

ENERGIJOS DETEKTORIAUS, NAUDOJAMO ŽODŽIO RIBOMS NUSTATYTI,  
ĮGYVENDINIMAS LAUKU PROGRAMUOJAMA LOGINE MATRICALiudas Stašionis<sup>1</sup>, Tomyslav Sledevič<sup>2</sup>

Vilniaus Gedimino technikos universitetas

El. paštas: <sup>1</sup>liudas.stasionis@dok.vgtu.lt, <sup>2</sup>tomyslav.sledevic@vgtu.lt

**Santrauka.** Pateikiamas žodžio ribų nustatymo modulis įgyvendinimas lauku programuojama loginė matrica (LPLM). Žodžio riboms nustatyti pasirinktas energijos detektorius, nes šis metodas, naudojant skaitmenines signalų apdorojimo priemones, įgyvendinamas efektyviai. Žodžio ribų nustatymo modulis buvo optimizuotas tiek, kad, įgyvendintas LPLM, jis užėmė 54 loginius elementus „Slice“ – tik 0,7 % „Spartan-6 LX45“ lusto išteklių. Eksperimentuojant nustatyta, kad esant 20 dB ir 15 dB signalo triukšmo santykiui, žodžio ribos nustatomos tiksliai, o kai šis santykis yra 10 dB ir 5 dB, žodžio ribos nustatomos 95 % ir 93 % tikslumu.

**Reikšminiai žodžiai:** lauku programuojama loginė matrica, žodžio ribų nustatymas, energijos detektorius, tylusis intervalas, signalo triukšmo santykis.

## Įvadas

Žodžio ribų nustatymo modulis – svarbi šių dienų kalbos signalo apdorojimo sistemos dalis. Žodžio ribas reikia nustatyti prieš atliekant kalbos atpažinimą, gryninimą ir kodavimą.

Kalbos ribų detektorius turi nustatyti, kada sistemos įėjime yra kalbos komponentės ir kada jų nėra. Ši uždavinį apsunkina tai, kad kalboje yra tylieji intervalai (netariamas žodis) ir vokalizautos pauzės (pauzė užpildyta mažos amplitudės kalbos komponentėmis) (Brotherton 1979). Šie kalbos intervalai negali būti maišomi, nes, priešingu atveju, riba gali būti nustatyta žodžio viduryje.

Žodžio riboms nustatyti galima taikyti įvairius metodus, tokius kaip: spektro lėkštumo nustatymas (Stejskal *et al.* 2010), ilgalaikio spektro divergavimo nustatymas (Ramirez 2004), daugiasluoksnio perceptrono naudojimas (Gevaert 2010), statistinio sutapatavimo testas (Ramirez 2005) ir energijos lygio nustatymas (Sakhnov *et al.* 2009). Įgyvendinamiems metodams taikomi keli kriterijai:

- Žodžio ribų nustatymo modulis yra tik tam tikros sistemos dalis, todėl šiai funkcijai negalima skirti daug aparatūros išteklių.
- Metodas turi būti pritaikytas aparatūrai (įterptinei sistemai), kurioje sistema bus įgyvendinta.

Pirmasis kriterijus yra pagrindinis veiksnys, kuris riboja sudėtingų žodžio ribos detektorių įgyvendinimą. Spektro analize, daugiasluoksnio perceptrono naudojimu ar statistinių sutapatavimo metodų taikymu grįsti metodai

sunkiai įgyvendinami, nes, pavyzdžiui, norint nustatyti signalo spektro parametrus, reikia atlikti transformaciją, o apmokant perceptroną (kaip ir taikant statistinius metodus) reikia atlikti daug aritmetinių operacijų. Šioms operacijoms atlikti reikia daug aparatūros išteklių. Įvertinus šiuos metodų įgyvendinimo ypatumus, žodžio ribų nustatymo moduliui įgyvendinti pasirinktas energijos detektorius.

Šis žodžio ribų nustatymo metodas nesunkiai įgyvendinamas LPLM, nes ši priemonė tinkama srautiniam duomenų apdorojimui ir gali atlikti kelis veiksmus vienu metu. Šiuo metu aparatūroje naudoti LPLM yra ekonomiškai naudinga, nes mažos galios LPLM (pvz., *Xilinx – Spartan* ir *Altera – Cyclone* šeimos) kainos susilygina su tos pačios klasės skaitmeninių signalų apdorojimo procesorių kainomis.

