



VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

ELEKTRONIKOS FAKULTETAS

TELEKOMUNIKACIJŲ INŽINERIJOS KATEDRA

Audrius Raškauskas

**PRARASTŲ PAKETŲ ĮTAKOS AUTOMATINIAM KALBOS
ATPAŽINIMUI TYRIMAS**

**RESEARCH OF IMPACT OF LOST PACKETS ON AUTOMATIC
SPEECH RECOGNITION**

Baigiamasis magistro darbas

Telekomunikacijų inžinerijos studijų programa, valstybinis kodas 62401T210

Telekomunikacijų technologijų specializacija

Elektros ir elektronikos inžinerijos mokslo kryptis

Vilnius, 2008

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

ELEKTRONIKOS FAKULTETAS

TELEKOMUNIKACIJŲ INŽINERIJOS KATEDRA

TVIRTINU
Katedros vedėjas

(Parašas)

(Vardas, pavardė)

(Data)

Audrius Raškauskas

**PRARASTŲ PAKETŲ ĮTAKOS AUTOMATINIAM KALBOS
ATPAŽINIMUI TYRIMAS**

**RESEARCH OF IMPACT OF LOST PACKETS ON AUTOMATIC
SPEECH RECOGNITION**

Baigiamasis magistro darbas

Telekomunikacijų inžinerijos studijų programa, valstybinis kodas 62401T210

Telekomunikacijų technologijų specializacija

Elektros ir elektronikos inžinerijos mokslo kryptis

Vadovas Magr. Aurimas Anskaitis
(Moksl. laipsnis, vardas, pavardė)

(Parašas)

(Data)

Konsultantas _____
(Moksl. laipsnis, vardas, pavardė)

(Parašas)

(Data)

Konsultantas _____
(Moksl. laipsnis, vardas, pavardė)

(Parašas)

(Data)

Vilnius, 2008

Vilniaus Gedimino technikos universitetas

Elektronikos fakultetas

Telekomunikacijų inžinerijos katedra

ISBN ISSN

Egz. sk. 2

Data 2008-....-....

Telekomunikacijų inžinerijos studijų programos baigiamasis magistro darbas

Pavadinimas **Prarastų paketų įtakos automatiniam kalbos atpažinimui tyrimas**

Autorius **Audrius Raškauskas**

Vadovas magr. **Aurimas Anskaitis**

Kalba

☒ X

lietuvių

☐

užsienio

Anotacija

Baigiamojo magistro darbo tikslas – ištirti lietuviškų žodžių balsių ir priebalsių reikšmingumą automatiniam kalbos atpažinimui. Darbe aptartas automatinio kalbos atpažinimo aktualumas, trumpai apžvelgta kalbos atpažinimo tyrimų istorija. Apžvelgtos automatinio kalbos atpažinimo sistemos. Išanalizuotas dinaminio laiko skalės kraipymo algoritmas, aptarti šio algoritmo naudojami požymių vektoriai. Sudaryti automatinio kalbos atpažinimo tiriamųjų ir etaloninių žodžių žodynai, sumodeliuoti paketų praradimai tiriamųjų žodžių balsėse ir priebalsėse. Atliktas prarastų paketų įtakos automatiniam kalbos atpažinimui tyrimas ir nustatytas lietuviškų žodžių balsių ir priebalsių reikšmingumas automatiniam kalbos atpažinimui. Prarastų paketų įtakos automatiniam kalbos atpažinimui tyrimo rezultatai palyginti su prarastų paketų įtaka kalbos suprantamumui.

Darbą sudaro 9 dalys: įvadas, automatinio kalbos atpažinimo istorija, automatinio kalbos atpažinimo sistemos, dinamiško laiko skalės kraipymo algoritmas, dinamiško laiko skalės kraipymo algoritme naudojami požymių vektoriai, automatinio kalbos atpažinimo tyrimo žodynai, prarastų paketų įtakos automatiniam kalbos atpažinimui tyrimas, darbo rezultatų apibendrinimas, literatūros sąrašas.

Darbo apimtis – 55 p. teksto be priedų, 11 iliustr., 17 lent., 23 bibliografiniai šaltiniai.

Atskirai pridedami darbo priedai.

Prasminiai žodžiai: AMR kodekas, automatinis kalbos atpažinimas, dinamiškas laiko skalės kraipymas, melų skalės dažnio kepstro koeficientai, kepstas, paketų praradimai.

Vilnius Gediminas Technical University
Electronics faculty
Telecommunications Engineering department

ISBN ISSN
Copies No. 2
Date 2008-....-....

Telecommunications Engineering study programme (master) thesis.

Title: **Research of Impact of Lost Packets on Automatic Speech Recognition**

Author **Audrius Raškauskas**

Academic supervisor **M. Sc. Aurimas Anskaitis**

Thesis language

☒ Lithuanian

☐ Foreign (English)

Annotation

The purpose of this work is research of impact of lost packets on automatic speech recognition. The actuality of automatic speech recognition are discussed. Concisely reviewed the history of automatic speech recognition. In the task analysis part speech automatic speech recognition systems are presented. We traversed dynamic time warping algorithm and we discussed about feature vectors used in this algorithm are reviewed. The Library of Lithuanian words for automatic speech recognition research and packet loss models are made. Research of impact of lost packets on automatic speech recognition is done and the results of investigation are presented. The results of automatic speech recognition research and the results of speech intelligibility research are compared, too.

Structure: introduction, history of automatic speech recognition, automatic speech recognition systems, dynamic time warping algorithm, feature vectors used in dynamic time warping, library of words used in automatic speech recognition, research of impact of lost packets on automatic speech recognition, conclusions, references.

Thesis consist of: 55 p. text without extras, 11 pictures, 17 tables, 23 bibliographical entries.

Appendixes included.

Keywords AMR codec, automatic speech recognition, cepster, dynamic time warping, mel frequency cepstral coefficients, packet loss.

Turinys

IVADAS.....	9
1. AUTOMATINIO KALBOS ATPAŽINIMO ISTORIJA	11
1.1. Pirmosios automatinio kalbos atpažinimo sistemos	12
1.2. Bendraujančios mašinos kūrimas	14
1.3. Reziumė ir perspektyvos	15
2. AUTOMATINIO KALBOS ATPAŽINIMO SISTEMOS.....	17
3. DINAMIŠKO LAIKO SKALĖS KRAIPYMO ALGORITMAS	20
4. DINAMIŠKO LAIKO SKALĖS KRAIPYMO ALGORITME NAUDOJAMI POŽYMIŲ VEKTORIAI	23
4.1. Signalo kepstras	23
4.2. Kepstro koeficientų esmė	27
4.3. Melų skalės dažnio kepstro koeficientai	28
4.4. Tiesinės prognozės kodavimas	29
5. AUTOMATINIO KALBOS ATPAŽINIMO TYRIMO ŽODYNAI.....	32
5.1. Žodžių su prarastais paketais modeliavimas	32
5.2. Tiriamųjų žodžių žodynas	34
5.3. Etaloninių žodžių žodynas	34
6. PRARASTŲ PAKETŲ ĮTAKOS AUTOMATINIAM KALBOS ATPAŽINIMUI TYRIMAS	35
6.1. 1 bandymas.....	35
6.2. 2 bandymas.....	36
6.3. 3 bandymas.....	44
7. DARBO REZULTATŲ APIBENDRINIMAS	52
8. LITERATŪRA	54
PRIEDAI.....	56
A priedas. Etaloninių ir tiriamųjų žodžių žodynų kompaktinė plokštelė.....	56

Iliustracijų sąrašas

1 pav. Kalbos atpažinimo ir suprantamumo technologijų raida	16
2 pav. Dviejų žodžių palyginimo naudojant DTW iliustracija.....	20
3 pav. Testinio ir etaloninio žodžių palyginimo rezultatai, kai ištartas tas pats žodis.....	21
4 pav. Testinio ir etaloninio žodžių palyginimo rezultatai, kai ištarti skirtingi.....	22
5 pav. „A“ balsio garso slėgio grafikas.....	24
6 pav. Signalų energijos tankis dažnių srityje.....	24
7 pav. Signalų spektro modulio kvadrato logaritmas	25
8 pav. Kepstro koeficientų grafikas	26
9 pav. Koduoto ir dekodeto žodžio „mažas“ oscilogramos su prarastais paketais.....	33
10 pav. Vidutinio atstumo tarp požymių vektorių priklausomybė nuo prarastų paketų skaičiaus	39
11 pav. Vidutinio atstumo tarp požymių vektorių priklausomybė nuo prarastų paketų skaičiaus	47

Lentelių sąrašas

5.1 lentelė. Tiriamųjų žodžių žodynas	34
5.2 lentelė. Etaloninių žodžių žodynas	34
6.1 lentelė. Tyrimo rezultatai etaloninių žodžių lyginant su tuo pačiu paketų praradimų paveiktu etaloniniu žodžiu.....	35
6.2 lentelė. Tyrimo rezultatai prarandant paketus pirmoje žodžio raidėje (priebalsėje).....	36
6.3 lentelė. Tyrimo rezultatai prarandant paketus vidurinėje žodžio raidėje (priebalsėje).....	37
6.4 lentelė. Tyrimo rezultatai prarandant paketus žodžio balsėje.....	38
6.5 lentelė. Tyrimo rezultatai prarandant paketus pirmoje žodžio raidėje (priebalsėje).....	40
6.6 lentelė. Tyrimo rezultatai prarandant paketus vidurinėje žodžio raidėje (priebalsėje).....	41
6.7 lentelė. Tyrimo rezultatai prarandant paketus žodžio balsėje.....	42
6.8 lentelė. 2 bandymo rezultatų palyginimas	43
6.9 lentelė. Tyrimo rezultatai prarandant paketus pirmoje žodžio raidėje (priebalsėje).....	44
6.10 lentelė. Tyrimo rezultatai prarandant paketus vidurinėje žodžio raidėje (priebalsėje).....	45
6.11 lentelė. Tyrimo rezultatai prarandant paketus žodžio balsėje.....	46
6.12 lentelė. Tyrimo rezultatai prarandant paketus pirmoje žodžio raidėje (priebalsėje).....	48
6.13 lentelė. Tyrimo rezultatai prarandant paketus vidurinėje žodžio raidėje (priebalsėje).....	49
6.14 lentelė. Tyrimo rezultatai prarandant paketus žodžio balsėje.....	50
6.15 lentelė. 3 bandymo rezultatų palyginimas	50

Santrumpų sąrašas

ACELP	<i>Algebraic Code Exited Linear Prediction</i> – kalbos dekodavimo algoritmas, naudojantis tiesinę prognozę
AMR	<i>Adaptive Multi-Rate</i> – adaptyvus daugiadažnis kalbos kodekas
ARPA	<i>Advanced Research Projects Agency</i> – pažangių mokslinių tyrimų projektų agentūra
CSR	<i>Continuous Speech Recognition</i> – rišlios kalbos atpažinimas
DTW	<i>Dynamic Time Warping</i> – dinamiško laiko skalės kraipymo algoritmas
EDGE	<i>Enhanced Data rates for GSM Evolution</i> – mobiliųjų telefonų technologija, skirta pagerinti duomenų persiuntimą, padidinti greitį ir kokybę
EGPRS	<i>Enhanced GPRS</i> – patobulintas visuotinis paketinis radijo ryšys
FSN	<i>Finite State Network</i> – baigtinės struktūros tinklas
GPRS	<i>General Packet Radio Service</i> – visuotinis paketinis radijo ryšys
GSM	<i>Global Standart for Mobile Communications</i> – globalus mobiliųjų telefonų ryšio standartas
HMM	<i>Hidden Markov Model</i> – paslėptasis Markovo modelis
HTK	<i>Hidden Markov Model Tool Kit</i> – paslėptojo Markovo modelio įrankių komplektas
ISR	<i>Isolated Speech Recognition</i> – atskirų žodžių atpažinimas
LPC	<i>Linear Prediction Coding</i> – tiesinės prognozės kodavimas
MFCC	<i>Mel Frequency Cepstral Coefficient</i> – melų skalės dažnio kepstro koeficientas
SUR	<i>Speech Understanding Research</i> – kalbos supratimo tyrimai
TTS	<i>Text-to-Speech</i> – teksto keitimas į kalbą
VoWLAN	<i>Voice over Wireless Local Area Network</i> – balso perdavimas per bevielio ryšio vietinį tinklą

IVADAS

Šiandien kalbos atpažinimo tyrimai susiję su įvairiomis ir labai skirtingomis mokslo sritimis: biologija, kompiuterija, elektrotechnika, kalbotyra, matematika, fizika, psichologija. Šiose mokslo srityse atliekami darbai susiję su akustika, dirbtiniu intelektu, kompiuterių algoritmais, informacijos teorija, tiesine algebra, tiesinių sistemų teorija, vaizdų atpažinimu, fonetika, fiziologija, tikimybių teorija, signalų apdorojimu. Sprendžiant kalbos atpažinimo klausimą įdėta daug pastangų ir pasiekta nemažai gerų rezultatų. Tačiau, nepaisant akivaizdžios pažangos, net ir geriausių automatinių atpažinimo sistemų efektyvumas yra blogesnis už žmogaus atpažinimo tikslumą. Taigi dar yra tikrai daug erdvės tokių sistemų tobulinimui.

Atpažinimo paskirtis yra automatiškai nustatyti, kas yra sakoma informacijos priėmimo sistemos. Tai gali būti atskiras žodis, žodžių seka ar rišlūs sakiniai. Informacijos priėmimo sistema, nustačiusi, kas jai buvo pasakyta, atlieka atitinkamus veiksmus. Kalbos atpažinimo technologijų taikymo pavyzdžių yra daug, pavyzdžiui, žmogaus atpažinimas pagal balsą, dialogas su kompiuteriu, balso įrašų stenografavimas, balsu valdoma Interneto naršyklė ar kita programa, balsu valdomi automatikos įrenginiai ir dar daug kitų.

Moksliniuose kalbinių technologijų tyrimuose kalbėtojo atpažinimas pagal jo balsą yra suprantamas kaip procesas, kurio metu iš asmens kalbos signalų išskiriami identifikaciniai požymiai, pagal kuriuos atpažįstamas konkretus asmuo. Kaip pavyzdį galima paminėti asmens balso palyginimą su balsais iš turimos balso bazės. Dialogas su kompiuteriu – tiesioginis bendravimas su kompiuteriu balso pagalba. Vartotojui ištarus vieną iš komandų, kompiuteris, priklausomai nuo atpažintos komandos, atitinkamai sureaguoja, pavyzdžiui, parodo duomenis, pasako kiek valandų, pagroja dainą ar pan. Programinė įranga, skirta balso įrašams stenografuoti naudojama konferencijų, interviu, paskaitų metu. Interneto naršymas balsu yra pakankamai patogus – tam skirtos programos gali atlikti įvairias komandas: paleisti naršyklę, padaryti atidaryti nurodytą interneto svetainę, uždaryti naršyklę ir kt. Taip pat labai patogus ir praktiškas buitinius elektros prietaisus valdant balsu juos įjungti, išjungti ar atlikti kitokius veiksmus.

Perduodant iš vienos sistemos į kitą ar perkoduojant balso informaciją galimi skaitmeninių paketų praradimai ar sugadinimai. Prarasti paketai gali padaryti tam tikrą įtaką kalbos supratimui ar automatiniam atpažinimui. Baigiamojo magistro darbo metu bus tiriama prarastų paketų įtaka automatiniam kalbos atpažinimui.

Baigiamojo magistro darbo tikslas – ištirti lietuviškų žodžių balsių ir priebalsių reikšmingumą automatiniam kalbos atpažinimui. Raidžių reikšmingumas bus tiriamas modeliuojant paketų praradimus vienoje ar kitoje žodžio balsėje ar priebalsėje. Baigiamajame bakalauro darbe [5] buvo ištirtas paketais perduotų lietuviškų žodžių balsių ir priebalsių

reikšmingumas kalbos suprantamumui. Lietuviškų žodžių suprantamumui didžiausią įtaką daro prarasti paketai priebalsėse, kurios yra pirmos žodžių raidės. Šiek tiek mažesnę įtaką daro paketų praradimai žodžių viduryje esančioje priebalsėje. O mažiausiai reikšmingos prarastų paketų atžvilgiu yra balsės bet kurioje žodžio vietoje. Panašius rezultatus tikimasi gauti ir tiriant prarastų paketų įtaką automatiniam kalbos atpažinimui. Tačiau rezultatai gali ir skirtis. Tiriant kalbos suprantamumą yra galimybė žodžius tiesiog atspėti, nes žodžių atpažinime dalyvauja žmonės. Automatinio kalbos atpažinimo proceso metu mašinos naudoja tam tikrus algoritmus ir skaičiavimus, nėra galimybės atspėti žodžių.

1. AUTOMATINIO KALBOS ATPAŽINIMO ISTORIJA

Mašinos, kuri veiktų kaip žmogus, galėtų kalbėti ir taisyklinga šnekamąja kalba atsakyti, kūrimas domino mokslininkus ir inžinierius jau prieš daug metų. Nuo 1930-ųjų metų, kai Homeras Dudley'į iš „Bell Laboratories“ pasiūlė kalbos analizės ir sintezės modelį, prie automatinio kalbos atpažinimo vis labiau priartėta: nuo paprastų, atpažįstančių tik keletą garsų, iki sudėtingų, sugebančių atpažinti rišlią kalbą, sistemų. Šiandien automatinio kalbos atpažinimo sistemos panaudojamos tokioms užduotims, kaip automatinis skambučių apdorojimas ar įvairios informacijos pateikimas, atlikti. Apžvelkime kai kuriuos kelių pastarųjų dešimtmečių pagrindinius ir įsimintiniausius automatinio kalbos atpažinimo tyrimų ir raidos etapus.

Noras automatizuoti paprastas užduotis atsirado daugiau nei prieš šimtą metų. 1881 metais Alexanderis Grahamas Bellas, jo pusbrolis Chichsteris Bellas ir Charlesas Summeris Taineris sukūrė rašymo prietaisą. Šio išradimo pagrindu 1888 metais jie įkūrė įmonę „Volta Graphophone Co“ ir pradėjo gaminti garso įrašymo ir atkūrimo mašinas. 1907 metais jų patentą įsigijo „American Graphophone Co“ ir įregistravo šį prekės ženklą kaip „Dictaphone“. Beveik tuo pat metu Thomas Edisonas išrado fonografą – garso įrašymo tam tikrame velenėlyje ir atkūrimo prietaisą. Šių prietaisų paskirtis buvo palengvinti sekretorių darbą. Kalba pirmiausiai buvo įrašoma, po to automatiškai išspausdinama. Taip sumažėjo brangiai apmokamų stenografistų poreikis. XX-ojo amžiaus pradžia dar buvo vadinamas „įstaigų mechanizacijos“ laikotarpiu. Panašus įstaigų automatizavimas įvyko ir po šimtmečio. Apie 1990-uosius metus buvo automatizuojami skambučių centrai. Automatinio kalbos atpažinimo sistemos valdė skambučius nukreipdamas juos į reikiamus įstaigos skyrius.

