

LIETUVIŲ KALBOS PAVIENIŲ ŽODŽIŲ ATPAŽINIMO ALGORITMO
IGYVENDINIMAS LAUKU PROGRAMUOJAMA LOGINE MATRICATomyslav Sledevič¹, Liudas Stašionis²

Vilniaus Gedimino technikos universitetas

El. paštas: ¹tomyslav.sledevic@vgtu.lt, ²liudas.stasionis@dok.vgtu.lt

Santrauka. Pateikiamas lietuvių kalbos pavienių žodžių atpažinimo algoritmo įgyvendinimas lauku programuojama loginė matrica (LPLM). LPLM įrenginys pasirinktas dėl lygiagrečiai veikiančių procesų įgyvendinimo galimybės taikant VHDL kalbą. Tai užtikrina spartų signalų apdorojimą esant taktiniam dažniui iki 50 MHz. Kalbos požymiams išskirti taikoma keps-trinė šnekos analizė. Požymiams palyginti taikomas dinaminis laiko skalės kraipymo (DSLK) metodas. Sudaryta 100 žodžių požymių biblioteka, kuri saugoma vidinėje LPLM BRAM atmintyje. Pasiiektas 94 % atpažinimo tikslumas priklausomai nuo kalbėtojo ir 58 % – nepriklausomai nuo kalbėtojo. Kepstro koeficientų skaičiavimas vienam žodžiui trunka 8,52 ms, esant 50 MHz taktiniam dažniui, ir šimtui DSLK – 66,56 ms, esant 25 MHz taktiniam dažniui.

Reikšminiai žodžiai: lauku programuojama loginė matrica, žodžio atpažinimas, keps-tras, dinaminis laiko skalės kraipymas.

Įvadas

Kalba – lengviausias ir greičiausias žmonių bendravimo ir informacijos keitimosi būdas. Todėl taip pat nesunkiai norima komunikuoti su kompiuterine technika ir kitais elektroniniais įtaisais, kurie per pastaruosius kelis dešimtmečius sparčiai tobulėjo ir vis dažniau tapo naudojami buityje. Greitas prietaisų valdymas kalba ypač palengvintų kasdienį žmonių gyvenimą, todėl daug šių dienų tyrėjų sprendžia žodžių atpažinimo tikslumo ir greitaveikos uždavinius tobulindami esamus kalbos atpažinimo metodus. Pagrindinės problemos, dėl kurių nėra sukurta patikimo metodo kalbai atpažinti, yra aplinkos poveikis kalbos signalui, kalbos kintamumas, netaisyklinga tartis, ribotas žodyno dydis, panašios fonetinės transkripcijos žodžiai (Tamulevičius 2008).

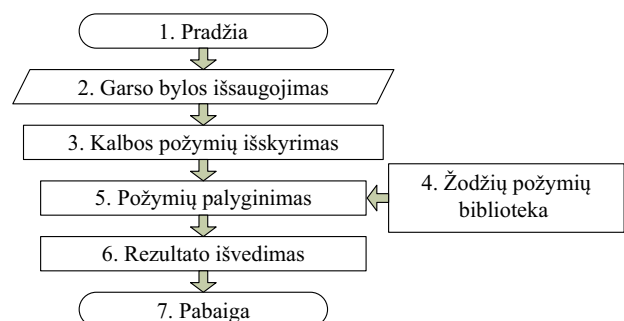
Kalbos požymiams išskirti dažniausiai taikomi metodai: spektrinė analizė, keps-trinė analizė (Tamulevičius *et al.* 2010), tiesinės prognozės kodavimas, melų-dažnių keps-triniai koeficientai. Kalbos požymiams klasifikuoti taikomi šie metodai: DSLK (Tamulevičius *et al.* 2010; Sart *et al.* 2010), paslėptieji Markovo modeliai (Veitch *et al.* 2011), dirbtinių neuronų tinklai (Amudha *et al.* 2008).

