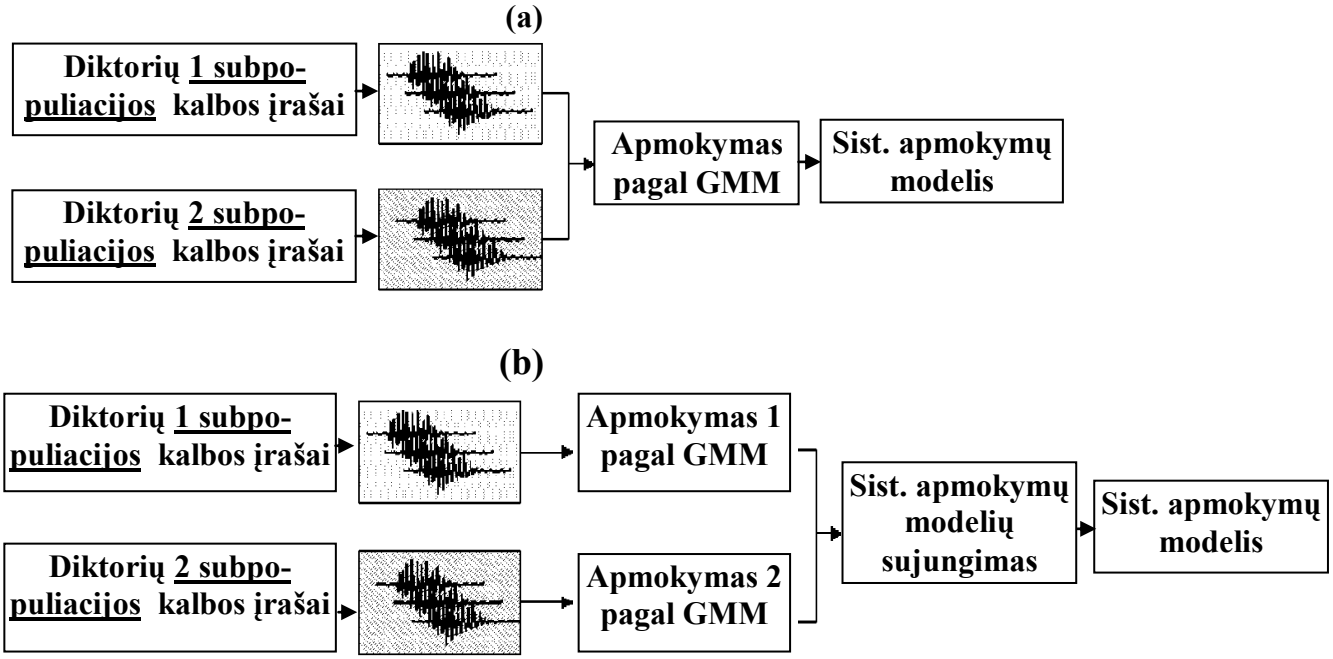
Bet kuriuo atveju naudojant Gausinius mišinius tikslas yra rasti kalbančiojo modelį, turintį stebimos požymių sekos maksimalią tikimybę ir gauti tikliausią atpažinimo rezultatą. Sitemoje sukaupiamas diktorių kalbos įrašų rinkinys, iš kurio sudaromas sitemos apmokymų modelis (angl.: Background model) (7 pav.)



7 pav. Kalbančiojo atpažinimo principinė schema naudojant Gausinių mišinių modelį

Šaltinis: sukurta autoriaus pagal KING, Josh (2008) *Speaker Verification Using Adaptad Gassian Mixture Model. Presentation.* Sk 11

Kitaip tariant, atliekamas sistemos apmokymas. Sitemos apmokymų modelis gali būti sudarytas pagal vieną iš 8 paveiksle pavaizduotų GMM sistemos apmokymo variantų.



8 pav. Du gausinių mišinių sistemos apmokymo būdai

Šaltinis: KING, Josh (2008) *Speaker Verification Using Adaptad Gassian Mixture Model. Presentation.* Sk 29

Kalbančiojo modelis lyginamas su apmokymo modelio duomenimis. „Kalbančiojo modelis“ - tai testuojamo kalbančiojo modelis, tikrinant nulinę hipotezę. Sistemos naudojimo metu atsiradus naujiems jos naudotojams (diktoriams) sistemos apmokymų modelis papildomas naujais duomenimis ir atliekamas požymių perskaičiavimas. (KING, J., 2008)

Turint gausinių mišinių požymiais peremtą sistemą ir tiriant kalbos įrašą Y, gatą iš galimai jį pasakiusio diktoriaus S, sitemoje įrašas nagrinėjamas pagal dvi hipotezes: 1**)** Y yra gautas iš S (nulinė hipotezė - H0); **2)** Y nėra gautas iš (atvirkštinė nulinei hipotezei H1) S. Paskaičiavus hipotezių tikimybes įvertinamas jų logaritmų skirtumas. (KING, J., 2008).

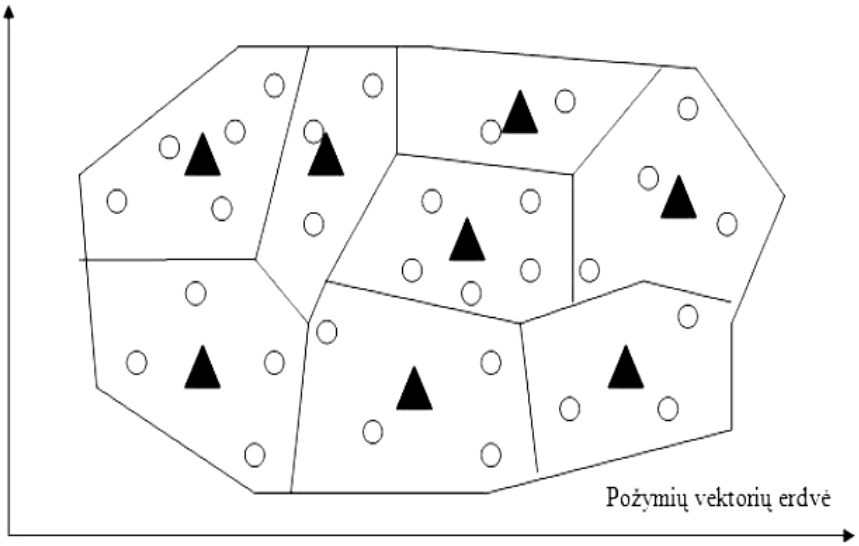
Kaip rašoma J. Kamarausko disertacijoje, turint S kalbėtojų S={1,2...,S}, apibūdinami atitinkamais GMM λ_1, λ_2,..., λ_s identifikavimo metu algoritmo tikslas yra rasti kalbančiojo modelį, turintį požymių sekos maksimalią aposteriorinę (paremtą tyrimu) tikimybę:

$$\hat{S} = \arg\max_{1 \le k \le S} \Pr(\lambda_k | k) = \arg\max_{1 \le k \le S} \frac{p(X | \lambda_k) \Pr(\lambda_k)}{p(X)} \quad (5)$$

Didžiausią tikimybę turintis kolbančiojo modelis laikoma kad yra sukurtas kalbančiojo S_n su kurio λ_n buvo gauta maksimali tikimybė.

1.6.3. Vektorinio kvantavimo modelis

Vektorinio kvantavimo (angl.: vector quantization – VQ) modelis naudojamas nepriklausančiame nuo ištarto teksto diktoriaus atpažinime. Tai gan nesunkiai realizuojamas algoritmas, pasižymintis greitu diktorių ištartų frazių palyginimu. Vektorinio kvantavimo esmė siekiant palyginti dvi frazes, visi kalbos signalo požymiai atvaizduojami mažesniu, iš anksto numatytu, požymių skaičiumi. Tam pradinių požymių aibė padalinama į pasirinktą skaičių grupių (angl. cluster) ir kiekviena sritis atvaizduojama požymių vidurkiu (centroidu).



9 pav. Galutinis požymių vektorių sugrupavimas ir centroidų suradimas

Šaltinis: KAMARAUSKAS, J (2009) Asmens pažinimas pagal balsą., p.73

Kaip savo darbe rašo J. Kamarauskas (2009), visi centroidai sudaro kodinę knygą, kuri ir yra kalbėtojo šablonas-modelis. Taigi, tokiu būdu sumažinamas duomenų kiekis, neprarandant esminės informacijos. Kodinių knygų kūrimui naudojami šie algoritmai:

• Apibendrintas Lloydo algoritmas (angl. GLA – Generalized Lloyd algorithm).

• Save organizuojantys žemėlapiai (angl. SOM – Self-organizing maps).
• Poriškai artimiausias kaimynas (angl. PNN – Pair-wise nearest neighbor).
• Pasikartojanti dalinimo technika, SPLIT.
• Atsitiktinė vietinė paieška (angl. RLS – Randomized local search).

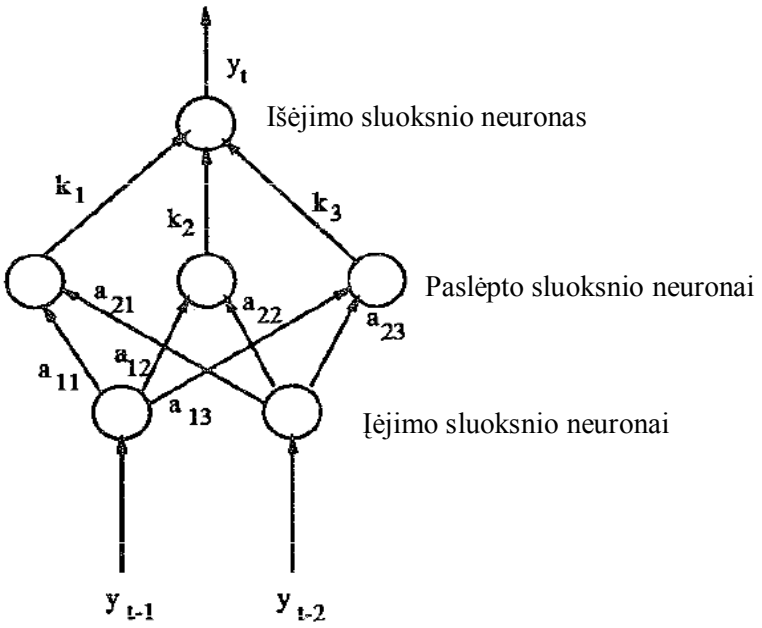
Kalbančiojo atpažinimo metu nustatoma vidutinė kvadratinė klaida lyginant testuojamojo balso signalo požymius su visų kalbėtojų kodinėmis knygomis. Asmuo, su kurio kodinės knygos pavidalu saugomais duomenimis palyginus testuojamo signalo požymius gaunama mažiausia paklaida, gali būti grąžinamas kaip indentifikavimo rezultatas. Ši algoritmą naudojo Povilas Treigys ir Antanas Lipeikas atlikdami savo mokslinį tyrimą, kurio tikslas buvo išskirti kalbantįjį iš kalbančiųjų minios.

1.6.4. Atraminių vektorių mašinos

Tai vienas iš naujesnių diktoriaus atpažinimui naudojamų metodų, kurį pasiūlė Vapnik 1995 metais. Modelis pirmiausiai buvo išbandytas veidų atpažinimo technologijose ir tik vėliau pastebėjus jo galimybes imtas taikyti ir diktoriaus atpažinimui. Modelis grindžiamas mokymo duomenų suskirstymu į dvi klases. O šios dvi klasės atskiriamos hiperplokštuma. Plokštuma sukuriama tarp dviejų artimiausių mokymo duomenų reikšmių iš skirtingų klasių. Modelis charakterizuojamas jo branduolio funkcija. Atraminių vektorių mašinos reikalauja, kad mokymo ir atpažinimo pavyzdžiai būtų atstovaujami fiksuoto ilgio vektoriais.

1.6.5. Dirbtiniai neuroniniai tinklai

Dirbtiniai neuroniniai tinklai - tai bene lanksčiausias tyrimams naudojamas metodas tinkantis įvairioms sritims atitinkamai juos apmokius. Neuroniniai tinklai gali būti naudojami visur, kur tiriamų įėjimų kiekis yra didesnis nei mūsų smegenys sugeba apdoroti ir suvokti. Ne išimtis ir kalbančiojo atpažinimas, kai naudojamų atpažinimui požymių yra ne venas ir ne du. Dirbtinio neuroninio tinklo su vienu paslėtu sluoksniu schema pateikiama 10 paveiksle.



10 pav. Neuroninis tinklas su vienu paslėptu sluoksniu

Šaltinis: sukurta autoriaus pagal LOVELL, C. Brian ir TSOI, Ah. Chung *Speacker Verification Using Artificial Neural Networks* p.300

Dirbtinio neurono išėjimas formuojamas kiekvieną įėjimo reikšmę padauginant iš atitinkamo koeficiento ir viską sumuojant. Iš gautos sumos atimama neurono paleidimo slenkstinė riba. Kai ta riba pasiekiama, neuronas aktyvuojamas ir paveikiamas aktyvavimo funkcijos. Neuronų grupės sprendžiančios vieną problemą sudaro sluoksnį. Neuroniniame tinkle sluoksniai siejasi tarpusavyje. Vidiniai neuroninio tinklo sluoksniai vadinami paslėptaisiais. Sluoksnis, į kurį paduodami įėjimo kintamieji, vadinamas įėjimo sluoksniu, galutinį rezultatą suformuojantys neuronai – išėjimo sluoksniu. Kiekviename neurone (n_{ij}) yra realizuotos funkcijos, kurios jį sužadinus įėjimų duomenis apdoroja ir paverčia išėjimu.

