

VILNIAUS UNIVERSITETAS
KAUNO HUMANITARINIS FAKULTETAS

INFORMATIKOS KATEDRA

Verslo informacijos sistemų studijų programa

Kodas 62603S108

IRUTĖ RAKAUSKAITĖ

MAGISTRO BAIGIAMASIS DARBAS

**KALBOS TECHNOLOGIJŲ PRODUKTŲ ADAPTAVIMO
GALIMYBIŲ KOMANDŲ ATPAŽINIMUI IR SINTEZEI TYRIMAS**

Kaunas 2009

VILNIAUS UNIVERSITETAS
KAUNO HUMANITARINIS FAKULTETAS

INFORMATIKOS KATEDRA

IRUTĖ RAKAUSKAITĖ

MAGISTRO BAIGIAMASIS DARBAS

**KALBOS TECHNOLOGIJŲ PRODUKTŲ ADAPTAVIMO
GALIMYBIŲ KOMANDŲ ATPAŽINIMUI IR SINTEZEI TYRIMAS**

Leidžiama ginti _____

Magistrantė _____

(parašas)

Darbo vadovas _____

(parašas)

Doc. dr. Vytautas Rudžionis

Darbo įteikimo data _____

Registracijos Nr. _____

Kaunas 2009

TURINYS

TURINYS	3
LENTELIŲ SĄRAŠAS	4
PAVEIKSLŲ SĄRAŠAS	5
SANTRUMPŲ SĄRAŠAS	6
SUMMARY	7
ĮVADAS	8
1. KALBOS TECHNOLOGIJŲ IR JŲ PRODUKTŲ APŽVALGA	10
1.1. <i>Garso samprata</i>	10
1.2. <i>Natūralios ir šnekamosios kalbos apdorojimo taikymo sritys</i>	12
1.3. <i>Svarbiausios balso technologijos, jų paskirtys ir panaudojimo sferos</i>	14
1.4. <i>Sudėtinės atpažinimo funkcijos</i>	19
1.5. <i>Kalbos sintezės metodai</i>	21
1.6. <i>Kalbos technologijų darbai stambiausiose pasaulio kompanijose ir Lietuvoje</i>	23
2. ESAMŲ KALBOS TECHNOLOGIJŲ PRODUKTŲ ANALIZĖ	28
2.1. <i>Pagrindinės šnekamosios kalbos atpažinimo taikymo sritys ir jose taikomi produktai</i>	28
2.1.1. <i>Kalbinis dialogas su kompiuteriu</i>	28
2.1.2. <i>Balso įrašų stenografavimas</i>	29
2.1.3. <i>Balsinė interneto naršyklė</i>	30
2.1.4. <i>Elektroninių prietaisų valdymas balsu</i>	31
2.1.5. <i>Asmens atpažinimas pagal jo balsą</i>	32
2.2. <i>Šnekamosios kalbos atpažinimo produktų palyginimas</i>	34
2.3. <i>Kalbos sintezei taikomi produktai ir jų palyginimas</i>	35
3. KALBOS TECHNOLOGIJŲ PRODUKTŲ ADAPTAVIMO GALIMYBIŲ TYRIMAS	37
3.1. <i>Lietuvių fonetikos transformavimo į anglų kalbos rašybą taisyklių sudarymo galimybių tyrimas</i>	37
3.2. <i>Anglišku žodžių atpažinimo tyrimas</i>	49
IŠVADOS	58
LITERATŪRA	59
PRIEDAI	63

LENTELIŲ SĄRAŠAS

1 lentelė.	Kalbos signalų atpažinimo taikymo ypatybės telefonijos ir multimedijos terpėse ..	18
2 lentelė.	„NaturallySpeaking Medical“ sistemos reikalavimai	29
3 lentelė.	„Toki Voki Voice Editor“ sistemos reikalavimai	29
4 lentelė.	„Windows Vista Home Basic“ sistemos reikalavimai	30
5 lentelė.	„Dragon NaturallySpeaking 10“ sistemos reikalavimai	31
6 lentelė.	„Via Voice“ sistemos reikalavimai	32
7 lentelė.	„Raktas“ sistemos reikalavimai.....	33
8 lentelė.	Šnekamosios kalbos atpažinimo produktų charakteristikos.....	34
9 lentelė.	Antrojo dalyvio žodžių aiškumo ir kokybės vertinimas	38
10 lentelė.	Trečiojo dalyvio žodžių aiškumo ir kokybės vertinimas	40
11 lentelė.	Antro ir trečio dalyvio vidutinis žodžių aiškumo ir kokybės vertinimas	42
12 lentelė.	Aukščiausias žodžių kokybės vertinimas.....	44
13 lentelė.	Transkripcijos taisyklių modelis	47
14 lentelė.	Žodžių atpažinimo skaičius.....	50
15 lentelė.	Klaidų skaičius	54
16 lentelė.	Priebalsinių garsų klaidų skaičius	55
17 lentelė.	Balsinių garsų klaidų skaičius.....	56

PAVEIKSLŲ SĄRAŠAS

1 pav.	Kalbos mechanizmas.....	10
2 pav.	Periodinės garso bangos.....	11
3 pav.	Apibendrinta nenutrūkstamos šnekamosios kalbos atpažinimo sistemos schema	13
4 pav.	Markovo grandinė su penkiomis būsenomis.....	19
5 pav.	Sintezės procesų diagrama.....	22
6 pav.	„ReadPlease 2003“ programos pagrindinis langas	38
7 pav.	Lietuviškų žodžių perskaitymo aiškumas	45
8 pav.	Lietuviškų žodžių perskaitymo kokybė	45
9 pav.	Geriausiai transkribuotų lietuviškų žodžių perskaitymo aiškumas.....	46
10 pav.	Geriausiai transkribuotų lietuviškų žodžių perskaitymo kokybė.....	46
11 pav.	„Dragon Naturally Speaking 10“ programos pagrindinis langas.....	49
12 pav.	Penkių diktorių žodžių atpažinimo efektyvumas	52
13 pav.	Diktorių žodžių atpažinimo efektyvumas kiekvienai žodžių grupei.....	53
14 pav.	Klaidos kiekvienai žodžių ilgio grupei	53
15 pav.	Klaidos žodžiuose su dusliaisiais ir skardžiais priebalsiais.....	56

SANTRUMPŲ SĄRAŠAS

- 3G** – Trečios kartos mobiliojo ryšio technologija (*angl. Third generation*);
- HTML** – Hiperteksto žymėjimo kalba (*angl. Hyper text Markup Language*);
- IBM** – Informacinių technologijų bendrovė (*angl. International Business Machines Corporation*);
- IP** – Interneto protokolas (*angl. Internet Protocol*);
- OLA** – Persidengimų sumavimo metodas (*angl. Overlap Add Method*);
- PMM** – Paslėptasis Markovo modelis (*angl. Hidden Markov Model*);
- PSOLA** – Sinchroninis persidengimų sumavimo metodas su pagrindiniu tonu (*angl. Pitch Synchronous Overlap Add Method*);
- SOLA** – Sinchroninis persidengimų sumavimo metodas (*angl. Synchronous Overlap Add Method*);
- VoIP** – IP telefonija (*angl. Voice over Internet Protocol*);
- XML** – Duomenų struktūrų bei jų turinio aprašomoji kalba (*angl. eXtensible Markup Language*).

RAKAUSKAITĖ, Irutė. (2009) *Research of possibilities to adapt speech technologies product for command recognition and synthesis*. MBA Graduation Paper. Kaunas: Vilnius University, Kaunas Faculty of Humanities, Department of Informatics. 67 p.

SUMMARY

Currently, the speech technologies develop and improve, increase the number of application areas, but is not widely available in Lithuanian language recognition and speech synthesis programs. Many speech technology products are adapted to the English language. The aim of this work - to examine of possibilities to adapt speech recognition and speech synthesis products for Lithuanian language.

To achieve the aim was researching two of the English language programs, „ReadPlease 2003“ and „Dragon Naturally Speaking 10“ the use of Lithuanian language possibilities.

The first test of the program „ReadPlease 2003“, established the Lithuanian words perusal clarity is only 35%. The same words were transcribed, then clarity of reading became 95%. It can be argued that the Lithuanian words recorded using the phonetic spelling of the English language and presentation of written words for „ReadPlease 2003“ perusal the words of clarity was 60% higher than Lithuanian words recorded using the phonetic spelling of the Lithuanian language.

The program „Dragon Naturally Speaking 10“ was used during the second test. This time was found that the best is to identify long words, at least - short. Most mistakes are made at the two vowels *a* and *i*, when its are shortened and extensioned, and with pronunciation of the sound [ə]. As well as more errors are made in the words with voiceless consonants than words with voiced consonants. There were calculated a general efficiency of identification of the program, according to the five speakers dictating the results, which is 84,8%. Therefore, it can be argued that dictate the words for the program, can not rely on the blind, however, under the supervision and correcting incorrectly entered words, it can be used as a great writing tool.

The work methodology: analysis and synthesis of the literature, empirical test - experimental and comparative analysis.

Working volume (not including list of tables and list of pictures, list of abbreviations, bibliography, and annexes) – 51 pages, there is 17 tables and 15 images in the work.

IVADAS

Kalbos technologijos yra viena iš galingiausių technologijų, kurią dvidešimt pirmajame šimtmečiuje kuria pasiekimai moksle. Tokių spartų kalbos technologijų vystymąsi lėmė platus šių technologijų pritaikymo sferų spektras: telekomunikacijos ir multimedija, internetinis bendravimas, medicina ir mokslo ugdymas.

Viena iš labiausiai išvystytų lietuviškų kalbos sintezės technologijų yra sintezatorius „Aistis“, kurį sukūrė P.Kasparaitis, iš Vilniaus universiteto, tačiau jos kokybė nėra pakankama, nes sukurta naudojant ne pačias tobuliausias šiai dienai sintezės technologijas. Lietuviškų kalbos atpažinimo produktų taip pat nėra daug. Tai „Raktas“ bei kelių telekomunikacinių paslaugų prototipai naudojantys šnekamosios kalbos atpažinimą.

Daugelis kalbos technologijų produktų yra pritaikyti anglų kalbai. Mažom kalbom, tokiom kaip lietuvių, reikia arba kurti savo kalbai skirtus atpažinimo ir sintezės modelius, arba stengtis panaudoti kitom kalbom sukurtus modelius atitinkamai juos adaptuojant. Todėl aktualu iširti anglų kalbai pritaikytų produktų adaptavimo lietuvių kalbai galimybes.

Darbo *objektas* – kalbos atpažinimo technologijas naudojantys produktai.

Darbo *tikslas* – išnagrinėti, kalbos atpažinimo ir kalbos sintezės produktų adaptavimo lietuvių kalbai galimybes. Norint pasiekti užsibrėžto tikslo iškelti tokie darbo *uždaviniai*:

1. apibrėžti kalbos technologijas, jų funkcijas, taikymo sritis;
2. apibūdinti kalbos technologijų pasiekimus pasaulinėje rinkoje ir Lietuvoje;
3. apžvelgti kalbos atpažinimui ir kalbos sintezei naudojamus produktus;
4. iširti programų „ReadPlease 2003“ ir „Dragon NaturallySpeaking 10“ adaptavimo lietuvių kalbai galimybes;
5. apibendrinti gautus rezultatus.

Darbe naudojami *metodai*: literatūros analizė, apibendrinimas, lyginamoji analizė ir eksperimentinis bandymas.

Pirmajame darbo skyriuje yra pateikiama garso samprata. Taip pat kalbos technologijų ir jų produktų apžvalga, liečianti natūralios ir šnekamosios kalbos apdorojimo taikymo sritis, svarbiausias kalbos technologijas, jų paskirtį ir panaudojimo sferas, kalbos atpažinimo uždavinius ir sintezės metodus bei atliktus darbus, kalbos technologijų srityje, Lietuvoje ir pasaulyje.

Antrajame – atlikta detali kalbos atpažinimo ir kalbos sintezės produktų analizė.

Trečiame skyriuje aprašytas kalbos technologijų produktų „ReadPlease 2003“ ir „Dragon NaturallySpeaking 10“ adaptavimo lietuvių kalbai galimybių tyrimas ir pateikiami tyrimo metu gauti rezultatai.

Dalis darbo rezultatų buvo publikuoti straipsnyje „Kalbos technologijų produktų adaptavimo galimybės komandų atpažinimui“, kuris išspausdintas 14-tosios tarpuniversitetinės magistrantų ir doktorantų mokslinės konferencijos „Informacinės technologijos“ pranešimų medžiagoje, p. 21 – 24, ISSN 2029 – 249X, 2009.

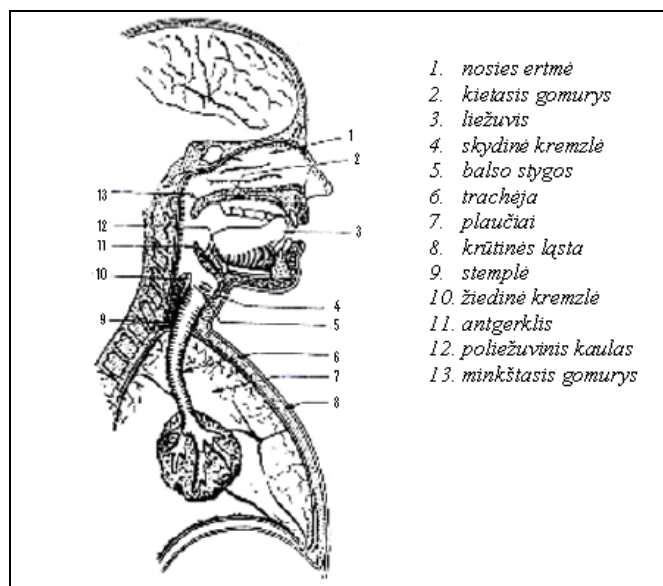
Darbą sudaro: lentelių, paveikslų ir santrumpų sąrašai, įvadas, 3 skyriai, išvados, naudotos literatūros sąrašas ir priedai. Darbo apimtis: 67 puslapiai, juose yra 17 lentelių, 15 paveikslų. Bibliografinį aprašą sudaro 42 šaltiniai.

1. KALBOS TECHNOLOGIJŲ IR JŲ PRODUKTŲ APŽVALGA

Šis skyrius yra pradedamas nuo garso sampratos teorijos. Toliau apžvelgiamos svarbiausios balso technologijos, jų paskirtis, apdorojimo taikymo sritys, panaudojimo sritys. Taip pat aprašomas pagrindinis balso atpažinimo algoritmas, kiti atpažinimo uždaviniai bei sintezės metodai. Pabaigoje pateikiami kalbos technologijų darbai stambiausiose pasaulio kompanijose ir Lietuvoje.

1.1. Garso samprata

Žmogaus kalbos generavimo mechanizmą sudaro plaučiai, trachėja, gerklos ir balso bei nosies traktai. Susikaupęs plaučiuose oras per trachėją patenka į gerklas (Lipeika, A., 2003). Balso stygos gerklose sukelia vibraciją. Gaunami impulsai, kurie toliau pereina į balso traktą. Balso traktas prasideda ties balsaskyle ir baigiasi ties lūpomis. Jį sudaro ryklė ir burnos ertmė. Vyro vidutinis balso trakto ilgis yra 17 cm. Balso trakto skersmenį įtakoja liežuvis, lūpų, minkštojo gomurio bei smakro padėtys, ir priklausomai nuo jų, balso traktas gali būti atviras arba visiškai uždarytas. Nosies traktas prasideda nuo gomurio ir baigiasi šnervėmis. Kai gomuris yra nuleidžiamas, nosies traktas yra akustiškai suporuojamas su balso traktu, kad sukurti nosinius garsus. Taigi impulsai iš balso stygų patenka į balso traktą, kur yra suformuojamas reikiamas garsas. Kalbos mechanizmas yra parodytas 1 paveiksle.



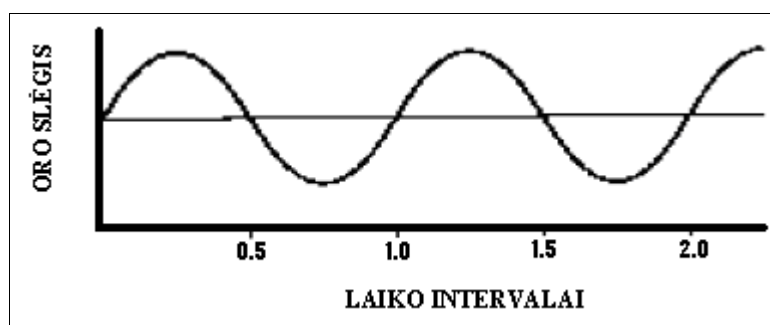
Šaltinis: Lipeika, A. (2003) *Kalbos organai, kalbos generavimas ir suvokimas*. [interaktyvus]. [žiūrėta 2008 m. gruodžio 20d.]. Prieiga per internetą: <<http://www.mif.vu.lt/~bastys/lipeika/>>.

1 pav. Kalbos mechanizmas

Kai balso stygos yra įtemptos, sudaromi vokalizuoti garsai. Kai balso stygos yra laisvos, oro slėgis arba eina per susiaurėjimą balso trakte ir sudaro nevokalizuosius garsus, arba yra suslegiamas ir staigiai išleidžiamas sukuriant trumpalaikį garsą.

Balso trakto dalys (liežuvis, lūpos, gomurys) juda beveik nepriklausomai viena nuo kitos. Įvairios šių dalių padėties balso trakte nulemia skirtingus garsus.

Garsas sklinda oru aukšto ir žemo slėgio tarpsniais (Lipeika, A., 2005). Paprasčiausias būdas pavaizduoti garsą diagrama yra nurodyti oro slėgį kiekviename laiko vienetu. Taip gaunamos periodinės garso bangos, kaip parodyta 2 paveiksle.



Šaltinis: Lipeika, A. (2005) *Formantiniai požymiai atpažįstant kalbą*. Informacijos mokslai. p. 215-219. ISSN 1392-0561.

2 pav. Periodinės garso bangos

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

Iš vieno šaltinio išleistos dvi skirtingos garso bangos susilieja į vieną. Ten, kur abiejų bangų slėgiai yra aukšti, jie papildo vienas kitą, sukurdami dar aukštesnį. Du žemi slėgiai atitinkamai pažemina bendrą slėgį, o aukštas ir žemas slėgiai neutralizuojasi. Sudėtingos garso bangos yra įvairių dažnių, amplitudžių bei ilgių paprastų bangų kombinacijos. Todėl kiekvieną garsą galima užrašyti kaip tam tikrame dažnių intervale esančių bangų kombinaciją.

Spektrinės diagramos apibūdina garsą tik labai trumpame laiko tarpe. Kalbančio žmogaus skleidžiamas garsas kinta laike, todėl kalbai pavaizduoti naudojamos dažnių ir laiko diagramos – spektrogramos. Spektrogramos abscisė yra laikas, ordinatė – dažnių intervalas. Spektrogramoje vaizduojamas kiekvieno dažnio bangų intensyvumas (amplitudė) tam tikru laiko tarpu. Šis intensyvumas vaizduojamas spalvos ryškumu. Kuo tam tikro dažnio bangos amplitudė yra didesnė, tuo tamsesnė intensyvumą nurodanti spalva.

Spektrinės diagramos viršūnės yra vadinamos formantėmis. Jos numeruojamos pradedant nuo pirmosios ir žymimos F_1 , F_2 , F_3 ir t.t. Formantės parodo intensyviausius taškus dažnių juostoje, jos pakankamai gerai atsispindi spektrogramose, ypač plačiajuostėse.

Kiekvieną kalbą sudaro baigtinė kalbos signalų arba garsų aibė. Neskaidomas elementarus kalbos signalas yra vadinamas fonema, o visa tam tikros kalbos fonemų aibė – fonemų abėcėlė. Kalbant, atskiros fonemos yra apjungiamos į didesnius vienetus. Taip gaunami žodžiai bei sakiniai.

Fonemos įrašinėjamos garso bangų pavidalu. Toks įrašas vadinamas analoginiu. Šiuolaikinė kompiuterinė įranga puikiai tinka fonemų įrašų analizei bei atpažinimui atlikti, tačiau gali dirbti tik su skaitmenine informacija. Todėl naudojamos diskretizavimo bei kvantavimo operacijos, kurių dėka, analoginis įrašas tampa skaitmeniniu.

1.2. Natūralios ir šnekamosios kalbos apdorojimo taikymo sritys

Natūrali kalba – kalba, kurios sandaros taisyklės remiasi esama vartoseną (Natural Language, 2001). Natūralios kalbos apdorojimas – lingvistikos ir informatikos sandūroje esanti disciplina, tirianti informacinių technologijų ir kompiuterinių programų pritaikymo visoms natūralios kalbos naudojimo sritims galimybes.

Natūralios kalbos apdorojimo taikymo sritys:

- teksto skaitymas (įgarsinimas);
- automatinis vertimas;
- automatinis teksto taisymas;
- informacijos paieška;
- automatinis tekstų kūrimas;
- automatizuotas teksto apibendrinimas;
- skaitomo teksto atpažinimas.

Pradinės sistemos, dirbančios „ribotame pasaulyje“ su ribotu žodynu, dirbo puikiai, suteikdamos mokslininkams optimizmo, kuris buvo nuslopintas, kai sistemos buvo praplėstos iki realesnių bei sudėtingesnių gyvenimiškų situacijų. Natūralios kalbos suvokimui priskiriamos labai sudėtingoms problemoms, kurias galėtų spręsti tik dirbtinio intelekto sistemos, dėl to, kad natūralios kalbos atpažinimas reikalauja gilių žinių apie išorinį pasaulį ir galimybės juo manipuliuoti. „Suvokimo“ apibrėžimas yra viena iš esminių natūralios kalbos apdorojimo problemų.