Vykdamas kalbos atpažinimo procesus atliekama daug aritmetinių, loginių, kreipimosi į atmintinę operacijų. Algoritmus perkeliama iš kompiuterių į įterptines sistemas mažinamas taktinis dažnis. Tokiu atveju vienintelis būdas išlaikyti greitaveiką yra atpažinimo procesų lygiagretinimas. Dėl galimybės lygiagrečiai apdoroti signalus esant taktiniam dažniui iki 100 MHz per pastaruosius 4 metus

tyrėjai vis dažniau kalbos atpažinimo algoritmus įgyvendina LPLM (Tamulevičius *et al.* 2010; Sart *et al.* 2010; Veitch *et al.* 2011; Amudha *et al.* 2008). Vienos loginės matricos (lusto) taikymas jos dydžio ir energijos suvartojimo (dėl sumažėjusio dažnio) atžvilgiu yra pranašesnis, palyginti su kelių branduolių procesoriais.

Pavienių žodžių atpažinimo algoritmas

Atsižvelgiant į anksčiau gautus rezultatus ir suformuluotas išvadas (Tamulevičius *et al.* 2010) siūlomas algoritmas (1 pav.) įgyvendintas taikant VHDL programavimo kalbą tik aparatinėmis priemonėmis. Įgyvendinant požymių biblioteka optimizuota iki 100 žodžių ir paspartintas DSLK metodas.



1 pav. Pavienių žodžių atpažinimo algoritmas

Fig. 1. Isolated word recognition algorithm

Igyvendintą pavienių žodžių atpažinimo algoritmą sudaro tokios dalys:

1. *Pradžia*. Aktyvuota RS-232 sąsaja. Atmintinė paruošta įrašymui. Laukiama iš kompiuterio garso bylos.
2. *Garso bylos išsaugojimas*. Viena *.wav formato garso byla iš kompiuterio per RS-232 sąsają siunčiama į LPLM vidinę atmintinę 115 200 bit/s greičiu. Kalbos signalas diskretizuotas 11 025 Hz dažniu, kvantuotas 256 lygiais. Pavienio žodžio trukmė – iki 1,5 s.
3. *Kalbos požymių išskyrimas*. Paspaudus mygtuką pradamas garso bylos skaitymas. Kalbos signalas suskirstomas į kadrus po 256 imtis, kurioms taikomas *Hanning* filtras. Kiekvieno kito kadro pradžia pastumta per 128 imtis pirmo atžvilgiu. 128 kalbos signalo kadrų sudaro pavienį žodį. Kalbos požymių vektoriumi laikomi 12 pirmųjų kepstro koeficientų, kurie kiekvienam kadru skaičiuojami pagal lygtį:

$$C = \text{real}(FFT(\log_2 |FFT(X)|)), \quad (1)$$

čia C – kepstro koeficientų vektorius; X – signalo vektorius; FFT – greitoji Furjė transformacija.

4. *Žodžių požymių biblioteka*. 100 žinomų žodžių požymių saugomi vidinėje LPLM atmintinėje. Biblioteka į LPLM įkeliama ją programuojant. Bet kurio žodžio požymių vektorius bibliotekoje gali būti pakeistas neperprogramuojant LPLM.

5. *Požymių palyginimas*. DLSK metodas taikomas norint rasti požymių vektorius, kurių sutapdinimo klaida mažiausia. Šiam tikslui pasiekti skaičiuojama 128×128 dydžio klaidų matrica ir ieškoma mažiausios klaidos kelio nuo pabaigos taško (127,127) iki pradžios taško (0,0). Klaidos matrica užpildoma taikant Euklido atstumo lygtį:

$$d = \sqrt{\sum_{i=1}^{12} (c_b(i) - c(i))^2}, \quad (2)$$

čia $c(i)$ – i -tasis kepstro koeficientas; $c_b(i)$ – i -tasis bibliotekoje saugomo kepstro koeficientas; d – Euklido atstumas tarp požymių vektorių.

6. *Rezultatų išvedimas*. Atpažinto žodžio numeris ir jo sutapdinimo klaida parodomi LCD monitoriuje. Pakoreguota arba naujai sudaryta požymių biblioteka gali būti nusiųsta į kompiuterį tam, kad kito LPLM programavimo metu būtų įkelta pagal numatytuosius parametrus.