Kalbos atpažinimo technologijomis domėjosi ir eiliniai gyventojai. Susidomėjimas atsirado dėl populiarių 1960-ųjų ir 1970-ųjų metų kino filmų. Stanley'io Kubricko kino filme „2001. Kosminė odisėja“ protingas kompiuteris „HAL“ kalbėjo natūraliu žmogaus balsu ir sugebėjo atpažinti bei suprasti rišlią kalbą. Šis kompiuterio antropomorfizmas supažindino plačiąją visuomenę su protingų mašinų potencialu. Garsiojoje žvaigždžių karų sagoje George'as Lucasas išplėtė protingų mašinų gabumus, padarydamas jas mobilias ir protingas, sugebančias atpažinti ir suprasti kalbą, judėti ir bendrauti su žmonėmis ir kitomis mašinomis.

Pastangos sukurti mašinas, galinčias imituoti žmonių bendravimą, prasidėjo XVIII-ojo šimtmečio antroje pusėje. Pirmiausiai buvo domimasi kalbėjimo mašinų kūrimu, o ne atpažinimo ar supratimo problema. 1773 metais Rusijos mokslininkui Christianui Kratzensteinui naudojant rezonansinius vamzdžius pavyko sudaryti balse. 1971 metais Wolfgangas von Kempelenas sukonstravo akustinę-mechaninę kalbos mašiną, 1800-ųjų metų viduryje Charlesas Wheatstone'as modifikavo Kempeleno kalbėjimo mašiną panaudodamas odinius rezonatorius.

Šioje mašinoje buvo galima keisti rezonatorių konfigūraciją ir sukurti skirtingus į kalbą panašius garsus.

Pirmojoje XX-ojo amžiaus pusėje Harvey'io Fletcherio ir kitų „Bell Laboratories“ darbuotojų tyrimai įrodė kalbos spektro (kalbos garso energijos pasiskirstymo dažnių srityje) ir jo garsinių savybių, tokių kaip suprantamumas, sąryšį. 1930 metais Homeris Dudley's sukūrė kalbos sintezatorių ir pavadino „VODER“. „VODER“ buvo Charleso Wheatstone'o mechaninės kalbėjimo mašinos elektrinis ekvivalentas su mechaniniu valdymu. 1939 metais „VODER“ buvo pademonstruotas pasaulinėje parodoje Niujorke ir buvo laikomas svarbiu mašininio kalbėjimo raidos įvykiu.

Harvey Fletcher'is ir Homer'is Dudley tvirtai nustatė signalo spektro svarbą fonetinės prigimties kalbos garsų identifikavimui. Po šių žymių mokslininkų tyrimų šiuolaikinės kalbos atpažinimo sistemos ir algoritmai pagrįsti kalbos signalo galios spektro matavimais. Su šiuolaikine skaitmenine signalų apdorojimo technika tai lengvai realizuojama.

1.1. Pirmosios automatinio kalbos atpažinimo sistemos

Pirmųjų bandymų sukurti automatinio kalbos atpažinimo sistemas metu daugiausia remtasi akustinės fonetikos teorija, kuri apibūdina fonetinius kalbos elementus (pagrindinius kalbos garsus) ir bando paaiškinti, kaip jie yra akustiškai įgyvendinti šnekamojoje kalboje. 1952 metais DAVISAS, BIDDULPHAS, ir BALASHEKAS iš „Bell Laboratories“, naudodami formantų¹ dažnius, įvertintus kiekvieno žodžio balsės srityse, sukūrė atskirų vieno kalbėtojo vienaženklių skaičių atpažinimo sistemą.

Apie 1950 metus HARRY'IS OLSONAS ir HERBERTAS BELARAS iš „RCA Laboratories“ sukūrė sistemą, kuri atpažino vieno kalbėtojo dešimt skiemenų, o „MIT Lincoln“ laboratorijoje JAMESAS ir CARMA FORGIE'IAI sukūrė nuo kalbėtojo nepriklausomą dešimties balsių atpažinimo sistemą. 1960 metais keletas Japonijos laboratorijų sukūrė specialią aparatinę įrangą kalbai atpažinti. Žinomiausios sistemos buvo J. SUZUKI ir K. NAKATA iš Tokijo radijo tyrimų laboratorijos sukurta balsių atpažinimo sistema, J. SAKAI ir S. DOSHITA iš Kioto universiteto sukurta fonemų atpažinimo sistema ir NEC laboratorijoje sukurta skaičių atpažinimo sistema. J. SAKAI ir S. DOSHITA savo darbe pirmą kartą kalbos analizei panaudojo kalbos segmentavimą. Kioto universiteto darbą galima laikyti rišlios kalbos atpažinimo sistemų pirmtaku.

DENISAS FRY'US ir PETERIS DENESAS iš Anglijos Universiteto sukūrė fonemų atpažinimo įrenginį, kuris atpažino keturias balses ir devynias priebalses. Įtraukę statistinę informaciją apie

¹ *Formantas* [lot. formans (kilm. formantis) – sudarantis] – žodžių darybos priemonė ar priemonių visuma, skirianti darinį nuo pamatinio žodžio.

galimas fonemų sekas, jie padidino fonemų atpažinimo tikslumą. Šiame darbe pirmą kartą panaudota statistinė sintaksė automatiniam kalbos atpažinimui.

1960 metais Tarybų Sąjungos mokslininkas Taras Vintsyukas laikui palyginti tarp dviejų sekų pasiūlė panaudoti dinamišką programavimą. Truputį vėliau H. Sakoe ir S. Chiba kalbos modelių palyginimui pasiūlė formalesnius metodus, žinomus kaip dinamiškas laiko skalės kraipymas (angl. *Dynamic Time Warping – DTW*). Nuo 1970-ųjų pabaigos, daugiausia dėl H. Sakoe ir S. Chiba publikacijų, dinamiškas programavimas tapo nepakeičiama automatinio kalbos atpažinimo technika.

1960 metų pabaigoje B. S. Atalas ir F. Itakura suformulavo pagrindines tiesinės prognozės kodavimo (angl. *Linear Prediction Coefficients* arba *Linear Prediction Coding – LPC*). Iki 1970-ųjų vidurio Itakura, Rabineris, Levinsonas ir kiti pateikė pagrindus, kaip kalbai atpažinti pritaikyti pavyzdžių atpažinimo technologiją, pagrįstą tiesinės prognozės kodavimu.

1970 metais Tomas Martinas įkūrė „Treashhold Technology Inc.“ ir sukūrė pirmą tikrą automatinio kalbos atpažinimo produktą „VIP-100“. Sistema buvo tik paprastiems darbams atlikti, pavyzdžiui, gaminių kokybės kontrolei ar jų rūšiavimui ant konvejerio. Tačiau „VIP-100“ sistema darė įtaką JAV gynybos departamento pažangių mokslinių tyrimų projektų agentūrai (angl. *Advanced Research Projects Agency – ARPA*), kad ši finansuotų kalbos suprantamumo tyrimų programą (angl. *Speech Understanding Research – SUR*). Viena iš sistemų, sukurtų ARPA rangovų, buvo „Harpy“. Ji sugebėjo naudodama 1011 žodžių žodyną priimtinu tikslumu atpažinti kalbą. „Harpy“ sistemoje pirmą kartą panaudotas baigtinės struktūros tinklas (angl. *Finite State Network – FSN*). Tačiau metodai, kurie naudojo baigtinės struktūros tinklus, nepaplitę iki 1990-ųjų metų.

1970 metais lygiagrečiai ARPA darbams atsirado dvi aiškios kalbos atpažinimo tyrimų kryptys, kurias vystė „IBM“ ir „AT&T Laboratories“. „IBM“ visą dėmesį sutelkė į balsu aktyvintos rašomosios mašinėlės, kurios svarbiausia funkcija paversti kalbą į raidžių ar žodžių seką, kūrimą. „AT&T“ tyrinėjimų tikslas – pateikti visuomenei automatizuotas telekomunikacijų paslaugas, tokias kaip numerio rinkimas balsu. „IBM“ ir „AT&T“ požiūris į kalbos atpažinimą padarė didelę įtaką žmogaus-mašinos komunikacijos technologijų raidai. Nepaisant skirtumų tarp šių požiūrių, buvo ir viena bendra kryptis: tyrimuose atsirado matematinis formalizmas ir tikslumas.

1980 metai išpopuliarino statistinius metodus, ypač paslėptąjį Markovo modelį (angl. *Hidden Markov Model – HMM*), kuris tapo vienu iš pagrindinių atpažįstant kalbą. Paslėptasis Markovo modelis yra dvigubai stochastinis procesas, kuris modeliuoja kalbos signalo kintamumą. Realus kalbos signalas iš prigimties yra kintamas. Žmogui ištarus tą patį žodį

akustiniai signalai gali stipriai skirtis. Paslėptasis Markovo modelis yra pagrindinė kalbos atpažinimo sistemos dalis, kuri vykdo statistinį pavyzdžių atpažinimo metodą.

Kita technologija, kuri buvo pristatyta 1980-ųjų metų pabaigoje, buvo dirbtiniai neuronų tinklai. Neuronų tinklai sukurti dar 1950 metais, bet iš pradžių nesuteikė reikiamų rezultatų. 1980 metais pasirodė lygiagrečiai paskirstyto apdorojimo modelis, kuris atgaivino mintį imituoti žmogaus nervų veikimo mechanizmą. Pirmiausiai neuronų tinklai buvo naudojami paprastoms užduotims, tokioms kaip kelių fonemų ar kelių žodžių atpažinimui. Vėliau, plėtojant tyrinėjimus neuronų tinklai buvo sujungti su paslėptuoju Markovo modeliu.

1990 metais padaryta didelė pažanga kuriant programinę įrangą. Pavyzdžiui, Kembridžo universitete buvo sukurtas paslėptojo Markovo modelio įrankių komplektas (angl. *Hidden Markov Model Tool Kit – HTK*). Jis buvo ir šiandien yra vienas iš plačiausiai naudojamų programinė įrangos įrankių automatiniam kalbos atpažinimui.

1.2. Bendraujančios mašinos kūrimas

Iki 1980-ųjų metų kalbos atpažinimo tyrimai nagrinėjo vieną pagrindinę problemą: kaip kalbos garsų bangos formą pakeisti į žodžius. Daugelis tyrėjų manė, kad kalbos pavertimas į tekstą turėjo būti pirmas žingsnis procese, kuris įgalintų mašiną suprasti žmogaus kalbą ir tinkamai į ją atsakyti. Analizuojant kalbos atpažinimo ir supratimo technologijas, buvo rastos dvi svarbios detalės žmogaus ir mašinos bendravimo procese. Pirma, kalbos atpažinimo sistemų vartotojai linkę kalbėti natūraliais sakiniais, kuriuose gali būti žodžių ne iš žodyno, negramatinių sandarų ar blogai suformuluotų sakinio dalių. Taip pat kalba dažnai iškraipoma aplinkos triukšmo, pašalinių garsų ar kitų žmonių kalbos. Antra, kaip ir žmonių bendravime, taip ir kalbant su mašina reikia, kad dialogas pasiektų tam tikrą supratimo lygį. Taigi buvo svarbu sukurti dialogų su mašiną valdymą. 1990-ųjų metų pradžioje sukurta daug sistemų, naudojančių dialogų valdymą. Taip tikėtasi galų gale sukurti mašiną, atkartojančią žmogaus bendravimą. Kaip pavyzdžius galima paminėti Victorio Zue'io iš Masačusetso technologijos universiteto sistemas „Pegasus“ ir „Jupiter“, bei Al Gorino iš „AT&T“ sistemą „HMIHI“ arba „How May I Help You“.

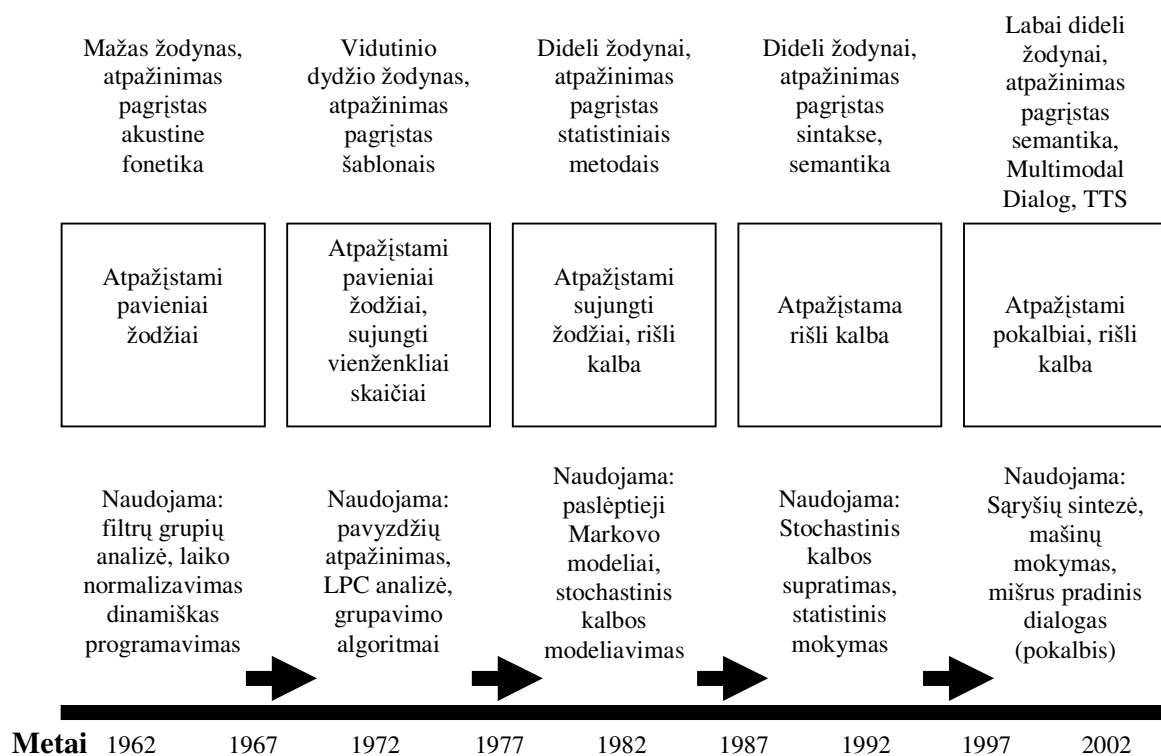
„Pegasus“ – tai pokalbių sistema, kuri klientams telefonu suteikdavo informaciją apie avialinijų skrydžius. „Jupiter“ panaši sistema, bet klientams suteikdavo informaciją apie orus. Šiose sistemose buvo realizuotas efektyvus dialogų valdymas ir vartotojai su mašina bendraudavo tarsi su kitu žmogumi. „HMIHY“ sistema buvo skirta skambučių pagalbos linijos vartotojų klausimams nukreipti. Šios automatinio kalbos atpažinimo ir supratimo sistemos nebuvo idealiai tikslios, tačiau jomis naudojosi daug vartotojų [2].

1.3. Reziumė ir perspektyvos

1 paveiksle pavaizduota kalbos atpažinimo ir suprantamumo technologijų raida. 1960 metais buvo galima atpažinti mažų žodynų (10 – 100 žodžių) pavienius žodžius. Žodžių atpažinimas pagrįstas akustinėmis-fonetinėmis kalbos garsų savybėmis. Tuo laiku buvo naudojami paprasti laiko normalizavimo metodai bei pradėti vystyti pažangesni dinamiško programavimo metodai. 1970 metais buvo kuriami vidutinio dydžio žodynų (100 – 1000 žodžių) šablonais pagrįsti atpažinimo modeliai. Taip pat šiuo laikotarpiu pradėti naudoti tiesinės prognozės kodavimo metodai (spektriniam vaizdavimui), pavyzdžių grupavimo metodai (nuo kalbančiojo nepriklausomai kalbai atpažinti) bei dinamiško programavimo metodai (atpažinti rišliai kalbai). 1980 metais pradėta dirbti su 1000 žodžių ir didesniais žodynais. Šio laikotarpio kalbos atpažinimo sistemų veikimas paremtas statistiniais metodais. Pradėti naudoti paslėptieji Markovo modeliai ir stochastiniai kalbos modeliai, kurie leido sukurti tikslus metodus atpažinti rišliai kalbai. 1990 metais sukurtos sistemos su neriboto žodžių skaičiaus žodynais, skirtos atpažinti ir suprasti rišlią kalbą. Šiuo laikotarpiu taip pat remtasi statistiniais metodais. XXI-ojo amžiaus pradžioje pradėtos kurti neriboto dydžio žodynų sistemos su pilnais semantiniais modeliais bei integruotomis teksto-į-kalbą (angl. *Text-to-Speech – TTS*) sintezės sistemoms. Atpažįstama ne tik rišli kalba, bet ir dialogai.

Po beveik penkiasdešimties tyrimų metų, kalbos atpažinimo sistemos pateko į rinką ir įvairiais būdais įvairiems vartotojams davė naudos. Pagrindiniai atpažinimo sistemų progreso etapai buvo 1960-1970 metai, kai pristatyti tiesinės prognozės ir kepsrinės analizės metodai, bei 1980 metai, kai pradėti taikyti statistiniai metodai, pagrįsti paslėptuoju Markovo modeliu. Visa tai pasiekta universitetų, privačių įmonių ir vyriausybės organizacijų tyrimų dėka. Tobulėjant technologijoms atsiras vis naujų kalbos atpažinimo sistemų pritaikymo būdų ir tai taps viena iš mūsų kasdienybės dalių, ypač jei bus sukurtos mašinos, galinčios imituoti žmogaus kalbą.

Taigi, šių laikų iššūkis – sukurti mašiną, kuri veiktų taip pat, kaip elgiasi protingas žmogus. Ligšioliniai bandymai tai tik pradžia, ir tikriausiai reikės dar daug metų, kol kas nors sukurs mašiną, savo gabumais prilygstančią žmogui.



1 pav. Kalbos atpažinimo ir suprantamumo technologijų raida

2. AUTOMATINIO KALBOS ATPAŽINIMO SISTEMOS

Kalbant apie automatinio kalbos atpažinimo istoriją buvo paminėtos įvairios atpažinimo metodikos. Šiame skyriuje bus aptarti kalbos atpažinimo metodai ir jų klasifikavimas.

Kalbos atpažinimo sistemos dažniausiai klasifikuojamos kaip pavienių žodžių ir tolydžios kalbos. Pavienių žodžių sistemos kiekvienam žodžiui, žodžių kombinacijai, ar frazėms turi atskirą akustinį modelį, dar vadinamą etalonu. Atskirų žodžių sistemoje po kiekvieno žodžio operatorius turi padaryti pauzę, kuri palengvina kalbos atpažinimo užduotį. Tai paprasčiausia atpažinimo forma, nes lengviau surandami žodžio pradžios ir galo taškai. Be to, žodžio tarimas nepaveikia kitų žodžių. Šios sistemos dar vadinamos izoliuotu žodžių atpažinimu (angl. *Isolated Speech Recognition – ISR*). Nepertraukiamos kalbos atpažinimo (angl. *Continuous Speech Recognition – CSR*) sistemos atpažinimo procese naudoja rišlią kalbą. Rišlios kalbos sistemoje žodžiai yra neatskirti pauzių. Tolydžią kalbą daug sudėtingiau traktuoti dėl daugybės įvairių efektų. Pirmiausia, sunku surasti žodžių pradžios ir pabaigos taškus. Kita problema yra ko-artikuliacija. Kiekviena fonema yra veikiamą kitų fonemų, žodžių pradžios ir pabaigos yra veikiamos gretimų žodžių. Rišlios kalbos atpažinimui taip pat daro įtaką kalbos greitis. Greita kalba sunkina atpažinimo procesą.

