

NEPRIKLAUSOMAS NUO KALBĖTOJO BALSO KOMANDŲ ATPAŽINIMO TYRIMAS

Daniel Bartoševič, Juozas Kamarauskas

Vilniaus Gedimino technikos universitetas

El. paštas: hellraise@gmail.com, juozas.kamarauskas@gmail.com

Santrauka. Straipsnyje aprašoma balso komandų atpažinimo tyrimas, panaudojant lietuvių kalbai sukurtą programą, taip pat kalbos atpažinimo sistemos struktūra, jos pagrindiniai elementai ir veikimo principai, sistemoje naudojamų akustinių modelių savybės. Sukurtas naujas 50 žodžių žodynas kuris buvo pritaikytas „Žodžių atpažintuvas“ programoje. Atlikti dviejų tipo eksperimentai nustatant atpažintuvo darbingumą. Pirmas, kai yra naudojamas 100 žodžių žodynas kuris platinamas kartu su programa ir antras naujai sukurtas. Rezultatai yra pateikti grafiškai ir palyginti.

Reikšminiai žodžiai: komandų atpažinimas, akustiniai modeliai, atpažintuvo darbingumas.

Ivadas

Lietuvių kalbos atpažinimui yra sukurtos dviejų tipų programos. Pirmoji programa „Atpažinimas“ yra grįsta dinaminio laiko skalės kraipymu ir priklauso atpažinimo, atliekant nežinomo ištartimo palyginimą su etalonais, grupei. Ši programa yra priklausoma nuo kalbėtojo. Antroji programa „Žodžių atpažintuvas“ yra nepriklausoma nuo kalbėtojo ir atpažinimui naudoja paslėptuosius Markovo modelius (Lipeika A., J. Lipeikienė 2006).

Programa „Žodžių atpažintuvas“ skirta automatinio atskirai tariamų žodžių atpažinimo analizei. Išskirtinė programos savybė yra jos gebėjimas vaizdžiai pateikti informaciją apie visą žodžių atpažinimo proceso eigą. „Žodžių atpažintuvas“ yra mažo žodyno (100 žodžių) paslėptais Markovo modeliais grįsta nepriklausoma nuo kalbėtojo atpažinimo sistema, kurioje naudojami nuo konteksto nepriklausomi fonemų (fonetinių vienetų, garsų) akustiniai modeliai.

Reikia pažymėti, kad programa „Žodžių atpažintuvas“ yra tik pradinėje kūrimo stadijoje ir vis dar yra tobulinama. Pradiniame jos variante dar nėra realizuotos visos joje numatytos funkcijos ir savybės.

Kalbos atpažinimo sistemos struktūra

Statistiniais šnekos atpažinimo metodais besiremiančių ir smulkesniais nei žodis kalbos vienetais grįstų kalbos atpažinimo sistemų struktūros yra bendros visoms kalboms. Atpažinimo sistemos struktūra pateikiama 1 paveiksle. Ši struktūra atspindi ištisinės šnekos atpažintuvą. Pagrindiniai atpažinimo sistemos elementai yra: požymių analizė,

akustinių ir kalbos modelių paruošimas – mokymas, žodžio ir sakinio atpažinimas.

Pateiktoje atpažinimo struktūroje reikia išskirti išteklius – akstinius ir kalbos modelius. Jeigu sistemos struktūra yra bendra visoms kalboms, tai akustinių ir kalbos modelių turinys yra priklausomi nuo konkrečios kalbos ir jos specifikos. Šie ištekliai kiekvienai kalbai yra ruošiami atskirai žmonių, turinčių žinių apie tą kalbą. Akstiniai ir kalbos modeliai yra paruošiami atskirai ir vėliau gali būti naudojami keliose ASA (automatinio šnekos atpažinimo) sistemose.