## Energijos lygio nustatymo modulis optimizavimas įgyvendinant LPLM

Energijos lygio nustatymo modulis įgyvendintas pagal trumpalaikio galingumo išraišką

$$E = \frac{1}{N} \sum_{n=0}^{N-1} x_n^2, \quad (1)$$

čia  $N$  – lango dydis, kuriame skaičiuojama energija;  $x_n$  –  $n$ -toji įėjimo signalo imtis. Pagal klasikinę energijos lygio nustatymo išraišką skaičiavimams atlikti reikia dviejų daugybos ir sumavimo elementų. Daugyba būtų naudojama kėlimo laipsniu ir dalybos veiksmams.

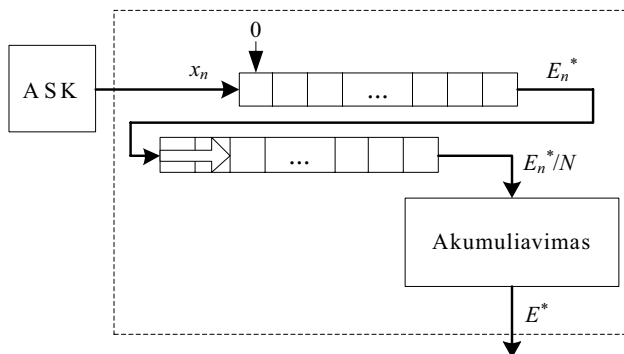
Įgyvendinant šiuos veiksmus LPLM, sumavimui reikia – atminties, o daugybai – specialių skaitmeninio signalo apdorojimo langelių (Xilinx LPLM atveju, tai DSP48 celės). Problema kyla įgyvendinant daugybos veiksmus, nes specialių langelių LPLM yra nedaug (Xilinx Spartan-6 atveju svyruoja nuo 8 iki 180, priklausomai nuo lusto galimumo). Todėl daugybos veiksmus reikia optimizuoti taip, kad specialių langelių šiam moduliui įgyvendinti nereikėtų.

Viena daugyba naudojama kėlimo laipsniu operacijai. Šį veiksmą galima pakeisti labai paprasta reikšminių bitų manipuliacija. LPLM testiniame modulyje yra integruotas LM4550 daugiakanalis garso apdorojimo modulis, kuriame yra analogo ir skaičiaus keitikliai (ASK ir SAK). LM4550 ASK išėjime gaunamas 18 bitų *signed* tipo signalas, kurio pirmasis bitas žymi įėjimo signalo poliarumą (1 pav.). Jei pirmoje skiltyje yra vienetas – tai ASK įėjime yra neigiamo poliarumo signalas, jei nulis – priešingai. Kėlimo laipsniu operacija keičiama reikšminio bito pakeitimo į nulį veiksmu (2 pav.). Tokiu būdu neigiamo poliarumo signalo dalis pakeičiama į priešingą, o teigiama dalis lieka nepakitusi.

Signalio būseną	Signalio kodas
+	0000...0001
0	0000...0000 1000...0000
-	1000...0001

1 pav. Signed duomenų tipo bitų vaizdas  
Fig. 1. Signed data type bit representation

Antra daugyba naudojama dalybos veiksmui įgyvendinti – šį veiksmą galima keisti bitų postūmiu į kairę. Postūmis per  $N$  bitų atitinka dalybą iš  $2^N$ , todėl naudojamas langas trumpalaikiam galimumui apskaičiuoti turi atitikti dalmenį.



2 pav. Trumpalaikio galimumo nustatymo modulis  
Fig. 2. Short-term power calculation module

Sumavimo operacija atliekama naudojant akumuliavimo elementą. Dalybos veiksmas atliekamas prieš sumuojant, nes tokiu būdu galima išvengti akumuliavimo elemento persipildymo.