Kalbos atpažinimo sistemos dar klasifikuojamos į nuo kalbėtojo priklausomas, nepriklausomas ar adaptyvias. Nuo kalbėtojo priklausomai sistemai svarbu, kad vartotojas prieš vykdydamas atpažinimo procesą įrašytų žodžio, sakinio ar frazės etaloną, kitaip tariant, vartotojas „moko“ sistemą atpažinti žodžius. Kai kurioms nuo kalbėtojo priklausomoms sistemoms užtenka, kad vartotojas įrašytų tik sistemos žodyno dalį, o sistema sugebėtų atpažinti visą žodyną. Nuo kalbėtojo nepriklausomai sistemai prieš ją naudojant nereikia nieko įrašyti. Šios sistemos sukurtos taip, kad veiktų su kiekvienu tam tikro tipo diktoriaus, pavyzdžiui, anglų kalba kalbančiu vartotoju. Adaptyvios atpažinimo sistemos sukurtos taip, kad jų veikimas prisitaiko prie naujo vartotojo kalbos bruožų.

Svarbus elementas kuriant kalbos atpažinimo sistemas yra žodyno dydis. Žodynas daro įtaką atpažinimo sudėtingumui, apdorojimo reikalavimams ir sistemos tikslumui. Akivaizdu, kad daug lengviau tiksliai atpažinti žodį, jei žodyne yra, pavyzdžiui, dvidešimt žodžių, o ne šimtas ar tūkstantis. Kalbos atpažinimo sistema žodyne ieško fonemų ir žodžių, o kuo didesnis žodynas, tuo sudėtingiau vykdyti paiešką.

Visiems galimiems žodžiams, panaudotiems sistemoje, žodynas apibūdina kiekvienam kalbos garsui leistinus akustinius modelius. Reikėtų pažymėti, kad žodyne gali būti skirtingai

ištartas tas pats žodis, pavyzdžiui, mēlas² ir mélas³. Paprastai yra du žodynų tipai: sistemos žodynas ir vartotojų žodynas. Sistemos žodyną sudaro pastovus žodžių sąrašas, kurį programinė įranga atpažįsta. Kartais galima paimti pilno sistemos žodyno dalį, kuri atpažinimo proceso metu bus greičiau apdorojama. Vartotojų žodynai paprastai kuriami atskirai, bet gali būti ir sistemos žodyno poaibiai.

Dauguma šiuolaikinių kalbos atpažinimo sistemų garsų sekų interpretavimui naudoja tikimybinis modelius. Tikimybiniai modeliai naudojami surinkti informacijai apie labiau tikėtinas žodžio kombinacijas. Taip gerinamas kontekstu pagrįstas palyginimas ir padidinamas žodžio tikslumas atpažįstant kalbą.

Akustinių modelių grupės gali būti mokomos daugelio vartotojų įrašyta kalba. Šios modelių grupės įvertina individualios diktoriaus tarties (dialekto), akcento ar panašius pokyčius. Svarbu etalonines modelių grupes mokyti aplinkoje, panašioje į tą, kurioje bus naudojamas atpažinimo procesas, t. y., negalima mokyti etaloninio komplekto naudojant ausinių mikrofoną, jei atpažinimo aplinka bus susijusi su balso perdavimu telefono linijomis. Kalbos atpažinimo algoritmas neapdorotus akustinius duomenis palygina su etalonų akustinių modelių sekomis.

Kalbos atpažinimo procesas prasideda nuo skaitmeninių pavyzdžių, išreikštų vartotojo žodžiais įvesties duomenyse, atrinkimo. Šie įvesties duomenys nebūtinai turi būti tiesiogiai įvedami vartotojo. Jie gali būti ir iš kito šaltinio, pavyzdžiui, iš kompiuterio atmintyje esančio žodyno ar iš kitų įvesties įrenginių.

Kitas žingsnis – akustinio signalo apdorojimas. Žodinė įvesties informacija, pakeista į skaitmeninę formą, yra skaidoma į stebėjimo sekas. Čia tikimasi, kad stebėjimo metu bus tiksliai atvaizduoti vartotojo žodžiais išreikšti įvesties duomenys. Didžioji dalis atpažinimo algoritmų naudoja spektrinę analizę, pavyzdžiui, tiesinės prognozės kodavimo analizę, melų skalės dažnio kepstro koeficientus (angl. *Mel Frequency Cepstral Coefficient* – MFCC), ausies sraigės modeliavimą ir kitus.

Atskiros stebėjimų sekos lyginamos su žinomų akustinių modelių komplektu. Kiekvienas modelis atvaizduoja fonemą. Modelių komplektas su žodžiu ar fraze yra suderinamas naudojant žodyną. Žodynas apibrėžia kiekvieno žodžio tarimą, kaip fonemų komplektą. Šis veiksmas gali būti atliktas naudojant įvairius procesus, tokius kaip paslėptieji Markovo modeliai, dinamiško laiko kraipymo algoritmas, neuronų tinklai ar šių procesų kombinacijos. Tačiau dažniausiai naudojamos paslėptaisiais Markovo modeliais pagrįstos sistemos.

Atpažinimo fazėje, jau egzistuojantys išmokyti akustiniai modeliai lyginami su apdorotais balso įvesties duomenimis. Dekoderis (pavyzdžiui, „Viterbi“ ar „Baum-Welch“) sutapdina balso

² Mēlas – nesisteminis garso aukščio vienetą, vartojamas muzikinėje akustikoje.

³ Mélas – tyčia sakoma neteisybė.

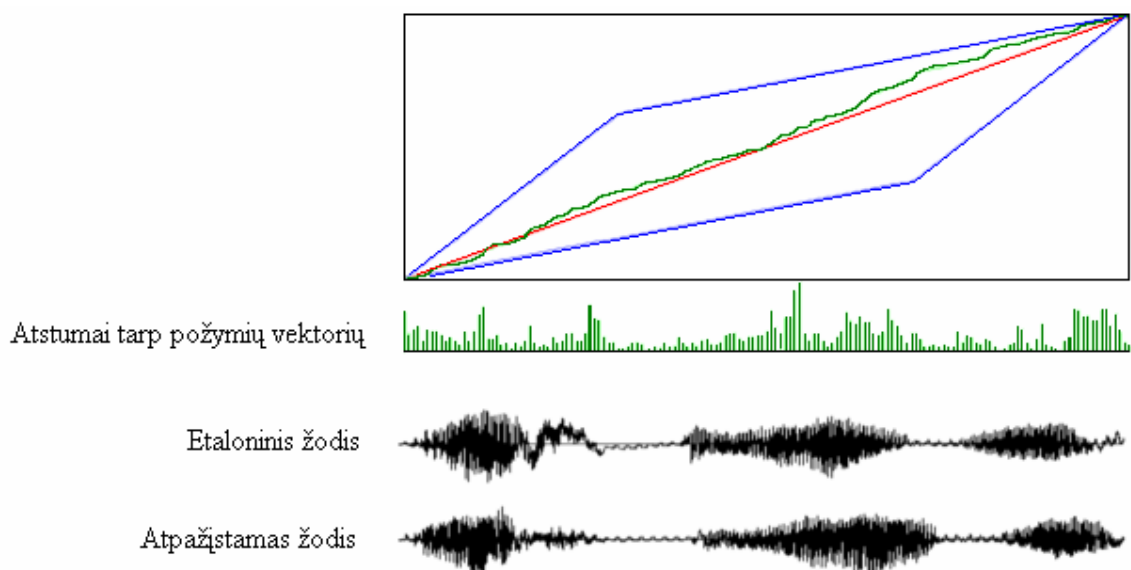
įvesties duomenis su labiausiai tikėtinais akustiniais modeliais. Dekoderis užrašo įvesties duomenis į tekstinių simbolių seką, kurias galima tiesiogiai apdoroti. Lyginant simbolius su akustiniais modeliais stengiamasi juos sugrupuoti į atpažįstamas grupes. Šios fazės rezultatas yra įvesties duomenų apytikslis įvertis, kuris panaudojamas atpažinimo procesui pabaigti.

Svarbu paminėti, kad automatinis kalbos atpažinimas blogai veikia triukšme, ypač kai trukdžiai yra kitų žmonių kalba. Žmonės yra pakantūs triukšmingoms aplinkoms, bet didėjant triukšmui automatizuoto kalbos atpažinimo rezultatai stipriai blogėja. Signalų iškraipymai dėl foninės žmonių kalbos sukelia daug sunkumų. Esant šnekamosios kalbos triukšmams tolerantiškiausios yra atpažinimo sistemos, kurios pagrįstos dirbtiniais neuronų tinklais.

Apibendrinant galima pasakyti, kad atskirų žodžių atpažinimo sistemos sukuria žymiai lengvesnes užduotis mašinoms, negu rišlios kalbos atpažinimo sistemos. Nuo kalbėtojo priklausomas sistemas lengviau sukurti ir naudoti, taip pat jos yra tikslesnės už nepriklausomas nuo kalbėtojo sistemas. Dėl šių priežasčių pirmosios kalbos atpažinimo sistemos buvo nuo diktorius priklausomos pavienių žodžių atpažinimo sistemos su riboto dydžio žodynu. To meto iššūkis buvo taip patobulinti kalbos atpažinimo technologijas, kad su jomis būtų galima pagerinti žmogaus ir kompiuterio sąveiką [11],[12].

3. DINAMIŠKO LAIKO SKALĖS KRAIPYMO ALGORITMAS

Sprendžiant pavienių žodžių atpažinimo uždavinį, reikia juos suvienodinti laiko atžvilgiu ir palyginti atpažįstamą žodį su etaloniniu. Atpažįstamo ir etaloninio žodžių trukmės ir atskirų garsų trukmės juose paprastai skiriasi, todėl jų palyginimas yra gana sudėtingas uždavinys. Vienas iš šio uždavinio sprendimo būdų yra kalbos pavyzdžių dinaminis laiko skalės kraipymas, kuris remiasi dinamiško programavimo metodu. Pabandykime suprasti dinaminio laiko skalės kraipymo algoritmo veikimą paanalizuodami keletą pavyzdžių. Dinaminis laiko skalės kraipymas pavaizduotas 2 paveiksle.

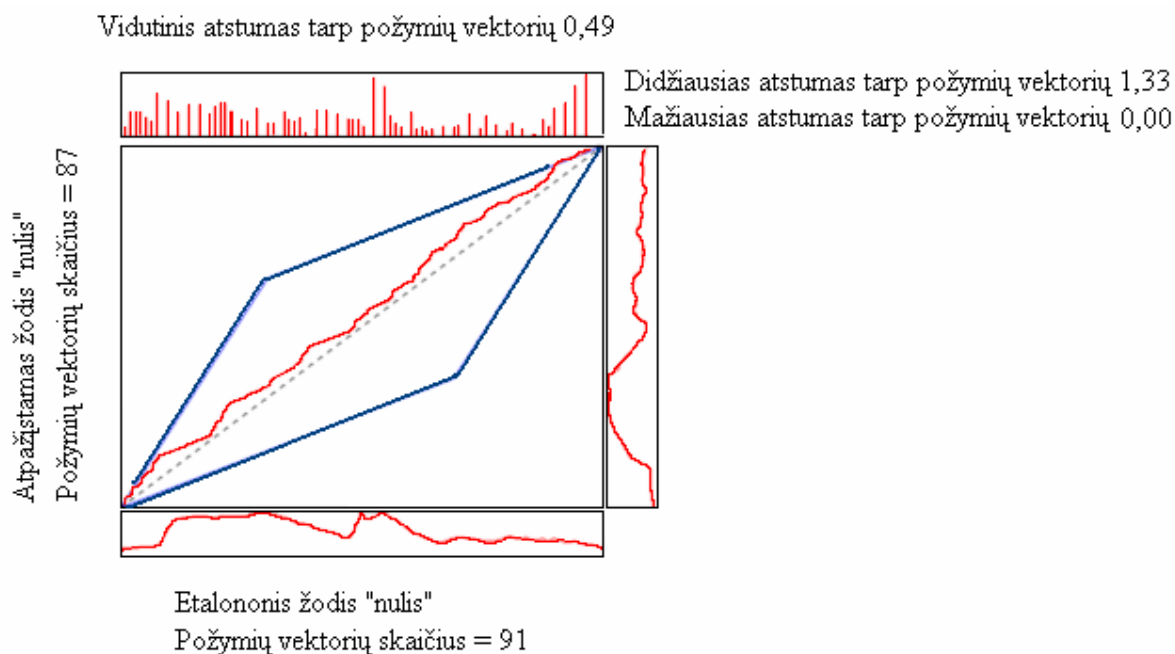


2 pav. Dviejų žodžių palyginimo naudojant DTW iliustracija

2 paveiksle esančio stačiakampio abscisių ašyje yra išdėstyti etaloninio žodžio požymiai, o ordinačių ašyje – testinio žodžio požymiai. Etaloninio žodžio ir atpažįstamojo žodžio signalų oscilogramos pavaizduotos apačioje. Lygiagretainis žymi geriausios požymių laiko skalės kraipymo trajektorijos paieškos sritį. Lygiagretainio įstrižainė žymi tiesinio laiko skalės kraipymo trajektoriją. Vingiuota kreivė žymi dinaminio laiko skalės kraipymo trajektoriją. Ši trajektorija atitinka geriausią etaloninio ir lyginamojo žodžių požymių sutapdinimą. Sutapdinimo kokybės rodiklis yra vidutiniai iškraipymai, paskaičiuoti esant geriausiai požymių laiko skalės kraipymo trajektorijai. Kuo vidutiniai iškraipymai mažesni, tuo sutapdinimas geresnis. Lyginant žodžius, kai ištartas tas pats tekstas, vidutiniai iškraipymai yra dažniausiai mažesni negu lyginant žodžius, kai ištartas skirtingas tekstas (3 ir 4 paveikslai). Norint atpažinti žodį naudojant

dinaminį laiko skalės kraipymo algoritmą, turime palyginti jį su visų žodyno žodžių etalonais. Etalonas, kuriam vidutiniai iškraipymai bus mažiausi ir bus atpažintas žodis.

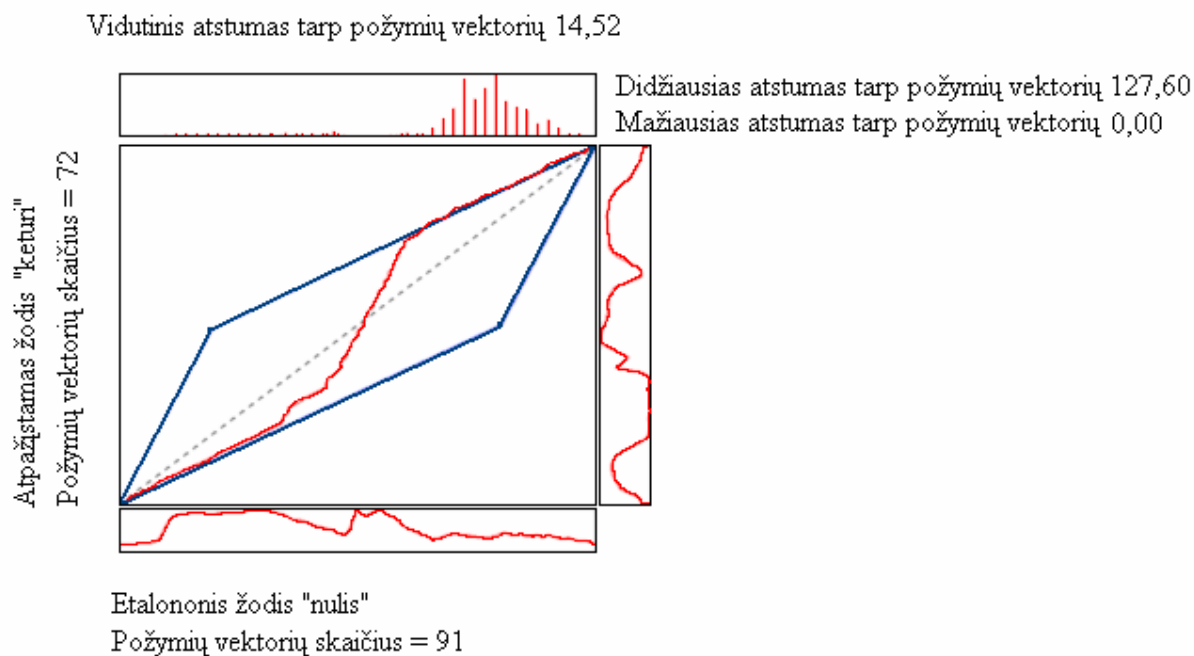
3 paveikslėlyje pavaizduotas testinio žodžio „nulis“ palyginimas su etaloniniu žodžiu „nulis“. Abscisių ašyje yra vaizduojamas etaloninio žodžio laikas, ordinačių – testinio. Taip pat užrašyta kiek požymių vektorių yra etaloniniame žodyje ir kiek testiniame. Šalia pagrindinio grafiko pavaizduotos etaloninio ir testinio žodžių energijos kreivės. Pagrindiniame grafike mėlyna spalva pavaizduotos optimalaus laiko skalės išlyginimo trajektorijos paieškos ribos. Punktyrinė linija žymi tiesinio laiko skalės išlyginimo trajektoriją. Šalia šios trajektorijos raudona spalva pavaizduota optimalaus laiko skalės išlyginimo trajektorija. Virš pagrindinio grafiko pavaizduoti atstumai tarp požymių vektorių, kurių indeksai yra taškai, kuriuos kerta pavaizduota optimalaus laiko skalės išlyginimo trajektorija. Jeigu šių atstumų reikšmių diapazonas yra nedidelis, tai vizualiai galime pastebėti žodžių panašumą. Šio grafiko dešinėje yra užrašytos šių atstumų minimali ir maksimali reikšmės. Jeigu žodžių pradžios ir (arba) galo taškai nustatyti neteisingai, tose vietose gaunami labai dideli atstumai, kurie gerai matomi grafike.



3 pav. Testinio ir etaloninio žodžių palyginimo rezultatai, kai ištartas tas pats žodis

Pagrindinis dviejų žodžių palyginimo rodiklis yra vidutinis atstumas tarp požymių vektorių lyginant du žodžius ir esant optimaliam laiko skalės išlyginimui. Šio pavyzdžio atveju jis yra lygus 0,49.

Jeigu yra lyginami skirtingi žodžiai, vidutinis atstumas gaunamas daug didesnis. Tai galime matyti 4 paveiksle.



4 pav. Testinio ir etaloninio žodžių palyginimo rezultatai, kai ištarti skirtingi

Šiame pavyzdyje testinis žodis yra „keturi“, o etaloninis žodis yra „nulis“. Kaip matome iš paveikslėlio, optimalaus laiko skalės išlyginimo trajektorijos gale atstumai tarp požymių vektorių yra labai dideli ir vidutinis atstumas tarp žodžių taip pat labai didelis (14,52). Išanalizavę 3 ir 4 paveiksluose pateiktus pavyzdžius, galime suprasti automatinio kalbos atpažinimo esmę. Lyginant testinį žodį su žodžių etalonais optimalaus laiko skalės išlyginimo trajektorijai vidutinis atstumas paprastai gaunamas mažiausias, kai testinis ir etaloninis žodžiai yra tas pats žodis, nors jis ir ištartas skirtingais laiko momentais ar skirtingų diktorių [10].