1 pav. Kalbos atpažinimo sistemos struktūra

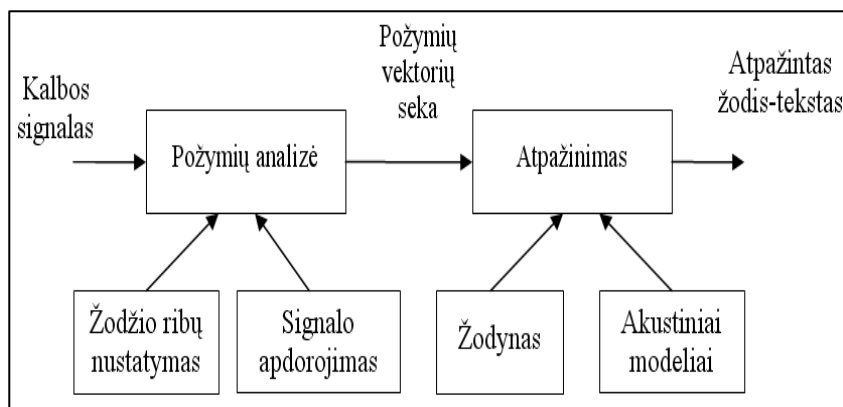


Fig. 1. Speech recognition system architecture

Kai kalbos signalas patenka į atpažinimo sistemą, t.y. kai vartotojas į mikrofoną pasako žodžius, tarp kurių turi būti ~1–2 s pauzės, automatiškai nustatomos žodžio ribos. Kalbos signalas dalijamas į 10–30 ms trukmės kadrus su 5–15 ms žingsniu tarp gretimų kadrų. Tada skaičiuojamas kiekvieno kadro normuotas signalo intensyvumo įvertis, tekste vadinamas energijos įverčiu. Nustačius, kad iš anksto parinktame gretimų kadrų skaičiuje įvertis yra didesnis už

tam tikrą iš anksto parinktą slenkstį, pirmojo kadro pradžioje fiksuojama žodžio pradžia. Tada analogiškai nustatoma žodžio pabaiga, tik šiuo atveju tikrinama, ar įvertis yra mažesnis už pasirinktą slenkstį.

Nustačius žodžio ribas, apdorojamas kalbos signalas, reprezentuojantis žodį. Atpažinimo sistemoje realizuoti standartiniai kalbos signalo apdorojimo ir požymių išskyrimo metodai (Rabiner, L. R.; Juang, B. H. 1993). Pirmiausia iš kalbos signalo pašalinama nuolatinė dedamoji – iš signalo atskaitų reikšmių atimamas jų vidurkis. Tada signalas dalijamas į 25 ms trukmės kadrus su 10 ms žingsniu tarp gretimų kadrų. Kiekvieno kadro signalas filtruojamas pirmos eilės ribotos impulsinės reakcijos filtru ir dauginamas iš Hemingo lango funkcijos. Sistemoje naudojamo filtro koeficientas – 0,97.

Vėliau atliekama kalbos signalo analizė – išskiriami akustiniai požymiai. Gaunama požymių vektorių seka, kitaip vadinama stebėjimų arba akustinių vektorių seka. Šioje atpažinimo sistemoje naudojami kalbos suvokimu, kai imituojamas žmogaus klausos aparato veikimas, grįsti požymiai – melų dažnių skalės kepsriniai koeficientai. Gauta seka pateikiama atpažinti ištartą žodį.

Kai atliekamas pavienių žodžių atpažinimas, kiekvieną k -ąjį duoto žodyno žodį atitinka atskiras modelis M_k . Žodžių atpažinimas atliekamas pagal formulę:

$$j = \arg \max_k P(M_k | O); \quad (1)$$

Čia M_k – k -ojo žodžio modelis, $O = \{o_1, o_2, \dots, o_T\}$ – kalbos signalo požymių vektorių seka, j – atpažinto žodžio modelio, su kuriuo gauta didžiausia aposteriorinė tikimybė $P(M_k | O)$, indeksas.