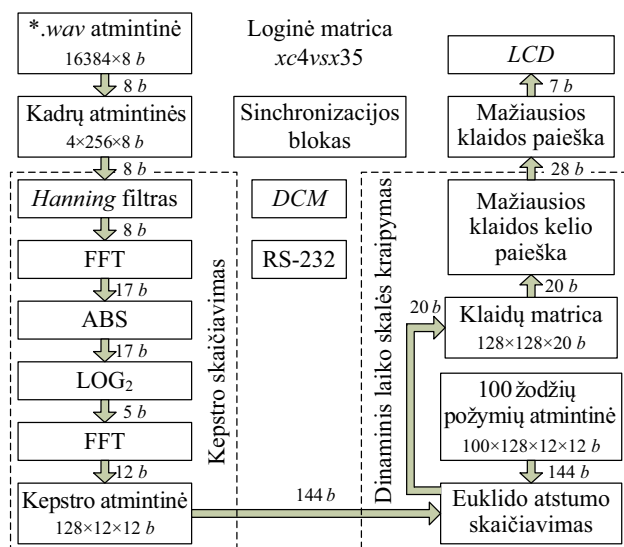
7. *Pabaiga*. Kol LPLM lieka įjungtas, laukiama kitos garso bylos įvedimo.

Algoritmo įgyvendinimas LPLM

Pavienių žodžių atpažinimo algoritmas įgyvendintas vienoje LPLM taikant VHDL programavimo kalbą (2 pav.).

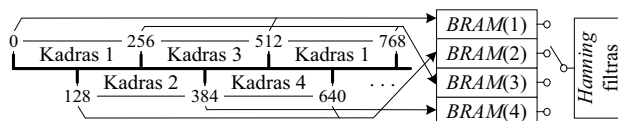
Algoritme naudojami moduliai pažymėti stačiakampiais blokais, įėjimo ir išėjimo duomenų magistralių pločiai nurodyti šalia rodyklių. Kiekvienas modulis yra sujungtas su sinchronizacijos bloku siekiant užtikrinti sinchronišką duomenų iš atmintinių skaitymą ir apdorojimą. Taikant programų paketą *Matlab* 8 bitų kalbos signalas siunčiamas į *.wav atmintinės bloką, kuriame saugomos 16 384 signalo vertės. Šios vertės įrašomos į vienas kitą iš dalies dengiančių kadrų atmintines BRAM (1–4) (3 pav.). Sinchronizacijos blokas nurodo, kuriame kadre esantis kalbos signalas apdorojamas *Hanning* filtru.

Kepstrui skaičiuoti taikomi du greitosios Furjė transformacijos moduliai FFT (Xilinx 2011), spektro absoliučiosios vertės skaičiavimo modulis ABS , logaritmo modulis LOG_2 ir kepstro atmintinės modulis, kuriame saugomi 1,5 s trukmės kalbos signalo pirmi 12 kepstro koeficientų. Šie koeficientai lyginami su atmintinėje saugomais požymiais skaičiuojant Euklido atstumus ir užpildant sutapdinimo klaidų matricą. Klaidų matricoje ieškoma mažiausios klaidos kelio (4 pav.). DLSK vykdomas 100 kartų, t. y.



2 pav. Pavienių žodžių atpažinimo algoritmas, įgyvendintas aparatinėmis priemonėmis

Fig. 2. Hardware implementation of isolated word recognition algorithm



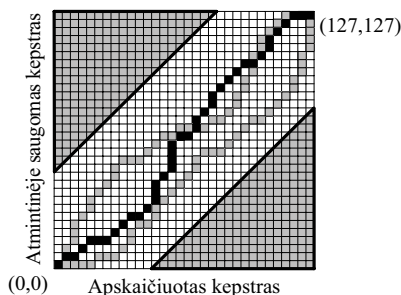
3 pav. Rašymo į kadrų atmintines diagrama