4. DINAMIŠKO LAIKO SKALĖS KRAIPYMO ALGORITME NAUDOJAMI POŽYMIŲ VEKTORIAI

Dinaminio laiko kraipymo algoritme naudojami požymių vektoriai: signalo kepstras, melų skalės dažnio kepstro koeficientai, tiesinės prognozės kodavimas. Šiame skyriuje aptarsime šiuos požymių vektorius.

4.1. Signalo kepstras

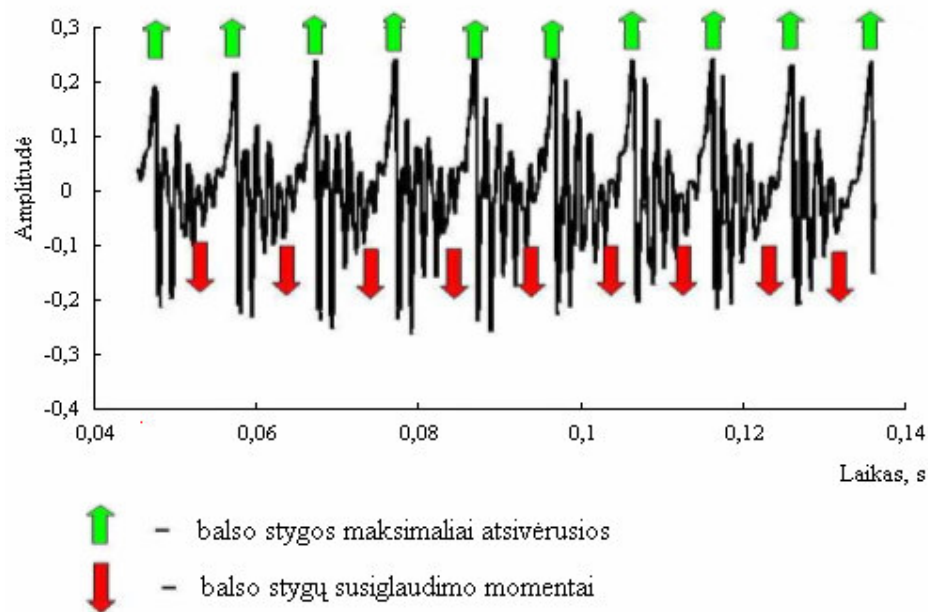
Kepstriniai metodai buvo sukurti plėtojantis spektrinės analizės metodams. Kepstro taikymo sritys įvairios: radiolokacija, hidrolokacija (sėkmingai naudojamas triukšmų mažinimui atspindėjusiuose signaluose), kalbos analizė (įvertinamas pagrindinis kalbančiojo dažnis), jūros ir sausumos seisminių reiškinių aptikimas ir t. t.

Kepstrinės analizės rezultatai labai priklauso nuo duomenų prigimties. Todėl remiantis empiriniais rezultatais, gautais vienoje srityje galima daryti ne tiek jau daug apibendrinimų teisingų ir kitoms sritims. Savo esme kepstriniai metodai tinka atspindžius turintiems signalams. Kepstrinės analizės privalumas yra informacijos apie fazę išsaugojimas. Tai leidžia atkurti signalo laiko struktūrą.

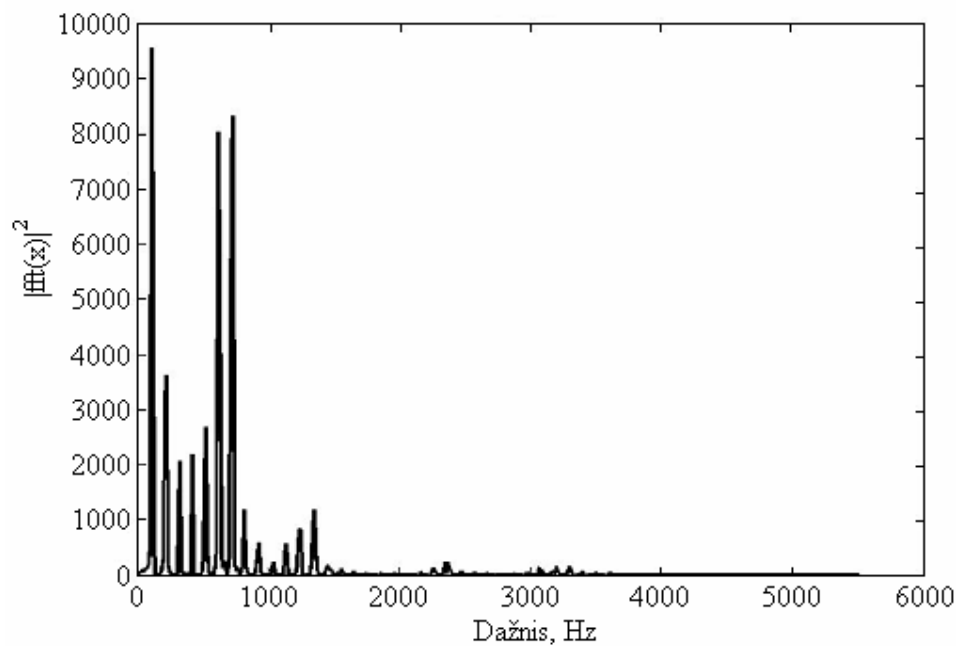
Kalbos signalo kepstro koeficientai teikia vertingą informaciją apie pagrindinio tono arba balso stygų svyravimų dažnį. Periodiškai virpant balso stygom garso slėgis tai sustiprėja, tai susilpnėja. Tik atsivėrus balso stygom garso slėgis pradeda stiprėti, ir atvirkščiai – kai balso stygos susispaudžia – garso slėgis sumažėja. Tai iliustruoja „a“ balsio garso slėgio grafikas, pavaizduotas 5 paveiksle.

Tačiau kaip įvertinti balso stygų susitraukimo dažnį kompiuteriu. Galima mėginti tiesiogiai vertinti garso slėgio lokalius maksimumo ir minimumo taškus ir kurti kokį nors algoritmą, kuris įvertintų ar eilinis lokalus ekstremumas žymi balso stygų maksimalaus atsivėrimo momentą. Tačiau tokį metodą sunku realizuoti bei duos nelabai tikslų ir patikimą rezultatą.

Tačiau galime apskaičiuokime to paties signalo diskrečiąją Furjė transformaciją ir pavaizduoti jos modulio kvadratą, kas reikštų energijos tankį dažnių srityje (6 paveikslas).



5 pav. „A“ balsio garso slėgio grafikas

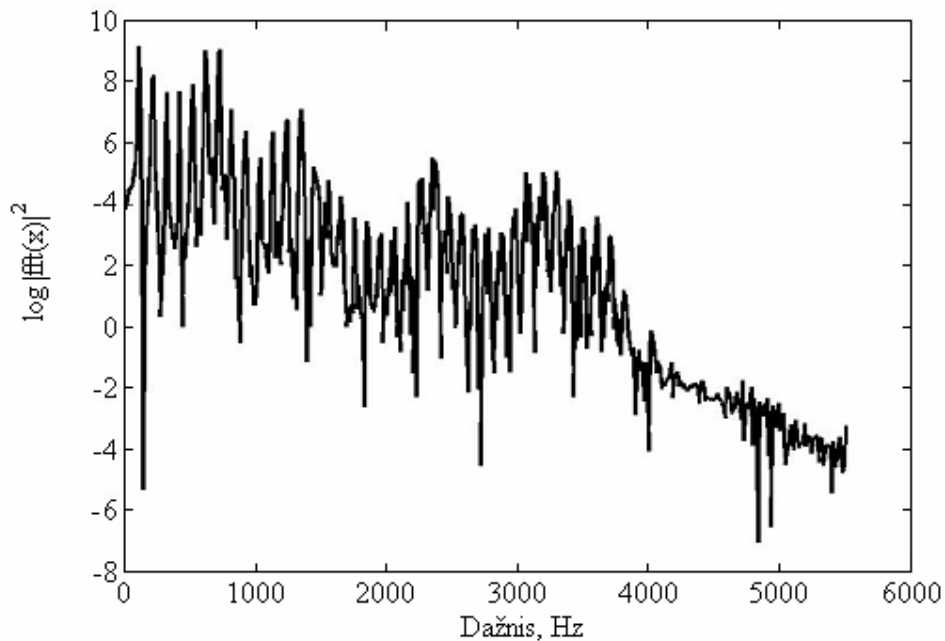


6 pav. Signalų energijos tankis dažnių srityje

6 paveiksle galime matyti periodiškai pasikartojančią teigiamų pusbangių seką spektre. Šis periodiškumas spektre atsiranda dėl periodinio balso stygų virpesio. Jeigu vidutinį laiko tarpą tarp gretimų stygų susitraukimo momentų pažymėtume T (laikas sekundėmis) ir analogišką periodą to paties balsio spektro F (1/sekundės), tai tarp jų bus paprastas ryšys:

$$T \cdot F = 1. \quad (1)$$

Jeigu apskaičiuotume spektro modulio kvadrato logaritmą, balsio periodiškumą dažnių srityje galėtume matyti dar aiškiau (7 paveikslas).



7 pav. Signalų spektro modulio kvadrato logaritmas

Taip atsitinka dėl to, kad realiai matuojami signalai yra glodūs, o glodžių signalų Furjė transformacija greitai artėja prie nulio. Logaritmo funkcija šį artėjimą panaikina. Yra ir kitų priežasčių, dėl ko yra naudinga apskaičiuoti spektro galingumo logaritmą.

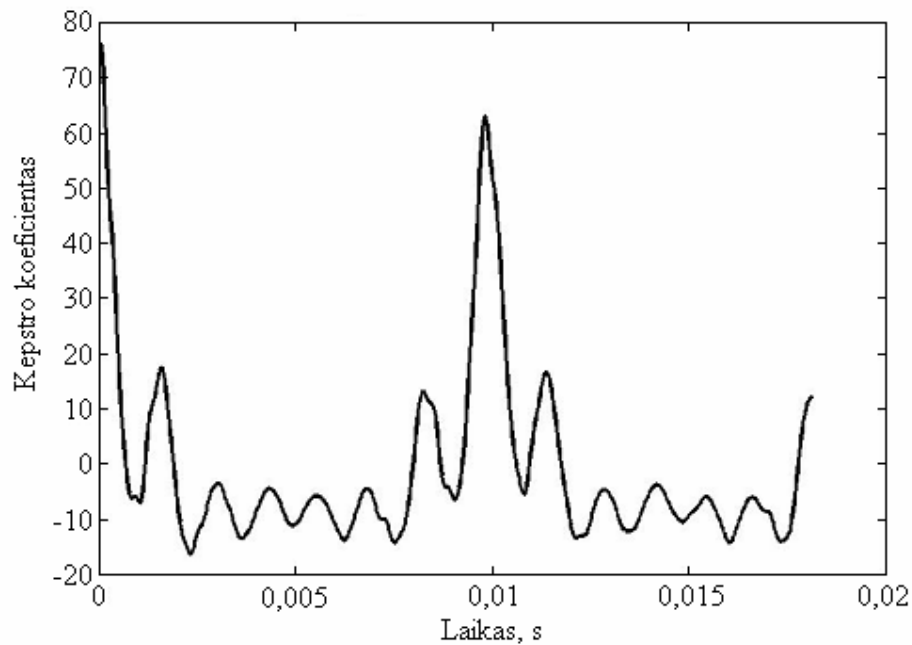
Dabar apibrėžkime, kas yra kepstro koeficientas. Literatūroje galima rasti tokį apibrėžimą: *diskrečiojo signalo x kepstro koeficientais c vadinama jo diskrečiosios Furjė transformacijos logaritmo atvirkštinė Furjė transformacija*, t.y.

$$c = \text{ifft}\left(\log\left(|\text{fft}(x)|^2\right)\right) \quad (2)$$

čia fft – diskrečioji Furjė transformacija, ifft – atvirkštinė Furjė transformacija.

Angliškas pavadinimas „cepster“ ir lietuviškas pavadinimas „keptras“ atspindi naudojamos procedūros eigą: Furjė transformacija – atvirkštinė Furjė transformacija (ir „spectr“

– atvirkščiai „cepster“ (spek-tras – atvirkščiai keps-tras)). Kepstro koeficientų grafikas pavaizduotas 8 paveiksle.



8 pav. Kepstro koeficientų grafikas

Kepstro koeficientų abscisių ašį žymime „t, [sek]“ nes pradinio signalo abscisių ašis yra laikas, Furjė transformacijos modulio kvadrato logaritmo – dažnis f , o apskaičiavę atvirkštinę Furjė transformaciją vėl grįžtame prie laiko skalės. Tai ir leidžia atkurti signalo laiko struktūrą.

Į kepstro koeficientų apskaičiavimo procesą galime žiūrėti kaip į konvejerinį (angl. *pipeline*) signalo apdorojimo būdą. Konvejerio sudedamosios dalys yra tokios [1], [17]:

- Pradinis signalas ($x = p(t)$);
- Diskrečioji Furjė transformacija (FFT);
- Logaritmas ($\log$);
- Atvirkštinė Furjė transformacija (IFFT);
- Signalų kepstros (c).

Signalų kepstros yra aukštos kokybės požymių vektorius, reprezentuojantis žmogaus balsą ir muzikinius signalus. Kalbos analizėje dažniausiai naudojamas melų skalės kepstros arba melų skalės dažnio kepstros koeficientai. Melų skalės dažnio kepstros koeficientai bus aptarti 4.3 poskyryje.

4.2. Kepstro koeficientų esmė

Natūraliai kyla klausimas kodėl kepstro koeficientai išryškina periodo reikšmę. Jei signalas $x = x(t)$ yra beveik periodinis, pavyzdžiui, garso slėgis atitinkantis balsę, tai jį galima išskaidyti į sąsūkos T periodo žadinimo signalo $p(t) = \sum_n \delta(t - nT)$ su vieno periodo impulso forma $a(t)$ sumą su paklaidos $\varepsilon(t)$ signalu. Todėl

$$x(t) = (p \cdot a)(t) + \varepsilon(t) \quad (3)$$

Kuo periodiškumas reguliaresnis tuo $\varepsilon(t)$ paklaida ir jos Furjė transformacija bus mažesni. Kadangi sąsūkos Furjė transformacija yra lygi sąsūkos dėmenų Furjė transformacijų sandaugai, tai galios apytikslė lygybė

$$|X(f)| \cong |P(f)A(f)|. \quad (4)$$

Sandaugos $P(f) \cdot A(f)$ logaritmas išsiskaido į logaritmų sumą ir tokiu būdu gausime:

$$\log|X(f)| \cong \log|P(f)| + \log|A(f)|. \quad (5)$$

T -periodinės funkcijos p Furjė transformacija yra $1/T$ periodinė, tiksliau

$$P(f) = \sum_n \delta\left(f - \frac{n}{T}\right), \quad (6)$$

todėl Furjė transformacijos modulio logaritmo grafike matyti periodiškai pasikartojantys maksimumai.

Reguliaraus periodo impulso spektro energija sukoncentruota arti nulinio dažnio $f = 0$ ir neturi išreikštos periodinės struktūros. Todėl atvirkštinėje Furjė transformacijoje su t artimais periodui T dominuoja periodinės dalies $\log(|P(f)|)$ komponentė.

Taigi kepstro koeficientų pirmojo arti 0 ryškaus lokalaus maksimumo abscisė gerai aproksimuoja beveik periodinio signalo periodo vidutinę reikšmę T .

Tarkime signalo imčių dažnis $F = 44100$ taškų per sekundę, o vidutinė periodo reikšmė $T = 0,01$ sekundės. Tuomet diskrečiojo kepstro, kuris gaunamas vietoje tolydžių Furjė

transformacijų naudojant diskrečiąsias, koeficientas c_n su indeksu $n = 44100 \cdot 0,01 = 441$ turėtų būti arti lokalaus maksimumo.

Pirmųjų kepstro koeficientų Furjė transformacija neblogai aproksimuoja suglodintą spektro logaritmą. Todėl pirmieji kepstro koeficientai gana dažnai laikomi kalbos signalo požymiais. Logika tokia: kadangi žmogus garsą atpažįsta pagal jo spektrą, o pirmieji 20-30 koeficientų netiesiogiai atspindi signalo spektro formą, tai šiuose kepstro koeficientuose glūdi pakankamai informacijos, kad iš jų būtų galima atpažinti garsą [16].

4.3. Melų skalės dažnio kepstro koeficientai

Skirtumas tarp kepstro ir melų skalės kepstro yra tas, kad dažnių ruožai išdėstomi logaritmiškai melų skalėje [7], [18]. Melų skalė, tai kalbos suvokimu pagrįsta dažnių skalė. Dažnis f keičiamas į melus m pagal formulę (7), kurią 1937 metais pasiūlė S. Stevensas, J. Volkmanas ir E. Newmanas [19].

$$m = 1127,01048 \log_e \left(1 + \frac{f}{700} \right) \quad (7)$$

Melų skalės dažnio kepstro koeficientai gaunami taip:

- Pradinis signalas ($x = p(t)$);
- Diskrečioji Furjė transformacija (FFT);
- Signalo dažnių skalė pakeičiama į melų skalę;
- Logaritmas ($\log$);
- Diskrečioji kosinuso transformacija (DCT);
- Melų skalės signalo keptras (c_m).

Melų skalės kepstro koeficientai atspindi kalbančiajam būdingą vidutinę balso trakto konfigūraciją. Taip pat atspindi ir pagrindinio tono harmonikas. Pagrindinio tono nesutapimas tarp mokymui ir atpažinimui skirtų duomenų taip pat gali sukelti atpažinimo klaidas. Vienas iš šios problemos sprendimų – užfiksuoti kiekvieną pagrindinio tono periodą ir naudoti kintamo ilgio normaliojo pasiskirstymo funkciją (Gauso langą), kurios paplitimas priklauso nuo pagrindinio tono. Šis būdas skirtas gauti spektrinę informaciją be pagrindinio tono dažnio [3], [4], [6].

4.4. Tiesinės prognozės kodavimas

Pagrindinė tiesinės prognozės idėja yra prognozuoti kalbos signalo atskaitą remiantis buvusiomis atskaitomis. Tarkim turime signalo atskaitų seką $\{s_1, \dots, s_N\}$. Elemento s_n M-osios eilės tiesinė prognozė formuojama kaip prieš jį einančių M atskaitų tiesinė kombinacija. Prognozuotą atskaitą pažymėję y_n , gauname

$$y_n = -\sum_{i=1}^M a_i s_{n-i} . \quad (8)$$

Koeficientai a_i , $i = 1, \dots, M$ šioje sąsūkoje ir yra ieškomieji prognozės koeficientai. Juos reikia parinkti taip, kad minimizuotume prognozės paklaidą. Vienos atskaitos prognozės paklaida e_n lygi

$$e_n = s_n - y_n , \quad (9)$$

o pilna kvadratinė paklaida apibrėžiama tokiu būdu:

$$\alpha = \sum_{n=n_0}^{n_1} e_n^2 = \sum_{n=n_0}^{n_1} \left[\sum_{i=0}^M a_i s_{n-i} \right]^2 = \sum_{n=n_0}^{n_1} \sum_{i=0}^{n_1} \sum_{j=0}^M a_i s_{n-i} s_{n-j} a_j \quad (10)$$

kur n_0 ir n_1 yra sumavimo rėžiai. Pažymėkime

$$c_{ij} = \sum_{n=n_0}^{n_1} s_{n-i} s_{n-j} . \quad (11)$$

Tada kvadratinė paklaida α užrašoma taip:

$$\alpha = \sum_{i=0}^{n_1} \sum_{j=0}^M a_i c_{ij} a_j . \quad (12)$$

Norint minimizuoti α , reikia apskaičiuoti α dalines išvestines a_k , $k = 1, \dots, M$ atžvilgiu, prilyginti jas nuliui ir išspręsti gautą lygčių sistemą. Atsižvelgiant į tai, kad $a_0 = 1$, gauname tokią tiesinių lygčių sistemą:

$$\sum_{i=0}^{M_i} a_i c_{ik} = -c_{0k}, \quad (13)$$

kur $k = 1, \dots, M$.