Norint, kad dirbtinis neuroninis tinklas funkcionuotų, jį reikia tinkamai apmokyti. Mokymas vykdomas nuosekliai paduodant įėjimo duomenis į tinklą ir derinant jungiančių neuronus komponentų svorius (koeficientus). Jei mokymo metu žinoma, kokie turimi gauti išėjimo duomenys su atitinkamais įėjimais, toks mokymas vadinamas su mokytoju. Jo metu mokytojas keičia komponentų svorius tol kol su atitinkamais įėjimais gaunami reikiami išėjimai (pasiekiama minimali arba užsibrėžta paklaida tarp sumodeliuotų reikšmių ir turimų gaut realių). Tinklai, kai nežinomi išėjimo duomenys vadinami save organizuojantys ir jie patys atlieka apsimokymą remiantis mazgo laimėtojo principu. „Išėjimo sluoksnio neuronai varžosi per rekurentinius ryšius siųsdami sau teigiamus signalus, o kaimyniniams šio sluoksnio neuronams – neigiamus, bei pagal pasirinktas matematines funkcijas keičiant neuronų svorinius koeficientus ir vykdant apmokymo iteracijas. Nusistovėjus pusiausvyrai, išėjime lieka vienas aktyvus neuronas“ (Kamarauskas, J.

2009, p 76). Save organizuojantys tinklai apmokomi tol kol pasiekiama nurodyta paklaidos riba arba įvykdomas užduotas apmokymo iteracijų skaičius.

1.7. Diktoriaus atpažinimui naudojami požymiai ir jų išskyrimas

Tobulinant kalbos ir kalbančiojo atpažinimo technologijas per visus tyrimų šioje srityje metus buvo sukurta ne vienas ir ne du matematiškai pagrįsti būdai kaip išskirti kalbos savybes ar jų rinkinius, kurių reikšmės būtų unikalios konkrečiam kalbančiajam. Tie matematiniai būdai tai - įvairūs skaičiavimai leidžiantys kalbos signalą išreikšti skaitinėmis išraiškomis (požymiais). Per jau keletą dešimtmečių trunkančius tyrimus kalbos atpažinimo srityje viso pasaulio mokslininkai sukūrė eilę požymių, kurie taikomi diktoriaus atpažinime segmentuojant signalą, nustatant žodžių, fonemų ar skiemenų tarimo ypatybes, bei identifikuojant kalbantįjį.

1.7.1. Tiesinis prognozavimas

Vienas pirmųjų skaitmeninei kalbos signalų analizei pradėtų naudoti metodų – tiesinis prognozavimas (Atal, Hanauer, 1971), kurio metu skaičiuojami požymiai – tiesinės prognozės koeficientai – gali būti naudojami kalbos signalų analizei kalbos atpažinimo sistemose. Tiesinis prognozavimo metodas tiksliai prognozuoja skardžius kalbos signalus, tačiau nėra efektyvus apdorojant dusliuosius. Pagrindinė tiesinio prognozavimo metodo idėja yra ta, kad kalbos signalas *Y* laiko momentu *i*, gali būti išreikštas prieš tai esančių *n* kalbos signalo reikšmių pagalba:

$$y(i) \approx \text{ı}_1 y(i - \ ) + \text{ı}_2 (i - \text{ː}) + \ .. + \text{ı}_n y(i - \text{ı}) \qquad \textbf{(6)}$$

Koeficientai a_1, a_2, a_n kalbos signalo lange yra laikomi pastoviais.

„Su prognoze susijusi paklaida vadinama prognozės paklaida arba žadinimo signalu“. (Kamarauskas J. 2009 p. 49.)

$$e(n) = \mathrm{s}[\mathrm{n}] - \hat{\mathrm{s}}[n] = \mathrm{s}[\mathrm{n}] - \sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{a}_{\mathrm{i}} s[n-i] \qquad \textbf{(7)}$$

Tiesinės prognozės modelio koeficientai gali būti rasti sprendžiant matricinę lygtį:

R=ar **(8)**

Matrica r - autokoreliacijos funkcijos koeficientai, matrica a – tiesinės prognozės koeficientai.

1.7.2. Spektrinė analizė

Spektrinė analizė yra kalbos signalo apdorojimo metodas leidžiantis išskirti akustines signalo charakteristikas skirtingos dažnių juostose. Dėl šios savybės ši požymių išskyrimo metodika yra tinkama kalbančiojo identifikavimui.

Spektrinė analizė remiasi prielaida, kad kalbos signalą galima suskaidyti į reikiamai trumpus intervalus, kuriuose signalas yra stacionarus arba kvazistacionarus. Trumpo intervalo signalo spektrinei analizei naudojami du metodai: filtrų bankų metodas ir greitosios Furjė transformacijos algoritmas. Filtrų banko naudojimo esmė – kalbos signalas perleidžiamas per tam tikrą apibrėžtą filtrų kiekį Q turintį filtrų banką, kuris perdengia aktualaus dažnio signalo diapazoną . Kiekvieno q-tojo filtro, kurio centrinis dažnis f_q, išėjimas yra tam filtrui tenkančios kalbos signalo dalies energija. Visų trumpo intervalo, kuriame kalbos signalas yra stacionarus, filtrų energijos sudaro trumpalaikį signalo spektrą. Filtrai gali būti nebūtinai vienodo dydžio. Pagal žmogaus klausos savybes žemuose dažniuose tikslinga naudoti mažesnio pločio filtrų juostas, nei aukštuose dažniuose.

Greitosios Furjė transformacijos algoritmas yra dažniau naudojamas, nes tai matematinis algoritmas, kuris paprasčiau realizuojamas kompiuterinėse sistemose nei filtrų bankas). Furjė transformacijos naudojamos ir Kepstrinėje analizėje.

1.7.3. Kepstrinė analizė

Pastaruosius 10 metų kalbos ir kalbančiojo atpažinimui labai plačiai naudojamas Melų dažnių kepstro MDK (Mel Frequency Cepstrum) požymiai. Šie požymiai turi keltą privalumų lyginant su kitais.

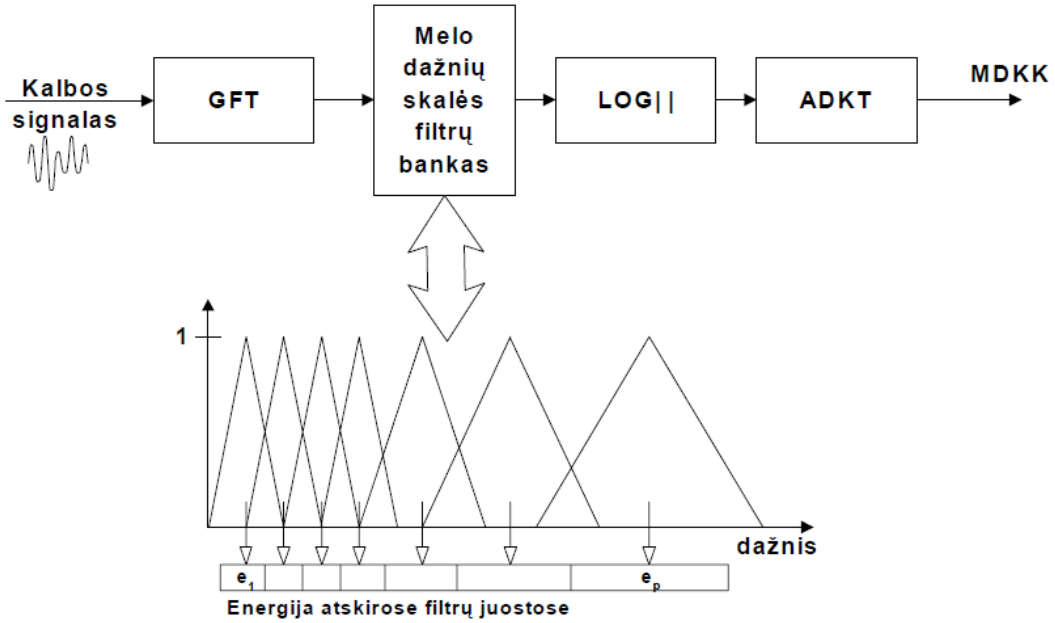
Pirmiausia, teoriškai, kalbos signalas gali būti modeliuojamas kaip kelių, turinčių skirtingas charakteristikas, šaltinių darinys. Tie kalbos šaltiniai - tai balso stygos, burna, gerklė, nosies ertmė, lūpos ir kt. Kepstrinė analizė leidžia šiuos kalbos šaltinius išskirti iš vieno signalo darant prielaidą, kad skirtingų kalbos signalo šaltinių parametrai turi savyje charakteristikas, kurios atskiriems garsams bei diktoriams yra skirtingos.

Kitas privalumas, kad kepstriniai koeficientai silpniau koreliuoja vienas su kitu, o tai žymiai supaprastina tolesnį analizės procesą.

Bene svarbiausias privalumas yra tas, kad kepstrinių koeficientų panaudojimas kalbos atpažinimui empyriškai demonstruoja ganėtinai gerus rezultatus jau daugelį metų.

Šiuo metu labiausiai paplitę kepstriniai koeficientai yra skaičiuojami naudojant Furjė ir diskrečiąją kosinusų transformacijas. Šiuo būdu yra skaičiuojami ir taip vadinami Melo dažnių kepstriniai koeficientai (MDKK). Skaičiavimo algoritmas susideda iš tokių etapų: Greitosios Furjė

transformacijos (angl.: fast Fourier transformation – FFT) koeficientų paskaičiavimas, logaritmavimas, atvirkštinės diskrečiosios kosinusinės transformacijos (ADKT) apskaičiavimas (11 pav.).



11 pav. Melo dažnių kepstrinių koeficientų apskaičiavimo algoritmo principinė schema

Šaltinis: Driaunys, Kęstutis (2000) *Lietuvių šnekamosios kalbos segmentavimo ir fonetinio atpažinimo tyrimas naudojant LTDIGITS garso įrašus.* Vilnius p. 30.

Pirmiausia yra apskaičiuojami kalbos signalo greitosios Furjė transformacijos koeficientai ir gaunamas signalo energetinis spektras.

$$S(l) = \frac{1}{N}\sum_{i=}^{N-} y\ i)v(i)e^{-j\frac{2\pi}{N}li}, \qquad (9)$$

kur l=0, 1, …, N-1 – spektro ataskaitos.

Gautas spektras yra filtruojamas panaudojant Melo dažnių filtrų banką, kuris išreiškiamas:

$$Mel(f) = :595\log_{10}(1+\frac{f}{/00}). \qquad (10)$$

Toliau atliekamas spektro S(k) logaritmavimas:

$$S(k) = \ \mathrm{g}_{10}\left|S(k)\right|^2. \qquad (11)$$

Tada iš logaritminio spektro yra skaičiuojamas kepstras atliekant atvirkštinę Furjė transformaciją. Kadangi skaičiuojama tik reali spektro dalis, tai kepstrą galima skaičiuoti panaudojant ADKT:

$$c(l) = \sum_{k=0}^{K-1} S(k)\cos(\pi\ (k-\frac{1}{2})/K), \qquad (12)$$

kur k = 0,1,2…K – spektro atskaitos, l = 0,1,2...L – kepstro koeficientų skaičius.

Pasinaudojus diskretine kosinusų transformacija gaunami c(l) – Melų dažnių kepstrinius koeficientus (MDKK) (Driaunys K., 2006).

1.7.4. Energija, delta ir akseleracijos koeficientai

Vienas iš pagrindinių požymių naudojamų kalbos atpažinime yra signalo energija. Paprastai naudojama normalizuota energija, kuri apskaičiuojama taip:

$$E = \log \sum_{i=1}^{N} y_i^2 \qquad \textbf{(13)}$$

Dažnai požymių vektorius papildomas dinaminiais kepstriniais koeficientais, arba vadinamaisiais delta koeficientais, kurie aprašo kepstrinių koeficientų kitimo greitį. Šie delta koeficientai padėjo pasiekti geresnių rezultatų daugelyje automatinio kalbos atpažinimo darbų. Delta koeficientai paskaičiuojamai pagal formulę:

$$\Delta_{..} = {}_{k}(l) - {}_{k-}(l), \qquad \textbf{(14)}$$

kur k – lango eilės numeris.

Kartais naudingi gali būti ir kepstro kitimo greičio, greičio koeficientai (delta-delta koeficientai), arba vadinamieji akceleracijos koeficientai. Delta-delta koeficientai apskaičiuojami panašiai kaip ir delta koeficientai, tačiau atimant dviejų gretimų delta koeficientų požymių vektoriaus narius:

$$\Delta\Delta_{..} = {}_{k}(l) - {}_{k-}(l), \qquad \textbf{(15)}$$

kur k – lango eilės numeris.