Problemos, darančios natūralios kalbos apdorojimą sudėtingu:

- žodžių ribų atpažinimas (kalbant, žodžiai dažnai nėra atskiriami vienas nuo kito; kur atskirti žodžius dažnai priklauso nuo to, kuris pasirinkimas tinka labiau gramatiškai ir pagal kontekstą);
- žodžių įvairiaprasmiškumas (daugelis žodžių turi daugiau kaip vieną reikšmę ir reikia pasirinkti tokią reikšmę, kuri labiausiai tinka pagal kontekstą);

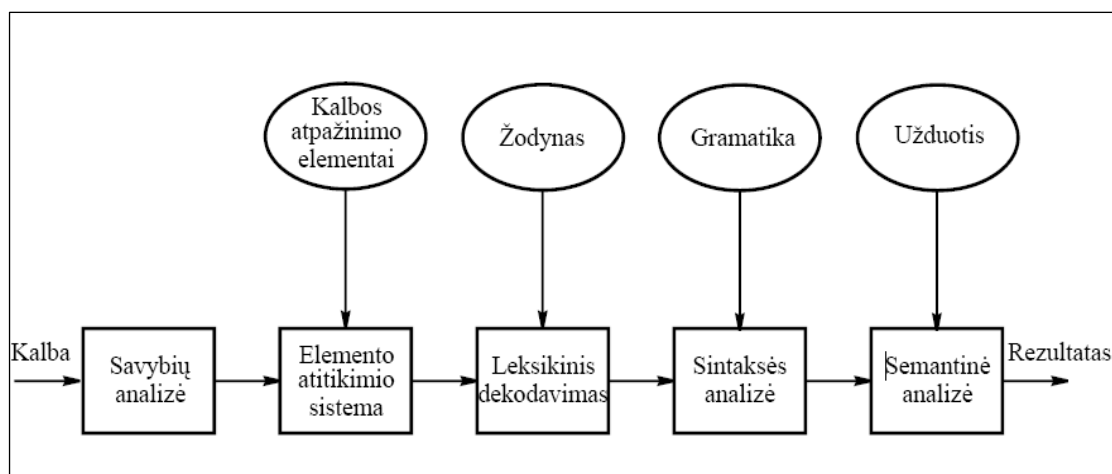
- sintaksinis įvairiaprasmiškumas (natūralios kalbos gramatika nėra vienareikšmiška. Tam tikram sakiniui gali būti sugeneruoti keli apdoravimo medžiai. Tinkamiausio pasirinkimas dažniausiai reikalauja semantinės bei kontekstinės informacijos);
- klaidingas arba nenumatytas įvedimas (užsieniečio akcentas arba regioninė tarmė, spausdinimo arba gramatinės klaidos; optinio ženklų atpažinimo klaidos);
- reiškinių konstatavimas ir sakymų planai (dažnai sakinių reikšmė yra netiesioginė).

Šnekamosios kalbos technologijos – tai žmogaus bendravimas su kompiuteriu balsu, kuris yra žymiai greitesnis bei patogesnis negu naudojantis klaviatūra ar pele. Žmogaus bendravimo su kompiuteriu balsu pagrindas yra šnekos atpažinimas ir sintezė.

Šnekamos kalbos apdoravimo taikymo sritys (Rudžionis, A. ir kt., 2001):

- kalbinis dialogas su kompiuteriu;
- balso įrašų stenografavimas;
- balsinė interneto naršyklė;
- elektroninių prietaisų valdymas balsu;
- asmens atpažinimas pagal jo balsą.

Nenutrūkstamos kalbos atpažinimo sistemoje signalui pirmiausiai atliekama savybių analizė. Sekančiai parenkama elemento atitikimo sistema, vykdomas leksinis dekodavimas, sintaksės analizė ir semantinė analizė (3 pav.).



Šaltinis: *PMM naudojimas kalbos atpažinime*. [interaktyvus]. Vilniaus Gedimino technikos universitetas (2002) [žiūrėta 2008 m. gruodžio 10d.]. Prieiga per internetą: <<http://www2.el.vgtu.lt/ssa/secB5.html>>.

3 pav. Apibendrinta nenutrūkstamos šnekamosios kalbos atpažinimo sistemos schema

Nenutrūkstamos šnekamosios kalbos atpažinimo sistemos schemos (3 pav.) aprašymas (PMM naudojimas kalbos atpažinime, 2002):

- Savybių analizė. Čia atliekama spektrinė ir/arba laikinė kalbos signalo analizė, kurios metu gaunami stebėjimo vektoriai, kurie aprašo įvairius kalbos garsus.
- Elemento atitikimo sistema. Čia, pirmiausiai, turi būti pasirinktas kalbos atpažinimo elementas. Tai gali būti įvairios lingvistinės žodžio dalys tokios, kaip fonemos, dvibalsiai, skiemenys ir pan., arba visas žodis ar net keli žodžiai. Bendru atveju, kuo paprastesnis yra pasirinktas elementas, tuo jų mažiau yra kalboje ir tuo sudėtingesnė yra jų struktūra nenutrūkstamame kalbos atpažinime. Didelio žodyno (daugiau kaip 1000 žodžių) kalbos atpažinime paprastai yra privalomas paprastų elementų (žodžio dalių) naudojimas, nes kitaip būtų sunku sudaryti tinkamą žodžių rinkinį. Tačiau tam tikrais specializuotais atvejais (kai žodyne yra mažai žodžių) abu variantai yra tinkami. Nepriklausomai nuo pasirinkto varianto, kalbos elemento požymiai randami apmokant modelį.
- Leksikinis dekodavimas. Šis procesas iškelia reikalavimus elemento atpažinimo sistemai, kad joje ištirti keliai atitiktų kalbos elementų sekai, kuri yra žodžių žodyne (leksikone). Šioje procedūroje atpažįstamų žodžių žodynas turi būti sudarytas pagal pasirinktus atpažinimo elementus. Jei elementai yra atskiri žodžiai ar žodžių junginiai, tai tada šis leksikinio dekodavimo žingsnis yra nereikalingas ir atpažinimo sistema supaprastėja.
- Sintaksės analizė. Šis procesas taip pat iškelia dar kelis reikalavimus elemento atitikimo sistemai, kad joje ištirti keliai ne tik atitiktų kalbos elementų sekai, kuri sudaro žodžius, bet ir šie žodžiai turi būti išsidėstę tokia tvarka, kuri atitinka gramatikos reikalavimus.
- Semantinė analizė. Šis procesas, taip kaip ir du prieš jį sekę procesai, toliau didina reikalavimus parenkant atpažinimo paieškos kelius. Vienas iš būdų kaip patenkinti semantinius reikalavimus yra naudoti dinaminį atpažinimo sistemos būsenos modelį. Priklausomai nuo atpažinimo sistemos būsenos tam tikri sintaksiškai teisingi žodžių junginiai yra atmetami. Tai palengvina atpažinimo uždavinį ir didina sistemos efektyvumą.

1.3. Svarbiausios balso technologijos, jų paskirtys ir panaudojimo sferos

Žmonės tarpusavio bendravime naudoja tiek gyvą, šnekamąją, tiek rašytinę kalbą. Istoriskai šnekamoji kalba išsivystė anksčiau už rašytinę. Žmonės kiekvieną dieną kalba ir klausosi daugybės

skirtingų kalbų bendraudami, perduodami savo mintis, norus, bei emocijas kitiems. Kalba yra pats svarbiausias žmonių bendravimo būdas, o šneka – pagrindinė bendravimo priemonė.

Kalbos technologijos kaip tik ir nagrinėja kalbos kaip pagrindinės bendravimo priemonės teoriją ir pagrindines mintis apie kalbos teoriją pateikia kaip kalbos generavimo, kalbos perdavimo ir kalbos suvokimo apžvalgą iš kalbinės pusės.

Skiriamos trys balso technologijų grupės (Rudžionis, A. ir kt., 2001):

- Balsu tariamų vienetų (žodžiai, jų sekos, frazės) automatinis nustatymas arba kalbos atpažinimas. Jo paskirtis yra automatiškai nustatyti, kas balsu sakoma informacijos priėmimo sistemai. Tai gali būti atskiras žodis (balso komanda), žodžių seka (PIN kodas), net rišlių sakinių skaitymas. Informacijos priėmimo sistema, nustačiusi, kas jai buvo pasakyta, atlieka adekvačius veiksmus, t.y. atpažinimas yra informacijos prieigos forma. Kalbos atpažinimas prasideda nuo garso sklidimo metu ore susidariusių bangų fiksavimo, tų bangų konvertavimo į fonemas - bazinius kalbos vienetus, žodžių formavimo iš gautų fonemų ir žodžių konteksto analizės, kad būtų galima užtikrinti teisingą panašiai skambančių žodžių tarimą (pvz. Kasti ir Kašti). Šiuo metu populiariausias kalbos signalų atpažinimo modelis – tolydinio tankio paslėptos Markovo grandinės.
- Teksto skaitymas balsu arba kalbos sintezė. Jos pagalba pagal reikiamą komandą balsu perskaitoma informacinėje sistemoje teksto pavidalu saugoma informacija. Sintzei priskiriamos ir paprastesnės informacijos pateikimo balsu formos, pavyzdžiui, iš anksto paruoštų žodžių ar jų sekų pateikimas balsu, esant tam tikram reikalavimui. Kalbos sintezė prasideda nuo mašininio pranešimo, t.y. žodžiais išreikšto teksto suformulavimo, žodžių skaidymo į fonemas, toliau seka specifinių teksto dalių (skaičiai, valiutos, žodžių intonacija, skyrybos ženklai) analizavimas, bei garsų generavimas.
- Kitos balso technologijos (asmens tapatybės vertinimas pagal jo balsą, kalbos signalų suspaudimas bei kodavimas, triukšmų slopinimas ir pan.). Asmens tapatybės vertinimas pagal jo balsą reikalingas teisėsaugoje ir komercinių operacijų vykdymui (asmens parašas). Pavyzdžiui, teisėsaugoje kalbos signalų technologijos naudingos fonoskopinių ekspertizių tobulinimui, kai pažeidėją reikia atpažinti iš balso įrašo, operatyvinėje veikloje ir panašiai. Plintant techninėmis naujovėmis plinta ir nauji garsų falsifikavimo būdai, todėl svarbu paruošti tinkamas priemones. Triukšmų šalinimas nuo kalbos signalų yra priemonė sukauptoms kultūros vertybėms restauruoti. Pavyzdžiui, per eilę dešimtmečių garso pavidalu yra sukaupta labai vertingos medžiagos, kurią reikia sutvarkyti šiuolaikinėmis priemonėmis. Triukšmų apdorojimo problemos iš esmės persmelkia beveik visas kitas kalbos technologijas. Balso signalų suspaudimas yra

taupaus balso įrašų saugojimo ar perdavimo priemonė, kurios poreikis jau jaučiamas ir Lietuvoje, operuojant balsu Internetu. Pavyzdžiui, teisėsaugoje kalbos signalų su asmenų tapatybiniais vertinimais neviešam saugojimui ar perdavimui. Kalbos kodavimo būdai dažniausiai yra skirstomi į tris plačias klases – kalbos signalo kodavimas, pirminis (šaltinio) kodavimas ir mišrus kodavimas. Kai bitų skaičius yra didelis dažniausiai naudojamas kalbos signalo kodavimas ir gaunamas geros kokybės garsas. Šaltinio kodavimas operatyvus turint mažą bitų skaičių, kai norima sugeneruoti sintetinį garsą. Mišrus kodavimas naudoja abiejų prieš tai minėtų kodavimo būdų technikas ir duoda geros kokybės garsą, panaudodamas vidutinį bitų skaičių.

Šiuolaikinės balso technologijos yra stipriai pažengusios į priekį, lyginant su buvusiomis prieš kelerius metus ar dešimtmečius. Tokių spartų balso technologijų vystymąsi lėmė platus šių technologijų pritaikymo sferų spektras. Šiuo metu balso technologijos labiausia taikomos:

- internetinio bendravimo sferoje;
- medicinos srityje;
- telekomunikacijų ir multimedijos sferose;
- mokslo ugdymo srityje.

Internetinio bendravimo sfera yra bene sparčiausiai besiplečianti balso technologijų pritaikymo sritis (How VoIP Works, 2008). Interneto vartotojų skaičius rekordiškai didėja, o tai lemia didelę paslaugų orientaciją būtent šiai žmonių grupei. Interneto bendravimo sritis yra betarpiškai susijusi su telekomunikacijų sritimi, visgi šią sritį vertėtų išskirti dėl skirtingos bendravimo platformos. Telekomunikacijų srityje bendravimas vyksta telefoninėmis linijomis, o interneto bendravimo srityje pagrindinis resursas yra internetinis ryšys. Be to telekomunikacijų srityje balso technologijos panaudojamos atkurti balsą, t.y. naudojama kalbos sintezės technologija, o internetinio bendravimo srityje balso technologijų paskirtis yra balso kodavimas ir dekodavimas.

Dar viena priežastis lemianti, ženkliai šios srities pažangą, yra šiuolaikinis interneto pralaidumas, kuris leidžia perduoti žmogaus balsą iš vienos vietos į kitą taip pat kokybiškai, kaip ir telefono linijomis, ir net dar kokybiškiau. Balso perdavimas internetu (VoIP) skiriasi nuo tradicinio telefoninio ryšio tuo, jog tradiciniu telefonu perduodamo garso dažnis svyruoja nuo 0,3 iki 3,4 kHz, o taikant VoIP technologiją kalbą galima perduoti normaliu 8 kHz dažniu. Taip yra išsaugoma aukšta perduodamos kalbos kokybė. Bendravimo įprastu telefonu būdas angliškai yra vadinamas „Circuit Switching“ (grandininis apsikeitimas). Tai reiškia, jog signalas yra nuolatos perduodamas pirmyn ir atgal, ir tiesiogiai priklauso nuo pokalbio trukmės, o ne nuo perduotos informacijos (kalbėjimo) kiekio. IP telefonijoje (VoIP) kalbos perdavimas angliškai yra vadinamas „Packet Switching“ (paketų apsikeitimas). Tai reiškia, jog signalas nepriklauso nuo pokalbio trukmės, o tiesiogiai priklauso nuo to, kiek paketų yra perduodama. Patį paketo dydį lemia tai, koku algoritmu

yra koduojamas perduodamas signalas. Algoritmas pasirenkamas pagal tai, kokia yra norima kokybė. Kaip minėta, galima perduoti ir 3,4 kHz kokybės signalą, tačiau esant šiuolaikinėms interneto galimybėms, be vargo galima perdavinėti ir 8 kHz signalą. Beje, šitokios kokybės signalo principas yra taikomas naujos kartos mobiliajame ryšyje „3G“.

Verslo atžvilgiu, tai taipogi labai patogi sritis vystyti savo verslą. Šiuo metu labai išpopuliarėjo taikomosios programos, leidžiančios bendrauti ne tik interneto vartotojui su kitu interneto vartotoju, tačiau jos turi galimybę kurti ištiesas keleto vartotojų konferencijas ir netgi tiesiogiai susisiekti su įprastų telefonų ir mobiliojo ryšio operatorių vartotojais. Už tokį sujungimą verslininkai gauna pelną, kadangi, dažniausiai, vartotojui yra pigiau skambinti iš savo asmeninio kompiuterio, nei skambinti iš įprasto ar mobiliojo ryšio telefono. Gamybos įmonės taip pat gamina įvairius telefoninius prietaisus, jungiamus tiesiogiai prie asmeninio kompiuterio, ar tarnybinės stoties. Interneto paslaugų tiekėjai siūlo IP telefonijos paslaugas, kurias noriai savo padaliniuose diegia didelės kompanijos.

Medicinos srityje balso technologijų pritaikymas labiausia sietinas su neįgaliesiems skirtais prietaisais ir paslaugomis (Rudžionis, A. ir kt., 2001). Šios priemonės padeda neįgaliesiems individams lengviau adaptuotis prie aplinkos, pritaikytos sveikiems žmonėms.

Balso technologijų panaudojimo spektras yra gana platus: ne tik akliems ir silpnaregiams, bet ir žmonėms su ribotomis vaikščiojimo galimybėmis. Bet kitaip, nei balso technologijų panaudojimo kitose srityse atveju, čia jos itin svarbios, nes yra vienintelis neįgalųjų integravimo į visuomenę būdas. Technologijos skirtos neįgaliesiems skirstomos į dvi dideles grupes:

- taikymai, orientuoti į informacijos pateikimą ir valdymą balsu (kompiuterio valdymas balsu, informacijos iš interneto ar kitų šaltinių skaitymas balsu ir t.t.).
- specializuoti įrenginiai skirti neįgaliesiems, į kuriuos integruoti balso technologijų komponentai (pvz. balsu valdomi invalido vežimėliai).

Į neįgaluosius orientuoti balso technologijų taikymai remiasi, jau adaptuotais balso apdorojimo technologijų pasiekimais: kalbos atpažinimu, kalbos sinteze ir kartais asmens atpažinimu. Tik kitaip nei sveiki žmonės, neįgalieji neturi alternatyvių, informacijos gavimo bei įrengimų valdymo, galimybių. Dėl didžiulės balso technologijų svarbos, neįgaliesiems yra priimtinas žemesnis balso technologijų lygis nei eiliniam vartotojui (didesnis atpažinime atsirandančių klaidų skaičius, prastesnis sintetinės kalbos lygis). Be to, paprastai, šiais neįgaliesiems pritaikytais įrengimais rūpinasi ne jų šeimininkai, o įvairios valstybinės ar nevalstybinės globos organizacijos.

Jau prieš keletą dešimtmečių susiformavo dvi natūralios atpažinimo plėtros pozicijos taikymo prasme, kurias sąlyginai pavadinsime telekomunikacine (ATT) ir multimedijine (IBM, vėliau Microsoft). Pirmuoju atveju dominavo daugiadiktorinis nedidelio žodyno atpažinimas

telefoniniuose kanaluose, antruoju – viendiktorinis labai didelio žodyno atpažinimas ramaus biuro aplinkoje.

1 lentelė

Kalbos signalų atpažinimo taikymo ypatybės telefonijos ir multimedijos terpėse

Faktoriai	Telekomunikacijos	Multimedija
Žodyno apimtis	Daugumoje keliasdešimt žodžių, kai kada iki 2000.	Nuo 5000 iki 60000 ir daugiau
Žodyno plėtimo galimybės	Reikia suformuoti duomenis kiekvieno naujo žodžio apmokymui.	Žodynas lengvai papildomas, naudojant naujų žodžių fonetines transkripcijas
Informacinis grįžtamas ryšys. Klaidų pastebėjimo ir jų koregavimo galimybės	Atsakymą gauname balsu. Galime jo nesulaukti. Dialogas turi būti labai kruopščiai suprojektuotas.	Rezultatą matome kompiuterio ekrane. Galima greita korekcija.
Diktoriaus adaptavimas	Veik nėra galimybių naujam diktoriui prisiderinti, jei jo balsas blogai atpažįstamas.	Interaktyvus apmokymas gali ženkliai padėti.
Vartotojai	Bet kuris diktorius.	Gali būti net ir vienas suinteresuotas diktorius.
Signalų kokybė	Labai didelės variacijos: diktorius, mikrofonas, triukšmai, signalo perdavimo kanalas.	Pakankamai gera.
Kontekstas	Naudojami skaičių pavadinimai ir valdymo komandos. Konteksto panaudojimas gana problematiškas.	Formalios ir statistinės gramatikos yra efektyvus įrankis.
Žodžių atpažinimo tikslumas	Reikia siekti kuo tikslesnio atpažinimo, kad nesutrikdytume dialogo.	Galima taikytis su ribotu klaidų skaičiumi jas ištaisant.

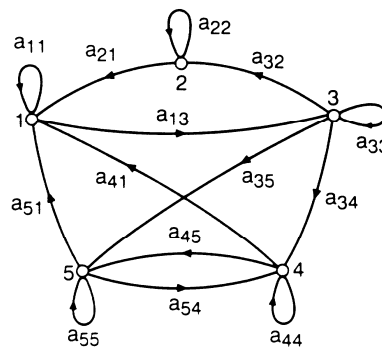
Šaltinis: Rudžionis, A. (2001) Balso technologijų taikymo lietuvių kalbai analizė ir perspektyvinių veiklos krypčių pagrindimas [interaktyvus]. [žiūrėta 2009 m. kovo 1 d.]. Prieiga per internetą: <http://www.likit.lt/frames/balso_tech/balsotech_st.htm>.

Mokslo ugdymo srityje, nuo 1990 m. asmeniniams kompiuteriams tampant vis galingesniais, atsirando galimybė sukurti tokius kompiuterius, su kuriais žmonės galėtų bendrauti žodžiu. Tokiu būdu bendravimas ir darbas su kompiuteriu taptų natūraliu ir realiu procesu. Tačiau šiai dienai kalbos technologijos dar nėra pakankamai išsivysčiusios, kad žmogaus ir kompiuterio bendravimas vyktų natūraliai. Kartu turi būti kuriami treniruokliai gimtosios bei užsienio kalbų mokymuisi, pritaikyti logopedijai ir kiti.

1.4. Sudėtinės atpažinimo funkcijos

Pagrindinis balso atpažinimo algoritmas yra paslėptasis Markovo modelis (Rabiner, L. et al., 1993). Paslėptojo Markovo modelio (toliau PMM) teoriniai pagrindai buvo suformuluoti L. E. Baum ir jo kolegų 7 – ajame dešimtmetyje. 8 – ajame dešimtmetyje J. K. Baker ir kolegos PMM pritaikė kalbos apdorojimui. Tai buvo kalbos atpažinimo sistema „The Dragon“. Šiuo metu PMM yra labiausiai paplitęs metodas kalbos atpažinime, ypač ištisinės kalbos. Be to PMM dažnai naudojami hibridinėse atpažinimo sistemose, t.y. sistemose, atpažinimui naudojančiose kombinuotus metodus.

Diskretinio laiko Markovo procesai:



Šaltinis: Rabiner, L.; Juang, B.-H. (1993) *Fundamentals of speech recognition*. Prentice Hall, New Jersey.