Aposteriorinė tikimybė $P(M_k | O)$ skaičiuojama naudojant Bejeso formulę:

$$P(M_k | O) = \frac{P(O | M_k) P(M_k)}{P(O)}; \quad (2)$$

čia $P(O | M_k)$ – tikėtumas, kad požymių vektorių seką O generavo modelis M_k , $P(M_k)$ – apriorinė modelio M_k tikimybė, $P(O)$ – požymių vektorių sekos O tikimybė. Vardiklyje esantis dydis $P(O)$ yra pastovus visiems modeliams ir neturi įtakos formulės (2) rezultatui, todėl jį tolesnius skaičiavimus neįtraukiamas. Apriorinė tikimybė

$P(M_k)$ vadinama kalbos modeliu, o tikėtumas $P(O | M_k)$ – akustiniu modeliu. Kadangi šioje sistemoje kalbos modelis nenaudojamas, t. y. sistemoje visų modelių apriorinės tikimybės vienodos, žodžio atpažinimas sistemoje priklauso tik nuo tikėtinumo $P(O | M_k)$, kuris apskaičiuojamas naudojant Viterbi algoritmą.

Sistema skaičiuoja kiekvieno akustinio modelio tikėtinumą, kad gauta seka buvo generuota to modelio, įvertį ir iš žodyno pateikia kaip atpažintą tą žodį, kurį reprezentuoja modelis, turintis didžiausią tikėtinumo įvertį.

Naujų žodžių, neesančių žodyne, sudarymas atpažinimui

Žodynas atspindi kalbos atpažinimo sistemos naudojamų žodžių sąrašą. Žodynas pateikia jame esančių žodžių transkripcijas (nuosekli kalbos vienetų seka), sudaromas tik iš apibrėžtos kalbos vienetų aibės elementų. Jis tuo pačiu atspindi ir akustinių modelių derinius, naudojamus žodžio atpažinime. Žodynas naudojamas tiek mokymo, tiek atpažinimo etapuose. Kalbos atpažinimo sistemų žodynai formuojami jau pagal esančius kalbos transkripcijų žodynus skaitmeniniame formate.

Žodyne esančios transkripcijos yra nepriklausomos nuo akustinių duomenų, kai pastarieji neįtakoja transkripcijų sudarymo. Tuo būdu atsiribojama nuo situacijos, kai analizuojant leksinį kintamumą reikia atsižvelgti ne tik į kalbos vienetų aibes, bet ir į atskiros žodžio tarimo ypatumus, atsispindinčius akustikoje.

Pirmas žingsnis norint sukurti naują žodį, reikia jį atitinkamai paversti į fonetinę transkripciją suprantamą programai. Žodžių transkripcijose ženklų /' / žymimas priebalsio minkštumas, ženklu /:/ – balsio ilgumas, skaičiais /0/, /1/ ir /2/ – atitinkamai kairinis, rištinis ir dešininis kirčiai. Pasinaudojant lietuvių kalbos žodynu (A. Kalėda, B. Kalėdienė, M. Niedzviecka 1991) ir garsyno LRN fonetinė sistema (Sigita Laurinčiukaitė 2008) yra sukuriami ir užrašomi žodžiai, pavyzdžiui: būti → b u:2 t' i, asmuo → a s m uo:l, kalbėti → k a l' b' e:2 t' i ir t.t. Tokiu būdu buvo sudarytas 50 žodžių žodynas, kuris yra pateiktas žemiau su atitinkamomis fonetinėmis transkripcijomis.

Antras žingsnis, visi naujai sukurti žodžiai yra užrašomi į .txt failą, esanti programos kataloge su „dictionary“ pavadinimu. Reikia atkreipti dėmesį, kad žodyno pavadinimas turi būti būtent „dictionary“, esant kitam pavadinimui bus rodoma klaida ir programa neveiks. Taip pat lietuviškas raides vartoti nepatartina, juos atitinkamai užkoduojamos: A-A1, Č-C2, Ė-E3, Į-I1, Š-S2, Ū-U4, Ž-Z2.

Kitas žingsnis, atlikti atitinkamus eksperimentus ir įsitikinti koks maksimalus tikslumas gali būti pasiektas.