1.7.5. Požymių vektorius

Atlikus kiekvieno signalo lango analizę ir apskaičiavus kelis kalbos signalo požymius iš jų galima sudaryti požymių vektorių, kuris bus naudojamas kaip automatinio kalbos atpažinimo sistemos įėjimas:

$$X_k = c_k \mid E_k \mid \Delta_{..} \mid \Delta_{k} \mid \Delta\Delta_{..} \quad \Delta\Delta_{..}], \qquad \textbf{(16)}$$

c_k – MDKK koeficientai arba kitų kepstriniai, spektriniai ar tiesinės prognozės koeficientai

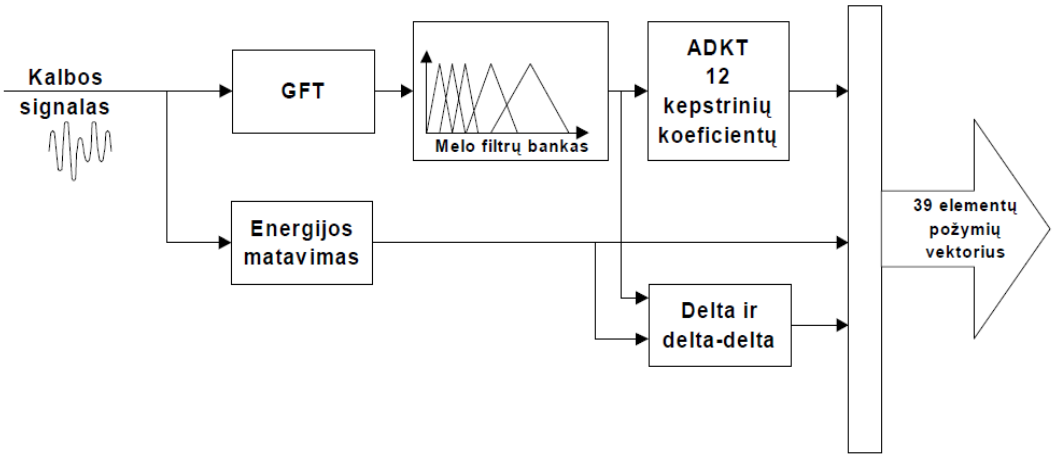
E_k - kalbos signalo energija.

Δ_k - MDKK delta koeficientai

ΔE_k - energijos delta koeficientai

$\Delta\Delta_k$ - MDKK akceleracijos koeficientai

$\Delta\Delta E_k$ - Energijos akceleracijos koeficientai.



12 pav. Požymių vektoriaus sudarymo blokinė schema

Šaltinis: Driaunys, Kęstutis (2000) *Lietuvių šnekamosios kalbos segmentavimo ir fonetinio atpažinimo tyrimas naudojant LTDIGITS garso įrašus.* Vilnius p. 32.

Požymių vektoriaus elementų skaičius priklauso nuo naudojamo MDKK koeficientų skaičiaus, pavyzdžiui, jei naudojame 12 MDKK, tai požymių vektorių sudarys 39 elementai. Požymių vektoriaus sudarymas pavaizduotas blokine schema 12 paveiksle.

Pirmiausia, prieš atliekant fonetinę signalo analizę, žodžių ribų radimą ir tų žodžių ar frazių atpažinimą, kalbos signalas turi būti segmentuojamas į pasirinkto įntervalo sekmentus, kurie gali apimti frazę, žodį ar jo dalį.

1.7.6. Požymių panaudojimas nagrinėtuose darbuose

Povilas Treigys ir Antanas Lipeika straipsnyje „Investigation of the speaker identification method based on clustered pseudostationary segments of voiced sounds“, nagrinėja kalbančiojo išskyrimą iš minios. Atstumų tarp mikrofono ir kalbančiųjų nustatymui naudojamas nuo teksto nepriklausantis atpažinimo algoritmas, kurio pagrindas iš pseudo stacionarių kalbos intervalų paimami kalbos signalai ir iš jų išskirti tiesinio prognozavimo koeficientai (angl: Linear Prediction Model), suskirstyti į klasterius panaudojus vektorinį kvantavimą. Tyrimo metu buvo nagrinėjamos 76 kalbančiųjų fonogramos. Iš kiekvinos fonogramos buvo išskiriam 10 tiesionio pronozavimo koeficientų. Identifikavimo procesas buvo atliekamas skaičiuojant vidutinį skirtumą tarp atitinkamų

koeficientų, suskirstytų į klasterius, iš skirtingų fonogramų. Eksperimento metu pozymiai buvo skirtomi nuo 1 iki 100 skirtingų bandymų metu. Atliktų ekspermentų rezultatas naudojant 60 ir 90 klasterių simetrinį skirtų paskaičiavimą yra 3 klaidos, o asimetrinio skaičiavimo 4 klaidos.

Haken Melin savo disertacijos „Automatic speaker verification on site and by telephone: methods, applications and assessment" tyrime ir kurtoje kalbančiojo atpažinimo sistemoje naudojo Melų dažnių kepstrinius koeficientus (MDKK) Mokslininko naudojamas 24 kanalų MDKK vektorius paremtas greitąja Furjė transformacija, logaritminės amplitudės filtrų banku, padengiančiu 300-3400 Hz signalo diapazoną, kosinusoidine transformacija ir kepstriniu filtravimu. Atpažinimui buvo naudojam skaitmenų pavadinimai Analizei naudotas visas diktoriaus ištartas žodis.

Claude Barras, Xuan Zhu ir kt. savo straipsnyje „System for AcousticSpeaker Identification in Seminars" aprašo atliktą kalbančiojo akustinio identifikavimo tyrimą. Kalbančiojo atpažinimas buvo vykdomas naudojantis požymių vektoriumi, kuris sudarytas iš 15 perseptroninių linijinio prognazavimo koeficientų, apskaičiuotų pagal Melų dažnių skalę, visų jų Delta ir Delta-Delta koeficientų ir Delta ir Delta-Delta energijos akceleracijos koeficientų. Testuojams sitemos treniravimosi modulio pagrindą sudarė GMM. Atliekant testavimus naudota 28 diktorių įrašai. Sistema buvo testuojama apmokius 15 ir 30 sek. Tiriamas atpažinimo tikslumas po 1 5 10 ir 20 testavimo sekundžių bei su skirtinga įrašimo technika (1 ir 4 kanalų mikrofonais). Atpažinimui naudojami požymiai išskiriami kas 10 ms naudojant 30ms langus. Taigi tyrimui naudojami žodžiai arba frazės, kurie skirstomi į 30ms langus. Geriausi atpažinimo rezultatai pasiekti apmokius sitemą 15 arba 30 sek po 20 sekunčių testavimo, Šie rezultatai ledžia pasiekti 1,8% paklaida.

Christoph Bregler, George Williams, Sally Rosenthal, Ian McDowall straipsnyje „Improving speaker verification with visual body language features" aprašo kaip kalbančiojo atpažimą galima pagerinti įvertinant žmogaus kūno kalbą, ypač veido išraiškas. Atliktame tyrime akustiniais požymiais paremtame atpažinime naudojami standartiniai Melų dažnių kepstriniai koeficientų vektorius (12 kepstrinės analizės reikšmių, energijos ir delta reikšmės). Ekspermentui naudojami sąlyginai ilgi kalbos segmentai iš kurių išskiriamas fiksuoto ilgio požymių vektorius. Eksperimantas atliktas analizuojant 4 val filmuotos medžiogos, kuriuoje nufimuoti kalbantys žmonės. Tyrimo metu nustatyta, naudojant minėtus požymius reikalingas geras garso įrašas. Esant dideliems triukšmams (17dB) paklaida siekia 20%, švarų akustinį signalą, atpažinimo tikslumas padidėja 5 kartus – paklaida 4,7%. Eksperimento metu nustatyta, kad identifikuojant vien akustiniu aspektu net esant žemai atpažinimo paklaidos tikimybei 4,7%, papildžius atpažinimą visualiniu atpažinimu paklaida sumažėja iki 4,1%.

S. Zribi Boujelbene, D. Ben Ayed Mezghani, N. Ellouze savo straipsnyje „Improved Feature Data for Robust Speaker Indentification Using Hybrid Gaussian Mixture Models- Sequentionl Mniimal Optimization System" aprašomame tyrime diktoriaus kalbos signalo analizei naudojo

Melo dažnių kepstrinius koeficientus, energiją, Delta ir Delta-Delta koeficientus. Požymiai apdorojami pagal gausinių mišinių modelį (GMM) suformuojant požymių vektorius. Sistemos treniravimo metu naudojamas nuoseklusis minimalusis optimizavimas (angl.: Sequential Minimal Optimization – SMO). Taigi, šio ekspermento esmė - GMM ir SMO apjungimas vienoje sistemoje. Buvo sudaryti du frazių bankai. Pirmasis bankas sudarytas iš aštuonių sakinių, o antrasis - iš tų pačių sakinių ir dar pridėti du sakiniai su „sa“ priesagomis. Eksperimente dalyvavo 47 diktoriai (16 moterų ir 31 vyras). 80 proc. įrašų buvo panaudota sistemos mokymui ir 20 proc. testavimui. Eksperimento metu požymiai buvo išskiriami susikirstant įrašus į langus. Požymiai išskirti iš kiekvieno sakinio viduriniojo lango. Šio tyrimo metu mokslininkams su pirmojo frazių banko duomenimis pavyko pasiekti vidutinį 91,49% atpažinimo tikslumą, o su antrojo banko – 85,11% tikslumą (su kiekvieno banko duomenimis buvo atliekami eksperimentai panaudojant skirtingą diktorių skaičių).

M. A. Fattah, F. Ren, S. Kuroiwa 2006 m savo straispnyje „Phoneme Based Speacker Modeling to Improve Speacker Identification“ aprašo diktoriaus atpažinimą panaudojant atrinktas fonemas arba jų junginius. Mokslinkai teigia, kad su tem tikromis fonemomis yra perduodamas itin ryškūs ir būdingi požymiai. Mokslininkai teigia, kad iškyrus atskirus fonemų segmentus ir naudojant juos atpažinimui, pats identifikavimo procesas ženkliai sutrumpėtų, nes požymiai išskiriami būtų ne iš viso kalbos segmento, o iš reikšminius diktoriaus požymius leidžiančių identifikuoti fonemų segmentų. Atpažinimui naudojami melų dažnių kepstriniai koeficientai gauti iš fonemų segmentų atskirų filtrų bankų.

J. Kamarauskas savo disertacijoje diktoriaus atpinimui sukūrė daugiafunkcijinę sistemą, kuri leido atlikti vokalizuotų garsų (priebalsių virtusių balsiais) išskyrimą taikant dirbtinius neuroninius tinklus (perceptronų ir atgalinio sklidimo) arba naudojant melų daznių koeficientus pagal kuriuos randamas pagrintinis kalbos signalo tonas. Mokslininkas pasiūlė atlikti melo dažnių kepstrinių koeficientų apjungimą su žadinimo signalo ir balso trakto požymiais. Sukurtos sitemos dėka mokslininkas palygino diktoriaus atpažinimo tikslumą naudojant vien MDKK ir juos apjungus su žadinimo signalo ir balso trakto požymiais. Pirmuoju atveju gautas atpažinimo klaidų lygis 5,86%, o naudojant pasiūlytą požynių sistemą -5,17%

Apibendrinant, pagal nagrinėtuose darbuose naudojamus algoritmus ir metodus, galima teigti, kad dažniausiai naudojama melų dažnių kepstriniai koeficientai ir GMM (juos naudoja Haken Melin disertacijos darbe „Automatic speaker verification on site and by telephone: methods, applications and assessment” , S. Zribi Boujelbene, D. Ben Ayed Mezghani, N. Ellouze savo eksperimente, kurį aprašė straipsnyje „Improved Feature Data for Robust Speaker Indentification Using Hybrid Gaussian Mixture Models- Sequentionl Minimal Optimization System“, Claude Barras ir Xuan Zhu siekdami sukurti kalbanciuju atpažinimą seminaruose, kurį aprašo „System for

AcousticSpeaker Identification in Seminars“, M. A. Fattah, F. Ren, S. Kuroiwa 2006m savo straispnyje „Phoneme Based Speacker Modeling to Improve Speacker Identification“). Diktoriaus atpažinimui daugiausiai išskiriami požymiai iš fonemų junginių, žodžių ar frazių, o kalbos atpažinime naudojami žodžio segmentai, fonemos. Kyla prielaida, kad stacionari fonemos dalyje yra užkoduotų požymių kurie atsiranda dėl individualių balso trakto savybių. Kadangi, analizės metu tokių tyrimų neaptikta, o pagal analizuotus darbus įžvelgiemos tendencijos, kad trumpėjant nagrinėjamam segmentui pasiekiami geresni diktoriaus atpažinimo rezultatai, nuspręsta atlikti diktoriaus atpažinimo tyrimą pagal stacionarią fonemos dalį.