4 pav. Markovo grandinė su penkiomis būsenomis

Būsenų aibė – $\{1, 2, \dots, N\}$;

Perėjimų laiko momentai – $t = 1, 2, \dots$;

Dabartinė būsena – q_t ;

Pirmos eilės diskretinio laiko Markovo grandinės tikimybinis aprašymas:

$$P[q_t = j \mid q_{t-1} = i, q_{t-2} = k, \dots] = P[q_t = j \mid q_{t-1} = i], \quad [1]$$

t.y. nagrinėjamos būsenos priklauso tik nuo praėjusios būsenos.

Nagrinėjant nuo laiko nepriklausančius Markovo procesus, perėjimų tikimybių aibę galima užrašyti:

$$a_{ij} = P[q_t = j \mid q_{t-1} = i], \quad 1 \leq i, j \leq N, \quad [2]$$

kur

$$a_{ij} \geq 0 \quad \forall j, i;$$

$$\sum_{j=1}^N a_{ij} = 1 \quad \forall i.$$

Tai stebimasis Markovo modelis. Šio proceso išėjimas – būsenų aibė kiekvienu laiko momentu, o kiekviena būsena atitinka stebimą įvykį. Paslėptojo Markovo modelio išėjimas kiekvienoje būsenoje nėra atsitiktinis. Tai apriboja jo panaudojimą realiems uždaviniams.

Paslėptojo Markovo modelio (PMM) esmė – stebėjimas (esant tam tikroje būsenoje) yra tikimybinė būsenos funkcija. Taigi šis modelis yra dvigubas atsitiktinis procesas su slypinčiu jame paslėptuoju procesu, kuris yra stebimas netiesiogiai.

PMM apibūdinti naudojami du modelio parametrai M ir N , bei trys tikimybiniai matai:

1. N – būsenų skaičius modelyje. Atskiros būsenos žymimos $\{1, 2, \dots, N\}$, o būsena laiko momentu $t - q_t$.

2. M – skirtingų stebėjimo simbolių būsenoje skaičius (diskretinio alfabeto dydis). Atskiri simboliai žymimi $V = \{v_1, v_2, \dots, v_M\}$.

3. $A = \{a_{ij}\}$ – būsenų perėjimo tikimybinis pasiskirstymas, kur

$$a_{ij} = P[q_{t+1} = j | q_t = i], \quad 1 \leq i, j \leq N.$$

4. $B = \{b_j(k)\}$ – stebėjimų simbolio tikimybinis pasiskirstymas, kuriame

$$b_j(k) = P[o_t = v_k | q_t = j], \quad 1 \leq k \leq M.$$

5. $\pi = \{\pi_i\}$ – pradinės būsenos pasiskirstymas, kuriame

$$\pi_i = P[q_1 = i], \quad 1 \leq i \leq N.$$

Kompaktiškumo dėlei naudojamas toks modelio parametrų aibės žymuo $\lambda = (A, B, \pi)$.

PMM – stebėjimų sekos generatorius. Gaunama stebėjimų seka $O = (o_1, o_2, \dots, o_T)$, čia o_T – simbolis iš alfabeto V , T – stebėjimų skaičius.

Sprendžiant PMM uždavinius susiduriama su trimis problemomis (Ceccarelli, M., 2000):

1. Kaip duotam modeliui $\lambda = (A, B, \lambda)$ ir stebėjimų sekai $O = (o_1, o_2, \dots, o_T)$ efektyviai apskaičiuoti stebėjimų sekos tikimybę $P(O/\lambda)$?

2. Kaip duotam modeliui $\lambda = (A, B, \lambda)$ ir stebėjimų sekai $O = (o_1, o_2, \dots, o_T)$ parinkti atitinkamą būsenų seką $q = (q_1, q_2, \dots, q_T)$, kuri yra kažkuria prasme optimali?

3. Kaip surasti modelio parametrus $\lambda = (A, B, \lambda)$, kurie maksimizuoja $P(O/\lambda)$?

Paprasčiausias būdas išspręsti išvardintas problemas – sunumeruoti visas galimas T ilgio sekas ir tiesiogiai paskaičiuoti tikimybę $P(O/\lambda)$.

Be paslėptojo Markovo modelio yra ir kitų balso atpažinimo uždavinių (Rudžionis, A. ir kt., 2001):

- Pašalinių garsų atmetimas. Klausydamas jį dominančios kalbinės informacijos, žmogus sugeba ignoruoti pašalinius pokalbius, triukšmus, muzikinius garsus ar panašiai, žinoma, jei pastarieji nėra pernelyg intensyvūs. Tai reiškia, kad reikia turėti galimybę atmesti įvairius akustinius garsus, kurių nėra kompiuterinio dialogo žodyne. Tinkamai parenkant atpažįstamų signalų panašumo slenkstį, tikrinama ar nagrinėjama komanda yra pakankamai panaši į kurią nors vieną iš leistinų komandų. Jei slenkstis pakankamai aukštas, tai atmetama ir dalis leistinų komandų, o jei šis slenkstis per žemas, atpažinimo

įtaisas beprasmiškai reaguoja į pašalinius garsus. Kuo tikslesnis yra atpažinimo algoritmas, tuo efektyviau veikia ši procedūra.

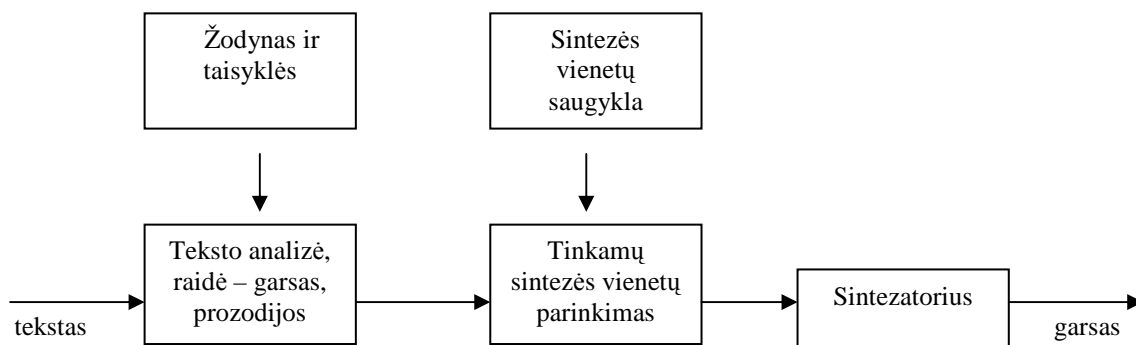
- Raktinių žodžių paieška. Tipinę žmogaus tariamą komandą galima pavaizduoti esminiais žodžiais: subjektas (kas vykdo), predikatas (kas vykdoma), objektas (kas yra vykdymo objektas), atributai (kaip vykdoma). Realioje frazėje esti funkcinė prasme neesminių žodžių (pvz. malonybės), nekalbinių intarpų (mikčiojimas, kostelėjimas), o funkciškai esminių žodžių tvarka gali keistis, nors prasmė išlieka ta pati. Raktinių žodžių paieška yra funkciškai esminių žodžių suradimas.
- Pasikliovimas atpažinimu. Vis labiau plinta procedūros, kai galutinis sprendimas priimamas pagal atpažįstamo vieneto panašumo mato reikšmę nebūtinai iš karto. Jei panašumas nėra pakankamai aukštas, automatas siūlo pakartoti užklausą, modeliuojant žmogaus elgesį, kai jis gali ne iš karto suprasti klausimo prasmę.
- Prisitaikymas prie akustinės aplinkos. Seniai pastebėta, kad laboratorijose paruoštos atpažinimo priemonės blogai veikia realaus taikymo sąlygomis. Pvz., laboratorijose gaunama mažesnė nei 1% skaičių pavadinimų sekų atpažinimo klaida gali išaugti iki 75% telefono kanaluose. „Texas Instrument“ surado efektyvų adaptavimosi būdą, kai efektingai išnaudojami laboratoriniai duomenis ir atpažinimo klaida nesumažėja. Didelės apimties laboratoriniai duomenys atspindi žmonių balsų įvairovę bei lingvistines ypatybes, o santykinai nedidelė adaptavimo medžiaga leidžia prisitaikyti prie triukšmų, kanalo, reverberacijų ir kitų faktorių.

1.5. Kalbos sintezės metodai

Kalbos sinteze vadinamas automatinis balsinio pranešimo generavimas iš pateikto teksto ar kitos simbolių sekos, t.y. tekstu pateiktos informacijos skaitymas balsu (Rudžionis, A. ir kt., 2001). Nekelia abejonių, kad sintezė labai nuo konkrečios kalbos savybių priklausanti kalbos technologijų sritis. Generavimui reikia naudoti konkrečiai kalbai paruoštus sintezės elementus (pastarieji dažnai vadinami sintezės vienetais) bei atsižvelgti į duotos kalbos gramatinės ypatybes (kirčiavimą, intonacines savybes ir pan.).

Norint tekstą paversti balsu, reikia nuosekliai atlikti eilę procedūrų. Visų pirma tekstą reikia tinkamai paruošti, atlikti jo transkripciją (pvz. lietuvišką žodį gąsdina mes tariame gazdina), nustatyti frazėje esančių žodžių kirčius, apibūdinti frazės tipą (paprasčiausia, klausiamoji ar šaukiamoji intonacija). Šis etapas paprastai vadinamas teksto normalizavimu. Galiausiai tekstas paverčiamas fonemų seka su prozodijomis. Šią seką reikia paversti tinkama sintezės vienetų seka, o pastarąją paversti balsu. Tipinis sintezatoriaus pavyzdys parodytas 5 pav.

Nors visos sintezatoriaus komponentės yra svarbios, tačiau kol kas labiausiai sintetinės kalbos kokybę nulemia sintezės vienetų kokybė. Jais gali būti skiemuo, skiemens dalis, fonema, jos dalis ar jų junginys - difonas. Sinteziės vienetai gali būti tiesiog iškirptos iš žmogaus balso signalo atkarpos arba tos pačios atkarpos gali būti paverstos požymiais (kepstras, filtrai, Furje transformacija, tiesinės prognozės parametrai). Pirmuoju atveju turime taip vadinamą signalo bangos konkatenaciją, o antruoju – parametrinę sintezę.



Šaltinis: Rudžionis, A. (2001) Balso technologijų taikymo lietuvių kalbai analizė ir perspektyvinių veiklos krypčių pagrindimas [interaktyvus]. [žiūrėta 2009 m. kovo 1 d.]. Prieiga per internetą: <http://www.likit.lt/frames/balSO_tech/balsotech_st.htm>.

5 pav. Sinteziės procesų diagrama

Tarp kalbos sintezės metodų istoriškai ilgai dominavo parametriniai sintezės metodai, nes sintezatorius buvo atskiras įrenginys, kurį valdydavo kompiuteris. Be to tuometinių kompiuterių techniniai parametrai, sintezės požiūriu, pirmiausia atmintis, buvo labai riboti. Maždaug prieš penkiolika metų buvo pasiūlytas, po to patobulintas, TD-PSOLA sintezės būdas.

PSOLA schemai būdingos trys fazės: signalo analizė, jo modifikavimas ir sintezė. Ypač reikšminga yra antroji, modifikavimo, fazė, kurios metu atliekami spektro pakeitimai, kad vienetų sandūrose prisiderinti prie reikiamų prozodinių parametru (pagrindinis tonas, energija, trukmė). Analizės metu nustatomas signalo pagrindinio tono žymės. Analizės langai parenkami taip, kad jie pakankamai persidengtų gretimoms pagrindinio tono žymėms (OLA). Kartu diskretinės Furje transformacijos pagalba išskiriama globali spektro gaubtinė ir spektro dalis, atitinkanti žadinimo funkciją. Sinteziės metu keičiamas analizės žingsnis (SOLA), jį sinchronizuojant su pagrindiniu tonu (PSOLA).

Pastaraisiais metais įsigali požiūris, kad PSOLA metodas išsisemia. Tai argumentuojama tuo, kad naudojamas santykinai nedidelis sintezės vienetų skaičius (1000 ar kiek daugiau), kurie neaprašo daugelio galimų fonetinių kontekstų. Nors sintezės vienetams saugoti pakanka gana nedidelės atminties (Mb eilės), tačiau sintetinė kalba nėra pakankamai natūrali. Todėl siekiama pakeisti klasikinę PSOLA schemą.

Vienas tokių bandymų yra sintezės vienetų pailginimas, t.y. pereinama nuo difono prie konstrukcijos balsis – priebalsis – balsis V-CV, tuo minimizuojant reikiamas pagrindinio tono modifikacijas bei sumažinant iškraipymus sandūrose. Taip pastebimai pagerinamas priebalsių aiškumas lyginant su difonine CV sinteze.

Kitais atvejais stengiamasi iš ribotos žodžių atsargos formuoti naujus. Pavyzdžiui yra sukaupta 2000 žodžių garsynas orų prognozės tikslams. Iš daugiaskiemenių žodžių imami skiemens ir sudaromi nauji žodžiai, juos pritaikant kitokiems kontekstams, nei jie buvo įrašyti.

Perspektyviu laikomas kombinuotos sintezės metodas, kai naudojami visų lygių sintezės vienetų kombinacija: difonai, skiemens ir žodžiai. Jis pritaikytas Vokietijoje kelionių planavimo dialogams automatizuoti. Sintezės modulis pirmoje eilėje ieško saugykloje pilnų žodžių, jei nerandamas pilnas žodis, tai antrame etape bandoma jį sudaryti iš skiemens. Jei ir tai nepavyko, atliekama difoninė sintezė. Kadangi natūraliau skamba ilgesni vienetai, tai šiuo atveju kalbos pranešimo kokybės prasme optimaliai išnaudojami sintezės vienetai

Geriausios šiandieninės sintezės sistemos naudoja konkatenacinę skirtingo tipo vienetų sintezę. Tokių sistemų pavyzdžiais galėtų būti „harmonika plus triukšmas“ modelis (HMN), japonų kalbos sintezės sistema, paremta kalbos modifikavimo algoritmu su harmonikų rekonstrukcija bei kinų kalbos konkatenacinės sintezės schema su automatiniu sintezės vienetų ir prozodinės informacijos generavimu. Pastaroji sistema orientuota į kinų kalbos mandarinų variantą (Chinese Mandarin). Tai toninė kalba, turinti 5 bazinius tonus. Kita ypatybė tai, kad šioje kalboje tėra apie 1300 fonologiškai leistinų skiemens, todėl natūralu skiemens pasirinkti baziniu sintezės vienetu. Siekiant kuo aukštesnio suvokiamumo ir natūralumo, autoriai pasiūlė automatinius vienetų parinkimo, jų trukmės nustatymo, prozodijos pritaikymo algoritmus.

1.6. Kalbos technologijų darbai stambiausiose pasaulio kompanijose ir Lietuvoje

Vienos iš stambiausių pasaulio kompanijų, kurios gamina balso atpažinimo technologijas yra „IBM“ (JAV), „Nuance“ (Jungtinė karalystė) ir „Microsoft“ (JAV).

„Nuance“ yra pasiekusi laimėjimų dokumentų atvaizdavimo, eForm dizaino ir asmeninių dokumentų valdymo srityje (Company overview, 2009). „Nuance“ priimti sprendimai tiksliai verčia kalbą tekstui iki 160 žodžių per minutę greičiu, leidžia balsu valdyti kompiuterio sistemą anglų, vokiečių, prancūzų ir italų kalbomis. „Nuance“ produktai (Dragon NaturallySpeaking 10 Product Suite, 2009):

- „Dragon NaturallySpeaking Profesional“ – galingiausias kalbos atpažinimo produktas profesionaliems darbalaukiu vartotojams. Ypač tikslus yra diktuojamos informacijos „Microsoft Office“ atkūrimas, „Windows“ programų valdymas, gautų elektroninių

laiškų perklausa ir dokumentų skaitymas garsu, paieška balsu internetu tariant internetinius adresus. Šis produktas idealus biurų darbuotojams, namų vartotojams ir profesionalams. Vartojant sutrumpinimus – žodines komandas, galima į tekstą įterpti tekstinius blokus arba paveikslėlius, tokius kaip pavardė, antraštė, parašas.

- „Dragon NaturallySpeaking Standart“ paleidžia programą, sukuria dokumentus ir leidžia valdyti darbalaukį balsu, kurti elektroninius laiškus, skubius pranešimus, naršyti po internetą. Verčia kalbą 160 žodžių per minutę greičiu, tikslumas 99%.
- „Network Speech Solutions“. Organizacijos pasitiki šiuo paketu, nes jis padeda tvarkyti vartotojų pokalbius, padidinti vartotojų patirtį, sumažinti darbo kainą, išplėsti pajamų galimybes. Žmonės gali vartoti kalbą bendravimui su skambučių centru. Skambinantys asmenys gali gauti informaciją ir tvarkyti sandorius naudodamiesi tik mikrofonu.
- „X/Mode Multimodal System“ sujungia automatinio kalbos atpažinimo ir teksto konvertavimo į kalbą technologijas mobiliajame internete ir multimedijos technologijose. „X/Mode Multimodal System“ leidžia teisingų multimodalinių programų, sudarytų iš daugelio modulių, greitą kūrimą ir išdėstymą, sujungiant balso, vaizdo ir garso aplinkas į vieną belaidį įrenginį naudojant jas visas vienu metu.
- „AudioMinig Development System“ leidžia XML kalbos integravimą užkoduojant komercinių, namų ir interneto programų galimybes. Sistema kuria XML kalbos kodą ir laiko sužymėtus duomenis kiekvienam žodžiui garso ar vaizdo faile. Programos įtraukia pridėtą kalbos kodą ir ieško atitikčių tinklo garso puslapiuose, įrašytuose skambučiuose, centro informacijoje, kompiuteriniame apmokyje, valdymo turinyje, garso ir vaizdo transliacijos redagavimo programose.

Neseniai „Microsoft“ paleido pirmą balsu valdomą programinę įrangą specialiai sukurtą „Windows Mobile™ 2003“ (Microsoft Gives Mobile Devices a Voice in Europe, 2009) – „Microsoft Voice Command“, kuri transformuoja „Pocket PC“ ir „Pocket PC Phone Edition“ įrangą į virtualius dialoginius asmeninius asistentus, kurie palengvina vartotojų darbą, leidžia valyti telefoną, kontroliuoti kontaktus, susitarimus ir „Windows Media Player“. Per kalbos komandas, kurios neturi būti iš anksto įrašytos, „Microsoft Voice Command“ siūlo vartotojui tokias pagrindines galimybes:

- Balsu valdomas „laisvų rankų telefonas“. Bet kur, mašinoje, oro uoste ar gatvėje, galima džiaugtis „laisvų rankų“ telefono pokalbiais be numerio rinkimo. Užtenka pasakyti vardą iš kontaktų sąrašo arba pasakyti, rinkti tam tikrą numerį.
- Skambinančiojo asmens informacijos pranešimas. „Pocket PC Phone Edition“ vartotojai gali gauti įeinančio skambučio informacijos pranešimą. Jei praleistas skambutis, galima

tiesiog paklausti „What call did I miss?“ ir „Voice Command“ išvardins skambučių laiką ir telefono numerius.

- Balsu valdoma kalendoriaus peržiūra. Nepriklausomai nuo vietos, vartotojas gali paklausti savo įrenginio „What’s my next appointment?“ ir susitikimo vieta, laikas, tema kiekvieno įrašo bus išstarta garsiai.
- Balsu valdomas „Windows Media Player“. Dabar balso komandos gali būti naudojamos kaip virtualus nuotolinis valdymas pasirenkant artistą, albumą ir paleidžiant groti muziką.
- Kitos balsu valdomos priemonės. Galima paliesti bet kurią programą iš „Start“ meniu ar programų katalogo, pvz., „Microsoft Internet Explorer“, „Outlook“ kalendorius, kontaktai ar skaičiuotuvas. Sužinoti nešiojamojo kompiuterio baterijos lygį, laiką, datą.

Kalbos atpažinimo priemonė yra visiškai integruota į „Windows Vista“ sistemą ir įdiegta naujausių „Microsoft“ kalbos technologijų pagrindu (Windows Vista Speech Recognition, 2008). Kalbos atpažinimas yra ypač tikslus ir naudojamas tobulėja, nes prisitaiko prie kalbančiojo kalbėjimo stiliaus ir leksikos.

Pirmą kartą „Microsoft“ kompanija balso atpažinimą integravo į „Office XP“ paketą (Microsoft Office versijų palyginimas, 2009). Panaudojant balsą teksto ir komandų įvedimui galima dirbti našiau – nereikia daryti nereikalingų judesių pelyte ar spausti specialių klavišu kombinacijų.

„IBM“ taip pat daug investavusi į balso atpažinimą kompanija. Vienas iš naujausių pasiekimų yra „IBM DirectTalk/2“ – techninė bazė, programinė įranga ir balsinės aplikacijos (IBM DirectTalk for Windows, 2001). „IBM DirectTalk/2“ balsinė aplikacija leidžia atlikti eilę operacijų vienu telefono skambučiu. Tai viena sistema, kuri atlieka daug įvairių funkcijų, tokių kaip atsakymas balsu, balso paštas, fakso aparatas – skambinantysis gali duoti užklausimus sistemai balsu, teksto konvertavimas į kalbą – informacija esanti kompiuteryje gali būti pasakyta žodžiais. „IBM DirectTalk/2“ dėka, žmonės gali gauti norimą informaciją ir paslaugą tiesiog paskambinę telefonu – ir tam nereikia žmogaus. Juo labiau – tai yra įmanoma 24 valandas per parą ir be jokių laukimų, pokalbio peradresavimo.

Kita taip pat populiori balso atpažinimo sistema „Embedded Via Voice“. Ši sistema padeda vairuotojui saugiau vairuoti ir visa dėmesį skirti keliui (Embedded ViaVoice, 2009).

Bendrovė „Pioneer Europe“ panaudojo „IBM“ sukurtą „Embedded ViaVoice“ garso atpažinimo programinę įrangą, kad jos gaminamą automobiliams skirtą garso ir navigacijos įtaisą „AVIC-HD1BT“ būtų galima paprastai valdyti balsu (IBM sukurta nauja balso atpažinimo sistema, 2006).