Priklausomai nuo sumavimo rėžių n_0 ir n_1 parinkimo galima išskirti kovariacinę ir autokoreliacinę parametrų c_{ik} išraišką. Tiesinių lygčių sistemą galima išspręsti bet kuriuo tiesinių lygčių sistemų sprendimo metodu, tačiau yra sukurti specialūs metodai, žymiai pagreitinantys sprendimą remiantis tam tikromis sistemos savybėmis. Pavyzdžiui, autokoreliacinės realizacijos atveju galima naudoti Levinsono-Durbino metodą.

Tiesinės prognozės koeficientai gali būti naudojami daugeliui tikslų: kalbos signalų suspaudimui, atpažinimui, sintezei.

Kalbos signalo sintezėje kartu su prognozės koeficientais dar naudojami pagrindinio tono periodo ilgis, požymis, nusakantis, ar signalas skardus, ar duslus, kvadratinė šaknis iš signalo atskaitų. Sintezuojant skardžius segmentus filtras sužadinamas kvaziperiodiniais impulsais, atstumas tarp kurių lygus vidutiniam pagrindinio tono periodui, o dusliams sužadinantį signalą sukuria balto triukšmo generatorius. Sužadinantis signalas dauginamas iš daugiklio (stiprinimo koeficiento), nusakančio pradinio ir sintezuoto segmento energijos santykį.

Tiesinė prognozė yra vienas galingiausių signalų analizės metodų. Ypač jis gerai užsirekomendavo kalbos analizėje. Tiesinė prognoze paremtas kalbos kodavimas duoda geriausius kalbos ir kompresijos kokybės parametrus. Šiuo metodu gaunami kalbą aprašantys parametrai, jų skaičius yra nedidelis ir jie apskaičiuojami greitai.

LPC metode daroma prielaida, kad kalbos signalas yra inicijuojamas garso šaltinio kintamo skerspjuvio vamzdelio viename iš galų. Žmogaus garso šaltinis (virpančios balso stygos) charakterizuojamas intensyvumu ir dažniu. Gerklės ir burnos kalbos traktas formuoja kintamo skersmens vamzdelį. Kalbos traktas charakterizuojamas rezonansiniais dažniais (formantėmis).

Tiesinės prognozės metodu yra įvertinamos formantės, atskiriant jas nuo kalbą generuojančio šaltinio, lemiančio kalbos garsumą ir toną.

Skaičiai aprašantys formantes ir liekamasis signalas gali būti saugojami ir siunčiami atskirai. Kalbos sintezavimas tiesinės prognozės metodu gaunamas apgręžiant metodą: liekamasis signalas tampa kalbos signalą generuojančiu šaltiniu, o iš formančių gaunamas filtras, atspindintis kalbos trakto geometriją. Filtruojant liekamąjį signalą ir gaunamas sintezuotas garsas.

Kadangi kalbą generuojantis traktas kinta laike, tiesinė prognozė atliekama trumpais laiko tarpais. Dažniausiai vienos sekundės trukmės kalbos signalas skaidomas į 30-50 langų.

Kalbos ir fonemų atpažinime dominuoja tiesinės prognozės koeficientais ir jų pagrindu gaunami požymiai. Tiesinės prognozės koeficientai glaustai perteikia informaciją apie kalbos fragmento spektro gaubiamąją. Kadangi žmogaus klausos aparatas taip pat kalbos signalo fragmentus transformuoja į dažninę sritį ir vėliau naudoja spektrinę informaciją atpažinimui, todėl ir tiesinės prognozės koeficientų naudojimas fonemų ir kalbos ir pasiteisina [22].

5. AUTOMATINIO KALBOS ATPAŽINIMO TYRIMO ŽODYNAI

Šio darbo metu atliktas aparatinio kalbos atpažinimo tyrimas ir paketų pradžios įtaka žodžių atpažinimui.

Žodžių atpažinimui panaudotas dinaminio laiko skalės kraipymo metodas – vienas plačiausiai naudojamų būdų kalbos atpažinimo sistemose. Nors yra ir kitų metodų, tokių kaip paslėptieji Markovo modeliai ar dirbtiniai neuronų tinklai, tačiau dinaminio laiko skalės kraipymo metodo populiarumą lėmė nesudėtinga ir greita „mokymo“ procedūra lyginant su paslėptaisiais Markovo modeliais ar dirbtiniais neuronų tinklais. Tai ir lėmė šio metodo pasirinkimą prarastų paketų įtakos automatiniam kalbos atpažinimui tyrime.

Atliktą tyrimą galima išskirti į dvi dalis. Pirmoji – kai etaloninių žodžių žodynas ir tiriamieji žodžiai yra to paties kalbėtojo. Antroji – kai etaloninių žodžių žodynas ir tiriamieji žodžiai yra skirtingų kalbėtojų. Pirmuoju atveju abiejų žodynų diktoriaus buvo vyras. Antruoju atveju etaloninių žodžių žodyno diktoriaus buvo moteris, o tiriamųjų žodžių – vyras.

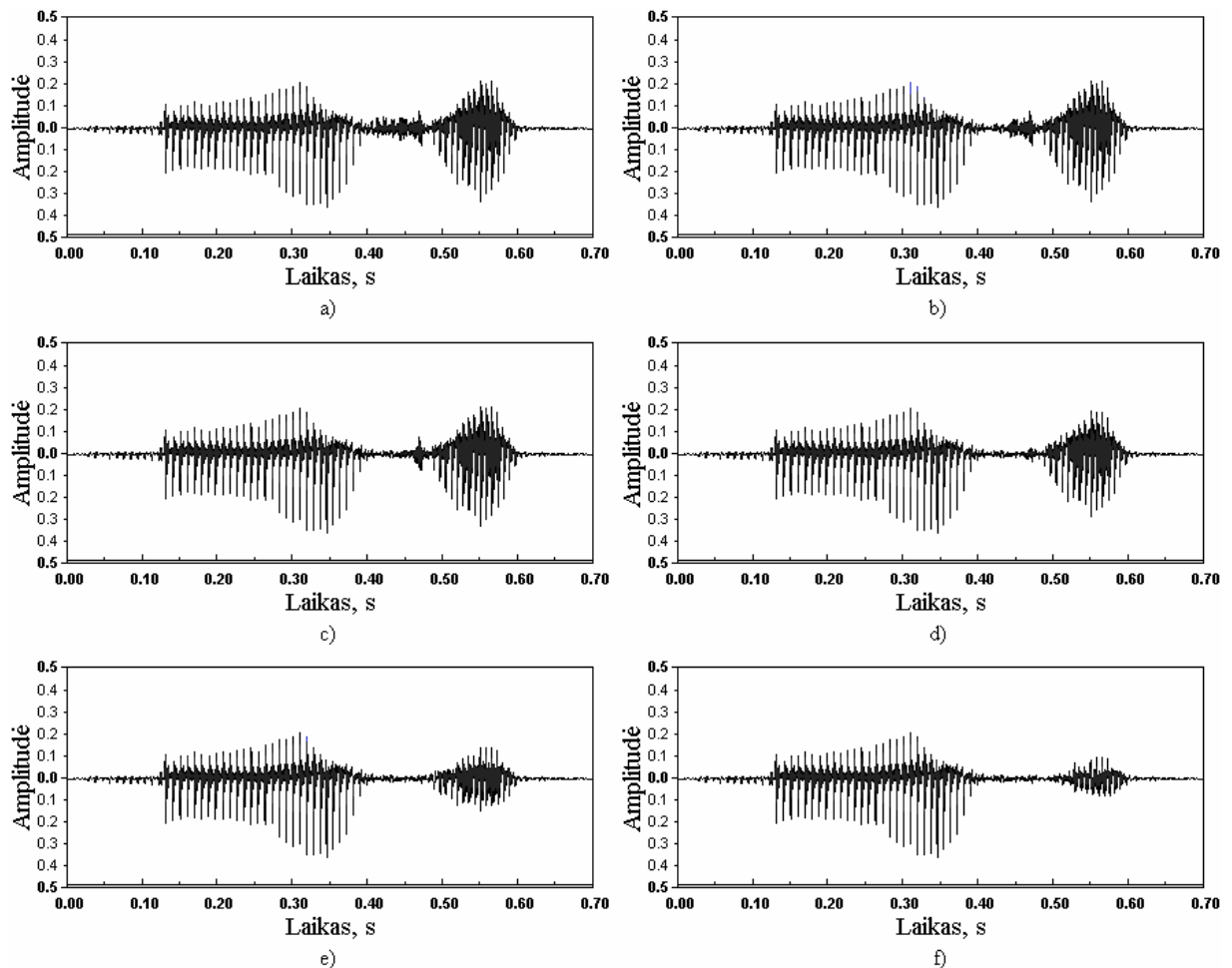
5.1. Žodžių su prarastais paketais modeliavimas

Paketingo duomenų perdavimo sistemos vis dažniau naudojamos balsui perduoti. Norint kalbą perduoti šiomis sistemomis, balso signalą reikia koduoti skaitmeniniu būdu. Tam naudojami įvairūs balso kodekai: G.723, G.728, G.729, SpeexN, GSM-06.10, GSM-FR, AMR ir kt. Paketų pradžios modelių sudarymui panaudotas adaptyvus daugiadažnis kalbos kodekas (angl. *Adaptive Multi-Rate – AMR*). Automatinio kalbos atpažinimo tyrimui pasirinktas AMR kodekas todėl, kad jis plačiai naudojamas mobiliojo ryšio technologijose (GSM, GPRS, EGPRS, EDGE), bevielio ryšio telefonuose (angl. *Voice over Wireless Local Area Network – VoWLAN*), plačiajuostėje IP telefonijoje, balso ir vaizdo konferencijose. Taip pat AMR kodekas naudojamas kalbos suglaudinimui trečios kartos 3G mobiliojo ryšio technologijoje.

Adaptyvus daugiadažnis kodekas turi aštuonis pagrindinius bitų srautus: 12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15 ir 4.75 kbit/s. Šis kodekas naudoja ACELP (angl. *Algebraic Code Excited Linear Prediction – ACELP*) principus visiems minėtiems bitų srautams. Bitų srautas paremtas 20 milisekundžių trukmės kadrais, kurių kiekviename yra 95, 103, 118, 134, 148, 159, 204 arba 244 bitai, atitinkamai 4.75, 5.15, 5.90, 6.70, 7.40, 7.95, 10.2 arba 12.2 kbit/s srautams. Paketų pradžios modelių sudarymui panaudotas AMR kodekas su 12.2 kbit/s srautu [9].

Žodžiai su prarastais paketais modeliuojami koduojant žodį AMR kodeku. Žodžiuose sumodeliuoti nuo vieno iki penkių AMR kodeko kadrų praradimų. Be to, kadrai priklauso tai pačiai raidei: balsei arba priebalsiui. Penkių sugadintų kadrų pakanka sugadinti vienai raidei. Įprasto prarastų kalbos kadrų dekodavimo pasekmė būtų labai nemalonus triukšmo efektas. Tam, kad pagerėtų subjektyvi kokybė, dekodudamas žodį AMR kodekas prarastus kalbos kadrus pakeičia prieš tai buvusio gero kalbos kadro (ar kelių kadrų) kopijomis arba ekstrapoliacijomis. Dėl šios AMR kodeko savybės ne taip juntamas kokybės pablogėjimas, tačiau prarastų kadrų efektas vis tiek juntamas, ypač kai prarandami keli kadrai paeiliui.

9 paveiksle pavaizduotos žodžio „mažas“ oscilogramos, kai žodis buvo koduotas ir dekoduetas AMR kodeku ir koduojant žodį buvo sugadinti kadrai: 9 pav. a – žodis tik perkoduotas AMR kodeku; 9 pav. b – koduojant sugadintas vienas kadras; 9 pav. c – koduojant sugadinti du kadrai; 9 pav. d – koduojant sugadinti trys kadrai; 9 pav. e – koduojant sugadinti keturi kadrai; 9 pav. f – koduojant sugadinti penki kadrai. Pateiktame paveiksle galime matyti kaip kinta kalbos signalas prarandant vis didesnę kadrų skaičių.



9 pav. Koduoto ir dekodueto žodžio „mažas“ oscilogramos su prarastais paketais

5.2. Tiriamųjų žodžių žodynas

Tiriamųjų žodžių žodynas sudarytas iš trisdešimties žodžių. Tiriamųjų žodžių žodynas pateiktas 5.1 lentelėje. Šioje lentelėje raudonai pažymėtos raidės, kuriose sumodeliuoti paketų praradimai. Tiriamųjų žodžių garso įrašai pateikti A priede.

5.1 lentelė. Tiriamųjų žodžių žodynas

Batas	Dėlė	Gama	Nagas
Matas	Dēmė	Guma	Namas
Patas	Dėžė	Lama	Naras
Ratas	Galas	Lema	Dyla
Vatas	Garas	Loma	Kyla
Dėlė	Mačas	Ratas	Šyla
Gėlė	Magas	Retas	Tyla
Lėlė	Maras	Rytas	Maras
Vėlė	Matas	Vėsa	Meras
Žėlė	Mažas	Visa	Mūras

5.3. Etaloninių žodžių žodynas

Etaloninių žodžių žodynas sudarytas iš penkiasdešimties žodžių. Etaloninių žodžių žodynas pateiktas 5.2. lentelėje. Žodyną sudaro dviskiemeniai žodžiai. Kiekvienas žodis turi bent vieną panašų žodį, kuris skiriasi tik viena raide. Etaloninių žodžių garso įrašai pateikti A priede.

5.2 lentelė. Etaloninių žodžių žodynas

Balà	Gālas	Lemà	Múras	Šyla
Bālas	Gamà	Lomà	Nāgas	Tyla
Bāras	Gāras	Māčas	Nāmas	Vagà
Bātas	Gėlė	Māgas	Nāras	Vatà
Bylà	Gėras	Māras	Natà	Vātas
Cāras	Gumà	Mātas	Notà	Vazà
Dėlė	Kāras	Māžas	Pātas	Vėlė
Dēmė	Kyla	Mēras	Rātas	Vēsà
Dėžė	Lamà	Mētas	Rētas	Visà
Dyla	Lėlė	Mitas	Rytas	Žėlė

6. PRARASTŲ PAKETŲ ĮTAKOS AUTOMATINIAM KALBOS ATPAŽINIMUI TYRIMAS

6.1. 1 bandymas

Pirmiausiai pabandyta ištirti, kaip automatinį kalbos atpažinimą veikia prarasti paketai, kai etaloninis žodis lyginamas su tuo pačiu etaloniniu žodžiu, bet su prarastais keliais paketais. Atlikus keletą bandymų pastebėta, kad nesvarbu kiek (vienas ar penki) ir kurioje žodžio vietoje prarasta 20 milisekundžių trukmės kadru, tai kalbos atpažinimui įtakos nedaro. Šio tyrimo metu panaudota dešimt testinių žodžių: *nagas*, *namas*, *naras*, *dyla*, *kyla*, *šyla*, *tyla*, *maras*, *meras*, *mūras*. Etaloninių žodžių žodynui taip pat panaudoti tie patys žodžiai. Tyrimo rezultatai pateikti 6.1 lentelėje.

6.1 lentelė. Tyrimo rezultatai etaloninį žodį lyginant su tuo pačiu paketų praradimų paveiktu etaloniniu žodžiu

Tiriamas žodis	Etalonas su neprarastais kadrų	Etalonas su prarastu 1 kadru	Etalonas su prarastais 2 kadrų	Etalonas su prarastais 3 kadrų	Etalonas su prarastais 4 kadrų	Etalonas su prarastais 5 kadrų
Nagas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
Namas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
Naras	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
Dyla	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
Kyla	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
Šyla	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
Tyla	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
Maras	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
Meras	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
Mūras	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta

Iš 6.1 lentelėje pateiktų rezultatų matome, kad paketų praradimai automatiniam kalbos atpažinimui įtakos nedarė, visi žodžiai buvo atpažinti. Tai galima būtų paaiškinti tuo, kad šiuo atveju atpažįstamo ir etaloninio žodžių trukmės vienodos. Taip pat beveik visi atstumai tarp etaloninio ir tiriamojo žodžių požymių vektorių lygūs nuliui, išskyrus tuos, kurie patenka į sugadintų kadru sritį. Tačiau vidutinis atstumas tarp požymių vektorių vis tiek yra daug mažesnis negu lyginant atpažįstamą žodį su kitais etaloniniais žodžiais, nors jie ir yra panašūs, skiriasi tik viena raide.

6.2. 2 bandymas

Toliau tyrimas atliktas, kai pasirinktas etaloninis žodis buvo lyginamas su kitu tokiu pačiu to paties diktoriaus pasakytu žodžiu. Atlikti trys bandymai: kai prarandami kadrai yra pirmoje žodžio raidėje (priebalsėje), kai parandami kadrai yra žodžio viduryje esančioje priebalsėje, ir kai prarandami kadarai yra žodžio bet kurioje balsėje. Kaip ir ankstesnio eksperimento metu, taip ir šio, panaudotas etaloninių žodžių žodynas, aprašytas 5.3 poskyryje.

Pirmiausiai atliktas bandymas kai etaloninis žodis lyginamas su kitu tokiu pačiu to paties diktoriaus pasakytu žodžiu su prarastais kadrtais pirmoje žodžio raidėje (priebalsėje). Tyrimui panaudota dešimt testinių žodžių: *batas, matas, patas, ratas, vatas, dėlė, gėlė, lėlė, vėlė, žėlė*. Tyrimo rezultatai pateikti 6.2 lentelėje.