2. TEORINIS RENGIAMO TYRIMO MODELIS

Remiantis K. Driaunio disertacijos aprašytais tyrimais, kuriuose buvo ištirta, kad žodžių fonemos turi joms būdingus požymius, buvo iškelta hipotezė, kad fonemų savybės leidžia atpažinti ne tik žodžius bet ir:

Atskiro diktoriaus ištartų žodyje fonemų stacionarios dalys pasižymi išskirtinėmis nuo diktoriaus priklausančiomis savybėmis, leidžiančiomis jį atpažinti.

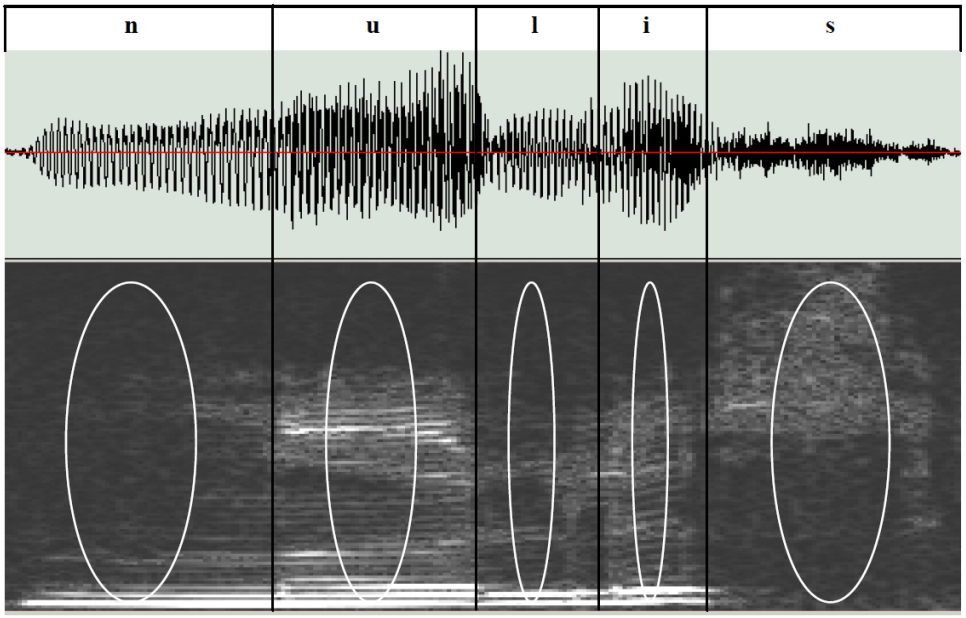
Šios hipotezės patikrinimo aktualumą padidina nagrinėtas ir trumpai anksteniame skyriuje aprašytas M. A. Fattah, F. Ren, S. Kuroiwa 2006m atliktas tyrimas. M. A. Fattah, F. Ren, S. Kuroiwa atlikti bandymai yra artimiausi tiems, kuriuos ruošiamasi atlikti šio mokslinio darbo metu. Mokslininkai savo tyrimuose išskyrė fonemas ir jų junginius, kurie geriausiai išreiškia unikalias diktoriaus balso trakto savybes. Eksperimentams buvo atrinktos šios fonemos ir jų junginiai: ah | ao |ay | eh |er | ey |f | ih | iy | k |n | r |s |t |th | uw | v | w. Tyrimui naudoasi YOHO garsyno duomenų baze kuria sudaro 138 kalbančiųjų -106 vyrų ir 32 moterų įrašai. Garsyno fazių turinį sudaro trijų dviženklių skaičių kombinacijos pvz.: 23-42-91. Frazių signalas eksperimento metu segmentuotas į 25ms trukmės langus, kurių poslinkis 10 ms. Tyrimo metu sistema buvo apmokoma. Apmokymų metu paskaičiuota, kuri fonema kiek kartų pasikartojo. Rečiausiai pasikartojo „r“ ir „er“ (<50 kartų), o dažniausiai „iy“ ir „t“ (>350 kartų). Geriausius rezultatus leido pasiekti fonemų junginys „iy“ tariamas kontekste.Su šiuo fonemų junginiu EER siekia tik 0,16%. Didžiausia ERR (20.67%) gaunama su „k” fonemos signalo segmentu naudojant atskiras fonemas.

Siekiant atlikti fonemos stacionarios dalies tinkamumo diktoriaus atpažinimui tyrimą, reikia pirmiausia pasirinkti, kokie požymiai bus išskiriami iš kalbos signalo segmentų ir naudojami eksperimentuose. Remiantis analizės rezultatais, pagal kuriuos matyti, kad nagrinėtuose tyrimiuose geriausius rezultatus mokslininkai pasiekia naudodami melų dažnių kepstrinių koeficientų požymių vektorius, numatyta juos naudoti rengiamame diktoriaus atpažinimo tyrime. Šis sprendimas priimtas taip pat atsižvelgiant į fonemos akustines savybes.

2.1. Akustinis fonemos modeliavimas

Remiantis K. Driaunio disertacijos aprašytais tyrimais, žinoma, kad kiekvieną žodžio fonemą sudaro trys jos fazės: santykinai stacionari fonemos dalis, kuri yra unikali kiekvienai fonemai (13 pav. pateikiamos LTDIGITS žodžio „nulis“ fonemų oscilograma ir spektrograma,

kurių stacionarios dalys apibrėžtos ovalais), ir ja supančios iš abiejų pusių ties fonemų ribomis labiau dinamiški ir tariamo žodžio dalių konteksto labiau sąlygoti fonemos fragmentai.



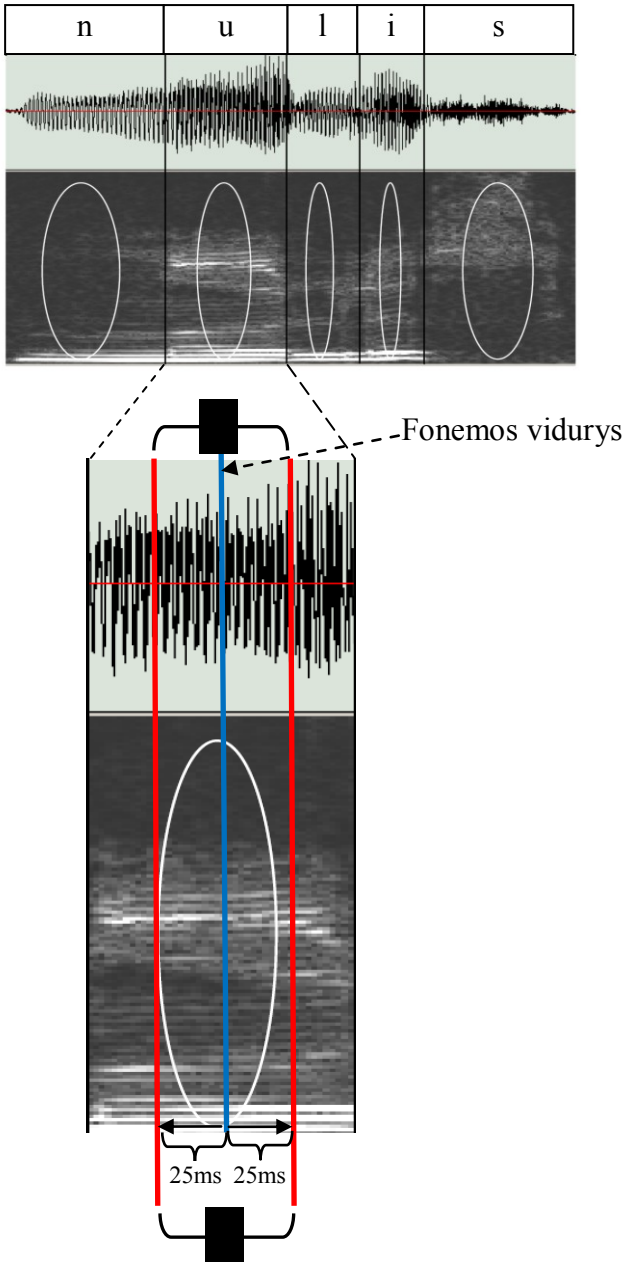
13 pav. Stacionarūs fonemų fragmentai

Šaltinis: Driaunys, Kęstutis (2000) *Lietuvių šnekamosios kalbos segmentavimo ir fonetinio atpažinimo tyrimas naudojant LTDIGITS garso įrašus.* Vilnius p. 67.

Pagal atliktą mokslinių tyrimų diktoriaus atpažinimo srityje analizę, kaip jau minėta, galima daryti prielaidą, kad stacionarioje dalyje sukaupta ne tik fonemos turinį atspindintys požymiai, bet ir požymiai nulemti individualių diktoriaus balso trakto savybių. Siekiant atlikti tyrimus šiai prielaidai patikrinti, reikia minėtas stacionarias fonemų dalis išskirti iš akustinio signalo.

2.2. Fonemos stacionarios dalies išskyrimas

Siekiant panaudoti stacionarias fonemų dalis diktoriaus atpažinimui, kaip minėta ankstesniame skyriuje, jas pirmiausia reikia išskirti. Tam reikalingas atskiras ruošiamos sistemos algoritmas. Stacionarios dalies išskyrimo principinė schema pavaizduota 14 paveiksle.

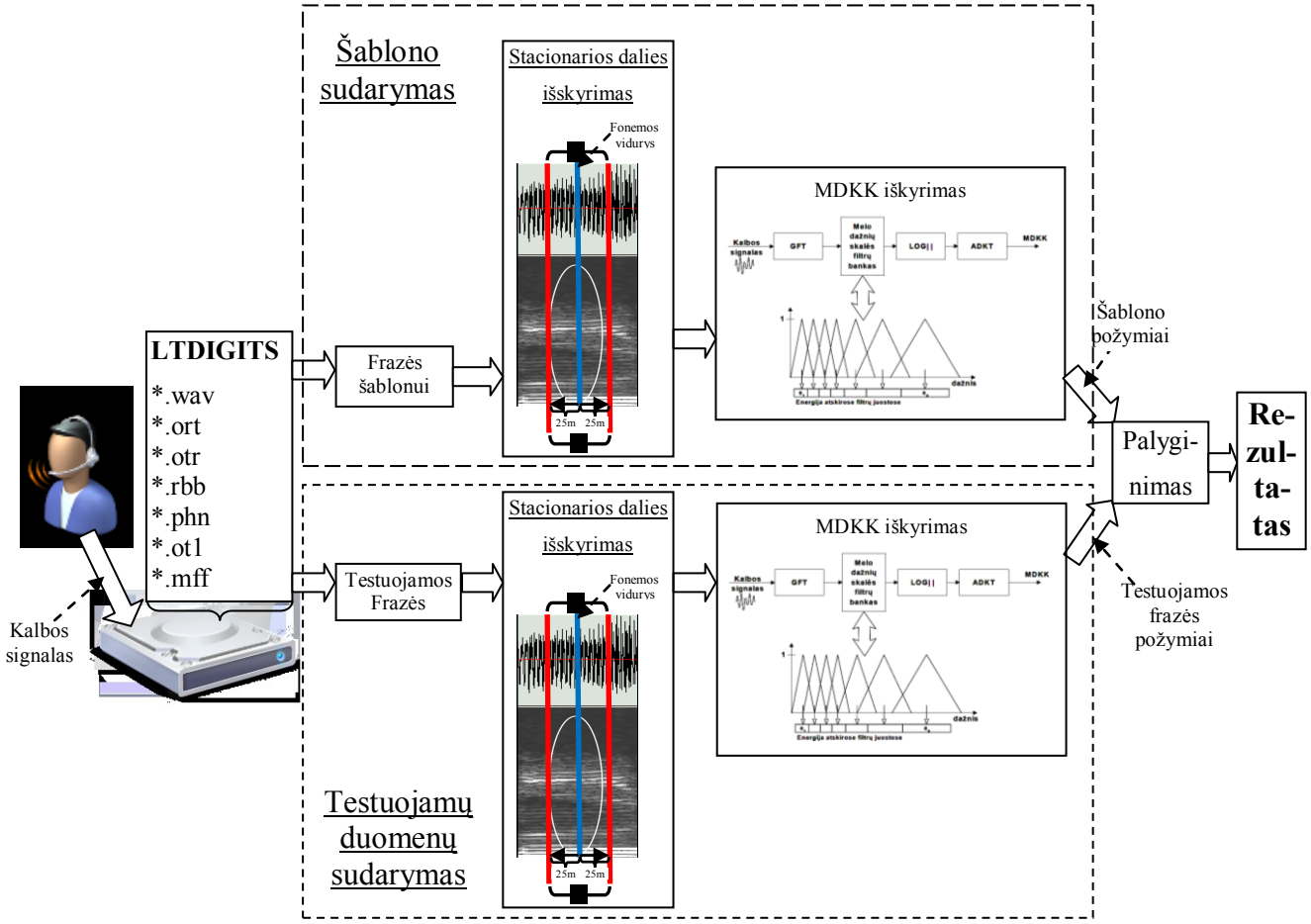


14 pav. Fonemos stacionarios dalies išskyrimas

Šaltinis: sukurta autoriaus

Kaip matome 14 pav. pateiktoje diagramoje, iš pasirinkto žodžio atrenkama norima fonema (šiuo atveju tai „u“) ir nustatomas fonemos centras. Tada paimama po vienodą signalo intervalą į abi puses (po 4 langus, t.y po 25ms (6,25ms x 4 langai)) nuo rasto fonemos centro. Taip gaunamas signalo segmentas, kuris laikomas stacionaria fonemos dalimi. Likę išskirtos fonemos signalo kraštai nenaudojami tolimesniame tyrime, nes laikoma, kad jie yra įtakoti greta esančių fonemų konteksto. Taigi, tokiu būdu stacionarioji fonemos dalis tiesiog „iškerpama“ ir naudojama tolimesniame tyrime.