Fig. 3. Writing to frame memories diagram

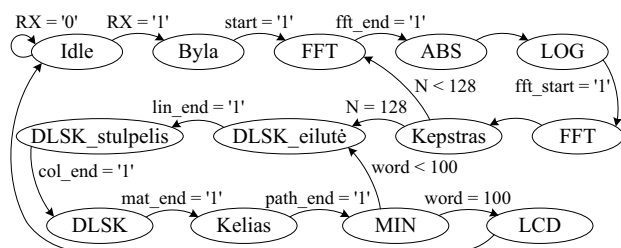
apskaičiuotas kepstras palyginamas su kiekvieno žodžio kepstru, saugomu požymių atmintyje. Žodis, kurio kelio klaida mažiausia tarp 100 tikrintų, laikomas atpažintu ir jo numeris išvedamas į LPLM makete esantį LCD monitorių.

Pilkos spalvos trikampiai klaidų matricos kampuose (4 pav.) žymi sritis, kuriose Euklido atstumų skaičiuoti nėra būtina, jeigu žinoma, kad atpažįstamo žodžio dinaminis iškraipymas laike yra minimalus, palyginti su žodžiu, kurio požymiai saugomi bibliotekoje. Kuo mažesnis dinaminis iškraipymas, tuo didesni gali būti trikampiai ir tuo mažesnę sutapdinimo klaidų sritį matricoje galima pildyti. Šiame darbe klaidų matricos apkarpymai netaikomi.

Sinchronizacijos bloke (2 pav.), kartu su kitais procesais veikia baigtinių būvių mašina, kuri aktyvuoja pavienius modulius, kontroliuoja kadrų sudarymą, atmintinių adresavimą (5 pav.). Idle būseną yra tol, kol negaunama duomenų iš kompiuterio per RS-232 sąsają. Kol *.wav byla yra įrašoma į vidinę LPLM atmintinę, yra būseną Byla. Paspaudus mygtuką maketo plokštėje pradedamas atpažinimo procesas. Taikant *Radix-2* algoritimą skaičiuojama greitoji Furjė transformacija. Kepstro koeficientai 1,5 s trukmės kalbos signalui skaičiuojami 128 kartus. Vėliau įjungiamas DLSK algoritmas, užpildoma pirmoji klaidų matricos eilutė, stulpelis ir likę matricos nariai. Tada pradeda mažiausios klaidos kelio paieška. Esant būsenai MIN nuosekliai lyginamos 100 žodžių kelių klaidos. Baigus palyginimą, minimalią klaidą atitinkančio žodžio numeris išvedamas į LCD ir grįžtama į Idle būseną, kuriai esant laukiama kito žodžio.



4 pav. Mažiausios klaidos kelio paieška
Fig. 4. Searching of minimum error path



5 pav. Žodžio atpažinimo sistemos baigtinių būvių mašina
Fig. 5. Finite state machine of word recognition system

LPLM įgyvendintas kepstro koeficiento skaičiavimo algoritmas veikia 50 MHz dažniu, DLSK – 25 MHz dažniu. Įgyvendintos algoritmų struktūros neleidžia padidinti jų taktinių dažnių $xc4v3x35$ loginėje matricoje.

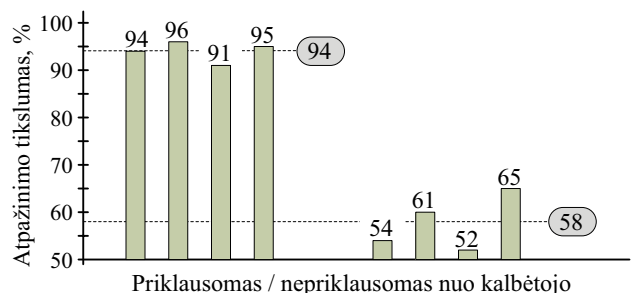
Rezultatai

Lietuvių kalbos žodžių atpažinimo sistemos tikslumas tikrinamas atpažįstant pavienius žodžius priklausomai ir nepriklausomai nuo kalbėtojo (6 pav.). Priklausomo nuo kalbėtojo (PK) testo metu LPLM vidinėje atmintinėje išsaugoti 100 žodžių požymių vektoriai lyginami su 100 to paties žmogaus ištartų žodžių kepstrų koeficientais. Eksperimentas atliktas keturis kartus (400 žodžių). Pasiektas vidutinis 94 % atpažinimo tikslumas. Nepriklausomo nuo kalbėtojo (NK) testo metu naudojama tų pačių žodžių požymių biblioteka, tačiau atpažinti taikomi kito žmogaus ištartų žodžiai. Eksperimentas atliktas naudojant keturių žmonių ištartus žodžių rinkinius (4×100 žodžių). Pasiektas vidutinis 58 % atpažinimo tikslumas.