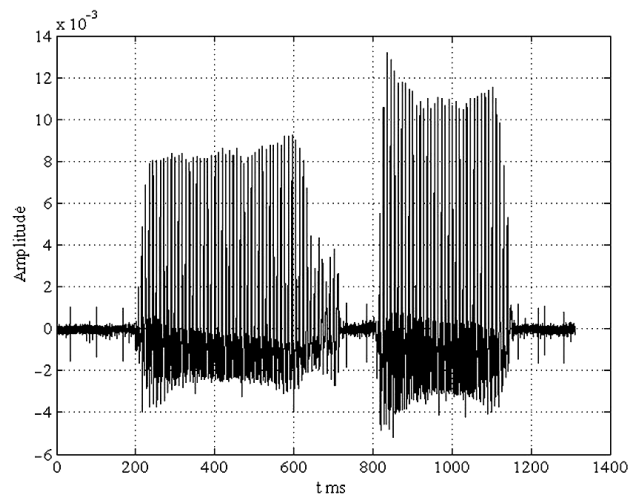
Gaunamas trumpalaikio galimumo įvertis  $E^*$  tiksliai neimituoja (1) išraiškos, tačiau naudojant šiuos veiksmus sutaupomi du DSP48 specialieji langeliai.

### Žodžio pauzės problema

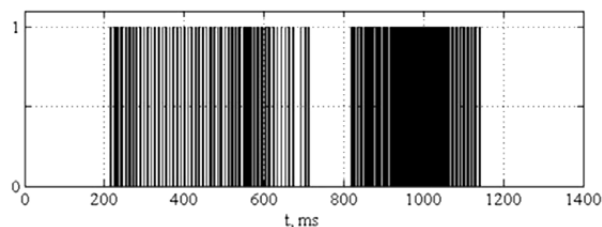
Vien trumpalaikio galimumo nustatymo modulis neužtenka žodžio riboms nustatyti, nes tariamuosiuose žodžiuose yra tylieji intervalai, kurie gali būti palaikyti žodžio pabaiga. Šios mažesnės energijos dalys susidaro dėl žodyje vartojamų priebalsių, pavyzdžiui, žodyje „darbas“.

Pavyzdiniame žodyje signalo amplitudė akivaizdžiai susilpnėja (3 pav.), tai trunka 100 ms. Bandant nustatyti šio žodžio ribas, taikant tik trumpalaikės galios metodą, gaunamas rezultatas, pateiktas 4 pav.

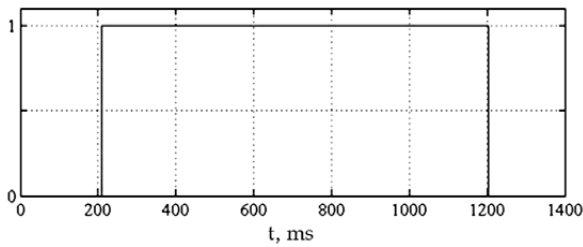
Iš gautų rezultatų matyti, kad detektorius ne vien neteisingai nustatė žodžio ribas 100 ms pauzės metu, bet kly-



3 pav. Žodžio „darbas“ laikinė diagrama  
Fig. 3. Word “darbas” time diagram



4 pav. Žodžio „darbas“ ribų nustatymo rezultatai, naudojant trumpalaikio galimumo modulį  
Fig. 4. Word “darbas” boundary detection results by using Short-term power module



5 pav. Žodžio „darbas“ ribų nustatymo rezultatai naudojant trumpalaikio galingumo modulį su skaitikliu

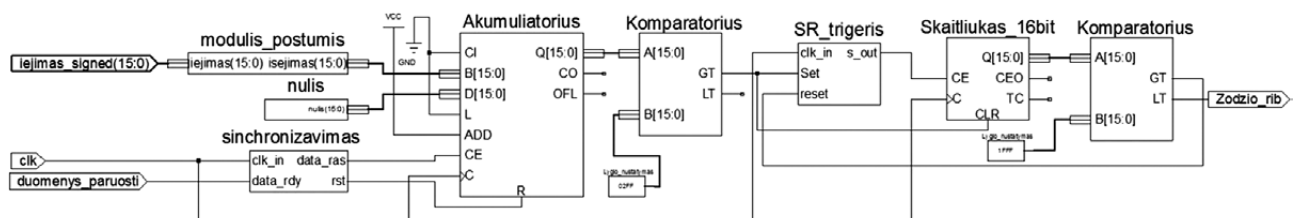
Fig. 5. Word “darbas” boundary detection results by using Short-term power module with counter

do ir kitose žodžio dalyse. Keičiant žodžio ribos nustatymo slenkstį, geresnių rezultatų nepasiekta. Ši problema sprendžiama naudojant skaitiklį, kuris perkraunamas energijos detektoriaus. Detektoriumi nustatoma, kad įėjime signalas viršijo nustatytą lygį, skaitiklis aktyvuojamas ir perkraunamas tol, kol įėjime signalo amplitudė viršija nustatytą lygį. Kai skaitiklio vertė mažesnė nei nustatyta, generuojamas aptikto žodžio impulsas. Skaitikliui viršijus nustatytą vertę, jis išjungiamas ir pažymima žodžio pabaiga.