Aprašytas fonemos stacionarios dalies išskyrimas numatytame eksperimente naudojamas diktoriaus atpažinimo sistemoje tiek sudarnat šabloną (mokymo duomenis), tiek apdorojant testuojamą diktoriaus pasakyto ir įrašyto žodžio signalą. Eksperimentui rengiamos diktoriaus atpažinimo sistemos veikimo principinė schema pateikiama 15 pav.



15 pav. Diktoriaus atpažinimo sistemos veikimo schema

Norint atpažinti diktorių tiek žmogaus smegenyse, tiek informacinėje sitemoje turi būti išsaugotas (atsimenamas) asmens, kurį norime atpažinti balso pavyzdys – šablonas (taikant informacines technologijas tai dažnai vadinama sistemos mokymo duomenimis). Taigi, kaip matome iš 15 pav. pavaizduotos schemos diktoriaus atpažinimas susideda iš dviejų pagrindinių etapų: šablono sudarymo (sistemos apmokymo) ir testuojamo duomenų (kalbos signalo) analizės. Šiuose atskiruose atpažinimo etapuose išanalizuoti duomenys vėliau lyginami tarpusevyje (15 pav. palyginimo blokas) ir pagal palyginimą gaunamas rezultatas (15 pav. rezultato blokas).

Planuojamame diktoriaus atpažinimo eksperimente naudojamos LTDIGITS garsyno frazės. Šio garsyno sudarymas ir analizė buvo atlikta ankstesnių mokslinių tyrimų metu. Šiuo metu jo kopijos saugomos skaitmeninėse laikmenose ir naudojamas kaip apdorotas diktorių frazių rinkinys naujiems tyrimams. LTDIGITS garsyne saugoma 800 garso įrašų, kuriuos padiktavo 100 diktorių (50 vyrų ir 50 moterų). Kiekvienam diktoriui priklauso 8 įrašai (Detaliau apie LTDIGIS

struktūrą aprašoma trečiajame darbo skyriuje.). Siekiant supaprastinti tyrimui kuriamos sistemos duomenų nuskaitymo algoritmą numatoma naudoti kiekvieno diktoriaus aštuonias frazes. LTDIGITS garsyne diktoriai sužymėti eilės numeriais (vyrai: M001-M050, moterys: F001- F050). Kiekvieno diktoriaus frazės sunumeruotos nuo T01 iki T08, jas sudarantys žodžiai taip pat sunumeruoti. Taigi, iš garsyno paimtas žodis identifikuojamas trijų numerių kombinacija. Pavyzdžiui, pirmo diktoriaus (vyro) trečios frazės antras žodis „*vienas*“ (M001; T03; 2).

Kaip jau minėta, reikia paruošti šabloninius duomenis ir imituojant realybę (sistemos apmokymų metu paskytos frazės identiškai vienodai diktorius po kurio laiko negali pakartoti) palikti testavimui dalį LTDIGITS kalbos signalo įrašų, kurie nenaudoti šablono sudarymui. Tai kuriamoje diktoriaus atpažinimo sistemojeniįgyvendinama naudojant minėtą diktorių, frazių ir žodžių numeraciją. Sudarant šabloną sistema įsimina šių trijų elemtų numerių kombinaciją. Vėliau, testavimo metu, paeiliui imant iš garsyno diktoriaus pasakytų frazių žodžius lyginami jų numeriai su šablone įsimintaisiais. Jei testavimui ruošiamos naudoti frazės ir šablono visų trijų elementų numeriai sutaps (Pvz., Šablonui naudotas žodis M001; T03; 2 ir testavimui norimas imti žodis M001; T03; 2) testavimui numatyta frazė bus praleista ir paimta kita, kurios elementų numeriai lyginami su šablono elemtų numeriais.

Tiek sudarant šabloną, tiek paimant testavimui kalbos signalą iš paimtų frazių išskiriamos lyginamų fonemų stacionarios dalys (15 pav) . Pagal išskirtų fonemų intervalų ribas atitinkamai atrenkami ankstenių tyrimų metu išskirti MDKK, kurių reikšmės saugomos LTDIGITS garsyno *.mff failuose Sistema kiekvieną „neatpažinto“ diktoriaus kalbos signalo segmentą (fonemos stacionarią dalį) lygina su kiekvieno diktoriaus šablone saugoma fonema. Šablono MDKK ir testinių frazių fonemų MDKK lyginami naudojant mažiausių skirtumų kvadratų metodą. Su kurio šablono diktoriaus fonemos stacionariąja dalimi lyginant testuojamąją gaunamas mažiausias skirtumas, tas diktorius ir itentifikuojams kaip testuojamos fonemos stacionarios dalies ir visos frazės, iš kurios buvo išskirta stacionarioji fonemos dalis, balso savininkas (15 pav.).

Remiantis šiuo teoriniu modeliu buvo parengta sistema ir atlikti 3 skyriuje aprašomi eksperimentai.

3. DIKTORIAUS ATPAŽINIMAS PAGAL STACIONARIUS FONEMOS FRAGMENTUS

Pagal antrajame skyriuje aprašytą diktoriaus atpažinimo modelį buvo sukurta sistema ir atlikti eksperimentai, kuriais siekama įsitikinti arba atmesti hipotezę, kad išskiriant fonemos stacionariąją dalį galima pagerinti diktoriaus atpažinimą.

3.1. Naudoti melų dažnių kepstriniai koeficientai

Šiame tyrime naudotasi K. Driaunio atliktų tyrimų ir eksperimentų metu apskaičiuotais LTDIGITS garsyno įrašų Melo dažnių kepstriniais koeficientais. Jie buvo apskaičiuoti HTK 3.0 paketo pagalba pasirinkus tokius parametrus:

Lango ilgis – 16 ms;

Žingsnis - 6,25 ms;

Melo filtrų banke 20 filtrų;

Požymių vektorius sudarytas pagal 1.6.5 skyriuje „Požymių vektorius“ aprašytą metodiką. Kiekviena kalbos signalo langą sudaro požymių vektorius iš 39 elementų.

3.2. Atlikti Eksperimentai

Eksperimentams atlikti naudotasi patobulintu LTDIGITS garsyno variantu. Šį garsyną kaip matome iš pirmos lentelės sudaro 7 tipų failai: frazės garso įrašas - signalo diskretos, frazės ortografija, transkripcija, žodžių ribos, fonemų ribos, pataisyta transkripcija, kurioje atskirų fonemų transkripcija išdėstyta stulpeliais, bei Melų dažnių kepstriniai koeficientai.

1 lentelė

LTDIGITS garsyno failų tipai ir jų turinys

Garsyno failo specifikacijos pavyzdys su plėtiniu	Garsyno failo turinys
...\LTDIGITS\MALE\M001\T01.**WAV**	Frazės garso failas - signalo diskretos
...\LTDIGITS\MALE\M001\T01.**ORT**	Frazės ortografija
...\LTDIGITS\MALE\M001\T01.**OTR**	Frazės žodžių transkripcija
...\LTDIGITS\MALE\M001\T01.**RBB**	Frazės žodžių ribos
...\LTDIGITS\MALE\M001\T01.**PHN**	Frazės žodžių fonemų ribos
...\LTDIGITS\MALE\M001\T01.**OT1**	Pataisyta frazės žodžių transkripcija
...\LTDIGITS\MALE\M001\T01.**MFF**	Frazės Melų dažnio kepstriniai koeficientai

Šaltinis: sudaryta autoriaus

Tyrimo metu atliekamas fonemų stacionarių dalių palyginimas. Siekiant užtikrinti palyginamumą į šabloną buvo įtraukiamos tik tos fonemos, kurios išskirtos iš tų diktorių frazių žodžių, kuriuos tas pats diktorius pakartoja kelis kartus toje pačioje ar skirtingose frazėse. Be to į šabloną įtraukiamos žodžių fonemos, kurios yra kitame diktoriaus tariamame žodyje. Pavyzdžiui, lyginant fonemas *e* šablonui jos paimamos iš kiekvieno diktoriaus žožio *penki*, o testavimui imamos tiek iš žodžių *penki*, kurių identifikacinis numeris (apie jį rašyta 2.2 skyriuje) neutampa su šablonui panaudotos fonemos žodžio numeriu, tiek iš žodžių *devyni*.

Šablonui žodžiai parinkti taip, kad būtų įtrauktos visos skirtingos LTDIGITS garsyne esančios fonemos (o, u, i, a, e). Siekiant surasti tinkamus šablonui ir tolimesniam palyginimui žodžius buvo peržiurėtas LTDIGITS frazių turinys.

2 lentelė

LTDIGITS frazių turinys

Frazės Nr	**Frazės turinys**						**Paaiškinimas**
1-5	nulis, vienas, du, trys keturi, penki, šeši, septyni, aštuoni, devyni						parenkami atsitiktinai šeši žodžiai
6	pradėti	baigti	sustoti	pauzė	laukti	tęsti	Sakomi žodžiai nurodyta lentelėje tvarka
7	pirmyn	atgal	į pradžią	į pabaigą	sekantis	perduoti	
8	taip	ne	pagalba	saugoti	start	stop	

Šaltinis: sudaryta autoriaus

Kaip matome iš antros lentelės tyrimui naudotame LTDIGITS garsyne kiekvienam diktoriui priklauso 8 įrašytos frazės. Kiekvieną ją sudaro 6 žodžiai. Pirmose penkiose frazėse atsitiktinai parenkant įrašyti po šešis lietuviškų skaitmenų pavadinimus nuo nulio iki devynių. Kadangi kiekvienam 1-5 frazės žodžiui galima parinkti bet kurį iš skaitmenų tai frazėje to paties skaitmens pavadinimas neretai kartojasi ir kelis kartus (pvz 44 vyro diktoriaus (m044) pirmoje frazėje (T01) „vienas“ tariamas du kartus). 6-8 frazėse yra diktorių padiktuotos lietuviškai tariamos valdymo komandos. Šios komandos yra tariamos visų diktorių vienoda tvarka, taip kaip surašytos antroje lentelė.

Peržiūrėjus LTDIGITS frazių turinį (2 lentelė) šablonui išskiriamos fonemos iš šių žodžių:

- *saug**o**ti*,
- *n**u**l**i**s*,
- ***a**tg**a**l*,
- *d**e**vyni*.

Testavimui fonemos atitinkamai išskiriamos iš tokių pat žodžių tik su kitu identifikaciniu numeriu. Taip pat atsižvelgiant struktūros panašumą su šablonui naudojamais žodžiais atitinkamų testuojamų fonemų išskyrimui buvo atrinkti žemiau pateiktieji:

- s**u**st**o**t**i**

- p**a**g**a**lb**a**
- p**e**nk**i**

Pasirinkus žodžius, kurie naudoti tyrime, buvo sukurti šablono sudarymo, o vėliau ir lyginamų žodžių fonemų nuskaitymo bei palyginimo su sudarytu šablonu algoritmai, kurie realizuoti MATLAB R2008a aplinkoje.

3.2.1. Šablono sudarymo algoritmas

Tyrimui rengiamas šablonas, kurį sudaro kiekvieno diktoriaus ištartų mus dominančių žodžių fonemų Melų dažnių kepstriniai koeficientai. Šablone diktoriaus, frazės, žodžio ir fonemos kombinacija yra unikali ir nesikartoja.

Šablono sudarymo algoritmas pirmiausiai pradedamas nuo kiekvienos frazės *.ORT failų peržiūros tikrinant ar pasirinktoje frazėje yra tuo metu ieškomas žodis. Jei žodis yra frazėje, jo ribos baitais nuskaitomos iš atitinkamos frazės *.RBB failų į kintamąjį „RibosB“. Nustačius reikiamo žodžio ribas atliekama jo analizė ir atskirame MATLAB m faile „Ribuskaitymas2.m“ nuskaitomos žodyje analizuojamų fonemų ribos iš LTDIGITS garsyno atitinkamos analizuojamos frazės *.PHN failo. Pagal analizuojamos frazės atrinktos fonemos ribas iš *.MFF failo nuskaitomi fonemos Melų dažnių kepstriniai koeficientai į kintamąjį „mas_mfcc“. Kaip jau minėta 2.1. skyriuje kalbos signalo langą sudaro požymių vektorius iš 39 elementų. Taigi paskaičiuojant kiek langų sudaro fonema, jos visų kepstrinių koeficientų elementų skaičius padalinamas iš 39. Tolimesnei analizei naudojamos tik tos fonemos, kurias sudaro daugiau nei 8 langai. Taip daroma todėl, kad nuo fonemos vidurio paimama po keturis langus į abi puses (viso 8 langai) – stacionari fonemos dalis. Kadangi langai yra persidengiantys, tai sudaro po 25 ms nuo fonemos vidurio į abi puses. Iškirpus fonemos stacionarios dalies kepstrinius koeficientus, jie rašomi į tekstinį šablono failą. Pirmoje kiekvieno įrašo eilutėje nurodoma: diktorius, frazė, žodis frazėje, fonema, fonemos vieta žodyje. Tada struktūrizuotai (8 eilutės – langai po 39 elementus) surašomi fonemos kepstriniai koeficientai.