1 lentelė. Naujai sudaryti žodžiai su atitinkama fonetinė transkripcija

Nr.	Žodis	Transkripcija			
1.	degalai	d'e g a l a i j'	26.	kainoraštis	k a i j' n o: 2 r a s' t' i s
2.	kategorija	k a t' e g o 0 r' i j' a	27.	knyga	k' n' i: g a 0
3.	nuotaika	n u: 2 o t a i j' k a	28.	krantas	k r a n l t a s
4.	pasiūla	p a s' u l a 0	29.	kursas	k u 0 r s a s
5.	romanas	r o m a: l n a s	30.	laikrodis	l a i: j' k r o d' i s
6.	sąveika	s a: 2 v' e i j' k a	31.	magistras	m a g' i 0 s t r a s
7.	sąvoka	s a: 2 v o k a	32.	metalas	m' e t a: l l a s
8.	skirtumas	s' k' i 0 r t u m a s	33.	natūralus	n a t u r a l u 0 s
9.	tvirtumas	t' v' i r t u 0 m a s	34.	pagroti	p a g r o: 2 t' i
10.	uždaryti	u z 2 d a r' i: 2 t' i	35.	pokalbis	p o: 2 k a l' b' i s
11.	vairuoti	v a i j' r u: 2 o t' i	36.	sekundė	s' e k u 0 n' d' e
12.	mąstyti	m a s' t' i: 2 t' i	37.	remontas	r e m o 0 n t a s
13.	aplinka	a p' l' i n k a 0	38.	alga	a l g a 0
14.	bauda	b a u d a 0	39.	baltas	b a: 2 l t a s
15.	blokuoti	b l o k u: 2 o t' i	40.	eglutė	e g l u 0 t e
16.	dėstytojas	d' e: 2 s' t' i: t o j' a s	41.	butas	b u 0 t a s
17.	detalus	d' e t a l u 0 s	42.	gamta	g a m t a 0
18.	druska	d r u s k a 0	43.	idėja	i d e: 2 j' a
19.	duona	d u: 2 o n a	44.	jėga	j' e g a 0
20.	elektra	e l' e k t r a 0	45.	dėkoti	d e k o: 2 t' i
21.	šeima	s e i m a 0	46.	išmokti	i s m o: 2 k' t' i
22.	galaktika	a l a: l k' t' i k a	47.	kambarys	k a m b a r i s
23.	granitas	g r a n' i 0 t a s	48.	laivas	l a i v a s
24.	įkvėpimas	i k' v' e p' i m a s	49.	literatūra	l i t e r a t' u r a 0
25.	įsakymas	i s a: l k' i: m a s	50.	motyvas	m o t' i v a s

Eksperimentas, kai atpažintuve naudojamas 100 žodžių žodynas

Šiame eksperimente yra naudojamas 100 žodžių žodynas kuris platinamas kartu su programa „Žodžių atpažintuvas“.

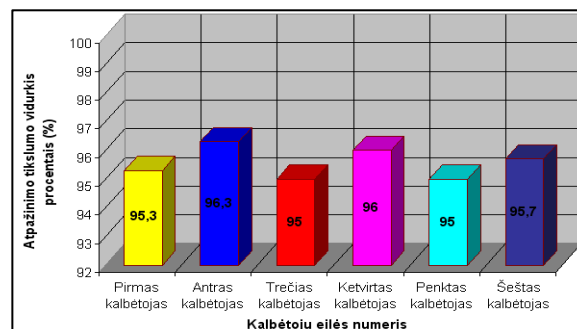
Atpažintuvo darbingumas yra vertinamas pagal žodžių atpažinimo tikslumą (T), skaičiuojamą pagal formulę (Lipeika A., J. Lipeikienė 2006):

$$T = 100\% - K, \text{ čia } K = \frac{D + S + I}{N} \cdot 100\%, \quad (3)$$

kur K – žodžių atpažinimo klaida, D – praleistų, S – pakeistų, I – įterptų žodžių skaičius, N – bendras atpažintamų žodžių skaičius testinėje imtyje.

Eksperimente dalyvauja 6 žmonės, 3 vyrai ir 3 moterys įvairaus amžiaus. Kiekvienas iš kalbėtojų ištaria po 3 sesijas, kurios kiekviena sudaryta iš 100 žodžių. Sesijos sudarymui buvo naudojama Cool Edit Pro 1.0 programa. Garso failas turi būti WAV formato su atitinkamais parametrais: diskretizacijos dažnis 11025 Hz, vienas įrašymo kanalas (mono), 16 bitų kvantavimo tikslumas.