6.2 lentelė. Tyrimo rezultatai prarandant paketus pirmoje žodžio raidėje (priebalsėje)

Tiriamas žodis	Žodis su neprarastais kadrtais	Žodis su prarastu 1 kadru	Žodis su prarastais 2 kadrtais	Žodis su prarastais 3 kadrtais	Žodis su prarastais 4 kadrtais	Žodis su prarastais 5 kadrtais
Batas	Atpažinta 10,7271	Neatpažinta 13,2078	Neatpažinta 13,1109	Neatpažinta 11,5342	Neatpažinta 11,6509	Neatpažinta 11,7790
Matas	Neatpažinta 10,3860	Neatpažinta 10,3954	Neatpažinta 10,9401	Neatpažinta 10,9798	Neatpažinta 10,8697	Neatpažinta 11,5146
Patas	Neatpažinta 11,0041	Atpažinta 11,2134	Atpažinta 11,2808	Neatpažinta 12,7917	Neatpažinta 14,5596	Atpažinta 17,5979
Ratas	Neatpažinta 12,1740	Neatpažinta 11,7438	Neatpažinta 11,4146	Neatpažinta 11,0218	Neatpažinta 11,1100	Neatpažinta 11,8671
Vatas	Neatpažinta 9,4808	Atpažinta 9,2753	Neatpažinta 9,4886	Neatpažinta 10,7433	Neatpažinta 14,8365	Neatpažinta 15,8587
Dėlė	Atpažinta 5,4045	Atpažinta 6,0306	Atpažinta 5,9344	Atpažinta 6,7914	Atpažinta 6,4953	Atpažinta 8,1571
Gėlė	Atpažinta 8,6641	Atpažinta 9,1232	Atpažinta 9,2270	Atpažinta 9,8710	Atpažinta 9,0630	Neatpažinta 10,2646
Lėlė	Atpažinta 6,9442	Atpažinta 7,5646	Atpažinta 7,9452	Atpažinta 7,9452	Atpažinta 9,9775	Atpažinta 11,5879
Vėlė	Atpažinta 8,1565	Atpažinta 8,0074	Atpažinta 8,1345	Atpažinta 8,4501	Atpažinta 11,6336	Atpažinta 12,2817
Žėlė	Atpažinta 5,9821	Atpažinta 6,0720	Atpažinta 6,2791	Atpažinta 6,2132	Atpažinta 6,6401	Atpažinta 6,9374

6.2 lentelėje pateikti duomenys apie sėkmingą žodžių atpažinimą, priklausomai nuo prarastų paketų skaičiaus. Taip pat lentelėje pateikti vidutiniai atstumai tarp etaloninio ir tiriamojo žodžių požymių vektorių. Po penkis sugrupuotų panašių žodžių automatinis atpažinimas labai skirtingas. Pirmi penki žodžiai (*batas, matas, patas, ratas, vatas*) atpažinti žymiai rečiau nei kiti (*dėlė, gėlė, lėlė, vėlė, žėlė*), nors paketų praradimai buvo modeliuojami pirmojoje žodžio raidėje. Testinį žodį tik perkodavus AMR kodeku atpažinti šeši žodžiai iš

dešimties. Praradus vieną 20 milisekundžių trukmės kadrą atpažinti septyni žodžiai, praradus du kadrus atpažinti šeši žodžiai, praradus tris, keturis ir penkis kadrus atpažinta po penkis žodžius. Analizuojant atpažinimo tyrimo rezultatus pastebėta, kad klaidingai atpažinti testiniai žodžiai žodyne turėjo daug panašių žodžių, kurie skyrėsi ne tik pirmąją žodžio raidę, bet ir viduriniąją, pavyzdžiui, žodžiui *matas* panašūs žodžiai yra ne tik pateikti 6.2 lentelėje, bet ir *mačas*, *magas*, *maras*, *mažas*, *metas*, *mitas*. Didėjant panašių žodžių skaičiui atpažinimo procesas tampa sudėtingesnis. Taigi, pirmųjų penkių testinių žodžių klaidingą atpažinimą stipriai įtakojo etaloninių žodžių žodyne esantys panašūs žodžiai, kurie skyrėsi ne tik pirmąją, bet ir kitomis raidėmis. Kitiems penkiems testiniams žodžiams etalonų žodyne buvo mažiau panašių žodžių, todėl ir atpažinta daugiau žodžių su prarastais paketais.

Kitas bandymas atliktas, kai etaloninis žodis lyginamas su kitu tokiu pačiu to paties diktoriaus pasakytu žodžiu su prarastais 20 milisekundžių trukmės kadrų žodžio viduryje esančioje priebalsėje. Tyrimui panaudota dešimt testinių žodžių: *dėlė*, *dėmė*, *dėžė*, *galas*, *garas*, *mačas*, *magas*, *maras*, *matas*, *mažas*. Tyrimo rezultatai pateikti 6.3 lentelėje.

6.3 lentelė. Tyrimo rezultatai prarandant paketus vidurinėje žodžio raidėje (priebalsėje)

Tiriamas žodis	Žodis su neprarastais kdrais	Žodis su prarastu 1 kadru	Žodis su prarastais 2 kdrais	Žodis su prarastais 3 kdrais	Žodis su prarastais 4 kdrais	Žodis su prarastais 5 kdrais
Dėlė	Atpažinta 8,4902	Atpažinta 9,1682	Atpažinta 9,0190	Atpažinta 9,1772	Atpažinta 9,3930	Atpažinta 8,7225
Dėmė	Atpažinta 11,0413	Atpažinta 11,5706	Neatpažinta 11,2943	Neatpažinta 11,3127	Neatpažinta 11,9129	Neatpažinta 11,7994
Dėžė	Atpažinta 8,9792	Atpažinta 8,7183	Atpažinta 8,3679	Atpažinta 8,2227	Atpažinta 8,6706	Atpažinta 8,9825
Galas	Atpažinta 10,2727	Atpažinta 9,5905	Atpažinta 10,7219	Atpažinta 10,3527	Atpažinta 10,6910	Atpažinta 10,6898
Garas	Neatpažinta 10,9387	Neatpažinta 11,3205	Neatpažinta 11,2528	Atpažinta 11,2274	Atpažinta 11,8558	Atpažinta 12,5369
Mačas	Atpažinta 13,3839	Atpažinta 13,1837	Neatpažinta 12,6062	Neatpažinta 12,0116	Neatpažinta 11,5713	Neatpažinta 11,9151
Magas	Atpažinta 8,7784	Atpažinta 9,5026	Atpažinta 9,7754	Atpažinta 10,7817	Atpažinta 10,9516	Atpažinta 11,0454
Maras	Atpažinta 6,8620	Atpažinta 7,2210	Atpažinta 7,4207	Atpažinta 7,3995	Atpažinta 7,4450	Atpažinta 8,1760
Matas	Neatpažinta 11,3358	Neatpažinta 11,0362	Neatpažinta 11,3368	Neatpažinta 10,4009	Neatpažinta 10,3918	Neatpažinta 10,0222
Mažas	Atpažinta 8,7426	Atpažinta 8,9463	Atpažinta 8,7555	Atpažinta 8,8003	Atpažinta 9,1666	Atpažinta 9,7353

Iš 6.3 lentelėje pateiktų rezultatų matome, kad modeliuojant paketų praradimus žodžio viduryje (priebalsėje), skirtingų žodžių atpažinimo rezultatai skirtingi. Vieni žodžiai buvo atpažinti idealiai, o kiti neatpažistami perkodavus žodį AMR kodeku ar praradus tik vieną 20

milisekundžių trukmės kadrą. Tik perkodavus AMR kodeku testinį žodį ir praradus vieną kadrą atpažinta po aštuonis žodžius. Praradus du kadrus atpažinti šeši žodžiai, praradus tris ir keturis kadrus atpažinta po septynis žodžius, o praradus penkis kadrus atpažinti šeši žodžiai. Klaidingo atpažinimo priežastys galėtų būti panašios kaip ir bandymo, kai prarandama pirma žodžio raidė. Sunkiau atpažinti žodžius, kuriems etaloninių žodžių žodyne yra daugiau panašių žodžių. Tačiau lyginant 6.2 ir 6.3 lentelėse pateiktus tyrimų rezultatus galime pastebėti, kad paketų praradimai žodžio viduryje esančioje priebalsėje daro mažesnę įtaką, negu paketų praradimai pirmoje žodžio raidėje.

Trečiasis bandymas atliktas, kai etaloninis žodis lyginamas su kitu tokiu pačiu to paties diktoriaus ištartu žodžiu su prarastais paketais balsėje. Paketų praradimai buvo modeliuojami tiek trumpose balsėse (pavyzdžiui, žodžio *guma* raidėje „u“), tiek ilgose (pavyzdžiui, žodžio *retas* raidėje „e“). Tyrimui panaudota dešimt testinių žodžių: *lama*, *lema*, *loma*, *ratas*, *retas*, *rytas*, *gama*, *guma*, *vėsa*, *visa*. Tyrimo rezultatai pateikti 6.4 lentelėje.

6.4 lentelė. Tyrimo rezultatai prarandant paketus žodžio balsėje

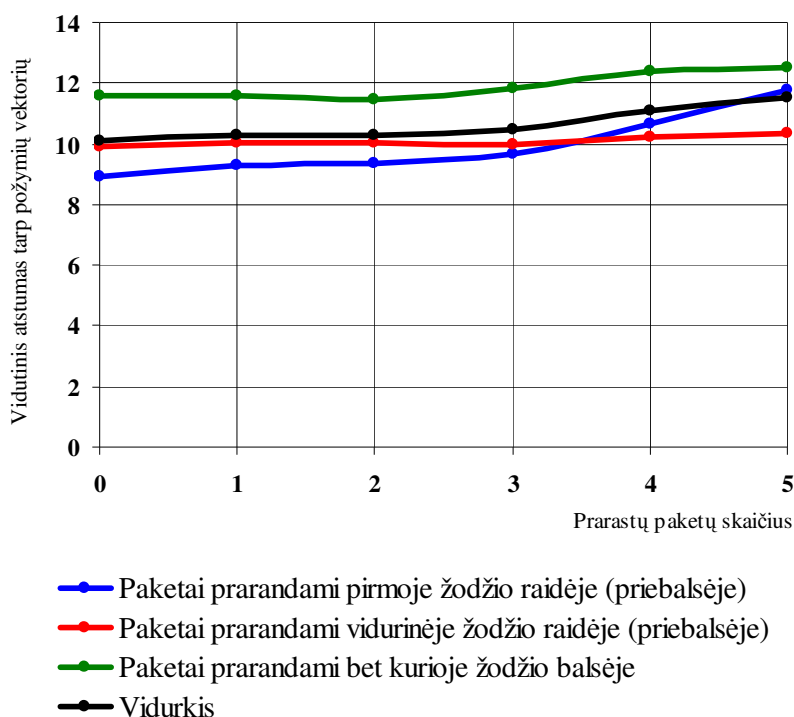
Tiriamas žodis	Žodis su neprarastais kdrais	Žodis su prarastu 1 kadru	Žodis su prarastais 2 kdrais	Žodis su prarastais 3 kdrais	Žodis su prarastais 4 kdrais	Žodis su prarastais 5 kdrais
Gama	Atpažinta 9,9027	Atpažinta 11,5681	Atpažinta 10,4806	Atpažinta 10,0144	Atpažinta 10,8340	Atpažinta 10,1850
Guma	Atpažinta 6,8905	Atpažinta 7,4357	Atpažinta 7,9685	Atpažinta 7,9921	Atpažinta 8,3131	Atpažinta 9,0239
Lama	Atpažinta 9,8802	Atpažinta 10,1579	Atpažinta 10,2542	Atpažinta 10,2017	Neatpažinta 12,0500	Neatpažinta 12,4736
Lema	Atpažinta 12,3414	Atpažinta 13,0123	Atpažinta 13,0432	Atpažinta 13,2177	Atpažinta 13,3363	Atpažinta 13,6579
Loma	Atpažinta 13,6706	Atpažinta 13,3418	Atpažinta 13,0507	Atpažinta 13,2636	Atpažinta 13,6904	Neatpažinta 13,6646
Ratas	Atpažinta 13,2626	Atpažinta 11,4501	Atpažinta 11,1162	Atpažinta 13,8020	Atpažinta 14,6663	Atpažinta 14,9725
Retas	Atpažinta 13,2985	Atpažinta 10,1205	Atpažinta 10,4403	Atpažinta 11,5932	Atpažinta 12,7244	Atpažinta 12,5055
Rytas	Atpažinta 12,2545	Atpažinta 13,5412	Atpažinta 14,5691	Atpažinta 14,4152	Atpažinta 14,0402	Atpažinta 14,2247
Vėsa	Atpažinta 10,6963	Atpažinta 10,9756	Atpažinta 10,4329	Atpažinta 10,6706	Atpažinta 11,4079	Atpažinta 11,5453
Visa	Neatpažinta 13,8372	Neatpažinta 14,3034	Neatpažinta 13,2929	Neatpažinta 13,1371	Neatpažinta 12,5977	Neatpažinta 12,6131

Paketų praradimai, atsirandantys bet kurioje žodžio balsėje, beveik nedaro įtakos automatiniam kalbos atpažinimui. Atpažinti ir tie žodžiai, kurie etaloninių žodžių žodyne turėjo daugiau panašių žodžių. Praradus iki penkių kadrų atpažinta po devynis žodžius, o kai prarasti penki kadrai atpažinti septyni žodžiai. Tik prarandant paketus žodyje *visa*, atpažinimo rezultatai

stipriai skyrėsi. Tai galima būtų paaiškinti tuo, kad raidė „i“ šiame žodyje yra labai trumpa ir prarandant paketus šiek tiek buvo sugadintos ir priebalsės „v“ bei „s“. Taip susigadino beveik visas žodis ir jis tapo nebeatpažįstamas.

Apibendrinant 6.2, 6.3 ir 6.4 lentelėse pateiktus rezultatus, galima pastebėti, kad paketų klaidos balsėse daro mažesnę įtaką automatiniam kalbos atpažinimui, nei paketų klaidos priebalsėse. Taip pat pastebėta, kad didesnę įtaką automatiniam kalbos atpažinimui turi žodžio pradžioje, o ne viduryje esančios priebalsės.

Paanalizuokim kaip kinta vidutiniai atstumai tarp etaloninio ir tiriamojo žodžių požymių vektorių, kai prarandami paketai skirtingose žodžio vietose. 10 paveiksle pavaizduotos vidutinio atstumo tarp požymių vektorių priklausomybės nuo prarastų paketų skaičiaus.



10 pav. Vidutinio atstumo tarp požymių vektorių priklausomybė nuo prarastų paketų skaičiaus

Kaip matome iš 10 paveikslo, didėjant prarastų paketų skaičiui vidutiniai atstumai tarp požymių vektorių yra linkę nestipriai didėti. Tik kai prarandami paketai pirmoje žodžio raidėje, vidutinis atstumas tarp požymių vektorių stipriai padidėjo praradus tris ir daugiau paketų (10 paveiksle mėlyna linija). Nedidelis vidutinio atstumo tarp požymių vektorių padidėjimas atsiranda dėl to, kad AMR kodekas prarastus kalbos kadrus pakeičia prieš tai buvusio gero kalbos kadro (ar kelių kadru) kopijomis arba ekstrapoliacijomis. O kai prarandami kadrai žodžio

pirmoje raidėje, AMR kodekas pakeičia prieš tai buvusiais kadrais, kuriuose nėra kalbos signalo, nes tiriami atskiri žodžiai. Todėl ir padidėja vidutinis atstumas tarp požymių vektorių. Taip pat iš 10 paveikslo matome, kad prarandant paketus žodžio balsėje, vidutiniai atstumai tarp tiriamųjų ir etaloninių žodžių požymių vektorių yra didesni, nei prarandant paketus žodžio priebalsėje. Tačiau gauti geresni kalbos atpažinimo rezultatai, kai paketai prarandami balsėje.

Išanalizavus 6.2, 6.3 ir 6.4 lentelėse pateiktus tyrimo rezultatus, pastebėta, kad automatiniam kalbos atpažinimui didesnę įtaką darė ne prarastų paketų skaičius, o panašių žodžių, esančių etalonų žodyne, kiekis. Dėl šios priežasties atliktas kitas bandymas. Prarastų paketų įtakos automatiniam kalbos atpažinimui tyrimas atliktas iš etaloninių žodžių žodyno pašalinus kai kuriuos žodžius. Žodyne palikti tik tie žodžiai, kurie panaudoti atpažinimo tyrimui. Atlikti taip pat trys bandymai: kai prarandami kadrai yra pirmoje žodžio raidėje (priebalsėje), kai prarandami kadrai yra žodžio viduryje esančioje priebalsėje, ir kai prarandami kadrai yra žodžio bet kurioje balsėje.

Pirmiausiai tirta, kokią įtaką aparatiniam kalbos atpažinimui daro prarastų paketų skaičius, kai paketai prarandami pirmoje žodžio raidėje (priebalsėje). Etaloninių žodžių žodynas sudarytas iš dešimties žodžių: *batas, matas, patas, ratas, vatas, dėlė, gėlė, lėlė, vėlė, žėlė*. Tiriamieji žodžiai pasirinkti tokie patys. Tyrimo rezultatai pateikti 6.5 lentelėje.

6.5 lentelė. Tyrimo rezultatai prarandant paketus pirmoje žodžio raidėje (priebalsėje)

Tiriamas žodis	Žodis su neprarastais kadrais	Žodis su prarastu 1 kadrais	Žodis su prarastais 2 kadrais	Žodis su prarastais 3 kadrais	Žodis su prarastais 4 kadrais	Žodis su prarastais 5 kadrais
Batas	Atpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	10,7271	13,2078	13,1109	11,5342	11,6509	11,7790
Matas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Neatpažinta	Neatpažinta
	11,0300	10,9672	11,7104	11,9994	13,0483	13,4006
Patas	Neatpažinta	Atpažinta	Atpažinta	Neatpažinta	Atpažinta	Atpažinta
	11,3002	11,2134	11,2808	15,5086	16,2655	17,5979
Ratas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	14,0400	13,1109	13,8077	14,2018	14,1963	14,1768
Vatas	Neatpažinta	Atpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	9,4808	9,2753	9,4886	10,7433	14,8365	10,7433
Dėlė	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	5,4045	6,0306	5,9344	6,7914	6,4953	8,1571
Gėlė	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Neatpažinta
	8,6641	9,1232	9,2270	9,8710	9,0630	10,2646
Lėlė	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	6,9442	7,5646	7,9452	7,9452	9,9775	11,5879
Vėlė	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	8,1565	8,0074	8,1345	8,4501	11,6336	12,2817
Žėlė	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	5,9821	6,0720	6,2791	6,2132	6,6401	6,9374

Šiuo atveju testinį žodį perkodavus AMR kodėku atpažinti aštuoni žodžiai iš dešimties. Praradus vieną 20 milisekundžių trukmės kadrą atpažinti devyni žodžiai, praradus du kadrus atpažinti aštuoni žodžiai, praradus tris ir keturis kadrus atpažinta po septynis žodžius, o praradus penkis kadrus atpažinti šeši žodžiai. Lyginant rezultatus su 6.2 lentelėje pateiktais rezultatais, automatinio kalbos atpažinimo rezultatai pagerėjo apie 20 procentų. Galima daryti išvadą, kad didesnis panašių žodžių skaičius etalonų žodyne daro nemažą įtaką automatiniam kalbos atpažinimui.

Kitas bandymas atliktas kai prarasti kadrai yra žodžio viduryje esančioje priebalsėje. Etaloninių žodžių žodynas sudarytas iš dešimties žodžių: *dėlė, dėmė, dėžė, galas, garas, mačas, magas, maras, matas, mažas*. Tiriamieji žodžiai pasirinkti tokie patys. Tyrimo rezultatai pateikti 6.6 lentelėje. Lyginant 6.3 ir 6.6 lentelėse pateiktus rezultatus matome, kad sumažinus panašių etaloninių žodžių skaičių, automatinio kalbos atpažinimo rezultatai pagerėjo apie 10 procentų. Testinį žodį tik perkodavus AMR kodėku ir praradus vieną kadrą atpažinta po devynis žodžius, praradus du, tris ir penkis kadrus atpažinta po septynis žodžius, o praradus keturis kadrus atpažinti aštuoni žodžiai.