3.2.2. Eksperimento bandymai ir modifikacijos.

Siekiant gauti aiškius atsakymus į keliamus klausimus neretai tenka patikrinti ne vieną galimą situacijos atvejį. Šio tyrimo metu taip pat buvo ne vienas eksperimento variantas išlaikant pamatinę jo dalį aprašytą 2.3. skyriuje ir 2.3.1 poskyryje. Toliau atskirai aprašomos kiekvieno eksperimento bandymo modifikacijos ir rezultatai.

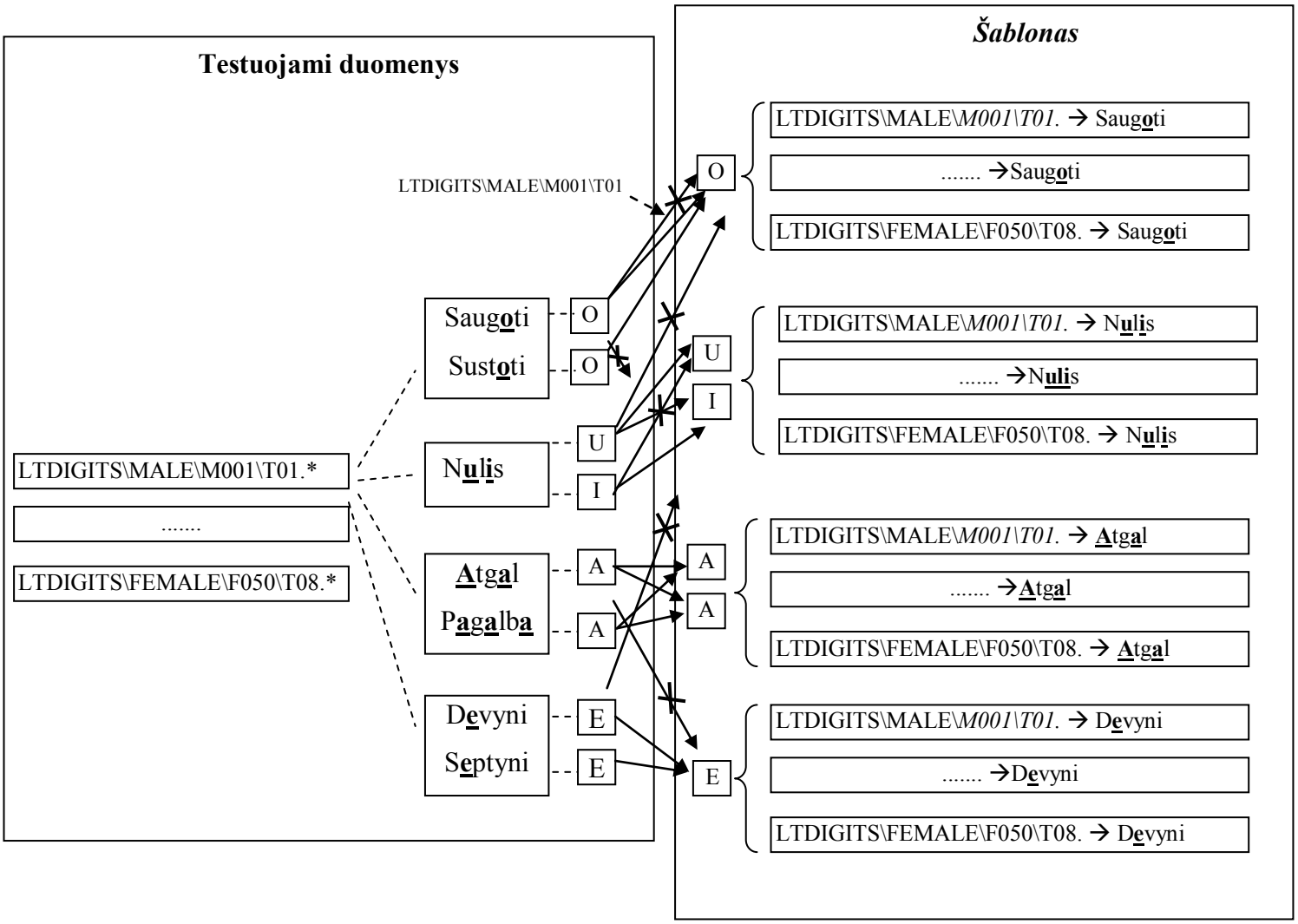
3.2.2.1. Diktoriaus atpažinimo algoritmas naudojant stacionarią fonemos dalį ir nelyginant greta esančių fonemų

Palyginimo algoritmas organizuojamas taip, kad lyginami būtų tik atitinkamos žodžių fonemos. Fonema(os) iš žodžio:

- *Saug**o**ti* lyginima su kitose to paties ir kitų diktorių pakartoto žodžio *sust**o**ti* fonema,
- *N**u**l**i**s* lyginama su tose pačiose frazėse to paties diktoriaus pakartoto ir kitose frazėse to paties ir kitų diktorių ištarto to paties žodžio fonemomis,
- ***A**tg**a**l* lyginima su kitose to paties ir kitų diktorių pakartoto žodžio *p**a**g**a**lb**a*** fonema,
- *D**e**vyni* lyginama su tose pačiose frazėse to paties diktoriaus pakartoto ir kitose frazėse to paties ir kitų diktorių ištarto to paties žodžio bei su žodžio *p**e**nki* fonema.

Palyginimo su šablonu algoritmas iki lyginamos fonemos kepstrinių koeficientų nuskaitymo vykdomas identiškai kaip ir šablono sudarymo algoritmo atveju aprašytu ankstesniame 2.3.1 skyriuje.

Nuskaičius koeficientus jie pertvarkomi ir išsaugomi į kintamąjį „mfcc_lyg_mas“ kaip matrica, kurią sudaro 8 eilutės ir 39 stulpeliai. Tada atliekamas šablonų failo „mfcc_SABL.txt“ nuskaitymas į kintamąjį „rec“. Taigi, turint nuskaitytus lyginamos fonemos kepstrinius koeficientus ir visą šabloną, atliekamas palyginimas kiekvieno šablone esančio įrašo su testuojamuoju. Analizuojant kiekvieną šablono įrašą lyginamos tik atitinkamos fonemos (pvz.: „a“ su „a“ ir pan.) Taip pat patikrinama ar iš LTDIGITS garsyno nuskaitytos diktoriaus frazės žodžio tiriamos fonemos stacionarioji dalis nėra įtraukta į šabloną. Patikrinimas atliekamas palyginant ar iš šablono paimto diktoriaus, frazės ir žodžio kombinacija sutampa su testuojama kombinacija. Grafiškai palyginimo schema pavaizduota 16 pav. Jei visos trys kombinacijos elementai paimti iš šablono sutampa su testuojamo žodžio (apie LTDIGITS garsyno žodžių identifikavimą rašyta šio darbo 2.2 skyriuje) - algoritmo tolimesnis vykdymas stabdomas ir nuskaitomas kitas žodis iš garsyno, jei palyginus bent vienas kombinacijos elementas nesutampa atliekamas kepstrinių koeficientų palyginimas ieškant mažiausios kepstro požymių vektorių paklaidos tarp testuojamos ir šablono žodžio fonemos. Paklaida skaičiuojama pagal 14 formulę. Šablono ir testuojamų signalų kepstrinių koeficientų palyginimo apribojimus 16 paveiksle žymi perbrauktos rodyklės.



16 pav. Testuojamų žodžio fonemų palyginimas su šablonu pirmo bandymo metu

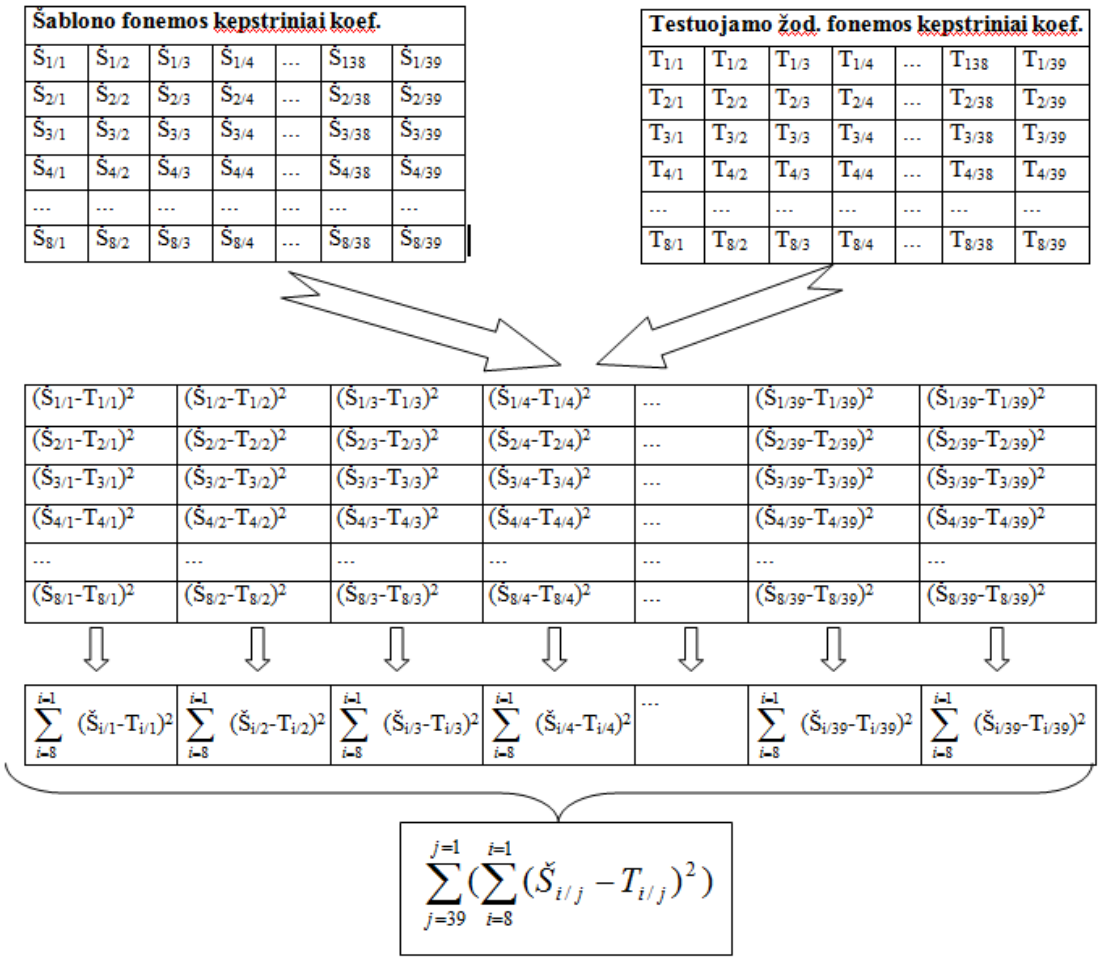
Šaltinis: sukurta autoriaus

Mažiausių skirtumų kvadratų suma išreiškiama bendrąja formule:

$$f(x,y)_{\min} = \min \sum_{i=}^{n} x_i - y_i)^2 \qquad \textbf{(17)}$$

kur x – testuojamo požymių vektoriaus elementas, y - šablono įrašo požymių vektoriaus elementas, i vektoriaus elemento numeris.

Šios formulės realizavimo algoritmas grafiškai pavaizduotas 17 paveiksle.



17 pav. Testuojamų ir šablono požymių vektorių skirtumų kvadratų sumos radimas

Šaltinis: sukurta autoriaus

Kaip vaizduojama 17 paveiksle pateiktoje schemoje, pirmiausiai išskiriami pasiringtų fonemų stacionarių dalių šablono ir testuojamo kalbos signalo melų dažnių kepstriniai koeficientai. Jų išskirama 312 (8 langai x 39 koeficientai vektoriuje) iš sistemoje saugomo šablono ir tiek pat iš testuojamos stacionariosios fonemos dalies. Išskyrus koeficientus skaičiuojami šablono ir testuojmo signalo atitinkamų reikšmių skirtumų kvadratai. Šių skirtumų kvadratai iš kiekvieno lango susumuojami su kito lango atitinkamo kepstrinių koeficiento reikšme (17 paveiksle tai žymi vertikalios rodyklės). Sumavimas kartojamas 8 kartus (kol visų langų reikšmės sudedamos)– gaunamos trisdešimt devynios šablono ir testuojamo segmento MDKK reikšmių skirtumų kvatratų sumos. Jos galiausiai susumuojamos ir gaunamas vienas reikšminis rezultatas, pagal kurį nuspredžiama ar šablono įrašo diktorius sutampa su testuojamo kalbos signalo diktoriumi.