Iš gautu rezultatų galime pasakyti, kad tikslumas svyruoja nuo 95% iki 96,3%. Kiekvienos sesijos metu buvo neatpažinta nuo 3 iki 6 žodžių. Dažniausiai pasitaikančios klaidos buvo tos, kurių žodžių skambėjimas buvo labai panašus. Aukščiausias tikslumas tarp vyrų siekė 96,3%, tarp moterų 96%. Išvada, kad tikslumas nepriklauso nuo žmogaus, amžiaus, bei lyties. Iš 1800 žodžių buvo neatpažinta 80, kas sudaro vidutinį atpažintuvo tikslumą apie 95,5%. Gauti rezultatai yra pateikti grafiškai 2 pav.



2 pav. Eksperimentas, kai atpažintuve naudojamas 100 žodžių žodynas dalyvaujant šešiems kalbėtojams

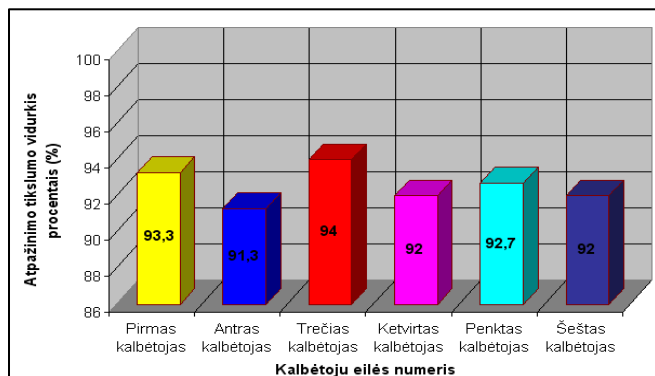
Fig. 2. Experiment, the recognition device used in the presence of 100 words in the vocabulary of six speakers

Eksperimentas, kai atpažintuve naudojamas 50 žodžių žodynas

Šiame eksperimente yra naudojamas naujai sukurtas 50 žodžių žodynas. Jis turi būti patalpintas į programos katalogą su pavadinimu „dictionary“. Eksperimente dalyvauja tie patys 6 žmonės, 3 vyrai ir 3 moterys, įvairaus amžiaus. Kiekvienas iš kalbėtojų ištaria po 3 sesijas, kurios kiekviena sudaryta iš 50 žodžių. Garso failas turi būti WAV formato su atitinkamais parametrais: diskretizacijos dažnis 11025 Hz, vienas įrašymo kanalas (mono), atskaitos koduotos 16 bitų. Tikslumas yra vertinamas pagal (3) formulę.

Iš gautu rezultatų matome, kad tikslumas svyruoja nuo 91,3% iki 94%. Kiekvienos sesijos metu buvo neatpažinta nuo 2 iki 5 žodžių. Atpažinime yra nemažai žodžių, kuriuos atpažindavo blogai, nepriklausomai nuo

kalbėtojo, pavyzdžiui: įkvėpimas, idėja, išmokti ir t.t. Taip pat yra daug žodžių, kurie buvo atpažinti labai gerai, pavyzdžiui: nuotaika, sąveika, kambarys ir t.t. Iš 900 žodžių buvo neatpažinta 67, kas sudaro vidutini atpažintuvo tikslumą apie 92,6%. Gauti rezultatai yra pateikti grafiškai 3 pav.

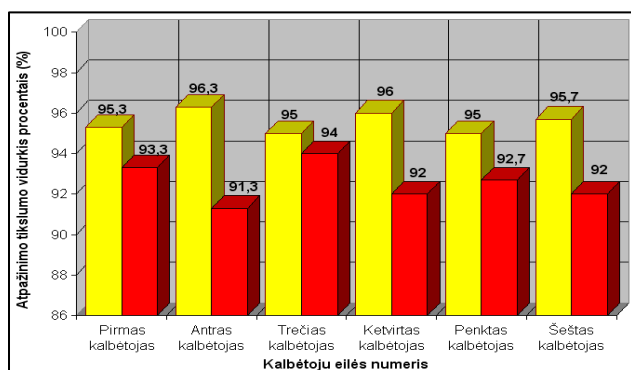


3 pav. Eksperimentas, kai atpažintuve naudojamas 50 žodžių žodynas dalyvaujant šešiams kalbėtojams

Fig. 3. Experiment, the recognition device used in the presence of 50 vocabulary words for six speakers

Eksperimentų išvados

Atlikus dviejų tipo eksperimentus galime padaryti tokias išvadas: žodyną kuri sudaro 100 žodžių tikslumas siekia apie 95,5%, tuo tarpu 50 žodžių žodyno tikslumas yra apie 92,6%. Skirtumas siekia 2,9%. Viena iš priežasčių kodėl taip yra, naujai sukurti žodžiai yra ilgesni ir sunkiau tariami, kas būtent įtakoja tikslumą. Rezultatų palyginimai pateikti 4 pav.