PK eksperimento metu pastebėta, kad klaidingai atpažįstami panašios fonetinės transkripcijos žodžiai: norėti → galėti, metas → metai, bendras → bankas, diena → vienas, laikas → vaikas, vyriausybė → valstybė. Todėl atpažinimo tikslumui didinti patartina požymių bibliotekos sudarymui taikyti fonetiškai skirtingus žodžius. NK atpažinimo tikslumas mažėja dėl žmonių kalbos akcentų, balso tembrų skirtumų.

Pavieniam iki 1,5 s trukmės žodžiui atpažinti, taikant 100 žodžių požymių biblioteką, reikia per 2 mln. taktinių impulsų, iš kurių 20 % – tai kepstrų koeficientų skaičiavimas ir 80 % – DLSK (1 lentelė). Bendra vieno žodžio atpažinimo trukmė – kiek daugiau negu 75 ms.

Tik aparatinėmis priemonėmis įgyvendintos atpažinimo sistemos tikslumas, dažnis ir trukmė palyginti su sistemos, kuri įgyvendinta iš dalies aparatinėmis ir programinėmis priemonėmis (2 lentelė). Pasiecta 262 kartus trumpesnė požymių išskyrimo trukmė, taktiniam dažniui



6 pav. Priklausomas ir nepriklausomas nuo kalbėtojo pavienio žodžio atpažinimo tikslumas, kai $f_d = 11\ 025$ Hz
Fig. 6. Speaker dependent and independent isolated word recognition rate at $f_d = 11\ 025$ Hz

esant du kartus mažesniai ir 145 kartus trumpesnė palyginimo trukmė (taikant DLSK), taktinį dažnį sumažinus keturis kartus. 4 % padidėjo priklausomo ir 5,3 % sumažėjo nepriklausomo nuo kalbėtojo atpažinimo tikslumas.

Esamos žodžių atpažinimo sistemos įgyvendinimas naudoja beveik perpus mažiau matricos loginės atmintinės (34,7 %), 7 % mažiau DSP celių ir trigubai daugiau blokinių RAM atmintinės, nes joje saugomi 100 žodžių požymių vektoriai (3 lentelė). Likusią laisvą LPLM atmintinės dalį planuojama panaudoti DLSK moduliui dubliuoti siekiant apie du kartus sumažinti požymių palyginimo laiką.

1 lentelė. Signalų apdorojimo trukmė

Table 1. Signal processing time

Modulis	Taktinių impulsų skaičius	Trukmė, ms	Taktinis dažnis, MHz
Kepstras	426 240 (20 %)	8,52	50
DLSK	1 664 000 (80 %)	66,56	25
Iš viso	2 090 240 (100 %)	75,08	

2 lentelė. Atpažinimo sistemų įgyvendinimo palyginimas, kai kalbos signalas diskretizuotas 11 025 Hz dažniu

Table 2. Comparison of recognition systems implementation, when speech signal sampled at 11 025 Hz

Atpažinimo etapas	Trukmė, ms	Dažnis, MHz	Atpažinimo tikslumas, %
Esamas įgyvendinimas (100 žodžių biblioteka)			
Požymių išskyrimas	8,52	50	PK = 94,0
Palyginimas	66,56	25	NK = 58,0
Buves įgyvendinimas (600 žodžių biblioteka)			
Požymių išskyrimas	2238	100	PK = 90,0
Palyginimas	57 952	100	NK = 63,3
Atpažinimo proceso paspartinimas (kartais)			
Požymių išskyrimas	262	–	PK = +4,0
Palyginimas	145	–	NK = -5,3