„IBM“ turi daugiau kaip 300 patentų balso technologijų srityje, kalbos laboratorijose įvairiose pasaulio šalyse daugiau kaip 300 kūrėjų ir 100 tyrinėtojų dirba su daugiau kaip 15 kalbų. Jau daugiau kaip 30 metų pirmaudama balso tyrinėjimų srityje, „IBM“ dalyvauja pritaikant balso technologijas skirtingose rinkose, taip keisdama visuomenės gyvenimą, darbo, vairavimo ir bendravimo įpročius.

Darbai balso technologijų srityje vykdomi Lietuvos universitetuose: Kauno technologijos universitete, Vilniaus universitete, Vytauto Didžiojo universitete, Šiaulių universitete. Taip pat ir Matematikos ir informatikos institute.

Kauno technologijos universiteto kalbos signalų tyrimo grupė vykdė darbus šiose srityse:

- Atvirkštinė filtracija (Noreika, S. ir kt., 1980). Sukonstruota nestacionarių signalų analizės sistema, kuri buvo pritaikyta atvirkštinės filtracijos uždavinių, t.y. balso stygų virpesių atstatymui, kai yra žinomas mikrofonu užfiksuotas įrašas.
- Fonemų diskriminacija ir konteksto reikšmė. Ši sritis visada buvo dėmesio centre. Jau nuo 1985 buvo palyginti klasifikatoriai (Euklido, Machalanobio matai, įvestas dichotominis klasifikatorius drauge su požymių erdvės optimizavimu) ir pastebėta fonemų diskriminavimo konkrečiame kontekste svarba (Domantas, A. ir kt., 1991; Domantas, A. ir kt., 1986; Rudžionis, A., 1986). Šie darbai buvo plečiami ir gauta daugiadiktorinio apie 2% dydžio sunkiai skiriamų fonemų prieš tris kontrastingiausias balsius A, U, I klaida. Pereinant nuo atviro balsio A iki uždaro I kontekstų (Rudžionis, A. ir kt., 1999; Rudžionis, V., 1995; Rudžionis, V., 1994).
- Diskretinis PMM modelis (Kalbos ir diktoriaus atpažinimas telefonijoje, 1995). Modeliuoti diskretiniai PMM, siekiant suformuoti žymiai efektyvesnius fonetinio klasifikavimo būdus.
- Pašalinių garsų atmetimas (Rudžionis, A., 1989). Sukurta pašalinių garsų atmetimo procedūra, kurios tikslas tinkamai parinkti atpažįstamų signalų panašumo slenkstį.
- Kalbos signalų aptikimas (Rudžionis, A., 2000). Patobulintas klasifikatorius „balsas – ne balsas“, išnaudojant šnekamosios kalbos skiemeninę struktūrą.

Vilniaus universitete daugiausia atliekami darbai susiję su kalbos sinteze ir jos taikymu akliesiems ir silpnaregiams. Sukurtas sintezatorius „Aistis“. Vilniaus universiteto tyrimo grupė vykdė darbus šiose srityse:

- Kalbos signalų apdorojimo priemonės (Girdenis, A. ir kt., 1996). Sukurta kalbos signalų apdorojimo programa „Kalbame“, leidžianti įvesti kalbos signalą, nupiešti jo grafą,

spektrogramą, pagrindinio tono grafiką, filtruoti ir karpyti signalą, pasiklausyti signalo ar jo segmento, sintezuoti signalą iš segmentų.

- Automatinis žodžių kirčiavimas (Kasparaitis, P., 2000; Kasparaitis, P., 2001). Sudarytos beveik 9000 veiksmožodžių kamienų, virš 53000 daiktavardžių ir būdvardžių kamienų, virš 2300 nekaitomų žodžių, daiktavardžių, būdvardžių ir veiksmožodžių galūnių, veiksmožodžių priešdėlių duomenų bazės. Sukurti žodžių darybos, kaitymo ir kirčiavimo algoritmai realizuoti kompiuterinių programų pavidalu.
- Automatinis teksto transkribavimas (Kasparaitis, P., 1999). Sudarytas 740 taisyklių rinkinys bei sukurtas šias taisykles interpretuojantis algoritmas.
- Fonetinių vienetų bazė. Sudaryta fonetinių vienetų bazė. Ji sudaroma iš žodžių ar sakinių, kuriuose būtų visi reikalingi vienetai, sąrašo. Sudarytą sąrašą perskaito diktorius ir jo balsas skaitmeniniu būdu įrašomas į kompiuterį. Po to iš šių įrašų iškerpami fonetiniai vienetai.
- Automatinis teksto skiemonavimas (Kasparaitis, P., 2001). Sukurtas skiemonavimo algoritmas, kuris leidžia beveik be klaidų skiemenuoti lietuvių kalbos tekstą.

Vytauto Didžiojo universitete atlikti tyrimai skirti šnekos sudedamųjų dalių – fonetinių vienetų savitų intonacinių savybių tyrimui (Talandytė, S. ir kt., 2004), atsiribojant nuo skiemenu, žodžių, frazių ir sakinio intonacijų, o intonacija suprantama kaip šnekos signalo – pagrindinio tono kitimas. Taip pat atlikti tyrimai balsių ir priebalsių trukmių, keičiantis šnekos tempui nuo lėto į vidutinį ir nuo vidutinio į greitą. Vykdamas Lietuvių kalbos informacinėje visuomenėje 2000 – 2006 metų programą, 2000 metais pradėti sistemingi lietuvių kalbos automatinio atpažinimo tyrimai. Šiuose tyrimuose dalyvavo Vytauto Didžiojo universiteto ir Lietuvos teismo ekspertizės centro mokslininkai. Buvo sukurta atskirai sakomų lietuvių kalbos žodžių automatinio atpažinimo, naudojant kalbos pavyzdžių dinaminį laiko skalės kraipymą, modeliavimo programinę įrangą.

Šiaulių universitete (Daunys, G., 2006) tiriamas balso trakto kitimas, tiriant garsus. Sukurtas naujas kalbos signalo trumpos atkarpos spektro įvertinimo metodas, kuris leidžia tiksliau įvertinti skardžių balsių garsų spektrą negu tradiciniai tiesinės prognozės metodai.

Matematikos ir informatiko institute sukurta keletas kalbančiojo identifikavimo ir verifikavimo pagal balsą metodų, kurie yra naudojami atliekant fonoskopines ekspertizes (Lipeika, A. ir kt., 2002).

2. ESAMŲ KALBOS TECHNOLOGIJŲ PRODUKTŲ ANALIZĖ

Šiame skyriuje analizuojamos šnekamosios kalbos atpažinimo taikymo sritys ir kiekvienoje srityje taikomi produktai bei analizuojami kalbos sintezei naudojami produktai. Tiek kalbos atpažinimo, tiek kalbos sintezės produktai palyginami tarpusavyje.

2.1. Pagrindinės šnekamosios kalbos atpažinimo taikymo sritys ir jose taikomi produktai

Šiame skyrelyje analizuojamos žmogaus bendravimo su kompiuteriu balsu taikymo sritys: kalbinis dialogas su kompiuteriu, balso įrašų stenografavimas, balsinė interneto naršyklė, elektroninių prietaisų valdymas balsu ir asmens atpažinimas pagal jo balsą. Taip pat pateikiami kiekvienoje srityje taikomų produktų pavyzdžiai.

2.1.1. Kalbinis dialogas su kompiuteriu

Tai yra tiesioginis bendravimas su kompiuteriu balso pagalba (Rudžionis, A. ir kt., 2001). Tam pritaikytose programose yra paruoštas komandų sąrašas ir, vartotojui ištarus vieną iš jų, kompiuteris, priklausomai nuo atpažintos komandos numerio, atitinkamai reaguoja. Komandos gali būti labai įvairios pvz.: „parodyk kiek valandų“, „pagrok Lietuvos himną“ ir t.t. programai nepavykus patikimai atpažinti balso komandos, išvedamas pranešimas apie tai ir prašymas pakartoti komandą. Atpažinus baigimo komandą, programa baigia darbą.

Šios srities kalbos technologijų taikymo produktų pavyzdžiai:

- „VOICE BANKING“ (Hello banking, 2009) – kalbos technologija, kuri suteikia klientams prieigą prie jų produktų ir paslaugų, naudodama telefoną. Klientai nėra priversti naudoti klaviatūrą. Balso bankininkystė leidžia skubiai padaryti užklausas, atlikti sandorius ir pan. Ja galima naudotis 24 valandas per parą, 7 dienas per savaitę. Prisijungti reikia telefonu. Prieiga naudoja balso autentiškumo nustatymą, kur pašnekovo telefonu balsas yra palygintas su turimais balso duomenimis. Balso bankininkystė turi didelį kalbos atpažinimo žodyną.
- „VOICE TRAVEL“ (Hello travel, 2009) – kalbos technologija, kuri įgalina avialinijas ir oro uostus automatizuoti klientų aptarnavimo funkcijas, naudojant telefoną. Klientai nėra priversti naudoti klaviatūrą. Klientai gali sužinoti kainas, rezervuoti bilietus. Ja galima naudotis 24 valandas per parą, 7 dienas per savaitę. Kaip ir balso bankininkystė turi didelį kalbos atpažinimo žodyną.

2.1.2. Balso įrašų stenografavimas

Programinė įranga, skirta balso įrašams stenografuoti naudojama interviu, posėdžių, forumų, sesijų, derybų, paskaitų, mitingų, konferencijų metu (Rudžionis, A. ir kt., 2001). Šiose programose kartu su tekstiniu redaktoriumi paleidžiama foninė garsinio išvedimo programa, valdoma iš teksto redaktoriaus aplinkos su makrofunkcijomis. Vartotojas gali sustabdyti išvedimą bet kuriuo laiko momentu bei tęsti nuo sustabdymo vietos, valdyti išvedimo greitį, pakartoti tam tikro įrašo fragmento išvedimą, automatiškai daryti pageidaujamo dydžio pauzes kas apibrėžtą laiko intervalą, keisti išvedimo programos parametrus nepertraukiant darbo.

Šios srities kalbos technologijų taikymo produktų pavyzdžiai:

- „DRAGON NATURALLY SPEAKING MEDICAL“ (Dragon NaturallySpeaking Medical, 2009) - kalbos technologija, kuri sukuria medicinos dokumentus, kalbant natūraliai ir tiesiogiai į kompiuterį ar į magnetofoną.

2 lentelė

„Dragon NaturallySpeaking Medical“ sistemos reikalavimai

PARAMETRAS	APRAŠYMAS
Procesorius	Intel Pentium 1 GHz
Sistemos atmintis	1 GB
Kietasis diskas	1 GB
Operacinė sistema	Windows XPHome/Professional/W2000/Vista
Garso plokštė	Palaikanti 16 bitų
CD-ROM diskas	
Triukšmą panaikinantis mikrofonas	
Interneto prieiga	
Palaikantis Bluetooth technologiją mikrofonas	

Šaltinis: *Dragon NaturallySpeaking Medical* [interaktyvus]. Nuance [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.nuance.com/naturallyspeaking/products/medical.asp>>.

- „TOKI VOKI VOICE EDITOR“ (Toki Voki, 2001) – tai balso redaktorius. Jis suteikia galimybę įrašyti savo balso žinutę ar pastabą, poredaguoti ją ar siųsti elektroniniu paštu.

3 lentelė

„Toki Voki Voice Editor“ sistemos reikalavimai

PARAMETRAS	APRAŠYMAS
Procesorius	Pentium 200Mhz
Operacinė sistema	Windows(TM): W95/W98/Me/NT4/W2000/XP/Vista
Mikrofonas	

Šaltinis: *Toki Voki* [interaktyvus]. Bluechillies. Atnaujinta 2001 liepos 29d. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.bluechillies.com/details/10112.html>>.

2.1.3. Balsinė interneto naršyklė

Interneto naršymas balsu yra pakankamai patogus – tam skirtos programos gali atlikti platų spektrą komandų: paleisti naršyklę, padaryti naršyklę matoma, užkrauti nurodytą interneto svetainę, uždaryti naršyklę, gauti tekstą, kuris yra užkrautoje interneto svetainėje, gauti užkrautos svetainės HTML kodą, stabdyti svetainės atidarymą, pakartotinai užkrauti svetainę, atidaryti namų svetainę, atidaryti sekančią svetainę, atidaryti prieš tai buvusią svetainę (Rudžionis, A. ir kt., 2001). Interneto svetainių skaitymo programos pagrindinis langas yra mažo formato, kad netrukdytų vartotojui stebėti interneto svetainių.

Šios srities kalbos technologijų taikymo produktų pavyzdžiai:

- „WINDOWS VISTA HOME BASIC“ (Windows žinynas ir instrukcijos, 2009). Kalbos atpažinimas operacinėje sistemoje „Windows Vista“ – kalbos technologija, kuri leidžia:

1) valdyti kompiuterį. Kalbėjimo atpažinimas klauso ir atsako į sakomas komandas. Kalbėjimo atpažinimas naudojamas programoms paleisti ir ryšiui su kompiuteriu užmegzti.

2) diktuoti ir redaguoti tekstą. Komandų atpažinimas naudojamas diktuoti žodžius žodžių apdorojimo programoms, tokioms kaip Microsoft Word, arba anketų pildymui tinkle. Kalbėjimo atpažinimas naudojamas ir tekstui kompiuteryje redaguoti.

Kalbos atpažinimo programa veikia anglų, prancūzų, ispanų, vokiečių, japonų, supaprastinta kinų ir tradicine kinų kalba.

4 lentelė

„Windows Vista Home Basic“ sistemos reikalavimai

PARAMETRAS	APRAŠYMAS
Procesorius	1 GHz 32-bit (x86), 64-bit (x64)
Sistemos atmintis	512 MB
Kietasis diskas	20 GB
Grafikos palaikymas	DirectX 9
Grafikos atmintis	32 MB
DVD-ROM diskas	
Audio išvestis	
Interneto prieiga	

Šaltinis: *Windows žinynas ir instrukcijos* [interaktyvus]. Microsoft [žiūrėta 2008 m. birželio 17d.]. Prieiga per internetą: <<http://windowshelp.microsoft.com/Windows/lt-LT/Help/55c37f8e-2b44-45df-bccb-e1938230b0041063.msp>>.

- „DRAGON NATURALLY SPEAKING 10“ (Dragon NaturallySpeaking 10, 2009) – tai technologija, kuri suteikia galimybę kalbėtis su kompiuteriu. Diktuojamas tekstas pasirodo dokumentuose, parašomi elektroniniai laišakai, greitosios žinutės. Programos tikslumas – 99 %.

5 lentelė

„Dragon NaturallySpeaking 10“ sistemos reikalavimai

PARAMETRAS	APRAŠYMAS
Procesorius	Intel Pentium 2.4 GHz
Sistemos atmintis	512 MB, 1 GB
Kietasis diskas	1 GB
Operacinė sistema	Windows Server 2000, Windows Server 2003, Windows XP SP2, Windows 2000 SP4, Windows Vista
Garso plokštė	Palaikanti 16 bitų
DVD-ROM diskas	
Triukšmą panaikinantis mikrofonas	
Interneto prieiga	
Palaikantis Bluetooth technologiją mikrofonas	

Šaltinis: *Dragon NaturallySpeaking 10* [interaktyvus]. Nuance. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.nuance.com/naturallyspeaking/products/preferred.asp>>.

2.1.4. Elektroninių prietaisų valdymas balsu

Labai patogiu ir praktiška buitinius elektros prietaisus įjungti ir išjungti balsu. Programinę įrangą sudaro personaliniame kompiuteryje veikianti komandų atpažinimo programa bei valdiklio programa, per nuoseklų prievadą iš kompiuterio priimanti atpažintos komandos numerį ir per I2C magistralę atliekanti elektrinių įrenginių įjungimą/išjungimą (Rudžionis, A. ir kt., 2001).

Šios srities kalbos technologijų taikymo produktų pavyzdžiai:

- „AVIC-HD1BT“ (įrenginys) (IBM sukurta nauja balso atpažinimo sistema, 2006). Bendrovė „Pioneer Europe“ panaudojo „IBM“ sukurta „Embedded ViaVoice“ garso atpažinimo programinę įrangą, kad jos gaminamą automobiliams skirtą garso ir navigacijos įtaisą „AVIC-HD1BT“ būtų galima paprastai valdyti balsu. „Pioneer“ navigacijos sistemai „IBM“ mokslininkai sukūrė kalbos algoritmus, kuriais atpažįstama kalba bei balsu skaitomas tekstas. Ši sistema praktiškai bet kurios mašinos vairuotojui leidžia saugiai ir lengvai valdyti GPS navigacijos sistemą, klausytis muzikos, gauti naujienas, rasti bankomatus, bankus, degalines, parduotuves, restoranus, išsirinkti važiavimo maršrutą. „Pioneer AVIC-HD1BT“ įtaisas gali būti įrengtas daugumoje mašinų, kur yra balso atpažinimo sistema navigacijai, garso takelių paieškai ir kitam

valdymui. Jame taip pat yra kietasis diskas, kuriame telpa 10 GB muzikos biblioteka, įrengta „Bluetooth“ sąsaja telefonui. Navigaciją, muziką ir telefoną galima kontroliuoti balsu naudojant „IBM Embedded ViaVoice“ technologiją. „AVIC-HD1BT“ laimėjo prestižinį Europos vaizdo ir garso asociacijos (EISA) apdovanojimą kaip 2006—2007 metų geriausia Europos navigacijos ir daugialypės terpės sistema.

- „EMBEDDED VIA VOICE“ (Embedded ViaVoice, 2009) - kalbos technologija mobiliems prietaisams ir automobilio komponentams. Ši technologija suteikia vartotojams balso prieigą prie informacijos bei užtikrina visišką kalbos atpažinimą.

6 lentelė

„Via Voice“ sistemos reikalavimai

PARAMETRAS	APRAŠYMAS
Operacinė sistema	Embedded Linux, Linux, QNX, Windows 2000, Windows CE, Windows Mobile, Windows XP Professional
Sistemos atmintis	700KB
Procesorius	Milijonai instrukcijų per sekundę (MIPs) reikalinga: 25 MIPs minimumas; 175 MIPS 100000 žodžių;
ROM ir flash atmintis	1100KB
Audio	16-bit sample, mono; 11/16/22KHZ sampling

Šaltinis: *Embedded ViaVoice* [interaktyvus]. IBM. Atnaujinta 2009 sausio 29d. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <http://www-306.ibm.com/software/pervasive/embedded_viavoice/>.

2.1.5. Asmens atpažinimas pagal jo balsą

Kaip jau minėta, tai labai svarbu kriminalistikoje. Moksliniuose kalbinių technologijų tyrinėjimuose kalbėtojo atpažinimas pagal jo balsą yra suprantamas kaip procesas, kurio metu iš asmens kalbos signalų išskiriami identifikaciniai požymiai, pagal kuriuos ir atpažįstamas konkretus asmuo (Rudžionis, A. ir kt., 2001). Kalbėtojo atpažinimas savo ruožtu skirstomas į identifikavimą ir verifikavimą. Tiek verifikavimas, tiek identifikavimas gali būti priklausomas arba nepriklausomas nuo teksto. Teismo ekspertizėje asmens atpažinimas pagal balsą vadinamas asmens identifikavimu ir dažniausiai yra nepriklausomas nuo teksto. Kriminalistikoje naudojami ir verifikavimo elementai, kai atliekamas asmens vertinimas pagal balsą. Kaip pavyzdį čia galima paminėti asmens balso palyginimą su balsais iš turimos balso bazės. Pagrindinis bruožas skiriantis teisminį asmens identifikavimą nuo kitų verifikavimo sistemų yra tai, kad turimas „nelinkusio bendrauti“ asmens balsas. 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 t.t.

Šios srities kalbos technologijų taikymo produktų pavyzdžiai:

- „RAKTAS“ (Balso atpažinimo modulis „RAKTAS“, 2009) – tai automatinio asmens identifikavimo pagal balsą sistema. Modulį „Raktas“ sudaro dvi dalys: mokymo ir atpažinimo posistemės. Modulio mokymo kursas vienam vartotojui užtrunka vos 30 – 35 sekundes, jei naudojamas slaptažodžio kartojimu 3 kartus. Vieno slaptažodžio frazės trukmė 2 – 3 sekundės. Slaptažodžio frazė gali būti pasakyta bet kokia kalba. Apmokymo rezultatų (vieno kalbančiojo balso 3 etalonų) saugojimui sunaudojama apie 30 kB atminties. Atpažinimo stadijoje, kalbantysis ištaria slaptažodį į mikrofoną. Jei kalbantysis yra tas, kuo jis sakosi esąs, modulis jį atpažįsta ir išduoda leidimą (žinutę). Jei asmuo bando apgauti modulį, tai jį uždraudžia priėjimą prie kompiuterių tinklo ar duomenų bazės, arba įėjimą į patalpas. Verifikavimas užtrunka tik 2 – 3 sekundes. Modulio „Raktas“ kaip biometrinės sistemos panaudojimo sritys: apsauga nuo nesankcionuoto priėjimo prie duomenų bazių, kompiuterinių tinklų ar patekimo į saugomas patalpas (pastatus, biurus, cechus ir pan.). „Raktas“ gali būti pritaikytas bankuose, dirbant su klientais telefonu. „Raktas“ kaip autonominis prietaisas gali būti naudojamas kaip „balso raktas“ butams, garažams, automobiliams, seifams ir pan. Speciali „Rakto“ panaudojimo sfera – kriminalistika. Programa, kartu su kitomis kalbančiojo identifikavimo sistemomis gali būti pritaikyta, kuriant automatizuotą asmens verifikavimo sistemą pagal balsą.

7 lentelė

„Raktas“ sistemos reikalavimai

PARAMETRAS	APRAŠYMAS
Personalinis kompiuteris	
Mikrofonas	
Garso atkūrimo sistema	
"Sound Blaster" tipo garso korta ir programa	

Šaltinis: *Balso atpažinimo modulis „RAKTAS“* [interaktyvus]. Technograma. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <http://www.technograma.lt/Kalbos/lt_duomenu.html>.