4 pav. Rezultatų palyginimas esant 100 ir 50 žodžių žodynams

Fig. 4. Comparison of results at 100 and 50-word dictionaries

Kitas svarbus dalykas kodėl nei pirmu nei antru atveju nebuvo pasiektas 100%, kadangi kiekvieno žmogaus balsas yra unikalus, todėl tai lemia, kad gauti tokie rezultatai. Be to galima pastebėti, kad atpažinimas tiek vyrų tiek moterų yra labai panašus, skirtumas siekia vos 1-2%, tai reiškia, kad programa yra visiškai nepriklausoma nuo kalbėtojo.

“Žodžių atpažintuvas” yra paslėptais Markovo modeliais grįsta atpažinimo sistema, kurioje naudojami nuo konteksto nepriklausomi fonemų (fonetinių vienetų, garsų) akustiniai modeliai. Tokia sistema sėkmingai veikia esant nedideliems žodynams (100 ar 50 žodžių). Esant didesnės apimties žodynams programos tikslumas ir efektyvumas prastės. Kitas dalykas, naujų žodynų sudarymas reikalauja daug darbo, kuo daugiau atliekama bandymų tuo didesnę tikslumą galima pasiekti.

Išvados

Visas balso atpažinimo programas galime suskirstyti į dvi grupes: priklausomos ir nepriklausomos nuo kalbėtojo. Visos nepriklausomos nuo kalbėtojo programos, remiasi vienu metodu - paslėptais Markovo modeliais.

Viena iš tokių yra „Žodžių atpažintuvas“ programa sukurta Matematikos ir informatikos institute.

Sukūrus naują 50 žodžių žodyną ir atlikus eksperimentus, gautas tikslumas siekia apie 92,6%, tai yra prastesnis rezultatas nei su 100 žodžių žodynu. Kadangi žodyne yra panaudoti ilgesni ir sunkiau tariami žodžiai kas būtent ir lemia tokį tikslumą. Taip pat dėl programos netobulumo kai kurių žodžių aptikimo ribos buvo klaidingos kas įtakojo į rezultatus.

Literatūra

- Lipeika A., J. Lipeikienė 2006. *Paslėptais Markovo modeliais grįsto atskirai tariamų žodžių atpažintuvo prototipas „Žodžių atpažintuvas“* P. 2-6
- Rabiner, L. R.; Juang, B. H. 1993. *Fundamentals of speech recognition*. Prentice-Hall, Inc, 496p.
- A. Kalėda, B.Kalėdienė, M. Niedzviecka 1991 *“Lietuvių – lenkų kalbos žodynas“*.
- Sigita Laurinčiukaitė 2008 *“Lietuvių šnekos atpažinimo akustinis modeliavimas“*. A Priedas. Garsyno LRN fonetinė sistema. P. 123-127
- Lipeika A., J. Lipeikienė 2006. *„At2006_IV_Ketv_3_Atpažintuvo Prototipas_061204.“* P. 19-29“

INVESTIGATION OF SPEAKER-INDEPENDENT VOICE COMMAND RECOGNITION

D. Bartoševič, J. Kamarauskas

Abstract

The article describes the Lithuanian language program designed for voice command recognition. Speech recognition system architecture, its main elements and principles of operation, the system used in the acoustic properties of models. A new dictionary of 50 words that have been adapted to “Žodžių atpažintuvas” program. Carry out two experiments in determining the type recognition device performance. First, when you use the 100 word vocabulary which is distributed with the program and the second newly created. Results are presented graphically and compared.

Keywords: command recognition, acoustic models, recognition device performance.