6.6 lentelė. Tyrimo rezultatai prarandant paketus vidurinėje žodžio raidėje (priebalsėje)

Tiriamas žodis	Žodis su neprarastais kdrais	Žodis su prarastu 1 kdrais	Žodis su prarastais 2 kdrais	Žodis su prarastais 3 kdrais	Žodis su prarastais 4 kdrais	Žodis su prarastais 5 kdrais
Dėlė	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	8,4902	9,1682	9,0190	9,1772	9,3930	8,7225
Dėmė	Atpažinta	Atpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	11,0413	11,5706	11,2943	11,3127	11,9129	11,7994
Dėžė	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	8,9792	8,7183	8,3679	8,2227	8,6706	8,9825
Galas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	10,2727	9,5905	10,7219	10,3527	10,6910	10,6898
Garas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	11,2711	11,5564	11,2707	11,2274	11,8558	12,5369
Mačas	Atpažinta	Atpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	13,3839	13,1837	12,6062	12,0116	11,5713	11,9151
Magas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	8,7784	9,5026	9,7754	10,7817	10,9516	11,0454
Maras	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	6,8620	7,2210	7,4207	7,3995	7,4450	8,1760
Matas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Atpažinta	Neatpažinta
	12,9869	13,1190	13,0874	13,0991	13,0158	12,4752
Mažas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	8,7426	8,9463	8,7555	8,8003	9,1666	9,7353

Trečias bandymas atliktas, kai prarasti kadrai yra žodžio balsėje. Etaloninių žodžių žodynas sudarytas iš dešimties žodžių: *lama, lema, loma, ratas, retas, rytas, gama, guma, vėsa*,

visa. Tiriamieji žodžiai pasirinkti tokie patys. Tyrimo rezultatai pateikti 6.7 lentelėje. Nesvarbu kiek kadrų prarasta, visais atvejais atpažinta po devynis žodžius. Klaidingai atpažintas žodis *visa*, nes prarandant paketus raidėje „i“, buvo sugadintos ir priebalsės „v“ bei „s“. Taip susigadino beveik visas žodis.

6.7 lentelė. Tyrimo rezultatai prarandant paketus žodžio balsėje

Tiriamas žodis	Žodis su neprarastais kdrais	Žodis su prarastu 1 kdrais	Žodis su prarastais 2 kdrais	Žodis su prarastais 3 kdrais	Žodis su prarastais 4 kdrais	Žodis su prarastais 5 kdrais
Gama	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	9,9027	11,5681	10,4806	10,0144	10,8340	10,1850
Guma	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	6,8905	7,4357	7,9685	7,9921	8,3131	9,0239
Lama	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	9,8802	10,1579	10,2542	10,2017	12,0609	12,4736
Lema	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	12,3414	13,0123	13,0432	13,2177	13,3363	13,6579
Loma	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	13,6706	13,3418	13,0507	13,2636	13,6904	14,0867
Ratas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	13,2626	11,4501	11,1162	13,8020	14,6663	14,9725
Retas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	13,2985	10,1205	10,4403	11,5932	12,7244	12,5055
Rytas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	12,2545	13,5412	14,5691	14,4152	14,0402	14,2247
Vėsa	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	10,6963	10,9756	10,4329	10,6706	11,4079	11,5453
Visa	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	13,8372	14,3034	14,0083	13,9826	13,8435	13,5806

6.8 lentelėje parodyta, kaip skyrėsi automatinio kalbos atpažinimo rezultatai, kai buvo naudojami skirtingo dydžio etaloninių žodžių žodynai. Naudojant mažesnę žodyną ir modeliuojant paketų pradžius pirmoje žodžio raidėje, atpažinimo rezultatai pagerėjo apie 20 procentų. Kai paketai prarandami žodžio viduryje esančioje priebalsėje, automatinio atpažinimo rezultatai pagerėjo apie 10 procentų. Ir modeliuojant paketų pradžius balsėje, atpažinimo rezultatai beveik nepasikeitė. Iš šioje lentelėje pateiktų rezultatų akivaizdžiai matome, kad žodyno dydis turi didelę įtaką automatiniam kalbos atpažinimui.

6.8 lentelė. 2 bandymo rezultatų palyginimas

Prarastų kadrų skaičius	Atpažintų žodžių skaičius, kai paketai prarandami pirmoje žodžio raidėje (priebalsėje), %		Atpažintų žodžių skaičius, kai paketai prarandami žodžio viduryje esančioje priebalsėje, %		Atpažintų žodžių skaičius, kai paketai prarandami brt kurioje žodžio balsėje, %	
	panaudotas 50 etaloninių žodžių žodynas	panaudotas 10 etaloninių žodžių žodynas	panaudotas 50 etaloninių žodžių žodynas	panaudotas 10 etaloninių žodžių žodynas	panaudotas 50 etaloninių žodžių žodynas	panaudotas 10 etaloninių žodžių žodynas
0	60	80	80	90	90	90
1	70	90	80	90	90	90
2	60	80	60	70	90	90
3	50	70	70	70	90	90
4	50	70	70	80	90	90
5	50	60	60	70	70	90

Apibendrinant atliktus tyrimus galima padaryti keletą išvadų.

1. Priebalsės daro didesnę įtaką automatiniam kalbos atpažinimui negu balsės. Be to, priebalsė, esanti žodžio pradžioje, daro didesnę įtaką automatiniam kalbos atpažinimui negu priebalsė, esanti žodžio viduryje. Tokios pat išvados buvo padarytos ir atlikus paketais perduotų žodžių suprantamumo tyrimą [5].

2. Didelę įtaką automatiniam kalbos atpažinimui daro panašių žodžių skaičius etaloninių žodžių žodyne. Kuo daugiau panašių žodžių, tuo sudėtingiau teisingai atpažinti tiriamuosius žodžius. Sumažinus etaloninių žodžių žodyne esančių žodžių skaičių nuo 50 iki 10, automatinio kalbos atpažinimo rezultatai pagerėjo nuo 10 iki 20 procentų.

3. Sunku įgyvendinti sąlygą, kad etaloniniai ir tiriamieji žodžiai būtų idealiai vienodi ar labai panašūs. Diktoriui sudėtinga ištarti tokius pačius ar panašius žodžius vienoda tonacija, vienodo ilgio, panašiu greičiu ir pan. O atliekant aparatinį kalbos atpažinimą remiamasi tik algoritmų skaičiavimais, nėra galimybės atspėti žodžius, kaip tai gali įvykti, kai žodžius bando atpažinti žmonės.

6.3. 3 bandymas

Kitas tyrimas atliktas, kai tiriamieji žodžiai buvo lyginami su kito diktoriaus (moters) pasakytais žodžiais. Kaip ir ankstesnio eksperimento metu, taip ir šio, panaudotas etaloninių žodžių žodynas, aprašytas 5.2 poskyryje. Atlikti trys bandymai: kai prarandami kadrai yra pirmoje žodžio raidėje (priebalsėje), kai parandami kadrai yra žodžio viduryje esančioje priebalsėje, ir kai prarandami kadarai yra bet kurioje žodžio balsėje.

Pirmiausiai atliktas bandymas kai prarasti kadrai yra žodžio pradžioje (priebalsėje). Tyrimui panaudota dešimt testinių žodžių: *batas, matas, patas, ratas, vatas, dėlė, gėlė, lėlė, vėlė, žėlė*. Tyrimo rezultatai pateikti 6.9 lentelėje. Iš lentelėje pateiktų rezultatų matome, kad esant skirtingiems etaloninių ir tiriamųjų žodžių diktoriams, automatinio kalbos atpažinimo rezultatai stipriai pablogėjo lyginant su tyrimo rezultatais, kai etaloniniai ir tiriamieji žodžiai yra to paties diktoriaus (6.2 lentelė). Perkodavus AMR kodėku testinį žodį atpažintas tik vienas žodis iš dešimties. Praradus vieną ir du 20 milisekundžių trukmės kadrus atpažinta po du žodžius, praradus tris ir keturis kadrus atpažinta po vieną žodį, o praradus penkis kadrus atpažinti trys žodžiai.

6.9 lentelė. Tyrimo rezultatai prarandant paketus pirmoje žodžio raidėje (priebalsėje)

Tiriamas žodis	Žodis su neprarastais kdrais	Žodis su prarastu 1 kadru	Žodis su prarastais 2 kdrais	Žodis su prarastais 3 kdrais	Žodis su prarastais 4 kdrais	Žodis su prarastais 5 kdrais
Batas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	13,0281	13,0266	13,1838	12,4527	13,1957	13,0093
Matas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	14,3670	14,0110	13,1314	12,7088	13,0505	12,2151
Patas	Atpažinta	Atpažinta	Atpažinta	Neatpažinta	Neatpažinta	Atpažinta
	13,5236	12,5529	12,6156	16,6122	16,5364	17,0815
Ratas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	16,2483	16,2195	15,7195	15,6864	15,5942	16,2958
Vatas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	13,9786	13,1924	12,7448	13,6277	18,1651	17,8166
Dėlė	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	8,5370	8,3846	7,9912	8,0481	7,9001	8,3389
Gėlė	Neatpažinta	Neatpažinta	Atpažinta	Atpažinta	Neatpažinta	Atpažinta
	10,2876	10,2648	10,4038	10,3039	10,2532	10,4623
Lėlė	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	12,2438	11,3912	11,6254	11,7557	11,8382	12,9431
Vėlė	Neatpažinta	Atpažinta	Neatpažinta	Neatpažinta	Atpažinta	Atpažinta
	10,2310	10,0042	9,7559	9,6477	11,3831	11,8871
Žėlė	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	9,5461	9,5492	9,4073	9,5150	9,3187	8,9770

Kitas bandymas atliktas, kai prarasti kadrai yra žodžio vidurinėje raidėje (priebalsėje). Tyrimui panaudota dešimt testinių žodžių: *dėlė, dėmė, dėžė, galas, garas, mačas, magas, maras, matas, mažas*. Tyrimo rezultatai pateikti 6.10 lentelėje. Šio bandymo metu neatpažintas nė vienas žodis. Akivaizdu, kad esant skirtingiems tiriamųjų ir etaloninių žodžių diktoriaus bei prarandant kadrus žodžio viduryje esančioje priebalsėje, automatinio kalbos atpažinimo rezultatai stipriai pablogėjo.

6.10 lentelė. Tyrimo rezultatai prarandant paketus vidurinėje žodžio raidėje (priebalsėje)

Tiriamas žodis	Žodis su neprarastais kadrais	Žodis su prarastu 1 kadru	Žodis su prarastais 2 kadrais	Žodis su prarastais 3 kadrais	Žodis su prarastais 4 kadrais	Žodis su prarastais 5 kadrais
Dėlė	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	9,7201	10,2833	10,1879	9,7861	9,1913	8,7253
Dėmė	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	11,1593	10,8712	10,4448	10,3698	10,7276	10,4640
Dėžė	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	11,6153	10,7753	10,7009	10,5926	10,3759	10,059
Galas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	14,4234	13,4288	13,0082	12,8797	11,5669	12,8543
Garas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	12,2661	12,4083	12,2301	12,2631	10,7887	11,3734
Mačas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	16,7037	16,1621	14,6952	14,1893	13,7948	13,8604
Magas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	12,3505	10,8458	10,4949	10,7362	11,4781	11,8338
Maras	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	12,5950	11,0792	12,6758	11,8718	11,8467	11,1165
Matas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	13,3804	13,3619	13,3519	13,4280	13,3756	13,2072
Mažas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	13,8134	13,5389	13,3519	12,9323	12,6530	12,6861

Trečias bandymas atliktas kai prarasti kadrai yra žodžio balsėje. Tyrimui panaudota dešimt testinių žodžių: *lama, lema, loma, ratas, retas, rytas, gama, guma, vėsa, visa*. Tyrimo rezultatai pateikti 6.11 lentelėje. Perkodavus AMR kodeku testinį žodį atpažintas vienas žodis iš dešimties. Praradus vieną ir du 20 milisekundžių trukmės kadrus neatpažintas nė vienas žodis, praradus tris kadrus atpažinti du žodžiai, o praradus keturis ir penkis kadrus atpažinta po vieną žodį. Taigi ir šiuo atveju automatinio kalbos atpažinimo rezultatai stipriai pablogėjo lyginant su atpažinimo rezultatais, kai ir tiriamieji, ir etaloniniai žodžiai yra pasakyti to paties diktoriaus.

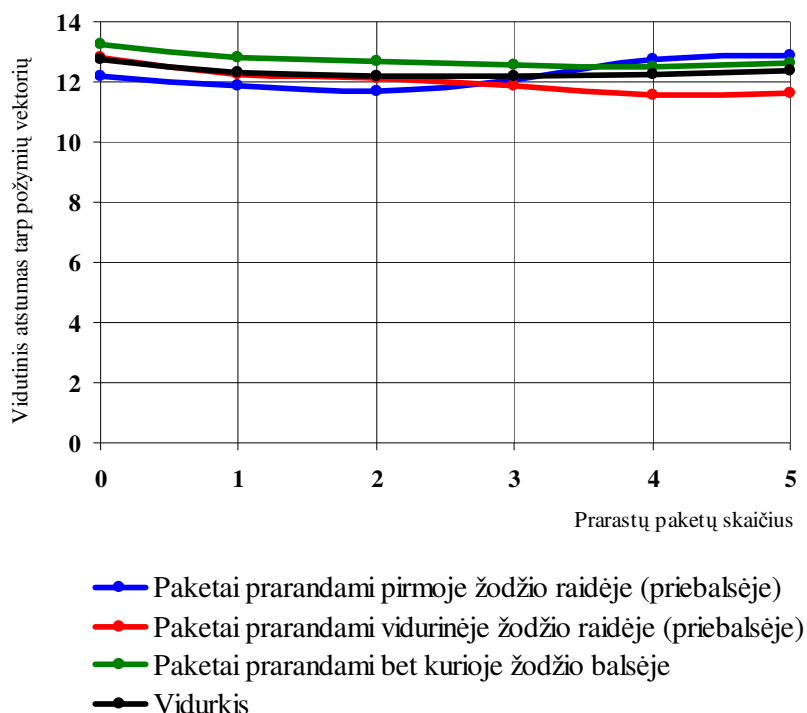
Esant skirtingiems etaloninių ir tiriamųjų žodžių diktoriaus negalime vienareikšmiškai pasakyti, kurios raidės (balsės ar priebalsės) yra reikšmingesnės automatiniam kalbos atpažinimui. Reikalingi papildomi tyrimai.

6.11 lentelė. Tyrimo rezultatai prarandant paketus žodžio balsėje

Tiriamas žodis	Žodis su neprarastais kadrais	Žodis su prarastu 1 kadru	Žodis su prarastais 2 kadrais	Žodis su prarastais 3 kadrais	Žodis su prarastais 4 kadrais	Žodis su prarastais 5 kadrais
Gama	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	11,7865	11,9137	11,9586	11,8080	11,3217	10,8062
Guma	Atpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	12,1502	11,6167	11,4319	11,4839	11,4966	11,7625
Lama	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	12,9077	11,7354	12,0035	11,8001	12,6432	12,8743
Lema	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	14,1195	14,2706	14,1171	13,9848	12,8973	13,5889
Loma	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	14,6635	14,4697	14,2138	14,3596	14,7198	15,1099
Ratas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	15,2830	14,3077	14,0767	13,6984	14,0003	14,2447
Retas	Neatpažinta	Neatpažinta	Neatpažinta	Atpažinta	Atpažinta	Atpažinta
	14,0307	13,2430	13,2517	12,7950	12,7140	12,1409
Rytas	Neatpažinta	Neatpažinta	Neatpažinta	Atpažinta	Neatpažinta	Neatpažinta
	13,6711	13,0417	12,9940	12,5831	12,6764	12,2133
Vėsa	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	11,4649	11,1872	11,3731	11,3963	11,2121	11,2542
Visa	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	12,2418	12,2631	11,7337	11,8277	11,6048	12,3274

Paanalizuokime kaip kinta vidutiniai atstumai tarp etaloninio ir tiriamojo žodžių požymių vektorių, kai prarandami paketai skirtingose žodžio vietose ir kai etaloniniai bei tiriamieji žodžiai yra skirtingų diktorių. 11 paveiksle pavaizduotos vidutinio atstumo tarp požymių vektorių priklausomybės nuo prarastų paketų skaičiaus. Iš paveikslo matome, kad šiuo atveju atstumai tarp požymių vektorių yra linę nestipriai mažėti didėjant prarastų paketų skaičiui. Tik kai prarandami paketai pirmoje žodžio raidėje, praradus tris ir daugiau AMR kodeko kadru, vidutinis atstumas tarp požymių vektorių truputį padidėjo (11 paveiksle mėlyna linija). Taigi, kai paketai prarandami pirmoje žodžio raidėje (priebalsėje), rezultatai panašūs kaip ir tuo atveju, kai ir tiriamųjų, ir etaloninių žodžių diktorius yra tas pats žmogus. Priežastys galėtų būti taip pat tokios pat: kai prarandami kadrai žodžio pirmoje raidėje, AMR kodekas pakeičia prieš tai buvusiais kadrais, kuriuose nėra kalbos signalo. Skirtingai negu tuo atveju, kai tiriamieji ir etaloniniai žodžiai buvo ištarti to paties diktoriaus, 11 paveiksle pavaizduotose priklausomybėse matome, kad vidutiniai atstumai tarp požymių vektorių yra panašūs prarandant paketus tiek balsėje, tiek priebalsėje. Lyginant su 10 paveiksle pateiktomis priklausomybėmis, matome, kad esant skirtingiems tiriamųjų ir etaloninių žodžių diktoriaus, vidutiniai atstumai tarp požymių vektorių yra didesni, nei tuo atveju, kai visi žodžiai yra to paties kalbančiojo. Kiekvieną kalbantįjį apibūdina taip vadinami intraindividualūs ir interindividualūs iškraipymai [20]. Intraindividualūs iškraipymai – tai to paties žmogaus skirtingų balso įrašų vidutinių atstumų tarp požymių vektorių

pasiskirstymas. Interindividualūs iškraipymai – tai skirtingų žmonių balso įrašų vidutinių atstumų tarp požymių vektorių pasiskirstymas. Skirtingų žmonių balso įrašams interindividualių iškraipymų pasiskirstymas yra pasislinkęs intraindividualių iškraipymų pasiskirstymo atžvilgiu, todėl atstumai tarp skirtingų žmonių balso įrašų yra didesni negu atstumai tarp to paties žmogaus balso įrašų.



11 pav. Vidutinio atstumo tarp požymių vektorių priklausomybė nuo prarastų paketų skaičiaus

6.2 poskyryje atlikto prarastų paketų įtakos automatiniam kalbos atpažinimui tyrimo metu pastebėta, kad teisingai atpažintų žodžių skaičius padidėjo, kai etaloninių žodžių žodyne buvo mažiau panašių žodžių. Dėl šios priežasties pabandyta sumažinti žodyne esančių žodžių skaičių kai etaloninių ir tiriamųjų žodžių diktoriai skirtingi.