3 lentelė. LPLM aparatinių išteklių sunaudojimas

Table 3. FPGA hardware utilization

Modulis	Slices	BRAMs	DSP48s
Esamas įgyvendinimas			
Kepstras	1454 (9,5 %)	22 (11,5 %)	9 (4,7 %)
DLSK	2129 (13,8 %)	118 (61,5 %)	12 (6,3 %)
Kiti	1757 (11,4 %)	0 (0,0 %)	0 (0,0 %)
Iš viso	5340 (34,7 %)	140 (73,0 %)	21 (11,0 %)
Buves įgyvendinimas			
Iš viso	10 187 (66,0 %)	46 (23,0 %)	35 (18,0 %)

Išvados

1. Tik aparatinėmis priemonėmis įgyvendintas pavienių žodžių atpažinimo algoritmas 4 % tiksliau atpažįsta žodžius priklausomai nuo kalbėtojo ir 5,3 % blogiau atpažįsta žodžius nepriklausomai nuo kalbėtojo. Šie skirtumai yra dėl kalbos signalo kvantavimo lygių sumažėjimo ir nevienodų bibliotekų dydžių. Mažesnė

biblioteka be dubliuojančių žodžių požymių pagerino to paties kalbėtojo ištartų žodžių atpažinimo tikslumą.

2. Kepstro koeficientų skaičiavimas paspartintas 262 kartus, DLSK – 145 kartus. Vieno žodžio atpažinimas 100 žodžių žodyne vykdomas per 75,08 ms.
3. DLSK užtrunka ilgiausiai, todėl paspartinimui siūloma ateityje didinti DLSK modulių skaičių, taktinį dažnį, riboti klaidos matricos dydį, mažinti kalbos signalo diskretizacijos dažnį. Atpažinimo tikslumui gerinti reikia didinti signalo kvantavimo bitų skaičių.

Padėka

Darbas remiamas Lietuvos mokslo tarybos (sutarties Nr. MIP-092/2012).

Literatūra

- Amudha, N., *et al.* 2008. FPGA implementation of isolated digit recognition system using modified back propagation algorithm, in *Int. Conference on Electronic Design*: 1–6.
- Sart, D., *et al.* 2010. Accelerating dynamic time warping subsequence search with GPUs and FPGAs, in *Proceedings of the International Conference on Data Mining*, 1001–1006.
- Tamulevičius, G. 2008. *Pavienių žodžių atpažinimo sistemų kūrimas*: daktaro disertacija. 150 p. VGTU/MII.
- Tamulevičius, G., *et al.* 2010. Hardware accelerated FPGA implementation of Lithuanian isolated word recognition system, *Electronics and Electrical Engineering* 99(3): 57–62.
- Veitch, R., *et al.* 2011. FPGA Implementation of a pipelined gaussian calculation for HMM-Based large vocabulary speech recognition, *International Journal of Reconfigurable Computing*, 1–10. <http://dx.doi.org/10.1155/2011/697080>
- Xilinx. Fast Fourier Transform. 2011. Prieiga per internetą: http://www.xilinx.com/support/documentation/ip_documentation/xfft_ds260.pdf

FPGA-BASED IMPLEMENTATION OF LITHUANIAN ISOLATED WORD RECOGNITION ALGORITHM

T. Sledevič, L. Stašionis

Abstract

The paper describes the FPGA-based implementation of Lithuanian isolated word recognition algorithm. FPGA is selected for parallel process implementation using VHDL to ensure fast signal processing at low rate clock signal. Cepstrum analysis was applied to features extraction in voice. The dynamic time warping algorithm was used to compare the vectors of cepstrum coefficients. A library of 100 words features was created and stored in the internal FPGA BRAM memory. Experimental testing with speaker dependent records demonstrated the recognition rate of 94%. The recognition rate of 58% was achieved for speaker-independent records. Calculation of cepstrum coefficients lasted for 8.52 ms at 50 MHz clock, while 100 DTWs took 66.56 ms at 25 MHz clock.

Keywords: field programmable gate array, word recognition, cepstrum, dynamic time warping.