2.2. Šnekamosios kalbos atpažinimo produktų palyginimas

Atlikus šnekamosios kalbos atpažinimo produktų analizę kiekvienoje taikymo srityje, sudaryta produktų palyginimo lentelė, išskiriant produktams pagrindinius bruožus: kalbėjimo stilių, žodyną, apmokymą, daugiakalbiškumą, prieiga per telefoną ir portabilumą. Gauti rezultatai pateikti 8 lentelėje.

8 lentelė

Šnekamosios kalbos atpažinimo produktų charakteristikos

Parametrai	Produktai								
	Voice banking	Voice travel	Dragon NaturallySpeaking Medical	Toki Voki Voice Editor	Windows Vista Home Basic	Dragon NaturallySpeaking 10	Via Voice	AVIC-HD1BT	Raktas
Kalbėjimo stilius	Komandos	Komandos	Laisvas	Laisvas	Komandos	Laisvas	Komandos	Komandos	Laisvas
Žodynas	>1000 žodžių	>1000 žodžių	>1000 žodžių	>1000 žodžių	>1000 žodžių	>1000 žodžių	>1000 žodžių	>1000 žodžių	>1000 žodžių
Apmokymas	+	+	+	+	+	+	+	+	+
Daugiakalbis	+	+	+	+	+	+	+	+	+
Telefonas	+	+							
Portabilumas			+		+	+	+		

Šaltinis: sudaryta autorės.

Visų analizuojamų paslaugų ir produktų žodynas yra didesnis nei 1000 žodžių, visos yra daugiakalbės ir turi apmokymo – naudotojo vadovus.

„Voice banking“ ir „Voice travel“ yra reikalingas telefonas kaip prieiga prie teikiamų produktų ar paslaugų. Parametras „Portabilumas“ parodo, kad produktas gali būti integruotas į kažkokį įrenginį, pavyzdžiui, navigacinę sistemą. Visuose produktuose naudojamas arba laisvas, arba komandinis kalbėjimo stilius.

„Voice banking“ ir „Voice travel“ paslaugos Lietuvos rinkoje nėra prieinamos. Pavyzdžiui, „baslo bankininkystė“ yra išvystyta Japonijos ir Šveicarijos šalių bankuose. Įrenginį „AVIC-HD1BT“ ir visus kitus likusius analizuotus produktus Lietuvoje galima laisvai įsigyti, tačiau nei vienas jų nėra nemokamas. Tik kai kurie produktai pateikia internete savo bandomąsias versijas, tačiau jos neturi pilno funkcionalumo. Nei vienas iš produktų nėra pritaikytas lietuvių kalbai, todėl toliau, eksperimentiniame skyriuje, kalbos atpažinimo tyrimui pasirinktas produktas „Dragon NaturallySpeaking 10“ ir ištirtos jo panaudojimo lietuvių kalbai galimybės.

2.3. Kalbos sintezei taikomi produktai ir jų palyginimas

Vienas kalbos sintezės produktų yra lietuviškai kalbantis sintezatorius „Aistis“. Jis gali vyrišku balsu perskaityti pateiktą tekstą. Sintezatorius sudarytas iš tokių pagrindinių bloků: žodžių skiemonavimas, žodžių kirčiavimas, transkribavimas ir kalbos signalo formavimas.

Prieš kirčiuojant ir transkribuojant lietuvių kalbos žodį, jį reikia suskiemuoti. Tai reikalinga todėl, kad kirčiavimo taisyklėse kirčio vieta nusakoma kaip kirčiuoto skiemens numeris, o be to transkribuojant reikia atsižvelgti į skiemenų ribas, nes dvibalsius ir mišriuosius dvigarsius gali sudaryti tik vienas skiemeniui priklausantis garsai (Kasparaitis, P., 2001) Skiemonavimui naudojamas algoritmas, kuris remiasi lietuvių kalbos skiemens struktūra, taip pat naudojamas priešdėlių atskyrimas bei balsių kombinacijos, kurios negali priklausyti vienam skiemeniui. Žodžių kirčiavimas remiasi žodžių dalių žodynais ir kaitymo bei kirčiavimo taisyklėmis.

Transkribavimui (teksto perrašymui į fonetinių vienetų seką) naudojamos formalios taisyklės, kurių kairėje pusėje nurodoma einamoji raidė, einamosios raidės požymiai (skiemens riba, kirtis ir priegaidė, kietumas/minkštumas), kairysis kontekstas ir dešinysis kontekstas, o dešinėje pusėje nurodomas fonetinis vienetas, kiek raidžių praleisti pritaikius taisyklę ir nuo kurios taisyklės pradėti kitos taisyklės paiešką (Kasparaitis, P., 1999). Transkribavimui iš viso naudojama apie 700 taisyklių.

Kalbos signalo formavimui pasirinktas konkatenuacinis metodas, kuriame jungiami natūralios diktoriaus kalbos segmentai. Sintezatoriaus „Aistis“ fonetinių vienetų bazei sudaryti, buvo naudojama programa „Kalbame“ (Girdenis, A. ir kt., 1996), kuri leidžia įvesti kalbos signalą, nupiešti jo grafą, pasiklausyti signalo ar jo segmento, sintezuoti signalą iš segmentų.

Lietuvių kalba turi unikalią ir gana sudėtingą kirčiavimo sistemą. Dėl to kirčiavimo procesas išskaidytas į visų gramatinių formų atpažinimą ir jų sukirčiavimą bei kirčiavimo varianto išrinkimą (Kasparaitis, P., 2000; Kasparaitis, P., 2001).

Kitas kalbos sintezės produktas yra programa „ReadPlease 2003“ (ReadPlease and ReadPlease Plus 2003, 2009). Skirtingai nei sintezatorius „Aistis“, kalbos sintezės programa yra nemokama. Ją galima parsisiųsti internete. Naudojantis kitomis panašiomis programomis, dažnai reikia jomis naudotis tiesiogiai internete, pavyzdžiui, „Oddcast TTS“ programa. Tai apsunkina programos veikimą. Programa „ReadPlease 2003“ neturi iš anksto nustatytų apribojimų įvedamų žodžių ar simbolių skaičiui. Pateiktą tekstą ji gali perskaityti keliais balsais. Taip pat galima reguliuoti skaitymo greitį bei mažinti ar didinti pateikto teksto šrifto dydį. Vartotojo sąsaja paprasta. Norint naudotis programa „ReadPlease 2003“ tereikia į ekraną suvesti norimą perskaityti tekstą ir paspausti mygtuką „Play“.

Pagrindiniai skirtumai tarp sintetatoriaus „Aistis“ ir programos „ReadPlease 2003“:

- sintetatorius „Aistis“ pritaikytas lietuvių kalbai, o programa „ReadPlease 2003“ anglų kalbai;
- programa „ReadPlease 2003“ yra laisvai prieinama ir galima parsisiųsti internetu, o sintetatoriui „Aistis“ reikalingos licencijos.
- programa „ReadPlease 2003“ skaito tekstą keliais balsais ir galima reguliuoti skaitymo greitį, o sintetatorius „Aistis“ skaito tekstą vienu balsu ir pastoviu tempu.

Dėl programos prieinamumo paprastumo, eksperimentiniame skyriuje, kalbos sintezės tyrimui pasirinktas produktas „ReadPlease 2003“ ir ištirtos jo panaudojimo lietuvių kalbai galimybės.

3. KALBOS TECHNOLOGIJŲ PRODUKTŲ ADAPTAVIMO GALIMYBIŲ TYRIMAS

Šiame skyriuje analizuojamos dvi programos, kurios veikia skirtingais principais. Vienai iš jų pateikiamas tekstas ir tada ji perskaito tekstą balsu (kalbos sintezės programa), kitai programai pateikiamas balsas ir tada ji padiktuotą balsą pateikia tekstu (kalbos atpažinimo programa).

Eksperto tikslas – ištirti programų panaudojimo lietuvių kalbai galimybes, t.y. lietuvišką žodį pateikti tokiais anglų kalbos atitikmenimis, kad pateiktas žodis būtų tinkamai ištartas lietuviškai bei sudaryti geriausios transkripcijos taisyklės ir įvertinti, kurie garsai, lietuviškai kalbančio diktorius, atpažįstami geriausiai. Taip pat nustatyti žodžių atpažinimo efektyvumus bei pateikti pasiūlymus komandų rinkinio sudarymui.

Tyrimui atlikti pasirinkta kalbos atpažinimo programa „Dragon NaturallySpeaking 10“ ir kalbos sintezės programa „ReadPlease 2003“.

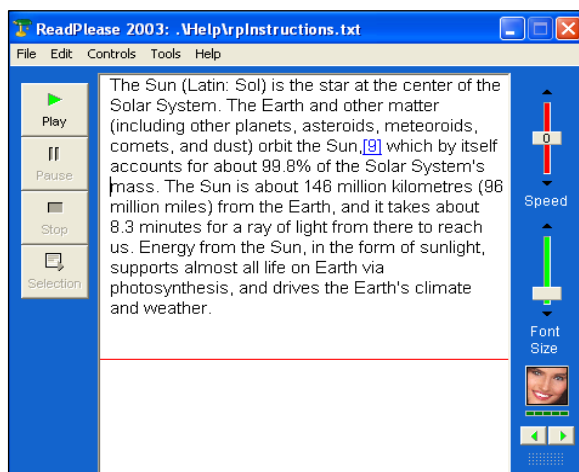
3.1. Lietuvių fonetikos transformavimo į anglų kalbos rašybą taisyklių sudarymo galimybių tyrimas

Tyrimui atlikti pasirinkta kalbos sintezės programa „ReadPlease 2003“. Atliktu tyrimu buvo siekiama nustatyti, ar įmanoma anglų kalba veikiančioje programoje „ReadPlease 2003“, lietuvišką žodį parašyti tokiomis anglų kalbos raidėmis, kad žodis skambėtų pagal lietuvišką tartį. Šiam tikslui buvo iškelti tokie uždaviniai:

- Išanalizuoti programos „ReadPlease 2003“ veikimo principus ir galimybes.
- Ištirti, kaip programa perskaito įvestus žodžius.
- Rasti žodžių aiškumą.
- Pagal tiriamus žodžius sudaryti geriausios transkripcijos taisyklių modelį.

Tyrimui atlikti naudojama programa „ReadPlease 2003“ perskaito bet kokią tekstą, kuris yra įvestas jos ekrane. 6 paveiksliuke pavaizduotas pagrindinis programos langas. Jame taip pat matomi visi parametrai, kurie buvo nustatyti atliekant bandymą.

Atliktame tyrime dalyvavo trys dalyviai. Pirmasis dalyvis išriko tokius lietuvių kalbos žodžius, kad būtų panaudotos visos lietuvių abėcėlės raidės, išskyrus balse su nosinėmis ir transkribavo išrinktus žodžius pagal anglų kalbos taisykles. Suvedus žodžius į programą (ir lietuviškus, ir transkribuotus) antrasis, po to trečiasis, dalyvis vertino kiekvieno žodžio aiškumą ir kokybę. Aiškumas buvo vertinamas „+“ arba „-“. Jei antrasis ir trečiasis dalyvis suprasedavo programos perskaitytą žodį, skirdavo „+“, priešingu atveju „-“. Žodžio kokybė buvo vertinama balais nuo 0 iki 5. Aukščiausias kokybės įvertinimas - „5“, žemiausias – „0“. Kokybės vertinimas priklausė nuo to kaip programos perskaityti žodžiai atitinka lietuviškąją tarimą.



6 pav. „ReadPlease 2003“ programos pagrindinis langas

9 lentelėje pateikiami lietuviški žodžiai ir sudaryti transkripcijos variantai bei antrojo dalyvio žodžių aiškumo ir kokybės vertinimas, 10 lentelėje – trečio dalyvio žodžių aiškumo ir kokybės vertinimas.

9 lentelė

Antrojo dalyvio žodžių aiškumo ir kokybės vertinimas

Nr.	Žodis	Aiškumas	Kokybė	Transkribuotas žodis	Aiškumas	Kokybė
1.	tupi	+	2	toopi	+	3
2.				toopee	-	0
3.	ant	-	0	unt	+	3
4.				ont	-	0
5.	anukas	-	0	unookas	+	3
6.				unuikas	-	0
7.				anookas	+	2
8.	bangos	+	1	bungos	+	5
9.				bongos	+	3
10.	galva	+	1	gulva	+	5
11.				golva	+	3
12.	vista	-	0	vishtaa	+	5
13.				veashta	+	1
14.				veeshta	+	1
15.	himnas	-	0	himnus	+	3
16.				heemnas	+	1
17.				heamnas	+	1

9 lentelės tęsinys

18.	vakcina	-	0	vukcina	-	0
19.				vuktsina	+	3
20.	dangus	-	0	dungoos	+	2
21.				dangoos	-	0
22.				dungus	-	0
23.	morka	+	5	morka	+	5
24.	ezys	-	0	ezhees	+	2
25.				azhees	+	0
26.				ezheas	-	0
27.	cesnakas	-	0	chesnukas	-	0
28.				chesnokas	+	1
29.				chesnakas	+	4
30.	ungurys	-	0	ungurees	-	0
31.				oongurees	-	0
32.				oongoorees	+	4
33.	fabrikas	+	3	fubrikas	+	3
34.				fubrikus	-	0
35.				fubreekas	+	2
36.	zyle	-	0	zeele	-	0
37.				zeale	-	0
38.				ziele	-	0
39.	instrumentas	+	2	instroomentas	+	3
40.				eenstroomentas	-	0
41.	yla	-	0	eala	+	2
42.				eela	-	0
43.				iela	-	0
44.	suris	-	0	sooris	+	3
45.				suiris	-	0
46.				soorees	-	0
47.	jura	-	0	ura	-	0
48.				yura	-	0
49.				yuiras	-	0

9 lentelės tęsinys

50.		yoora	+	4
-----	--	-------	---	---

Šaltinis: sudaryta autorės.

10 lentelė

Trečiojo dalyvio žodžių aiškumo ir kokybės vertinimas

Nr.	Žodis	Aiškumas	Kokybė	Transkribuotas žodis	Aiškumas	Kokybė
1.	tupi	+	0	toopi	+	3
2.				toopee	-	0
3.	ant	-	0	unt	+	4
4.				ont	-	0
5.	anukas	-	0	unookas	+	3
6.				unuikas	-	0
7.				anookas	+	1
8.	bangos	+	1	bungos	+	5
9.				bongos	+	4
10.	galva	+	1	gulva	+	5
11.				golva	+	4
12.	vista	-	0	vishtaa	+	5
13.				veashta	+	1
14.				veeshta	+	1
15.	himnas	-	1	himnus	+	4
16.				heemnas	+	1
17.				heamnas	+	1
18.	vakcina	-	0	vukcina	-	0
19.				vuktsina	+	3
20.	dangus	-	0	dungoos	+	3
21.				dangoos	-	0
22.				dungus	-	0
23.	morka	+	5	morka	+	5
24.	ezys	-	0	ezhees	+	4
25.				azhees	+	3
26.				ezheas	-	0

10 lentelės tęsinys

27.	cesnakas	-	0	chesnukas	-	0
28.				chesnokas	+	1
29.				chesnakas	+	5
30.	ungurys	-	0	ungurees	-	0
31.				oongurees	-	0
32.				oongoorees	+	4
33.	fabrikas	+	1	fubrikas	+	4
34.				fubrikus	-	0
35.				fubreekas	+	1
36.	zyle	-	0	zeele	-	0
37.				zeale	-	0
38.				ziele	-	0
39.	instrumentas	+	2	instroomentas	+	3
40.				eenstroomentas	-	0
41.	yla	-	0	eala	+	3
42.				eela	-	0
43.				iela	-	0
44.	suris	-	0	sooris	+	4
45.				suiris	-	0
46.				soorees	-	0
47.	jura	-	0	ura	-	0
48.				yura	-	0
49.				yuiria	-	0
50.				yoorra	+	3

Šaltinis: sudaryta autorės.

Apžvelgus vertintojų vertinimų rezultatus pastebėta, kad jie beveik nesiskiria, todėl tyrimui atlikti pasirinkta dviem vertintojais, darant prielaidą, kad didinant vertintojų skaičių rezultatai keistųsi nežymiai.

Pagal 9 ir 10 lentelių duomenis sudaryta vidurkinė, abiejų vertintojų, žodžių aiškumo ir kokybės vertinimo lentelė. Gauti rezultatai pateikti 11 lentelėje.

11 lentelė

Antro ir trečio dalyvio vidutinis žodžių aiškumo ir kokybės vertinimas

Nr.	Žodis	Aiškumas	Kokybė	Transkribuotas žodis	Aiškumas	Kokybė
1.	tupi	+	1	toopi	+	3
2.				toopee	-	0
3.	ant	-	0	unt	+	3,5
4.				ont	-	0
5.	anukas	-	0	unookas	+	3
6.				unuikas	-	0
7.				anookas	+	1,5
8.	bangos	+	1	bungos	+	5
9.				bongos	+	3,5
10.	galva	+	1	gulva	+	5
11.				golva	+	3,5
12.	vista	-	0	vishtaa	+	5
13.				veashta	+	1
14.				veeshta	+	1
15.	himnas	-	0,5	himnus	+	3,5
16.				heemnas	+	1
17.				heamnas	+	1
18.	vakcina	-	0	vukcina	-	0
19.				vuktsina	+	3
20.	dangus	-	0	dungoos	+	2,5
21.				dangoos	-	0
22.				dungus	-	0
23.	morka	+	5	morka	+	5
24.	ezys	-	0	ezhees	+	3
25.				azhees	+	1,5
26.				ezheas	-	0
27.	cesnakas	-	0	chesnukas	-	0
28.				chesnokas	+	1

10 lentelės tęsinys

29.				chesnakas	+	4,5
30.	ungurys	-	0	ungurees	-	0
31.				oongurees	-	0
32.				oongoorees	+	4
33.	fabrikas	+	2	fubrikas	+	3,5
34.				fubrikus	-	0
35.				fubreekas	+	1,5
36.	zyle	-	0	zeele	-	0
37.				zeale	-	0
38.				ziele	-	0
39.	instrumentas	+	2	instroomentas	+	3
40.				eenstroomentas	-	0
41.	yla	-	0	eala	+	2,5
42.				eela	-	0
43.				iela	-	0
44.	suris	-	0	sooris	+	3,5
45.				suiris	-	0
46.				soorees	-	0
47.	jura	-	0	ura	-	0
48.				yura	-	0
49.				yuirā	-	0
50.				yoora	+	3,5

Šaltinis: sudaryta autorės.

Pagal abiejų vertintojų vidutinį žodžių aiškumo ir kokybės vertinimą, išrinkti žodžiai, kuriems buvo skirtas aukščiausias kokybės įvertinimas, ir surašyti 12 lentelėje.

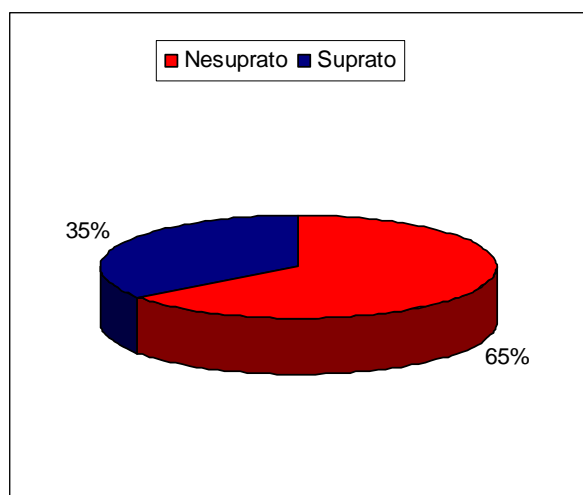
12 lentelė

Aukščiausias žodžių kokybės vertinimas

Nr.	Žodis	Kokybė	Geriausia žodžio transkripcija	Kokybė
1.	tupi	1	toopi	3
2.	ant	0	unt	3,5
3.	anukas	0	unookas	3
4.	bangos	1	bungos	5
5.	galva	1	gulva	5
6.	vista	0	vishtaa	5
7.	himnas	0,5	himnus	3,5
8.	vakcina	0	vuktsina	3
9.	dangus	0	dungoos	2,5
10.	morka	5	morka	5
11.	ezys	0	ezhees	3
12.	cesnakas	0	chesnakas	4,5
13.	ungurys	0	oongoorees	4
14.	fabrikas	1,5	fubrikas	3,5
15.	zyle	0	zeele	0
16.	instrumentas	2	instroomentas	3
17.	yla	0	eala	2,5
18.	suris	0	sooris	3,5
19.	jura	0	yooraa	3,5

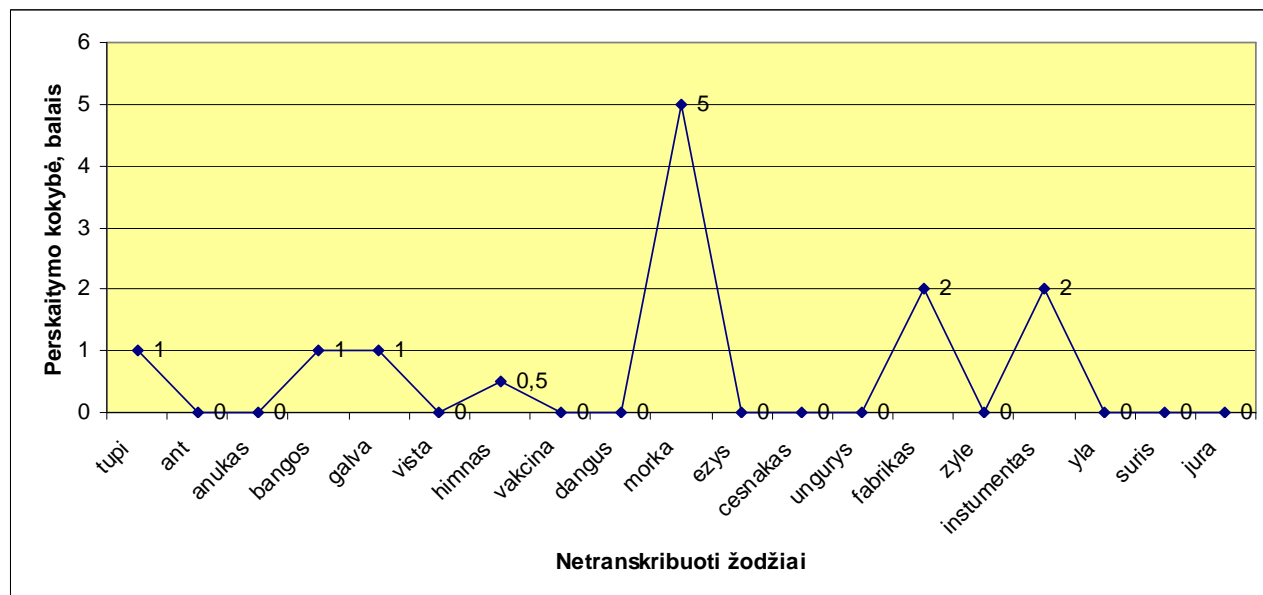
Šaltinis: sudaryta autorės.