Mažiausią apskaičiuotą skirtumą kvadratų sumų sumą turintis šablono įrašas pripažįstamas kaip to paties diktoriaus, kuris ištarė ir testuojamą žodį.

3.2.2.2. Diktoriaus atpažinimo rezultatai naudojant stacionarią fonemos dalį ir nelyginant greta esančių fonemų

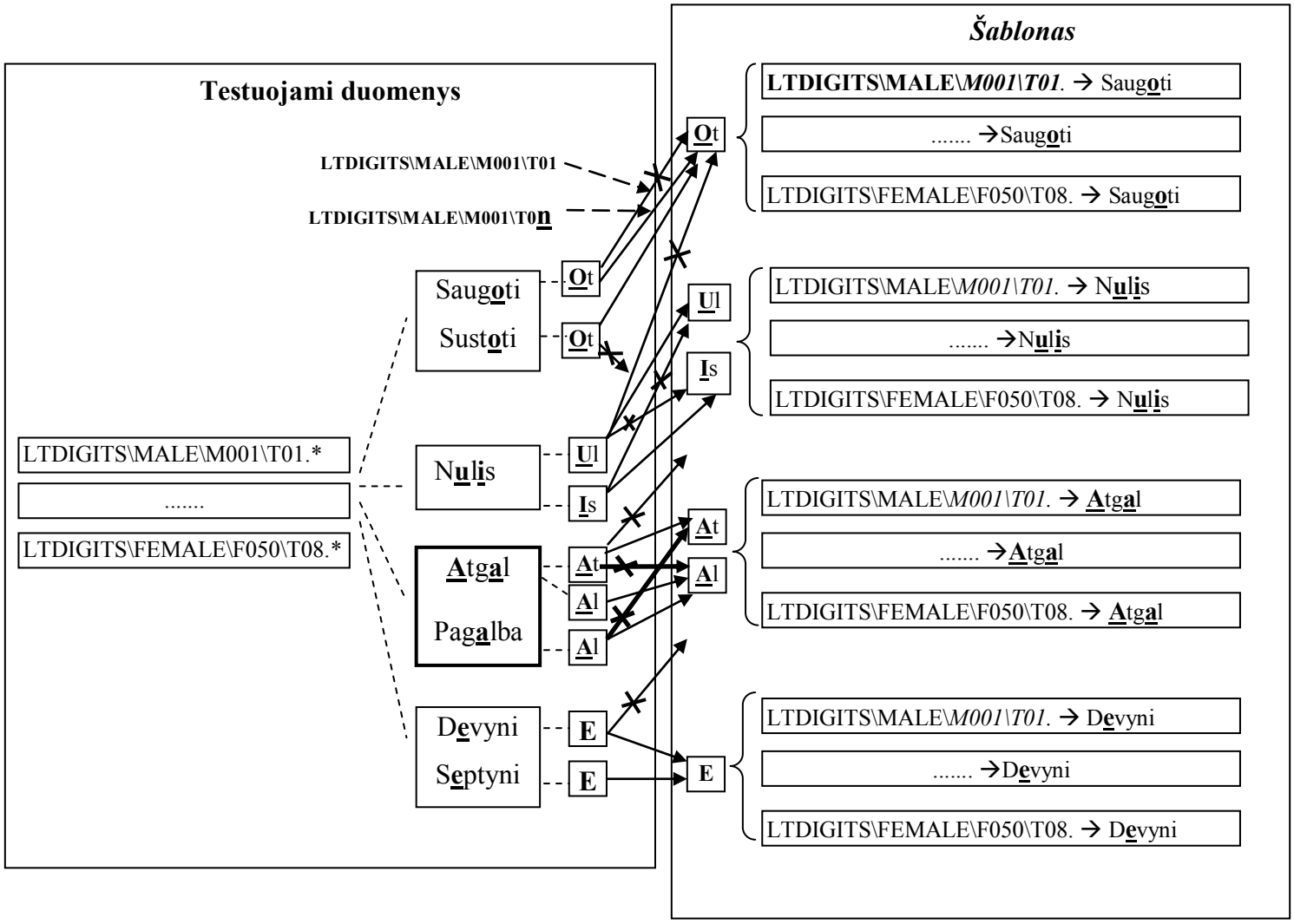
Palyginimų rezultatai (testuojamo ir šabloninio įrašo rekvizitai: diktorius, frazė, žodis, fonema) įrašomi į tekstinį failą. Tam, kad būtų galima apskaičiuoti teisingų diktorių atpažinimų procentą iš karto palyginama ar panašiausio šablono, kurio požymių vektorių skirtumų kvadratų suma (skaičiuojama kaip parodyta15 pav.) su testuojamuoju yra mažiausia, diktorius sutampa su testuojamo žodžio diktoriumi. Jei taip, įrašoma eilutės pabaigoje „T“, jei ne – „N“

Suskaičiavus atpažinimų statistiką gautas labai žemas teisingų diktoriaus atpažinimų procentas. Teisingai atpažinti diktoriai vos **26,3 %** visų atvejų.

Tačiau paanalizavus rezultatų failą buvo pastebėta, kad diktoriai pagal lytis supainiojami gana retai. Atlikus statistinę analizę gauta, kad diktoriai pagal lytį naudojant fonemų stacionarią dalį atpažįstami net **91,8 %** tikslumu

3.2.2.3. Diktoriaus atpažinimo algoritmas naudojant stacionarią fonemos dalį ir lyginant greta esančių fonemų sutapimą.

Antro bandymo metu naudotas tas pats šablonas, kaip ir pirmo bandymo metu. Šiame bandyme buvo patikslintas palyginimo algoritmas – uždėti papildomi apribojimai, kurie leistų lyginti tik tas vienodas testines ir šablono fonemas, po kurių einančios sekančios fonemos taip pat sutampa. Pavyzdžiui.: lyginant fonemas *o* po jų seka tokios pat fonemos - *t*: saug***ot***i ir sust***ot***i. Žodžio *saugoti* fonema *o* panaudojama šablonui, vėliau žodžio *sustoti* atitinkama fonema panaudojama testavimui.) Taip tikimasi, kad bus lyginami panašiausia kalbos maniera (tai, kaip buvo rašyta šio darbo 1.3 skyriuje, itin akcentavo M.C. McDermott, T. Owen, F.M. McDermott savo darbe „Voice indentifikation:Aural/Spectrographic Method“ (1996m) ištartos fonemos ir tai leis padidinti atpažinimo tiklumą. Šio palyginimo principinė schema pateikta 18 paveiksle.



18 pav. Testuojamų žodžio fonemų palyginimas su šablonu antro bandymo metu

Šaltinis: sukurta autoriaus

Palyginimo algoritmo pakeitimai labiausiai įtakoję testinių žodžių *atgal* ir *pagalba* palyginimą su šablonu schemoje pažymėti paryškintomis rodyklėmis.

3.2.2.4. Diktoriaus atpažinimo rezultatai naudojant stacionarią fonemos dalį ir lyginant greta esančių fonemų sutapimą

Kaip ir pirmo bandymo palyginimų rezultatai (testuojamo ir šabloninio įrašo rekvizitai: diktorius, frazė, žodis, fonema) įrašomi į tekstinį failą. Tam, kad būtų galima apskaičiuoti teisingų diktorių atpažinimų procentą iš karto palyginama ar panašiausio šablono, kurio požymių vektorių skirtumų kvadratų suma (skaičiuojama kaip parodyta 17 pav.) su testuojamuoju yra mažiausia,

diktorius sutampa su testuojamo žodžio diktoriumi. Jei taip, įrašoma eilutės pabaigoje „T“, jei ne – „N“

Suskaičiavus atpažinimų statistiką gautas priešingai nei tikėtasi dar žemesnis teisingų diktoriaus atpažinimų procentas. Teisingai atpažinti diktoriai vos **8,5 %** visų atvejų.

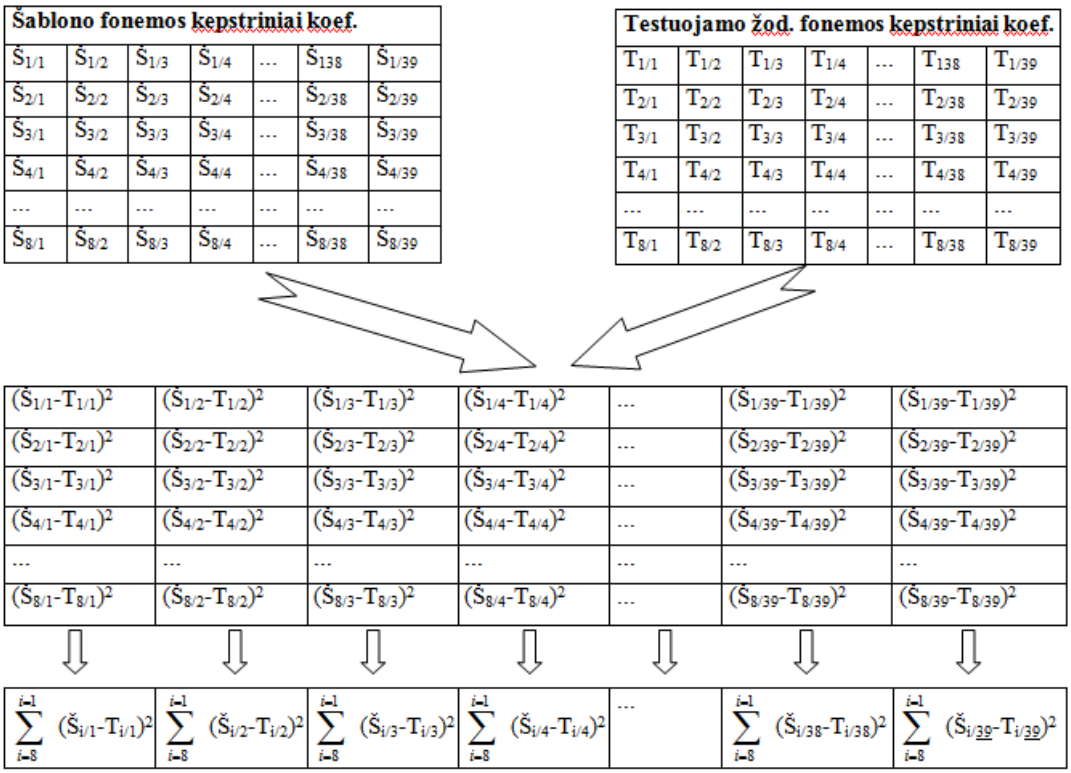
Tačiau paanalizavus rezultatų failą buvo pastebėta, kad diktoriai pagal lytis supainiojami kur kas rečiau, kaip ir ankstesnio bandymo metu. Atlikus statistinę analizę gauta, kad šio bandymo metu diktoriai pagal lytį naudojant fonemų stacionarią dalį atpažįstami **62,7%** tikslumu.

Šio bandymo rezultatai, turėdami priešingą tinslumo poslinkį nei tikėtasi, paskatino iškėlė klausimą kodėl taip nutiko. Buvo nuodugniai patikrintas algoritmas, jame neradus klaidų nutarta atlikti trečią bandymą.

3.2.2.5. Diktoriaus atpažinimo algoritmas naudojant stacionarią fonemos dalį ir vertinant kiekvieną MDKK atskirai.

Dėl netikėtai sumažėjusio atpažinimo tikslumo antro bandymo metu buvo ieškoma to priežasčių atliekant trečiąjį bandymą. Šio bandymo metu siekiama nustatyti, kurie kepstriniai koeficientai duoda atpažinimui didžiausią teigiamą arba neigiamą įtaką. Tikėtasi, kad nustačius didžiausią įtaką darančius koeficientus, juos galės eliminuoti ir pakartoti atpažinimo testavimą pagal antro bandymo algoritmą.

Šio bandymo metu, užuot susumavus testuojamos ir šablono fonemos stacionarios dalies skirtingų kepstrinių koeficientų požymių skirtumų kvadratus ir gavus vieną skaičių (kaip tai daryta pirmo ir antro bandymo metu), atliekamas stebėjimas kaip identifikuojamas diktorius pagal atskirų požymių skirtumų kvadratus. Tai pavaizduota 19 pav. atlikus atitinkamas korekcijas schemoje lyginant su 17 pav. Šiuo atveju gaunamos 39 šablono ir testuojamos kalbos signalo MDKK reikšmių skirtumų kvadratų sumos.



19 pav. Testuojamų ir šablono požymių vektorių skirtumų kvadratų sumos radimas

Šaltinis: sukurta autoriaus

Ieškant kepstrinių koeficientų, kurie galimai turi svertinę teigiamą arba neigiamą įtaką bendram atpažinimui, pagal kiekvieną melų dažnių kepstrinį koeficientą, identifikuotą atskiruose signalo diskretų languose (priminsime, kad tyrime naudojami aštuoni kalbos signalo langai žr. 2.3.1. skyrių), gaunamas atskiras atpažinimo rezultatas. Kiekvienam kepstriniam koeficientui generuojamas ir saugomas atskiras atpažinimo rezultatų failas.

Taigi, šio bandymo pabaigoje turime 39 failus su atpažinimo rezultatais pagal kiekvieną melų dažnių kepstrinį koeficientą.