Kaip ir anksčiau, pirmiausiai pabandyta išsiaiškinti kokią įtaką automatiniam kalbos atpažinimui daro žodyno dydis, kai paketai prarandami pirmoje žodžio raidėje (priebalsėje). Etaloninių žodžių žodynas sudarytas iš dešimties žodžių: *batas, matas, patas, ratas, vatas, dėlė, gėlė, lėlė, vėlė, žėlė*. Tiriamieji žodžiai pasirinkti tokie patys. Tyrimo rezultatai pateikti 6.12 lentelėje. Kaip matome iš lentelėje pateiktų rezultatų, automatinis kalbos atpažinimas pagerėjo tik truputį. Lyginant su 6.9 lentelėje pateiktais rezultatais, atpažintų žodžių skaičius padidėjo tik

praradus du, tris ir keturis 20 milisekundžių trukmės kadrus atitinkamai vienu, dviem ir vienu žodžiu. Sumažinus etaloninių žodžių skaičių žodyne, automatinio kalbos atpažinimo rezultatai beveik nepagerėjo.

6.12 lentelė. Tyrimo rezultatai prarandant paketus pirmoje žodžio raidėje (priebalsėje)

Tiriamas žodis	Žodis su neprarastais kdrais	Žodis su prarastu 1 kdrais	Žodis su prarastais 2 kdrais	Žodis su prarastais 3 kdrais	Žodis su prarastais 4 kdrais	Žodis su prarastais 5 kdrais
Batas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	15,2005	13,6627	13,8818	12,4527	13,1957	13,0093
Matas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	14,3670	14,0110	13,1314	12,7088	13,0505	12,2151
Patas	Neatpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	13,5236	12,5529	12,6156	17,6150	16,8793	17,0815
Ratas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	16,2483	16,2195	15,7195	15,6864	16,1985	16,6327
Vatas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	14,0402	13,2628	12,7448	13,6277	18,1651	17,8166
Dėlė	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	8,5370	8,3846	7,9912	8,0418	7,9001	8,3389
Gėlė	Neatpažinta	Neatpažinta	Atpažinta	Atpažinta	Neatpažinta	Atpažinta
	10,5922	10,7864	14,3257	14,4427	10,2532	14,7483
Lėlė	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	12,2438	11,3912	11,6254	11,7557	11,8382	12,9431
Vėlė	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	10,3454	10,0042	10,1015	9,8993	11,3831	11,8871
Žėlė	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	9,8463	9,8464	9,6190	9,5459	9,3187	8,9770

Kitas bandymas atliktas kai prarasti kadrai yra žodžio viduryje esančioje priebalsėje. Etaloninių žodžių žodynas sudarytas iš dešimties žodžių: *dėlė, dėmė, dėžė, galas, garas, mačas, magas, maras, matas, mažas*. Tiriamieji žodžiai pasirinkti tokie patys. Tyrimo rezultatai pateikti 6.13 lentelėje. Šio bandymo metu pastebėta, kad automatinio kalbos atpažinimo rezultatai stipriai pagerėjo, ypač kai prarandami tik keli paketai. Tik perkodavus žodį AMR kodeku atpažinti šeši testiniai žodžiai, praradus vieną ir du kadrus atpažinta po penkis žodžius, praradus tris kadrus atpažinti keturi žodžiai, praradus keturis kadrus atpažinti trys žodžiai ir praradus penkis kadrus atpažinti tik du žodžiai. Lyginant 6.12 ir 6.13 lentelėse pateiktus rezultatus matome, kad žodžio pradžioje esanti priebalsė yra reikšmingesnė negu žodžio viduryje esanti priebalsė.

6.13 lentelė. Tyrimo rezultatai prarandant paketus vidurinėje žodžio raidėje (priebalsėje)

Tiriamas žodis	Žodis su neprarastais kdrais	Žodis su prarastu 1 kdrais	Žodis su prarastais 2 kdrais	Žodis su prarastais 3 kdrais	Žodis su prarastais 4 kdrais	Žodis su prarastais 5 kdrais
Dėlė	Atpažinta	Atpažinta	Atpažinta	Neatpažinta	Neatpažinta	Atpažinta
	12,9070	12,9908	12,4987	11,6917	10,7534	9,7431
Dėmė	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Neatpažinta	Neatpažinta
	13,9698	13,6718	11,5245	15,8568	12,2678	12,3044
Dėžė	Atpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	14,7527	14,8386	14,4801	12,7955	13,3912	12,2742
Galas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Neatpažinta
	15,9377	14,1989	13,5079	13,0510	13,0572	16,2193
Garas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	13,4756	13,6006	13,4736	13,1647	12,5484	12,8201
Mačas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	19,1528	18,2317	16,6320	16,4132	15,8865	16,0425
Magas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	12,8926	10,8458	10,4949	10,7362	11,4781	13,6529
Maras	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Neatpažinta
	13,5280	12,8641	13,5670	12,9368	12,3883	12,0616
Matas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	15,2584	14,5115	15,2849	14,8265	14,6916	14,1161
Mažas	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	15,7423	14,1623	14,3240	12,9323	12,6530	12,6881

Ir trečias bandymas atliktas kai prarasti kadrų yra žodžio balsėje. Etaloninių žodžių žodynas sudarytas iš dešimties žodžių: *lama, lema, loma, ratas, retas, rytas, gama, guma, vėsa, visa*. Tiriamieji žodžiai pasirinkti tokie patys. Tyrimo rezultatai pateikti 6.14 lentelėje. Šiuo atveju atpažinimo rezultatai taip pat pagerėjo, tačiau ne taip stipriai, kaip tuo atveju, kai paketai prarandami žodžio viduryje esančioje priebalsėje. Testinį žodį tik perkodavus AMR kodeku, praradus vieną, keturis ir penkis kadrus atpažinta po keturis žodžius, praradus du kadrus atpažinti trys žodžiai, o praradus tris kadrus atpažinti penki žodžiai. Automatinio kalbos atpažinimo rezultatai pagerėjo apie 30 procentų. Iš šių rezultatų matome, kad skirtingai nei tuo atveju, kai ir etaloniniai, ir tiriamieji žodžiai yra to paties diktorius, balsės yra reikšmingos žodžių dalys automatiškai atpažįstant žodžius.

6.14 lentelėje parodyta, kaip skyrėsi automatinio kalbos atpažinimo rezultatai, kai buvo naudojami skirtingo dydžio etaloninių žodžių žodynai. Naudojant mažesnę žodyną ir modeliuojant paketų praradimus pirmoje žodžio raidėje, atpažinimo rezultatai pagerėjo apie 10 procentų. Kai paketai prarandami žodžio viduryje esančioje priebalsėje, automatinio atpažinimo rezultatai pagerėjo iki 60 procentų (priklausomai nuo prarastų paketų skaičiaus). Ir modeliuojant paketų praradimus balsėje, atpažinimo rezultatai pagerėjo apie 30 procentų. Iš šioje lentelėje pateiktų rezultatų akivaizdžiai matome, kad žodyno dydis turi didelę įtaką automatiniam kalbos atpažinimui.

6.14 lentelė. Tyrimo rezultatai prarandant paketus žodžio balsėje

Tiriamas žodis	Žodis su neprarastais kadrtais	Žodis su prarastu 1 kadrtais	Žodis su prarastais 2 kadrtais	Žodis su prarastais 3 kadrtais	Žodis su prarastais 4 kadrtais	Žodis su prarastais 5 kadrtais
Gama	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	16,3992	14,5539	15,5233	16,5792	15,2146	14,4991
Guma	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Neatpažinta
	12,1502	12,3006	12,4865	12,2498	12,1017	13,6205
Lama	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	12,9077	11,7354	12,0035	11,8001	12,6432	12,8743
Lema	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	14,8243	14,7244	14,6568	14,6132	13,8774	14,3379
Loma	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	16,0478	16,2210	15,9825	15,9576	15,8255	15,6507
Ratas	Neatpažinta	Neatpažinta	Neatpažinta	Atpažinta	Atpažinta	Atpažinta
	16,9401	16,0776	16,0661	16,0111	16,0812	15,1857
Retas	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta	Atpažinta
	14,5016	13,3238	13,5100	12,7950	12,7140	12,1409
Rytas	Atpažinta	Atpažinta	Neatpažinta	Atpažinta	Neatpažinta	Neatpažinta
	14,0015	13,0597	13,0920	12,5831	12,6764	12,7548
Vėsa	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta
	12,7890	12,4098	12,8122	12,7594	12,6227	12,8338
Visa	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Neatpažinta	Atpažinta
	13,8239	14,4296	13,4990	13,4404	13,2527	13,4728

6.15 lentelė. 3 bandymo rezultatų palyginimas

Prarastų kadrų skaičius	Atpažintų žodžių skaičius, kai paketai prarandami pirmoje žodžio raidėje (priebalsėje), %		Atpažintų žodžių skaičius, kai paketai prarandami žodžio viduryje esančioje priebalsėje, %		Atpažintų žodžių skaičius, kai paketai prarandami birt kurioje žodžio balsėje, %	
	panaudotas 50 etaloninių žodžių žodynas	panaudotas 10 etaloninių žodžių žodynas	panaudotas 50 etaloninių žodžių žodynas	panaudotas 10 etaloninių žodžių žodynas	panaudotas 50 etaloninių žodžių žodynas	panaudotas 10 etaloninių žodžių žodynas
0	10	10	0	60	10	40
1	20	20	0	50	0	40
2	20	30	0	50	0	30
3	10	30	0	40	20	50
4	10	20	0	30	10	40
5	30	30	0	20	10	40

Apibendrinant 6.3 poskyryje atliktus tyrimus galima padaryti keletą išvadų:

1. Priebalsė, esanti žodžio pradžioje, daro didesnę įtaką automatiniam kalbos atpažinimui negu priebalsė, esanti žodžio viduryje. Tokios pat išvados padarytos ir atlikus paketais perduotų žodžių suprantamumo tyrimą [5].

2. Šiuo atveju taip pat didelę įtaką automatiniam kalbos atpažinimui daro panašių žodžių skaičius etaloninių žodžių žodyne. Kuo daugiau panašių žodžių, tuo sudėtingiau teisingai atpažinti tiriamuosius žodžius. Sumažinus etaloninių žodžių žodyne esančių žodžių skaičių nuo 50 iki 10, automatinio kalbos atpažinimo rezultatai prarandant skirtingą paketų skaičių skirtingose raidėse pagerėjo nuo 10 iki 60 procentų.

7. DARBO REZULTATŲ APIBENDRINIMAS

Baigiamojo magistro darbo metu apžvelgtas automatinio kalbos atpažinimo aktualumas, trumpai apžvelgta atpažinimo tyrimų istorija bei aptartos automatinio kalbos atpažinimo sistemos. Aptartas dinamiško laiko skalės kraipymo algoritmas ir šiame algoritme naudojami požymių vektoriai: kepstras, melų skalės dažnio kepstro bei tiesinės prognozės koeficientai. Kalbos signalo kepstro koeficientai teikia vertingą informaciją apie pagrindinio tono arba balso stygų svyravimų dažnį. Kepstrinės analizės privalumas yra tas, kad išsaugoma informacija apie signalo fazę ir tai leidžia atkurti signalo laiko struktūrą. Tiesinės prognozės koeficientai gali būti naudojami daugeliui tikslų: kalbos signalų suspaudimui, atpažinimui, sintezei. Tiesinė prognozė yra vienas galingiausių signalų analizės metodų. Šiuo metodu gaunami kalbą aprašantys parametrai, jų skaičius yra nedidelis ir jie apskaičiuojami greitai. Kalbos atpažinime dominuoja tiesinės prognozės koeficientais ir jų pagrindu gaunami požymiai. Koeficientai glaustai pateikia informaciją apie kalbos fragmento spektro gaubiamąją. Kadangi žmogaus klausos aparatas taip pat kalbos signalo fragmentus transformuoja į dažninę sritį ir vėliau naudoja spektrinę informaciją atpažinimui, todėl ir tiesinės prognozės koeficientų naudojimas atpažinime pasiteisina.

Atlikus prarastų paketų įtakos automatiniam kalbos atpažinimui tyrimą nustatytas lietuviškų žodžių balsių ir priebalsių reikšmingumas. Padarytos išvados:

1. Priebalsės daro didesnę įtaką automatiniam kalbos atpažinimui negu balsės. Be to, priebalsė, esanti žodžio pradžioje, daro didesnę įtaką automatiniam kalbos atpažinimui negu priebalsė, esanti žodžio viduryje. Tokios pat išvados buvo padarytos ir atlikus paketais perduotų žodžių suprantamumo tyrimą [5].

2. Priebalsės yra reikšmingesnės automatiniam kalbos atpažinimui tiek tuo atveju, kai ir tiriamieji, ir etaloniniai žodžiai yra pasakyti to paties diktorius, tiek tuo atveju, kai tiriamieji ir etaloniniai žodžiai yra pasakyti skirtingų diktorių.

3. Didelę įtaką automatiniam kalbos atpažinimui daro panašių žodžių skaičius etaloninių žodžių žodyne. Kuo daugiau panašių žodžių, tuo sudėtingiau teisingai atpažinti tiriamuosius žodžius. Sumažinus etaloninių žodžių žodyne esančių žodžių skaičių nuo 50 iki 10, automatinio kalbos atpažinimo rezultatai stipriai pagerėjo. Kai kuriais atvejais net iki 60 procentų.

4. Dinaminio laiko skalės kraipymo algoritmas skirtas automatinio kalbos atpažinimo tyrimams, kai naudojamas nedidelis žodynas. Todėl geresni rezultatai gauti, kai etaloninių žodžių žodynas buvo mažesnis. Taip pat yra žinoma, kad geriau atpažįstami ilgi žodžiai, o atliktame tyrime naudoti trumpi, dviskiemeniai žodžiai. Tai taip pat turėjo įtakos automatinio kalbos atpažinimo rezultatams.

5. Sunku įgyvendinti sąlygą, kad etaloniniai ir tiriamieji žodžiai būtų vienodi ar labai panašūs. Diktoriui sudėtinga taip pat ar panašiai ištarti žodžius: vienoda tonacija, vienodo ilgio, panašiu greičiu ir pan. Todėl dažniausiai tiriamasis ir lyginamasis garso įrašai stipriai skiriasi. Šis skirtumas susidaro dėl skirtingų garso įrašų darymo sąlygų, asmens skirtingų emocinių būsenų, triukšmo įtakos, garso įrašymo kanalų nesutapimo ir pan. Atliekant automatinį kalbos atpažinimą remiamasi tik algoritmų skaičiavimais, todėl garso įrašų skirtumai turi didelę įtaką blogiems automatinio kalbos atpažinimo rezultatams. Tuo tarpu kalbos atpažinime dalyvaujant žmonėms, yra galimybė žodžius tiesiog atspėti.

8. LITERATŪRA

1. Girčys R., Stankus A. Pulsinės bangos dedamųjų tyrimas dekompozicijos metodu. Prieiga per Internetą <http://www.elin.ttu.ee/BME-Wg/R&D/Confinfo/BME97/SEKCIJOS/4SEKC-7.HTM>, žiūrėta 2007 05 05.
2. Juang B.H., Rabiner L. R. Automatic Speech Recognition – A Brief History of the Technology Development. Georgia Institute of Technology, Atlanta.
3. Kamarauskas J. Pagrindinių tonų sinchronizuotas kalbos signalo apdorojimas (PSSP) kalbančiojo atpažinime. Konferencijos medžiaga. 2005.
4. Lipeika A. Formantiniai požymiai kalbos atpažinime. Konferencijos medžiaga. 2005.
5. Raškauskas A. Paketais perduotų žodžių suprantamumo tyrimas. Baigiamasis bakalauro darbas. Vilnius, 2006.
6. Rudžionis A. Computer recognition of isolated words in fixed length feature space. // Proc. of the XIth International Congress of Phonetic Sciences.- Tallinn, Estonia, 1987.-Vol.5.
7. Song Z., Zhang G., Zheng F. Comparison of Different Implementations of MFCC. Center of Speech Technology, State Key Lab of Intelligent Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, P.R. China.
8. Vibro B., Vibro K. Cepstrum Analysis. Prieiga per Internetą <http://www.bkvibro.com/Main/Monitoring+Solutions+&+Applications/Measurement+Techniques/Cepstrum+Analysis>, žiūrėta 2007 05 05
9. 3rd Generation Partnership Project. Adaptive Multi-Rate (AMR) codec. 3GPP Specifications Series. Prieiga per Internetą <http://www.3gpp.org/ftp/Specs/html-info/26-series.htm>, žiūrėta 2007 10 27.
10. Automatinis atskirai sakomų lietuvių kalbos žodžių atpažinimas, naudojant kalbos pavyzdžių dinaminį laiko skalės kraipymą. Prieiga per Internetą http://www.likit.lt/all/balso_tech/btech_mii.htm, žiūrėta 2007 04 17.
11. Automatic Speech Recognition (ASR) Speech / Voice Recognition. Prieiga per Internetą <http://www.globalsecurity.org/intell/systems/asr.htm>, žiūrėta 2008 03 30.
12. Automatic Speech Recognition Techniques. Prieiga per Internetą <http://www.globalsecurity.org/intell/systems/asr-tech.htm>, žiūrėta 2008 03 30.
13. Cepstrum. From Wikipedia, the free encyclopedia. Prieiga per Internetą <http://en.wikipedia.org/wiki/Cepstrum>, žiūrėta 2007 10 12.
14. Discrete cosine transform. From Wikipedia, the free encyclopedia. Prieiga per Internetą http://en.wikipedia.org/wiki/Discrete_cosine_transform, žiūrėta 2007 10 12.

15. Dynamic Time Warping. From Wikipedia, the free encyclopedia. Prieiga per Internetą http://en.wikipedia.org/wiki/Dynamic_time_warping, žiūrėta 2007 01 20.
16. Kepstro koeficientai ir konvejerinis (pipeline) signalų apdorojimas <http://www.mif.vu.lt/~bastys/academic/ATE/Kepstras/kepstras.htm#kepstras>, žiūrėta 2007 04 14.
17. Mark Huckvale University College London. SFS/ESection Help File. Prieiga per Internetą <http://www.phon.ucl.ac.uk/resource/sfs/esection/esection.htm>, žiūrėta 2007 05 06.
18. Mel-frequency cepstrum. From Wikipedia, the free encyclopedia. Prieiga per Internetą http://en.wikipedia.org/wiki/Mel_frequency_cepstral_coefficient, žiūrėta 2007 10 12.
19. Mel scale. From Wikipedia, the free encyclopedia. Prieiga per Internetą http://en.wikipedia.org/wiki/Mel_scale, žiūrėta 2007 10 12.
20. SIVE ir fonoskopinės ekspertizės. Kalbos signalų apdorojimo metodų bei balso suvokimo modelių, naudojamų SIVE programinės įrangos pakete, aprašymas. Prieiga per Internetą http://www.technogama.lt/Kalbos/lt_sive.html, žiūrėta 2008 04 01.
21. Speech Recognition Based on DTW 1.0 - Matlab source code. Prieiga per Internetą <http://www.advancedsourcecode.com/dtwspeech.asp>, žiūrėta 2007 10 11.
22. Tiesinės prognozės koeficientai. Prieiga per Internetą http://www.mif.vu.lt/~bastys/academic/ATE/LPC/LPC_Kodavimas.htm#2, žiūrėta 2007 06 01.
23. Welcome to the Speaker Recognition Homepage. Prieiga per Internetą <http://speaker-recognition.info>, žiūrėta 2007 10 09.

PRIEDAI

A priedas. Etaloninių ir tiriamųjų žodžių žodynų kompaktinė plokštelė