Pagal aukščiausią žodžių kokybės vertinimą (12 lentelė) sudaryta lietuviškų žodžių perskaitymo aiškumo diagrama (7 pav.) ir kokybės diagrama (8 pav.) bei geriausiai transkribuotų lietuviškų žodžių perskaitymo aiškumo diagrama (9 pav.) ir kokybės diagrama (10 pav.)



Šaltinis: sudaryta autorės.

7 pav. Lietuviškų žodžių perskaitymo aiškumas



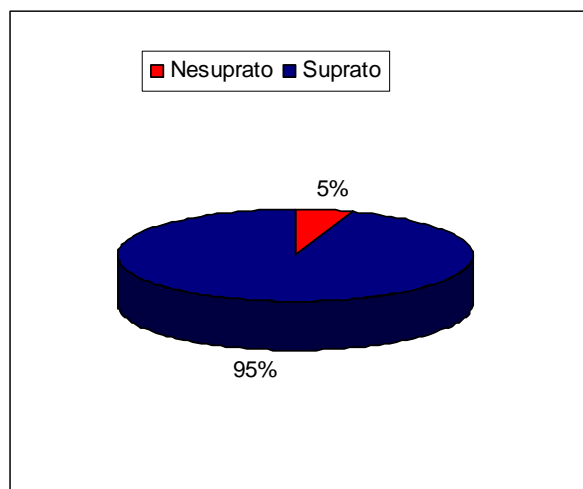
Šaltinis: sudaryta autorės.

8 pav. Lietuviškų žodžių perskaitymo kokybė

Iš visų pateiktų žodžių tik vieno žodžio *morka* lietuviška pateiktis buvo perskaityta pakankamai aiškiai. Jis gavo 5 balų kokybės įvertinimą ir nebuvo transkribuotas. Tarptautiniai žodžiai *fabrikas* ir *instrumentas* taip pat buvo perskaityti aiškiai, tačiau ne pakankamai ir jų kokybė įvertinta tik po 2 balus. Žodžiai *tupi*, *bangos* ir *galva* įvertinti po 1 balą, o žodžiui *himnas* skirtas 0,5 balo kokybės įvertinimas.

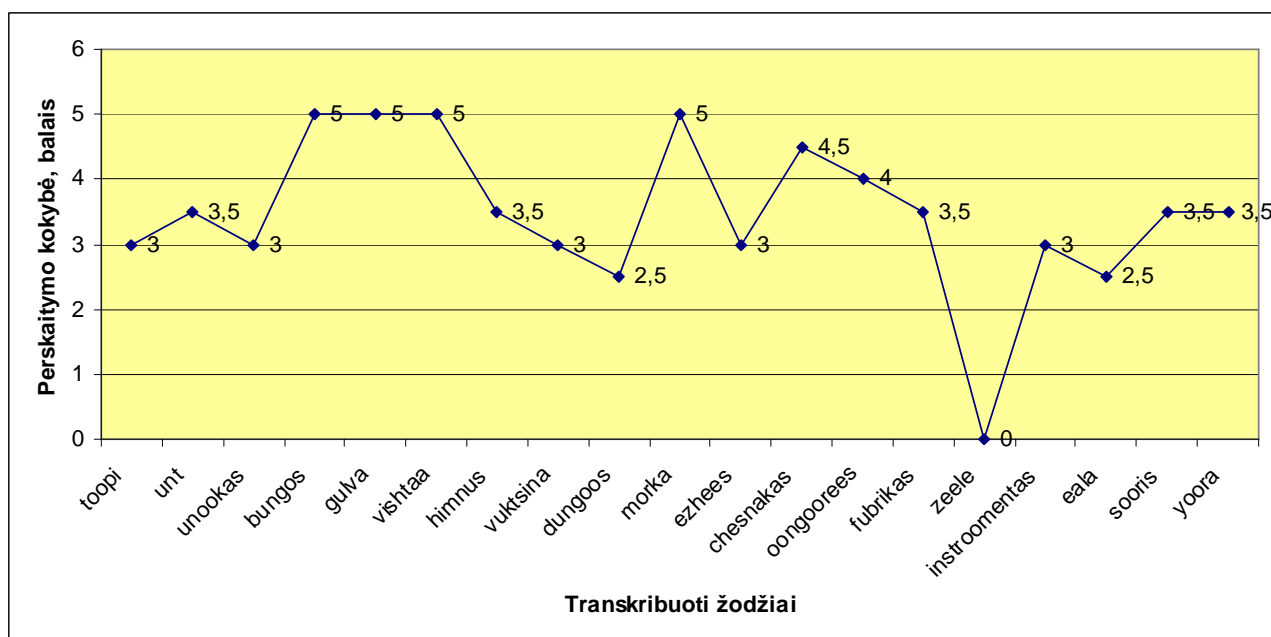
Iš programos „ReadPlease 2003“ perskaitytų lietuviškų žodžių, vertintojas suprato tik septynis žodžius (8 pav.). Apskaičiuotas lietuviškų žodžių supratimo tikslumas siekia vos 35% (7

pav.). Siekiant padidinti šį tikslumą, lietuvių kalbos žodžiai buvo užrašomi pagal anglų kalbos fonetikos taisykles (t.y. transkribuoti).



Šaltinis: sudaryta autorės.

9 pav. Geriausiai transkribuotų lietuviškų žodžių perskaitymo aiškumas



Šaltinis: sudaryta autorės.

10 pav. Geriausiai transkribuotų lietuviškų žodžių perskaitymo kokybė

Po transkribavimo, tik vieno žodžio *zylė* kokybė buvo įvertinta 0, kadangi šiam žodžiui aiški transkripcija nebuvo užrašyta, nes *ë* raidės anglų kalboje nėra ir neįmanoma išgauti jos tarimo. Net keturiems žodžiams skirtas 5 balų kokybės įvertinimas. Tai *bangos*, *galva*, *višta* ir *morka*.

Iš programos „ReadPlease 2003“ perskaitytų lietuviškų žodžių, vertintojas suprato 18 žodžių (10 pav.). Apskaičiuotas transkribuotų lietuvių kalbos žodžių aiškumo tikslumas 95% (9 pav.). Lyginant transkribuotų ir netranskribuotų žodžių aiškumo tikslumus, galima apskaičiuoti, kad skirtumas tarp jų yra 60%.

Pagal aukščiausios žodžių kokybės vertinimo lentelę (12 lentelė) yra matoma, jog pavyzdžiui, anglų kalboje pučiamąjį garsą „č“ žymi dvi raidės – „ch“, tuo tarpu lietuvių kalboje šis garsas žymimas viena raide „č“. Atsižvelgiant į visiems žodžiams užrašytus geriausius transkripcijos variantus sudarytas taisyklių modelis, kuris pateiktas 13 lentelėje.

13 lentelė

Transkripcijos taisyklių modelis

Nr.	Lietuviškos taisyklės	Geriausios transkripcijos taisyklės	Pavyzdžiai	Pastabos
Balsiai:				
1.	axx	uxx	ant, anūkas	
2.	xax	xux	galva, bangos, dangus	
3.	xxa	xxa	morka	
4.	xxa	xxaa	višta	Kai –a kirčiuotas.
5.	uxx	ooxx	ungurys	
6.	xux	xoox	tupi	
7.	xūx	xoox	sūris	
8.	xyx	xeex	žvyras, ežys	
9.	yxx	eaxx	yla	
10.	ė	-	zylė	Vietoj –ė, sakoma e.
Priebalsiai:				
11.	ž	zh	ežys	
12.	š	sh	višta	
13.	č	ch	česnakas	
14.	c	ts	vakcina	
15.	j	y	jūra	

Šaltinis: sudaryta autorės.

Kombinacijos, pavyzdžiui *axx*, reiškia balsės poziciją žodyje. Šiuo atveju balsė *a* yra pirmoji raidė. Pagal nagrinėtus pavyzdžius trumpųjų balsių *e* ir *i* tarimas lietuviškame žodyje nesiskyrė nuo tarimo transkribuotame žodyje. Balsės *o* tarimas taip pat daugelyje žodžių atitiko, tik kaip kur ji buvo tariama panašiai į raidę *a*, pavyzdžiui žodžiui *bangos* parašytos dvi transkripcijos *bongos* ir *bungos*, siekiant išgauti panašiausia garso atitikmenį žodžio šaknyje. Pirmajai transkripcijai vertintojai skyrė 3,5 balus, o antrajai - 5. Remiantis tuo, užrašyta 7 geriausios transkripcijos taisyklė.

Visų kitų priebalsių, kurių nėra paminėta 13 lentelėje, tarimas dažniausiai nesiskyrė nuo atitinkamų lietuvių kalbos priebalsių tarimo.

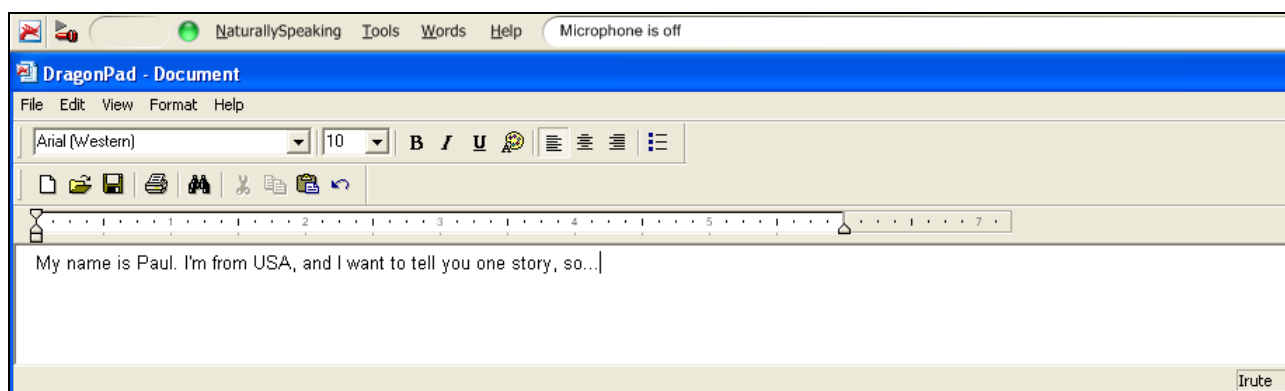
Apibendrinat tyrimo rezultatus, galima teigti, kad lietuvių kalbos žodžiams užrašyti naudojant anglų kalbos fonetinę rašybą ir pateikus užrašytus žodžius programai „ReadPlease 2003“ perskaityti, žodžių aiškumas tapo 60% procentų didesnis nei užrašius pagal lietuvių kalbos fonetinę rašybą.

3.2. Anglišku žodžių atpažinimo tyrimas

Tyrimui atlikti pasirinkta programa „Dragon NaturallySpeaking 10“. Atliktu tyrimu buvo siekiama nustatyti, kaip efektyviai pasirinkta programa atpažįsta lietuviškai kalbančio diktoriaus diktuojamus žodžius. Šiam tikslui buvo iškelti tokie uždaviniai:

- Išanalizuoti programos „Dragon NaturallySpeaking 10“ veikimo principus ir galimybes.
- Iširti, ar programa atpažįsta diktuojamus žodžius.
- Rasti žodžių atpažinimo efektyvumą.
- Nustatyti kur programa dažniausiai daro klaidas.
- Pateikti pasiūlymus komandų rinkinio sudarymui.

Tyrimui atlikti naudojama programa „Dragon NaturallySpeaking 10“ suteikia galimybę kalbėti su kompiuteriu. Ji leidžia rašyti tekstus tik šnekant per mikrofoną, o tekstą ji išveda ten kur nurodomas žymeklis (pvz. Skype, MSN Messenger, MS Word ir pan.). 11 paveiksle pavaizduotas programos pagrindinis langas, kai diktuojamas tekstas užrašomas pačios programos redaktoriuje.



11 pav., „Dragon NaturallySpeaking 10“ programos pagrindinis langas

Prieš atliekant tyrimą, programa buvo pratinama prie tyrime dalyvaujančių dalyvių balso ir tarimo. Tai buvo atlikta iš programoje pateiktų testų išsirinkus vieną ir perskaičius. Pripratinti programą prie balso yra rekomenduojama siekiant geresnio žodžių atpažinimo efektyvumo, t.y. kuo tikslesnio padiktuotų žodžių užrašymo. Kiekvienam diktoriui buvo sukurtas dalyvio profilis.

Atliktame tyrime dalyvavo penki lietuviškai kalbantys diktoriai. Visi diktoriai mokėjo anglų kalbą vidutiniškai. Prieš atliekant tyrimą, diktoriai buvo supažindinti su programos veikimo principu, pademonstruotas programos veikimas. Visi dalyviai pratę dirbti kompiuteriu, ir kalbėjo natūraliai. Jie programai diktavo po penkiasdešimt anglų kalbos žodžių. Eksperimento metu praktiniu būdu buvo patikrintas programos veikimas, kiekvieną žodį diktuojant po dešimt kartų.

1 priede atvaizduojama kaip programa „Dragon NaturallySpeaking 10“ savo redaktoriuje išveda pirmojo diktoriaus padiktuotus žodžius.

Diktoriams pateikti žodžiai buvo laisvai pasirinkti, tačiau suskirstyti į tris grupes pagal simbolių skaičių: trumpi (iki 3 simbolių), vidutiniai (4 – 5 simboliai) ir ilgi (6 ir daugiau simbolių).

Eksperto metu gauti rezultatai pateikiami 14 lentelėje.

14 lentelė

Žodžių atpažinimo skaičius

Nr.	Žodis	Bendras diktavimų skaičius	Bendras atpažinimo skaičius	Atpažinimo skaičius				
				Pirmas diktorius	Antras diktorius	Trečias diktorius	Ketvirtas diktorius	Penktas diktorius
<i>Trumpi žodžiai</i>								
1.	age	50	50	10	10	10	10	10
2.	all	50	47	10	10	10	9	8
3.	bad	50	50	10	10	10	10	10
4.	bee	50	0	0	0	0	0	0
5.	big	50	50	10	10	10	10	10
6.	car	50	49	10	10	9	10	10
7.	eat	50	42	8	9	8	7	10
8.	fun	50	50	10	10	10	10	10
9.	gas	50	49	10	10	9	10	10
10.	hat	50	31	10	1	10	3	7
11.	yes	50	50	10	10	10	10	10
12.	jar	50	10	10	0	0	0	0
13.	lie	50	50	10	10	10	10	10
14.	run	50	8	0	2	0	0	6
15.	sky	50	38	10	10	0	10	8
16.	sun	50	30	5	8	2	8	7
17.	zoo	50	0	0	0	0	0	0
<i>Vidutinio ilgumo žodžiai</i>								
18.	baby	50	43	9	7	9	10	8
19.	book	50	50	10	10	10	10	10
20.	escape	50	50	10	10	10	10	10
21.	father	50	0	0	0	0	0	0
22.	import	50	50	10	10	10	10	10
23.	mind	50	47	10	10	9	8	10
24.	muscle	50	50	10	10	10	10	10

14 lentelės tęsinys

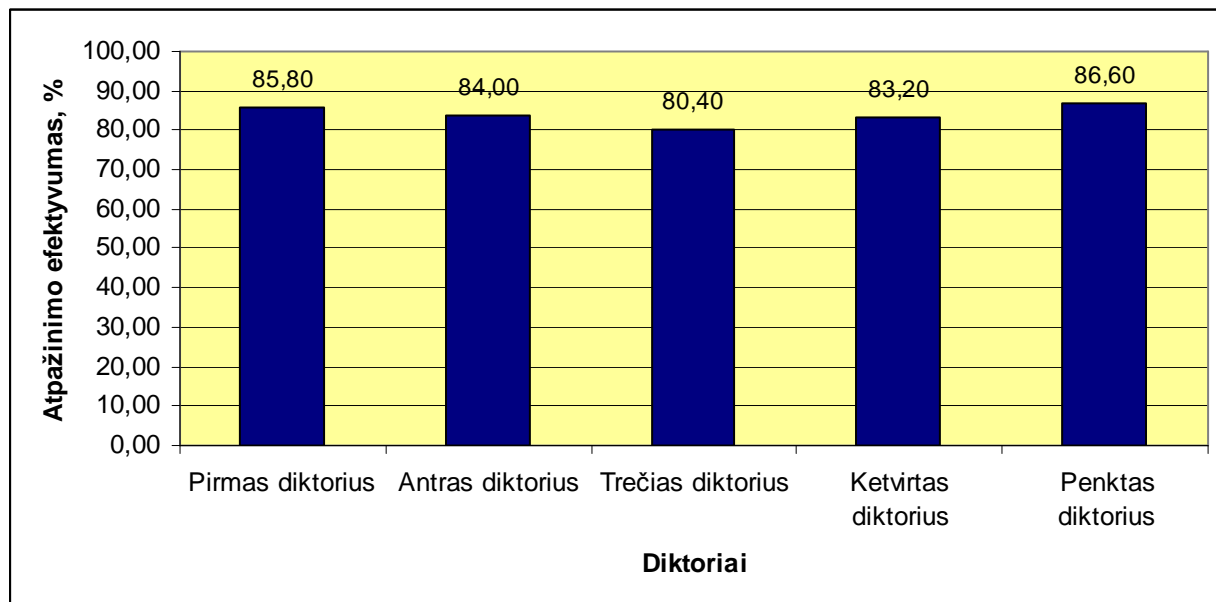
25.	next	50	50	10	10	10	10	10
26.	people	50	50	10	10	10	10	10
27.	scream	50	50	10	10	10	10	10
28.	shower	50	0	0	0	0	0	0
29.	silver	50	46	6	10	10	10	10
30.	story	50	50	10	10	10	10	10
31.	strong	50	50	10	10	10	10	10
32.	sweet	50	41	10	7	9	5	10
33.	unit	50	50	10	10	10	10	10
34.	voice	50	50	10	10	10	10	10
<i>Ilgi žodžiai</i>								
35.	advantage	50	50	10	10	10	10	10
36.	crossroads	50	50	10	10	10	10	10
37.	description	50	50	10	10	10	10	10
38.	education	50	50	10	10	10	10	10
39.	hundred	50	47	10	8	9	10	10
40.	introduction	50	37	1	10	10	6	10
41.	yesterday	50	50	10	10	10	10	10
42.	kilometer	50	50	10	10	10	10	10
43.	kitchen	50	50	10	10	10	10	10
44.	lighthouse	50	50	10	10	10	10	10
45.	mistake	50	50	10	10	10	10	10
46.	silence	50	50	10	10	10	10	10
47.	terrible	50	38	10	9	0	10	9
48.	Thursday	50	50	10	10	10	10	10
49.	umbrella	50	47	10	9	8	10	10
50.	vegetable	50	50	10	10	10	10	10

Šaltinis: sudaryta autorės.

14 lentelėje nuo 1 iki 17 numerio esantys žodžiai priklauso trumpų žodžių grupei, nuo 18 iki 34 – vidutinio ilgumo ir nuo 35 iki 50 – ilgų. Stulpelyje „Bendras diktavimų skaičius“ pateiktas kiekvieno žodžio visų diktorių bendras diktavimų skaičius, o stulpelyje „Bendras atpažinimo

skaičius“ – visų diktorių bendras atitikusių žodžių skaičius. Stulpelyje „Atpažinimo skaičius“ pateiktas atskirai kiekvieno diktoriaus, dalyvavusio tyrime, atitikusių žodžių skaičius.

Pagal žodžių atpažinimo skaičių (12 lentelė) apskaičiuotas bendras žodžių atpažinimo efektyvumas kiekvienam iš penkių diktorių atskirai. Gauti rezultatai pateikti 12 pav.



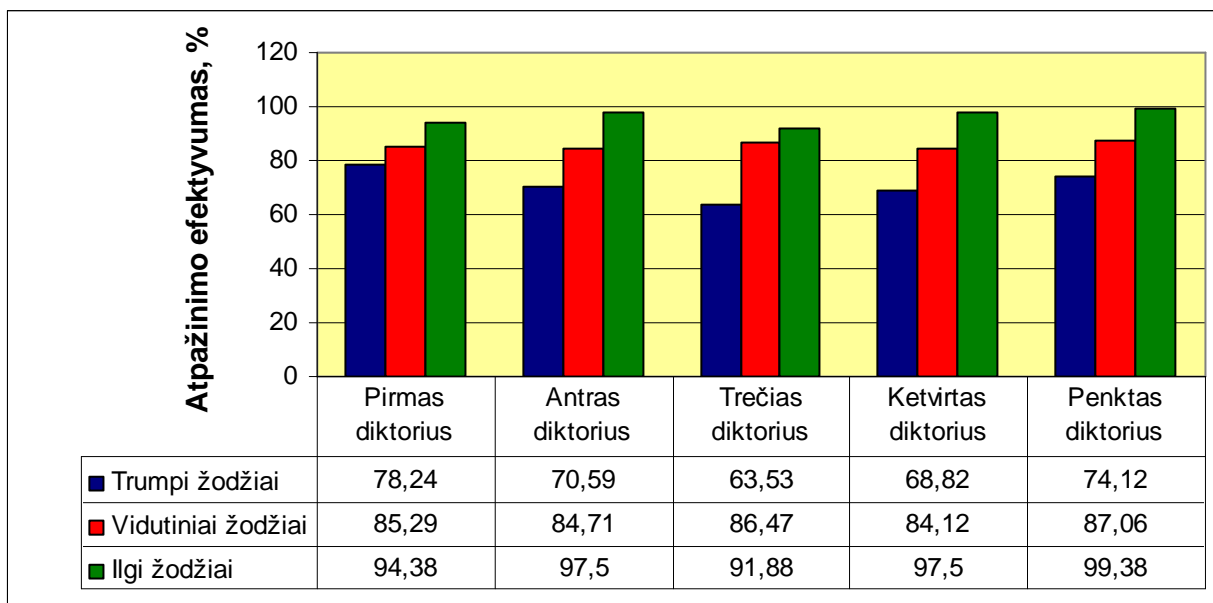
Šaltinis: sudaryta autorės.