2.1.1.1. Diktoriaus atpažinimo rezultatai naudojant stacionarią fonemos dalį ir vertinant kiekvieną MDKK atskirai.

Šiame bandyme, kaip ir pirmame bei antrame bandymuose palyginimų rezultatai įrašomi. Tačiau kaip jau minėta, skirtingai nei ankstesnių bandymų atvejais mes turime ne vieną, o 39 rezultatų tekstinius failus, kuriuose yra testuojamo ir šabloninio įrašo rekvizitai: diktorius, frazė, žodis, fonema. Taip pat, kaip pirmo ir antro bandymo metu tam, kad būtų galima apskaičiuoti teisingų diktorių atpažinimų procentą iš karto palyginama ar panašiausio šablono, kurio požymių

vektorių skirtumų kvadratų suma (skaičiuojama kaip parodyta 17 pav.) su testuojamuoju yra mažiausia, diktorius sutampa su testuojamo žodžio diktoriumi. Jei taip, įrašoma eilutės pabaigoje „T“, jei ne – „N“

Suskaičiavus atpažinimų statistiką gauti teisingų atpažinimų rezultatai naudojant vieną iš 39 požymių rezultatai svyruoja nuo **0,1947%** (16 požymis) iki **1,6553%** (9 požymis) atpažinimų iš visų testuojamų diktorių įrašytų žodžių (3 lentelė).

3 lentelė

Diktorių atpažinimo tikslumas naudojant pasirinktą požymį

Požymis	Tikslumas, %	Požymis	Tikslumas, %	Požymis	Tikslumas, %	Požymis	Tikslumas, %	Požymis	Tikslumas, %
1	1,3359	**9**	*1,6553*	**17**	1,1685	**25**	0,5842	**33**	0,8763
2	1,3632	**10**	1,2658	**18**	0,2921	**26**	0,6816	**34**	1,0711
3	1,4606	**11**	1,7527	**19**	0,5842	**27**	0,6816	**35**	0,7790
4	1,3632	**12**	1,5579	**20**	0,4869	**28**	0,9737	**36**	0,4869
5	0,7790	**13**	1,1685	**21**	0,7790	**29**	0,2921	**37**	0,7790
6	1,2658	**14**	0,5842	**22**	0,8763	**30**	1,2658	**38**	0,7790
7	1,1685	**15**	1,2658	**23**	0,7813	**31**	0,7790	**39**	1,2658
8	1,2658	**16**	*0,1947*	**24**	0,6816	**32**	0,7790	-	-

Šaltinis: sudaryta autoriaus

Pagal lytį teisingų atpažinimo tikslumas svyruoja nuo **49,6592 %** (19 požymis) iki **57,1568% (4 požymis)** (4 lentelė).

4 lentelė

Diktorių atpažinimo pagal lytį tikslumas naudojant pasirinktą požymį

Požymis	Tikslumas, %	Požymis	Tikslumas, %	Požymis	Tikslumas, %	Požymis	Tikslumas, %	Požymis	Tikslumas, %
1	51,9084	**9**	52,1908	**17**	50,3408	**25**	51,8987	**33**	51,4119
2	52,9698	**10**	53,5540	**18**	50,1461	**26**	49,5618	**34**	51,3145
3	52,0935	**11**	56,3778	**19**	49,6592	**27**	49,7566	**35**	51,2171
4	*57,1568*	**12**	51,5093	**20**	51,3145	**28**	49,7566	**36**	51,2171
5	51,2171	**13**	50,5355	**21**	50,9250	**29**	*49,1723*	**37**	50,4382
6	50,6329	**14**	50,0487	**22**	51,7040	**30**	51,8014	**38**	50,0487
7	52,8724	**15**	52,3856	**23**	52,3438	**31**	50,0487	**39**	50,5355
8	54,9172	**16**	50,4382	**24**	49,8539	**32**	51,0224	-	-

Šaltinis: sudaryta autoriaus

Kaip matome bandymo metu iš gautų rezultatų pateikiamų 4 lentelėje atpažinimo tikslumas siekia maksimaliai 1,6553%. Tikslumas 5,11 karto mažesnis lyginant su antruoju bandymu, kai diktoriaus atpažinimas atliktas pagal stacionarią fonemos dalį ir įverinamas greta esančių fonemų sutapimas . Tuo tarpu lyties atpažinimo tikslumas (antro bandymo metu pasiektas 62,7 %) taip ženkliai nekrenta atliekant palyginimą pagal atskirus požymius. Jis sumažėja tik 1,09 karto lyginant su pasiektu atpažinimo tikslumu naudojant 9 požymį ir 1,26 karto naudojant 26 požymį.

Taigi, diktoriaus atpažinimas naudojant stacionarią fonemos dalį ir vertinant kiekvieną MDKK bandymas deja nepadėjo atskleisti kritusio atpažinimo tiklsumo priežasties ir rasti vieno ar

keleto MDKK, kurie teikia ženkliai mažesnį atpažinimo tikslumą nei kiti koeficientai ir neigiamai veikia bendrą atpažinimą.

IŠVADOS IR PASIŪLYMAI

1. Kalbos signalo apdorojimo tyrimai atliekami nuo 1930m kalbos ir kalbančiojo atpažinimo tikslais. Klabos atpažinime pasiekti rezultatai, kurie leidžia sukurti sistemas tinkamas komerciniam platinimui, o kalbančiojo atpažinime dėl žmogaus balso kintamumo laiko eigoje ir kitų neapibrėžtumų tokie atpažinimo tikslumo rezultatai dar nepasiekti.

2. Atlikus atliktų mokslininkų tyrimų analizę, paaiškėjo, kad pastarąjį dešimtmetį dažniausiai atpažinimui naudojami Gausiniai mišinių modeliai ir įvairios jų modifikacijoje remiantis iš kalbos signalo išskiriamais Melo dažnių kepstriniai koeficientais (energija, Delta ir Delta-Delta).

3. Atlikus atliktų mokslininkų tyrimų analizę, paaiškėjo, kad mažinant analizuojamo kalbos signalo intervalą gaunami tikslesni diktoriaus atpažinimo rezultatai.

4. Diktoriaus atpažinimas pagal stacionarius fonemos fragmentus išskiriant kepstrinius koeficientus pirmo bandymo metu laido teisingi atpažinti diktorius vos **26,3 %** visų atvejų, o pagal lytį net **91,8 %** tikslumu, taigi kelta hipotezė patvirtinta tik lyties nustatymui.

5. Diktoriaus atpažinimas, pagal identiškame kontekste ištartų fonemų stacionarius fragmentus išskiriant kepstrinius koeficientus laido teisingai atpažinti diktorius vos **8,5%** visų atvejų, o pagal lytį **62,7%** tikslumu.

6. Diktoriaus atpažinimo tikslumas naudojant stacionarią fonemos dalį ir vertinant kiekvieną MDKK atskirai siekia maksimaliai 1,7% (9-as). Lyties atpažinimo didžiausais tikslumas 52,2% (9 – as MDKK) o mažiausias 49.6% (26 –as MDKK).

7. Diktoriaus atpažinimasnaudojant stacionarią fonemos dalį ir vertinant kiekvieną MDKK nepadėjo rasti vieno ar keleto MDKK, kurie teikia ženkliai mažesnį atpažinimo tikslumą nei kiti koeficientai ir neigiamai veikia bendrą atpažinimą.

LITERATŪRA

1. Atal B.S., Hanauer S.L. (1971) *Speech Analysis and Synthesis by Linear Prediction of the Speech Wave.* J. Acoust. Soc. Am. 50 (2). pp.637-655.

2. BARRAS Claude, ZHU Xuan, LEUNG Cheung-Chi, GAUVAIN Jean-Luc, ir LAME Lori. *The CLEAR'07 LIMSI System for AcousticSpeaker Identification in Seminars* [interaktyvus]. Orsey: Spoken Language Processing Group LIMSI-CNRS, [Žiūrėta 2010 m. sausio 16 d.]. Prieiga per internetą: http://clear-evaluation.org/downloads/papers/clear07_barras_draft.pdf

3. BOUJELBENE S. Zribi, MEZGHANI D. Ben Ayed, ELLOUZE N (2009). *Improved Feature Data for Robust Speaker Indentification Using Hybrid Gaussian Mixture Models- Sequentionl Miimal Optimization System* International Review on Computers and Software.

4. BREGLER Christoph, WILLIAMS George, ROSENTHAL Sally, MCDOWALL Ian. *Improving speaker verification with visual body language feature* [interaktyvus]. New York: Courant Institute of Mathematical Sciences of New York University, [Žiūrėta 2010 m. sausio 16 d.]. Prieiga per internetą: http://www.cims.nyu.edu/~bregler/ICASSP09/bregler_icassp09.pdf

5. DRIAUNYS, Kęstutis (2006) *Lietuvių šnekamosios kalbos segmentavimo ir fonetinio atpažinimo tyrimas naudojant LTDIGITS garso įrašus.* Vilnius p. 28. –cit. pagal Deller J.R., Hansen J.H.L., Proakis .G. (2000) *Discrete-Time Processing of Speech Signals.* IEEE Press, Second ed. New York.171 p.

6. DRIAUNYS, Kęstutis (2006) *Lietuvių šnekamosios kalbos segmentavimo ir fonetinio atpažinimo tyrimas naudojant LTDIGITS garso įrašus.* Vilnius. 171 p.

7. FATTAH, Mohamed Abdel; REN, Fuji; KUROIWA,Shingo (2006) *Phoneme Based Speacker Modeling to Improve Speacker Identification* AIML 06 International Conference, June 13-15. Sharm El Sheikh, Egypt. [žiūrėta: 2009 lapkričio 12]. Prieiga per internetą: < http://www.icgst.com/con06/aiml06/Final_Articles/P1120608101.pdf >

8. JIN, Qin (2007) *Robust Speaker Recognition.* Pittsburgh. CMU-CS-07-00. 3, 5, p. [žiūrėta: 2009 lapkričio 12]. Prieiga per internetą: < http://www.lti.cs.cmu.edu/Research/Thesis/QinJin.pdf > 155 p.

9. JUANG B.H. ir RABINER L. R. (2005) *Automatic Speech Recognition – A Brief History of the Technology Development.* NewJersey. 2 p. [žiūrėta 2009 lapkričio 10] . Prieiga per internetą: < http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf > 24p.

10. KAMARAUSKAS, Juozas (2009) *Asmens pažinimas pagal balsą.* Vilnius. Technika 17-18, [žiūrėta: 2009 lapkričio 12]. Prieiga per internetą: < http://www.mii.lt/files/mii_dis_09_kamarauskas.pdf>. 150 p.

11. KING, Josh (2008) *Speaker Verification Using Adaptad Gassian Mixture Model. Presentation* [interaktyvus]. [Žiūrėta 2010 m. sausio 15 d.]. Prieiga per internetą: http://www.cse.ohio-state.edu/~kingjo/presentations/speakerrec.pdf 56sk.

12. LIU, Roulun *Speech signal processing. Lecture 16* School of Information Engineering of Shandong University. Weihai. 26sk.

13. LOVELL, C. Brian ir TSOI, Ah. Chung *Speacker Verification Using Artificial Neural Networks* [interaktyvus]. Queensland: Departament of Electrical Engineering of University of Queensland, [Žiūrėta 2010 m. sausio 16 d.]. Prieiga per internetą: http://www.assta.org/sst/SST-90/cache/SST-90-Chapter11-p8.pdf .298-303 p.

14. *Medicinos enciklopedija (1 dalis)* (1991). Vilnius: Valstybinė enciklopedijų leidykla 910 p. ISBN 5-89950-006-9

15. MELIN, Haken (2006) *Automatic speaker verification on site and by telephone: methods, applications and assessment* Stockholm. 331 p.

16. McDermott, Michael ; Owen, Tom; McDermott, Frank . (1996) *Voice indentification. The Aural/Spectrographic Method* [interaktyvus]. Owl Investigations, Inc., [Žiūrėta 2011 m. gegužės 09 d.]. Prieiga per internetą: http://www.tapeexpert.com/pdf/voiceidauralspectro.pdf

17. National Science and Technology Council (2006) *Speaker Recognition.* [interaktyvus] Washington, DC. [žiūrėta: 2009 lapkričio 10]. Prieiga per internetą: < http://www.biometrics.gov/Documents/ SpeakerRec.pdf >.

18. Rudžionis, A.; Ratkevičius, K.; Dumbliauskas, K.; Rudžionis, V. (2008) Control of Computer and Electric Devices by Voice. *System Engineering, computer technology.* Kaunas. Kauno technologijos unversitetas, p. 11-16 ISSN 1392-1215.

19. TREIGYS, Povilas; LIPEIKA, Antanas (2006) *Investigation of the speaker identification method based on clustered pseudostationary segments of voiced sounds*. ŪKIO TECHNOLOGINIS IR EKONOMINIS VYSTYMAS. Vilnius. Vol XII, No 1. 50-55 p.

20. VAPNIK, Vladimir; CORTES, Corinna (1995) *Support-Vector Networks.*MACHINE LEARNING.Boston. Vol XX, No 3. 273-297p.