12 pav. Penkių diktorių žodžių atpažinimo efektyvumas

Pirmojo diktoriaus atpažinimo efektyvumas yra 85,80%, antrojo – 84,00%, trečiojo – 80,40%, ketvirtojo – 83,20%, penktojo – 86,60%. Atpažinimo efektyvumo svyravimų amplitudė yra 6,2%. Apskaičiuotas vidutinis atpažinimo efektyvumas yra 84,8%.

Visų penkių diktorių žodžių atpažinimo efektyvumas yra apylygis, todėl tyrimui atlikti pasirinkta penkiais diktoriais, darant prielaidą, kad didinant diktorių, kurie moka anglų kalbą vidutiniškai, skaičių rezultatai keistųsi nežymiai.

Pagal žodžių atpažinimo skaičių (12 lentelė) apskaičiuotas bendras žodžių atpažinimo efektyvumas kiekvienai žodžių ilgio grupei atskirai ir atvaizduotas 13 pav.



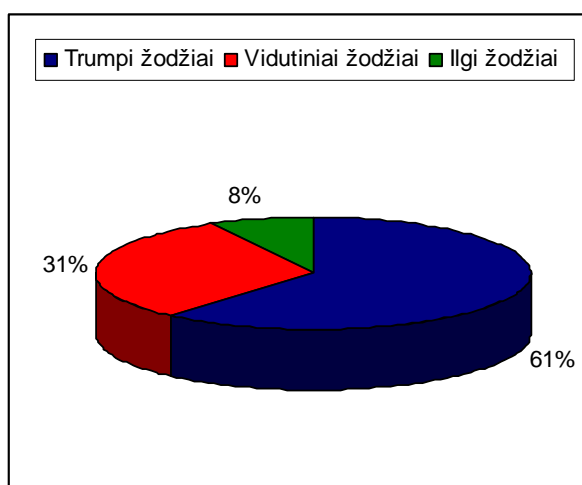
Šaltinis: sudaryta autorės.

13 pav. Diktorių žodžių atpažinimo efektyvumas kiekvienai žodžių grupei

Nei vienos žodžių grupės programa neatpažino 100% tikslumu. Geriausiai atpažįstami ilgi žodžiai. Visų diktorių ilgų žodžių atpažinimo efektyvumas viršija 90%. Vidutinio ilgumo žodžių atpažinimo efektyvumas svyruoja intervale nuo 84,12% iki 87,06%, trumpų žodžių – nuo 63,53% iki 78,24%.

Mažiausiai atpažįstami trumpi žodžiai. Trečiojo diktoriaus trumpų žodžių atpažinimas siekia vos 63,53%. Tai žemiausias rodiklis lyginant kiekvieno diktoriaus atpažinimo efektyvumus skirtingose žodžių grupėse.

Pagal tyrimo rezultatus rastas klaidų skaičius pateiktas 15 lentelėje ir, pagal joje surašytas klaidas, procentiškai atvaizduotas kiekvienai žodžių ilgio grupei (14 pav.).



Šaltinis: sudaryta autorės.

14 pav. Klaidos kiekvienai žodžių ilgio grupei

Klaidų skaičius

Nr.	Žodis	Klaidos	Klaidų skaičius
<i>Trumpi žodžiai</i>			
1.	all	old	3
2.	bee	be, beat	50
3.	car	can't, our	1
4.	eat	yet, each	8
5.	gas	guys, get	1
6.	hat	had, that, pants	19
7.	jar	John, Jenna, Jonah	40
8.	run	from, room, gun, then	42
9.	sky	try	12
10.	sun	some, fun, son	20
11.	zoo	you, you'll, do	50
<i>Vidutinio ilgumo žodžiai</i>			
12.	baby	David, I be, bit	7
13.	father	55 to, five, other, either	50
14.	mind	find, my	3
15.	shower	show, shaula, showwen, xiao, sholem, sharla, trowel	50
16.	silver	Silva, seal, sin	4
17.	sweet	street, sleep	9
<i>Ilgai žodžiai</i>			
18.	hundred	wouldn't, you when it	3
19.	introduction	in production, in the abduction	13
20.	terrible	tenable, edible	12
21.	umbrella	under an, on the other	3

Šaltinis: sudaryta autorės.

Stulpelyje „Klaidos“ pateikti visi klaidų variantai, kurie buvo užrašyti. Matome, kad programa dažniausiai klydo užrašydama panašaus tarimo žodžius (*sun – some, hat – had, mind – find*).

Pagal klaidas kiekvienai žodžių grupei (14 pav.) matoma, kad daugiausiai klaidų programa darė trumpuose žodžiuose, net 61% bendro klaidų skaičiaus. Mažiausiai – ilguose žodžiuose, tik 8% bendro klaidų skaičiaus.

Rastas klaidų skaičius žodžiuose pagal priebalsinius ir balsinius garsus. Gauti rezultatai pateikti 16 ir 17 lentelėse.

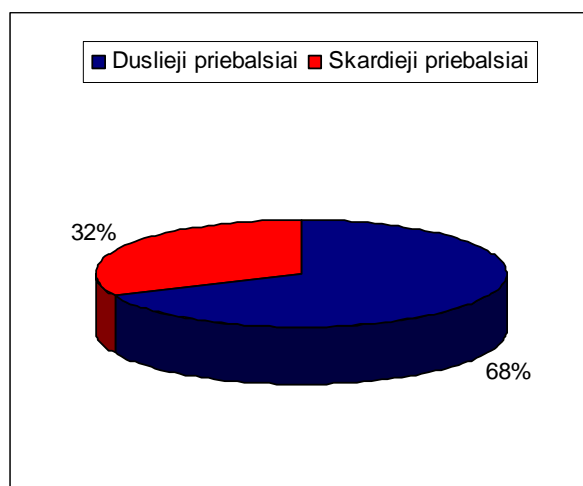
16 lentelė

Priebalsinių garsų klaidų skaičius

Trankskripcija	Klaidų skaičius	Pavyzdžiai
<i>Duslieji priebalsiai:</i>		
[l]	4	all, silver, terrible, umbrella
[b]	4	bee, baby, terrible, umbrella
[g]	1	gas
[h]	4	hat, father, shower, hundred
[j]	1	jar
[r]	9	jar, run, father, shower, silver, hundred, introduction, terrible, umbrella
[n]	5	run, sun, mind, hundred, introduction
[z]	1	zoo
[m]	2	mind, umbrella
[d]	3	mind, hundred, introduction
<i>Skardieji priebalsiai:</i>		
[c]	2	car, introduction
[t]	4	eat, hat, father, introduction
[s]	5	sky, sun, shower, silver, sweet
[k]	1	sky
[t]	4	father, sweet, introduction, terrible

Šaltinis: sudaryta autorės.

Priebalsiai suskirstyti į dvi grupes: skardžiuosius ir dusliuosius. Skardieji priebalsiai – *c, t, s, k, t*, duslieji – *l, b, g, h, j, r, n, z, m, d*. Iš 16 lentelės duomenų, apskaičiuotas klaidų skaičius skardiesiems ir dusliesiems priebalsiams pavaizduotas 15 paveikslėlyje.



Šaltinis: sudaryta autorės.

15 pav. Klaidos žodžiuose su dusliaisiais ir skardžiais priebalsiais

Žodžiuose su skardžiais priebalsiais yra daroma 36% mažiau klaidų nei žodžiuose su dusliaisiais priebalsiais. Iš to seka, kad skardžiuosius priebalsius programa atpažįsta geriau.

17 lentelė

Balsinių garsų klaidų skaičius

Trankskripcija	Klaidų skaičius	Pavyzdžiai
<i>Duslieji priebalsiai:</i>		
[i:]	3	bee, eat, sweet
[i]	3	baby, silver, introduction
[æ]	2	gas, hat
[a:]	3	car, jar, father
[ʌ]	5	run, sun, hundred, introduction, umbrella
[u:]	1	zoo
[ə]	7	father, shower, silver, hundred, introduction, terrible, umbrella
[e]	2	terrible, umbrella
[o]	1	all
[ai]	2	sky, maind
[ei]	1	baby

Šaltinis: sudaryta autorės.

Pagal balsinių garsų klaidų skaičių (17 lentelė) matyti, kad daugiausia klaidų daroma ties:

1. balsių *a* ir *i* trumpinimu bei ilginimu;
2. garso [ə] tarimu.

Pirmąją klaidą gali lemti kelios priežastys. Pirmą yra tai, kad diktorius neteisingai išlaiko reikiamą balsės ilgumą, antrą – žodžiai turintys daugiau balsių yra mažiau suprantami ir ne tokie aiškiai tariami kaip kad turintys daugiau priebalsių, todėl programa ir neužrašo teisingai tokių diktuojamų žodžių.

Ties garso [ə] tarimu, žodžiuose fiksuota daugiausiai klaidų. Lietuvių kalboje tokio garso nėra, todėl lietuviakalbiams diktoriams yra pakankamai sudėtinga jį išstarti.

Apibendrinant, visus gautus šio tyrimo rezultatus galima teigti, kad diktorių žodžių atpažinimo efektyvumas svyruoja nuo 80,40% iki 86,60%, todėl aklausiai pasikliauti programa diktuojant žodžius negalima, tačiau prižiūrint ir pataisant neteisingai įvestą žodį, ji gali būti naudojama kaip puiki rašymo priemonė.

Sudarinėjant komandų rinkinį programai „Dragon NaturallySpeaking 10“ siūloma naudoti ilgus žodžius, t.y. žodžius, kurie turi šešis ar daugiau simbolių, kadangi juose klaidos tikimybė yra mažiausia lyginant su trumpais ar vidutinio ilgumo žodžiais. Rekomenduojama vengti žodžių, kurie turi dusliųjų priebalsių, nes jų atpažinimas yra kur kas mažesnis nei žodžių su skardžiais priebalsiais. Taip pat rekomenduojama vengti žodžių kurie turi garsą [ə], kadangi jo tarimas lietuviakalbiams diktoriams kelia daugiausia keblumų tarp visų balsinių garsų tarimo.

IŠVADOS

1. Svarbiausios kalbos technologijų grupės yra kalbos atpažinimas ir kalbos sintezė. Kalbos atpažinimo paskirtis yra automatiškai nustatyti, kas balsu sakoma informacijos priėmimo sistemai, o kalbos sintezės pagalba pagal reikiamą komandą balsu perskaitoma informacinėje sistemoje teksto pavidalu saugoma informacija.

2. Šiuo metu kalbos technologijos vystosi ir tobulėja, daugėja jų taikymo sričių, tačiau nėra plačiai prieinamų lietuvių kalbos atpažinimo ir sintezavimo programų. Didžiosios kompanijos pirmiausia orientuojasi į didžiąsias pasaulio kalbas, todėl lietuvių kalba, dėl mažo rinkos potencialo, jas mažai domina.

3. Apžvelgus eilę kalbos atpažinimo ir kalbos sintezės produktų, dėl programų prieinamumo ir paprasto integravimo, eksperimentiniam tyrimui atlikti pasirinkta kalbos atpažinimo programa „Dragon Naturally Speaking 10“ ir kalbos sintezės programa „ReadPlease 2003“.

4. Eksperimentinio tyrimo metu iširtos pasirinktų anglų kalba veikiančių programų „ReadPlease 2003“ ir „Dragon NaturallySpeaking 10“ panaudojimo lietuvių kalbai galimybės.

5. Pirmuoju tyrimu su programa „ReadPlease 2003“, nustatytas lietuviškų žodžių perskaitymo aiškumas yra vos 35%. Tuos pačius žodžius transkribavus perskaitymo aiškumas gautas 95%. Galima teigti, kad lietuvių kalbos žodžiams užrašyti naudojant anglų kalbos fonetinę rašybą ir pateikus užrašytus žodžius programai „ReadPlease 2003“ perskaityti, žodžių aiškumas tapo 60% didesnis nei užrašius pagal lietuvių kalbos fonetinę rašybą.

6. Antruoju tyrimu su programa „Dragon Naturally Speaking 10“, nustatyta, kad geriausiai yra atpažįstami ilgi žodžiai, mažiausiai – trumpi. Daugiausia klaidų yra daroma ties dviejų balsių *a* ir *i* trumpinimu bei ilginimu, ir ties garso [ə] tarimu. Taip pat daugiau klaidų yra daroma žodžiuose su dusliaisiais priebalsiais nei žodžiuose su skardžiais priebalsiais. Apskaičiuotas bendras programos atpažinimo efektyvumas, pagal penkių diktorių diktavimo rezultatus, yra 84,8%. Todėl galima teigti, kad akiai pasikliauti programa diktuoju žodžius negalima, tačiau prižiūrint ir pataisant neteisingai įvestą žodį, ji gali būti naudojama kaip puiki rašymo priemonė.

LITERATŪRA

Mokslinės literatūros sąrašas:

1. Ceccarelli, M. (2000) *International Symposium on History of Machines and Mechanisms Proceedings HMM*. p. 404. ISBN 978-0-7923-6372-9.
2. Daunys, G. (2006) *Spektro įvertinimas iš trumpos kalbos signalo atkarpos*. Kaunas: Technologija. p. 58-63. ISBN 9955-09-993-3.
3. Domatas, A.; Rudžionis, A. (1991) *Towards more reliable automatic recognition of the phonetic units*. Proc. of the XIIth International Congress of Phonetic Sciences. Aix-En Provence, France, p. 478-481.
4. Domatas, A.; Rudžionis, A.; Ezerskis, N. (1986) *Dichotomijų panaudojimas fonemų skyrimui*. ARSO medžiaga. Kaunas. p. 35 - 35.
5. Girdenis, A.; Kasparaitis, P.; Pečeliūnaitė, A.; Skirmantas, P.; Undzėnas V. (1996). *Lietuvių kalbos bei jos tarmių prozodinių reiškinių ir fonemų alofonų analizė*. Moksl. ataskaita. Lietuvos valst. mokslo ir studijų fondas, registr. Nr. 94–081/3 G–1994 01 06. VU BKK. Vilnius.
6. *Kalbos ir diktoriaus atpažinimas telefonijoje*. (1995) VSMF remiamo mokslinio KTU darbo ataskaita. Reg. Nr. BK 57E112, Kaunas.
7. Kasparaitis, P. (1999) *Transcribing of the Lithuanian Text Using Formal Rules*. Informatica, 10(4), p. 367-376.
8. Kasparaitis, P. (2001) *Automatic Stressing of the Lithuanian Nouns and Adjectives on the Basis of Rules*. Informatica, 12(2), p. 315-336.
9. Kasparaitis, P. (2001) *Lietuvių kalbos kompiuterinis sintezatorius „Aistis“*. Garso korta 2001, KTU Technologija.
10. Kasparaitis, P. (2000) *Automatic Stressing of the Lithuanian Text on the Basis of a Dictionary*. Informatica, 11(1), p. 19-40.
11. Lipeika, A. (2003) *Kalbos organai, kalbos generavimas ir suvokimas*. [interaktyvus]. [žiūrėta 2008 m. gruodžio 20d.]. Prieiga per internetą: <<http://www.mif.vu.lt/~bastys/lipeika/>>.
12. Lipeika, A. (2005) *Formantiniai požymiai atpažįstant kalbą*. Informacijos mokslai. p. 215-219. ISSN 1392-0561.
13. Lipeika, A.; Lipeikienė, L.; Telksnys, L. (2002) *Development of isolated word speech recognition system*. Informatica, 13(1).
14. Noreika, S.; Rudžionis, A. *Žadinimo šaltinio analizė naudojant atvirkštinę filtraciją*. ARSO medžiaga, Jerevanas, 1980, p. 139 -142.

15. *PMM naudojimas kalbos atpažinime*. [interaktyvus]. Vilniaus Gedimino technikos universitetas (2002) [žiūrėta 2008 m. gruodžio 10d.]. Prieiga per internetą: <<http://www2.el.vgtu.lt/ssa/secB5.html>>.

16. Rabiner, L.; Juang, B.-H. (1993) *Fundamentals of speech recognition*. Prentice Hall, New Jersey.

17. Rudžionis, A. (1986) *Fonetinių vienetų skyrimo analizė keičiant apmokymo duomenis ir diskriminantes*. ARSO medžiaga. Kaunas. p. 26 - 27.

18. Rudžionis, A.; Rudžionis, V. (1999) *Phoneme discrimination analysis*. Moscow: Speech Technology Today, IX Session of the Russian Acoustical Society. p. 62 - 65.

19. Rudžionis, A. *Detection of speech signals and endpointing in noise*. (2000) Vilnius: Senamiesčio Spaustuvė. p. 143-145. ISBN 9986-738-13-X.

20. Rudžionis, A. (1989) *Izoliuotų žodžių, vaizduojamų fiksuoto ilgio vektoriais, atpažinimas*. Kalbos signalų informatika. Kijevas: Kalbos signalų informatika. p. 65 – 69.

21. Rudžionis, A. (2001) *Balso technologijų taikymo lietuvių kalbai analizė ir perspektyvinių veiklos kryptių pagrindimas* [interaktyvus]. [žiūrėta 2009 m. kovo 1 d.]. Prieiga per internetą: <http://www.likit.lt/frames/balso_tech/balsotech_st.htm>.

22. Rudžionis, V. (1995) *Įvairių klasifikatorių ir analizatorių palyginimas daugiadiktoriniame fonemų atpažinime*. Informacinės technologijos - 95, Kaunas, KTU. p. 87-93.

23. Rudžionis, V. (1994) *Klasifikatorių palyginimas viendiktoriniam fonemų klasifikavimui*. Informacinės technologijos - 94, Kaunas, KTU.

24. Talandytė, S.; Raškinis, A.; Raškinis, G. (2004) *Nuo konteksto nepriklausomų lietuvių šnekos fonetinių vienetų intonacijos (pagrindinio tono) modeliai ir jų tyrimas*. Konferencijos Informacinės technologijos 2004 pranešimų medžiaga. p. 217–221.

Informacijos šaltinių sąrašas:

25. *Balso atpažinimo modulis „RAKTAS“* [interaktyvus]. Technograma. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <http://www.technograma.lt/Kalbos/lt_duomenu.html>.

26. *Company overview* [interaktyvus]. Nuance. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.scansoft.co.uk/company/>>.

27. *Dragon NaturallySpeaking 10* [interaktyvus]. Nuance. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.nuance.com/naturallyspeaking/products/preferred.asp>>.

28. *Dragon NaturallySpeaking 10 Product Suite* [interaktyvus]. Nuance. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.nuance.co.uk/naturallyspeaking/products/default.asp>>.

29. *Dragon NaturallySpeaking Medical* [interaktyvus]. Nuance [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.nuance.com/naturallyspeaking/products/medical.asp>>.
30. *Embedded ViaVoice* [interaktyvus]. IBM. Atnaujinta 2009 sausio 29d. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <http://www-306.ibm.com/software/pervasive/embedded_viavoice/>.
31. *Hello banking* [interaktyvus]. Emerging Technologies. Atnaujinta 2008 rugsėjo 15d. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.em-t.com/Banking-Voice-Recognition-Solutions-s/119.htm>>.
32. *Hello Travel* [interaktyvus]. Emerging Technologies. Atnaujinta 2008 rugsėjo 15d. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.em-t.com/Arabic-Travel-Voice-Recognition-Solutions-s/57.htm>>.
33. *How VoIP Works* [interaktyvus]. HowStuffWorks. Atnaujinta 2008 spalio 30d. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://communication.howstuffworks.com/ip-telephony1.htm>>.
34. *IBM DirectTalk for Windows* [interaktyvus]. IBM. Atnaujinta 2001m. gegužės 10d. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.redbooks.ibm.com/abstracts/sg245985.html?Open>>.
35. *IBM sukurta nauja balso atpažinimo sistema* [interaktyvus]. IBM Lietuva. Atnaujinta 2006 spalio 23d. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.elektronika.lt/reviews/theme/350/5781/>>.
36. *Microsoft Gives Mobile Devices a Voice in Europe* [interaktyvus]. Microsoft. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.microsoft.com/windowsmobile/en-us/default.aspx>>.
37. *Microsoft Office versijų palyginimas* [interaktyvus]. Microsoft. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.microsoft.com/lietuva/officexp/evaluation/compare.aspx>>.
38. *Natural Language* [interaktyvus]. AAAI. Atnaujinta 2008m. gruodžio 13d. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.aaai.org/AITopics/pmwiki/pmwiki.php/AITopics/NaturalLanguage>>.
39. *ReadPlease and ReadPlease Plus 2003*. [interaktyvus]. ReadPlease Corporation. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.readplease.com/english/readplease.php>>.
40. *Toki Voki* [interaktyvus]. Bluechillies. Atnaujinta 2001 liepos 29d. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.bluechillies.com/details/10112.html>>.

41. *Windows Vista Speech Recognition* [interaktyvus]. Microsoft. Atnaujinta 2008 vasario 14d. [žiūrėta 2009 m. gegužės 25d.]. Prieiga per internetą: <<http://www.microsoft.com/enable/products/windowsvista/speech.aspx>>.

42. *Windows žinynas ir instrukcijos* [interaktyvus]. Microsoft [žiūrėta 2008 m. birželio 17d.]. Prieiga per internetą: <<http://windowshelp.microsoft.com/Windows/lt-LT/Help/55c37f8e-2b44-45df-bccb-e1938230b0041063.msp>>.

PRIEDAI

1 PRIEDAS. Pirmojo diktoriaus diktavimo rezultatai.



The screenshot shows a window titled "DragonPad - Document" with a menu bar (File, Edit, View, Format, Help) and a toolbar (NaturallySpeaking, Tools, Words, Help, Normal mode). The main text area contains the following list of words and phrases, each repeated multiple times:

be be be be be be be be be be
big big big big big big big big big big
car car car car car car car car car car
age age age age age age age age age age
bad bad bad bad bad bad bad bad bad bad
yet each eat eat eat eat eat eat eat eat
fun fun fun fun fun fun fun fun fun fun
gas gas gas gas gas gas gas gas gas gas
hat hat hat hat hat pants had had had had
yes yes yes yes yes yes yes yes yes yes
lie lie lie lie lie lie lie lie lie lie
jar jar jar jar jar jar jar jar jar jar
all all all all all all all all all all
John Graham and then come down some of them come from term
sky sky sky sky sky sky sky sky sky sky
some sun sun sun some sun sun some some some
so do you do you'll do you do you'll you'll you'll you'll
be baby baby baby baby baby baby baby baby baby
book book book book book book book book book book
55 to five to 55 55 five to five to five to five to 55 to 55 five to
import import import import import import import import import import
escape escape escape escape escape escape escape escape escape escape
muscle muscle muscle muscle muscle muscle muscle muscle muscle muscle
mind mind mind mind mind mind mind mind mind mind
next next next next next next next next next next
people people people people people people people people people
scream scream scream scream scream scream scream scream scream scream
Shaula show when Shaula show show show show show show Shaula
story story story story story story story story story story
strong strong strong strong strong strong strong strong strong strong
Sweet Sweet Sweet Sweet Sweet Sweet Sweet Sweet Sweet Sweet
Silva silver seal Silva sin silver silver silver silver silver silver
unit unit unit unit unit unit unit unit unit unit
voice voice voice voice voice voice voice voice voice voice
mistake mistake mistake mistake mistake mistake mistake mistake mistake mistake
silence silence silence silence silence silence silence silence silence silence
hundred hundred hundred hundred hundred hundred hundred hundred hundred
Thursday Thursday Thursday Thursday Thursday Thursday Thursday Thursday Thursday Thursday
terrible terrible terrible terrible terrible terrible terrible terrible terrible terrible
yesterday yesterday yesterday yesterday yesterday yesterday yesterday yesterday yesterday yesterday
lighthouse lighthouse lighthouse lighthouse lighthouse lighthouse lighthouse lighthouse lighthouse lighthouse
vegetable vegetables vegetable vegetable vegetable vegetable vegetable vegetable vegetable vegetable
kitchen kitchen kitchen kitchen kitchen kitchen kitchen kitchen kitchen kitchen
kilometer kilometer kilometer kilometer kilometer kilometer kilometer kilometer kilometer kilometer
education education education education education education education education education education
crossroads crossroads crossroads crossroads crossroads crossroads crossroads crossroads crossroads crossroads
umbrella umbrella umbrella umbrella umbrella umbrella umbrella umbrella umbrella umbrella
description description description description description description description description description description
introduction in production in production in production in production in production in production in production in production
advantage advantage advantage advantage advantage advantage advantage advantage advantage advantage

2 PRIEDAS. Straipsnis „Kalbos technologijų produktų adaptavimo galimybės komandų atpažinimui“.

KALBOS TECHNOLOGIJŲ PRODUKTŲ ADAPTAVIMO GALIMYBĖS KOMANDŲ ATPAŽINIMUI

Irutė Rakauskaitė

Vilniaus universitetas, Kauno humanitarinis fakultetas, Muitinės g.8, Kaunas, Lietuva

Santrauka. Straipsnyje nagrinėjamos lietuvių fonetikos transformavimo į anglų kalbos rašybą taisyklių sudarymo galimybės, panaudojant šnekamosios kalbos sintezės programą „ReadPlease 2003“ ir lietuviškai kalbančio diktoriaus anglišku žodžių atpažinimas, panaudojant kalbos atpažinimo programą „Dragon Naturally Speaking“. Aptariamieji tyrimai bei pateikiami gauti rezultatai.

Raktiniai žodžiai: kalbos technologijos, adaptavimas, kalbos atpažinimas, kalbos sintezė.

1. Įvadas

Kalbos technologijos yra viena iš galingiausių technologijų, kurią dvidešimt pirmajame šimtmečiuje sukūrė pasiekimai moksle. Tokių spartų balso technologijų vystymąsi lėmė paltus šių technologijų pritaikymo sferų spektras.

Daugelis kalbos technologijų produktų yra pritaikyti anglų kalbai, todėl aktualu iširti anglų kalbai pritaikytų produktų adaptavimo lietuvių kalbai galimybes.

Mažom kalbom, tokiom kaip lietuvių, reikia arba kurti savo kalbai skirtus atpažinimo ir sintezės modelius, arba stengtis panaudoti kitom kalbom sukurtus modelius atitinkamai juos adaptuojant. Eilė tokių adaptavimo modelių buvo sukurta slovėnų, vengrų, slovaku, danų kalboms [1], tačiau šiuose tyrimuose nebuvo bandyta pritaikyti adaptavimo lietuvių kalbai. P.Kasparaitis, iš Vilniaus universiteto, yra pasiūlęs pradinę lietuviškų fonetinių vienetų konvertavimo sistemą ir tai vienintelis toks atliktas darbas šioje srityje.

Šio straipsnio tikslas yra parodyti, kaip efektyviai pasirinkta kalbos atpažinimo programa „Dragon Naturally Speaking“ atpažįsta lietuviškai kalbančio diktoriaus diktuojamus žodžius ir, ar įmanoma anglų kalba veikiančioje sintezės programoje „ReadPlease 2003“, lietuvišką žodį parašyti tokiomis anglų kalbos raidėmis, kad žodis skambėtų pagal lietuvišką tartį. Atliektais tyrimais nustatytas transkribuotų ir netranskribuotų lietuviškų žodžių aiškumo tikslumas ir lietuviškai kalbančio diktoriaus žodžių atpažinimo efektyvumas.

Straipsnis organizuotas sekančiai: antrame skyrelyje pateikiamos svarbiausios kalbos technologijų grupės ir taikymo sritys, trečiame – atlikti tyrimai, ketvirtame – tyrimų išvados.

2. Svarbiausios kalbos technologijos

Šnekamosios kalbos technologijos yra skiriamos į tris grupes [2]:

1. balsu tariamų vienetų (žodžiai, jų sekos, frazės) automatinis nustatymas arba kalbos atpažinimas;
2. teksto skaitymas balsu arba kalbos sintezė;
3. kitos balso technologijos (asmens tapatybės vertinimas pagal jo balsą, kalbos signalų suspaudimas bei kodavimas, triukšmų slopinimas ir pan.).

Žmogaus bendravimo su kompiuteriu balsu pagrindas yra šnekos atpažinimas ir sintezė.

Balsu tariamų vienetų (žodžių, jų sekų, frazių) automatinio nustatymo arba kalbos atpažinimo paskirtis yra automatiškai nustatyti, kas balsu sakoma informacijos priėmimo sistemai [3]. Tai gali būti atskiras žodis (balso komanda), žodžių seka (PIN kodas), net rišlių sakinių skaitymas. Informacijos priėmimo sistema, nustačiusi, kas jai buvo pasakyta, atlieka adekvačius veiksmus, t.y. atpažinimas yra informacijos priėmimo forma. Kalbos atpažinimas prasideda nuo garso sklaidimo metu ore susidariusių bangų fiksavimo, tų bangų konvertavimo į fonemas - bazinius kalbos vienetus, žodžių formavimo iš gautų fonemų ir žodžių konteksto analizės, kad būtų galima užtikrinti teisingą panašiai skambančių žodžių tarimą (pvz. Kasti ir Kašti). Šiuo metu populiariausias kalbos signalų atpažinimo modelis – tolydinio tankio paslėptos Markovo grandinės.

Kalbos sinteze vadinamas automatinis balsinio pranešimo generavimas iš pateikto teksto ar kitos simbolių sekos, t.y. tekstu pateiktos informacijos skaitymas balsu [4]. Nekelia abejonių, kad sintezė labai nuo konkrečios kalbos savybių priklausanti kalbos technologijų sritis. Generavimui reikia naudoti konkrečiai kalbai paruoštus sintezės elementus (pastarieji dažnai vadinami sintezės vienetais) bei atsižvelgti į duotos kalbos gramatinės ypatybes (kirčiavimą, prozodines, intonacines savybes ir pan.).

Šnekamosios kalbos technologijų taikymo sritys [2]:

1. kalbinis dialogas su kompiuteriu;
2. balso įrašų stenografavimas;
3. balsinė interneto naršyklė;
4. elektroninių prietaisų valdymas balsu;
5. asmens atpažinimas pagal jo balsą.

Nors šiuo metu kalbos technologijos vystosi ir tobulėja, daugėja jų taikymo sričių, tačiau nėra plačiai prieinamų lietuvių kalbos atpažinimo ir sintezavimo programų. Didžiosios kompanijos pirmiausia orientuojasi į didžiąsias pasaulio kalbas, todėl lietuvių kalba, dėl mažo rinkos potencialo, jas mažai domina.

Viena iš labiausiai išvystytų lietuviškų kalbos sintezės technologijų yra sintezatorius „Aistis“, tačiau jos kokybė nėra pakankama, nes sukurta naudojant ne pačias tobuliausias šiai dienai sintezės technologijas – difonų jungimą naudojant MBROLA algoritmus.

Lietuviškų kalbos atpažinimo produktų taip pat nėra daug. Tai „Raktas“ bei kelių telekomunikacinių paslaugų prototipai naudojantys šnekamosios kalbos atpažinimą. Dėl šių priežasčių yra aktualu iširti kalbos technologijų produktų adaptavimo galimybes komandų atpažinimui.

3. Pasirinktų produktų adaptavimo galimybių tyrimas

Analizuojamos dvi programos, kurios veikia skirtingais principais. Vienai iš jų pateikiamas tekstas ir tada ji perskaito tekstą balsu (sintezės programa), kitai programai pateikiamas balsas ir tada ji padiktuotą balsą pateikia tekstu (atpažinimo programa). Eksperimento tikslas – iširti programų panaudojimo lietuvių kalbai galimybes, t.y. lietuvišką žodį pateikti tokiais anglų kalbos atitikmenimis, kad pateiktas žodis būtų tinkamai ištartas lietuviškai bei nustatyti geriausios transkripcijos taisykles ir įvertinti, kurie garsai, lietuviškai kalbančio diktoriaus, atpažįstami geriausiai; nustatyti žodžių atpažinimo efektyvumus.

3.1 Lietuvių fonetikos transformavimo į anglų kalbos rašybą taisyklių sudarymo galybės

Tyrimas atliktas su programa „ReadPlease 2003“. Tyrime dalyvavo du dalyviai. Pirmasis dalyvis išrinko tokius lietuvių kalbos žodžius, kad būtų panaudotos visos lietuvių abėcėlės raidės, išskyrus balse su nosinėmis ir transkribavo išrinktus žodžius pagal anglų kalbos taisykles. Suvedus žodžius į programą (ir lietuviškus, ir transkribuotus) antrasis dalyvis vertino kiekvieno žodžio aiškumą ir kokybę. Aiškumas vertinamas „+“ arba „-“. Jei antrasis dalyvis suprasedavo programos perskaitytą žodį, skirdavo „+“, priešingu atveju „-“. Žodžio kokybė vertinama balais nuo 0 iki 5. Aukščiausias kokybės įvertinimas - „5“, žemiausias – „0“. Kokybės vertinimas priklausė nuo to kaip programos perskaityti žodžiai atitinka lietuviškąjį tarimą. Lentelėje 1 pateikiami lietuviški ir geriausiai transkribuoti žodžiai.

Lentelė 1. Žodžių aiškumo ir kokybės vertinimas

Žodis	Aiškumas	Kokybė	Žodis	Aiškumas	Kokybė
tupi	+	2	toopi	+	3
ant	-	0	unt	+	4
anukas	-	0	unookas	+	3
bangos	+	1	bungos	+	5
galva	+	1	gulva	+	5
vista	-	0	vishtaa	+	5
himnas	-	0	himnus	+	4
vakcina	+	3	vukcinaa	+	3
dangus	-	0	dungoos	+	3
morka	+	4	morka	+	4
ezys	-	0	ezhees	+	3
cesnakas	-	0	chesnakas	+	5
ungurys	-	0	oongoorees	+	4
fabrikas	+	1	fubrikas	+	4
zyle	-	0	zeele	-	0
instrumentas	+	2	instroomentas	+	3
yla	-	0	eala	+	3
suris	-	0	sooris	+	4
jura	-	0	ura	-	0

Šaltinis: sudaryta autorės.

Iš visų pateiktų žodžių tik vieno žodžio *morka* lietuviška pateiktis buvo perskaityta pakankamai aiškiai. Todėl šis žodis nebuvo transkribuotas. Apskaičiuotas lietuviškų žodžių supratimo tikslumas vos 36%. Siekiant padidinti šį

tikslumą, lietuvių kalbos žodis buvo užrašomas pagal anglų fonetikos taisyklės (t.y. transkribuotas). Apskaičiuotas transkribuotų lietuvių kalbos žodžių aiškumo tikslumas 89,5%. Lyginant transkribuotų ir netranskribuotų žodžių aiškumo tikslumus, matome, kad skirtumas tarp jų yra 53,5%.

Pagal žodžių vertinimo lentelę (Lentelė 1) sudarytos tokios geriausios transkripcijos taisyklės:

Lentelė 2. Geriausios transkripcijos taisyklės

Lietuviškos taisyklės	Geriausia transkripcija	Pavyzdžiai	Pastabos
Priebalsiai:			
ž	zh	ežys	
š	sh	višta	
č	ch	česnakas	
c	-	vakcina	Vietoj c, sakoma s.
j	-	jūra	Vietoj j, sakoma dž.
Balsiai:			
axx	uxx	ant, anūkas	
xax	xux	galva, bangos, dangus	
xxa	xxa	morka	
xxa	xxaa	višta	Kai –a kirčiuotas.
uxx	ooxx	ungurys	
xux	xoox	tupī, dangus, ungurys	
xūx	xoox	sūris	
xyx	xeex	žvyras, ežys	
yxx	eaxx	yla	
ė	-	zylė	Vietoj –ė, sakoma e.

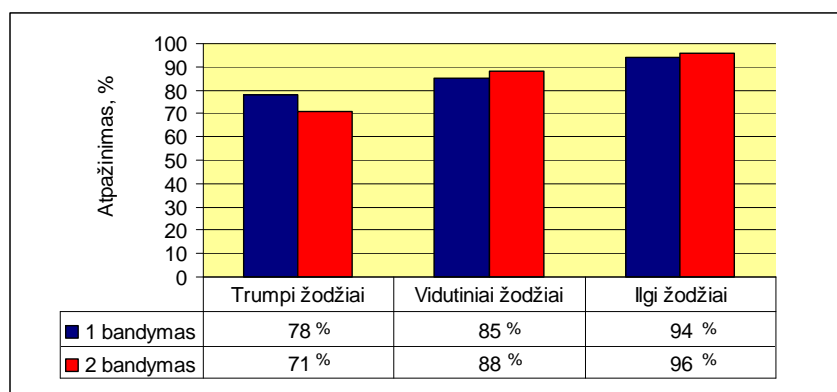
Šaltinis: sudaryta autorės.

Kombinacijos, pavyzdžiui *axx*, reiškia balsės poziciją žodyje. Šiuo atveju balsė *a* yra pirmoji raidė. Pagal nagrinėtus pavyzdžius trumposios balsės *e* ir *i* tarimas lietuviškame žodyje nesiskyrė atitiko nuo transkribuoto. Balsė *o* taip pat atitiko. Visų kitų priebalsių, kurių nėra paminėta 2 lentelėje, tariamas nesiskyrė nuo lietuvių kalbos tarimo.

3.2 Anglišku žodžių atpažinimas

Tyrimas atliktas su programa „Dragon Naturally Speaking“. Tyrime dalyvavo vienas lietuviškai kalbantis diktoriaus, kuris programai diktavo penkiasdešimt anglų kalbos žodžių. Eksperimento metu praktiniu būdu buvo patikrintas programos veikimas, kiekvieną žodį diktuoju po dešimt kartų. Eksperimentas buvo pakartotas antrą kartą.

Diktuojami žodžiai buvo laisvai pasirinkti, tačiau suskirstyti į tris grupes pagal simbolių skaičių: trumpi (iki 3 simbolių), vidutiniai (4 – 5 simboliai) ir ilgi (6 ir daugiau simbolių). Pavyzdžiui, trumpų žodžių grupei priklausė tokie žodžiai kaip *sky*, *big*; vidutinių – *mind*, *silver*; ilgų – *advantage*, *education*.



Paveikslas 1. Atpažinimas (%) kiekvienai žodžių grupei pirmuoju ir antruoju bandymu

Šaltinis: sudaryta autorės.

Nei vienos žodžių grupės programa neatpažino 100% tikslumu. Pirmuoju bandymu trumpus žodžius atpažino 78% tikslumu, vidutiniškus - 85%, ilgus - 94%. Antruoju bandymu trumpus žodžius atpažino 71% tikslumu, vidutiniškus - 88%, ilgus - 96%.

Programa dažniausiai klydo užrašydama panašaus tarimo žodžius (*sun* – *some*, *hat* – *had*). Taip pat daugiausia klaidų darė ties balsių *a* ir *i* trumpinimu bei ilginimu. Klaidos suskirstytos ir pagal balsių fonetinę transkripciją:

Lentelė 3. Balsinių garsų klaidų skaičius

Trankskripcija	Klaidų skaičius
[i:]	2
[æ]	1
[a:]	2
[ʌ]	4
[u:]	1
[ei]	1
[ə]	4
[i]	4
[e]	1
[o]	1

Šaltinis: sudaryta autorės.

Tai gali lemti dvi priežastys: pirma yra tai, kad diktoriaus neteisingai išlaiko reikiamą balsės ilgumą, antrą – žodžiai turintys daugiau baslių yra mažiau suprantami ir ne tokie aiškiai tariami kaip kad turintys daugiau priebalsių, todėl programa jų ir neužrašo teisingai.

Klaidos taip pat buvo suskirstytos ir pagal žodžiuose esančius priebalsius (dusliuosius, skardžiuosius). Su dusliaisiais priebalsiais (*t, s, f*) buvo fiksuota 29% visų klaidų, su skardžiaisiais (*b, h, j, r, n, z, v, l, m, d*) - 71%. Matome, kad daugiau klaidų daroma tuose žodžiuose, kurie turi skardžiųjų priebalsių.

4. Išvados

Atliktais tyrimais buvo nustatytos dviejų programų „ReadPlease 2003“ ir „Dragon Naturally Speaking“ panaudojimo lietuvių kalbai galimybės.

Pirmuoju tyrimu su programa „ReadPlease 2003“, nustatytas lietuviškų žodžių aiškumo tikslumas yra vos 36%. Tuos pačius žodžius transkribavus aiškumo tikslumas gautas 89,5%. Galima teigti, kad lietuvių kalbos žodžiams užrašyti naudojant anglų kalbos fonetinę rašybą ir pateikus užrašytus žodžius programai „ReadPlease 2003“ perskaityti, žodžių aiškumas tapo 53,5% procentais didesnis nei užrašius pagal lietuvių kalbos fonetinę rašybą.

Antruoju tyrimu su programa „Dragon Naturally Speaking“, nustatyta, kad lietuviškai kalbančio diktoriaus žodžių atpažinimo efektyvumas yra 85 - 86%. Geriausiai yra atpažįstami ilgi žodžiai, mažiausiai - trumpi. Daugiausia klaidų yra daroma ties dviejų balsių *a* ir *i* trumpinimu ir ilginimu. Taip pat daugiau klaidų yra daroma tuose žodžiuose, kurie turi skardžiųjų priebalsių daugiau nei dusliųjų. Todėl galima teigti, kad aklaai pasikliauti programa diktuojant žodžius negalima, tačiau prižiūrint ir pataisant neteisingai įvestą žodį, ji gali būti naudojama kaip puiki rašymo priemonė.

Tolimesniu tyrimu bus siekiama tirti garsų transkribavimo taisykles sintezatoriui ir jomis naudojantis sudaryti valdymo komandų rinkinį programai „Dragon Naturally Speaking“, kuris būtų efektyviausias lietuviškai kalbančiam diktoriui. Taip pat bus tikslinamas lietuviškai kalbančio diktoriaus žodžių atpažinimo efektyvumas, išplečiant diktorių skaičių.

Literatūra

- [1] **Andrej Zgank, Zdravko Kacic, Frank Diehl, Jozef Juhar, Slavomir Lihan, Klara Vicsi, Gyorgy Szaszak.** Graphemes as Basic Units for CrosslingualSpeech Recognition (2005), In ASIDE-2005, paper 24.
- [2] **A. Rudžionis.** Balso technologijų taikymo lietuvių kalbai analizė ir perspektyvinių veiklos kryptių pagrindimas (2001) [interaktyvus]. [žiūrėta 2008 m. birželio 1 d.]. Prieiga per internetą: http://www.likit.lt/frames/balzo_tech/balsotech_st.htm.
- [3] **A.Rudžionis.** Pagrindinių kalbos signalų technologijų plėtros ypatumai: atpažinimas. Garso korta – 2001, Kaunas, Technologija, 2001.
- [4] **A.Rudžionis.** Pagrindinių kalbos signalų technologijų plėtros ypatumai: sintezė. Informacinės technologijos'2001, Kaunas, Technologija, 2001, 460 - 464 p.