Naudojant energijos detektorius kartu su skaitikliu žodžio ribos nustatytos teisingai (5 pav.). Detektorius nesuklydo tiek žodžio pauzės srityje, tiek kitose dalyse.

### Žodžio ribų nustatymo modulio įgyvendinimas LPLM

Žodžių ribų nustatymo modulis turi tris įėjimo signalus (6 pav.): *iejimas\_signed(15:0)*, *clk* ir *duomenys\_paruosti*. Pirmas įėjimas yra skaitmeninis garso signalas, gaunamas iš ASK, kurio pirmasis bitas žymi poliarumą. Kiti du įėjimai skirti sinchronizuoti: *clk* skirtas blokams sinchronizuoti laike, o *duomenys\_paruosti* žymi naujos skaitmeninio garso signalo imties pasirodymą. Esant *clk* signalo impulso kylančiam frontui, atliekami tokie veiksmai: akumuliacija, palyginimo, būsenos išstatymo ir inkrementavimo. Šio signalo impulsų pasirodymo dažnis turi būti daugiau nei keturis kartus didesnis nei duomenų iš ASK pasirodymo dažnis. Priešingu atveju veiksmai bus atliekami nekorektiškai.



6 pav. Žodžio ribų nustatymo modulio programos schema

Fig. 6. Word boundary detection module program schematic

Pagal *duomenys\_paruosti* įėjimo signalo kylantį frontą sinchronizavimo modulis aktyvuoja ir perkrauna akumuliacijos elementą. Sinchronizavimo blokas, pasirodžius naujiems duomenims iš ASK, aktyvuoja kaupimo operaciją. Įėjimo signalui pasiekus ribinę energijos detektoriaus apdorojimo langą, akumuliacijos elementas perkraunamas.

Gautas skaitmeninis signalas iš ASK *iejimas\_signed(15:0)* apdorojamas *modulis\_postumis* bloke, kuris kartu su akumuliacijos elementu skaičiuoja trumpalaikį galingumą. Gautą trumpalaikio galingumo vertę komparatorius lygina su nustatyta lygiu ir, jei nustatyta vertė viršijama, *GT* (angl. *Greater Than*) išėjime generuojamas loginis vienetas. Sugeneruotas loginis vienetas reiškia, kad aptiktos žodžio komponentės. Komparatoriaus sugeneruotas loginis vienetas perkrauna ir aktyvuoja skaitiklį, kuris inkrementuojasi ties kylančiu *clk* frontu. Toliau, kol skaitiklio vertė neviršija nustatytos, generuojamas žodžio buvimo požymis (loginis vienetas) išėjime *Zodzio\_rib*. Jei skaitiklio vertė viršija nustatytą, minėtame išėjime generuojamas loginis nulis ir skaitiklis išjungiamas.

Įgyvendinus šį energijos detektorius „Spartan-6 LX45“ LPLM buvo panaudoti tik 54 konfigūruojami loginiai blokai „Slice“ – tik 0,7 % sistemos išteklių.

### Eksperimentinis žodžio ribų nustatymo modulio tyrimas

Norint nustatyti žodžių ribų nustatymo modulio efektyvumą, buvo parinkta 20 įvairių pavienių žodžių ir 10 sakinių įrašų. Šie įrašai buvo paveikti Gauso normuoto skirstinio atsitiktinio triukšmo. Bandymų metu buvo keičiamas S/T santykis nuo 20 dB iki -10 dB, mažinant santykį kas 5 dB. Įrašai apdoroti realiuoju laiku. Apdorojimas atliktas naudojantis „ATLYS“ bandomuoju LPLM moduliu, kuriame integruotas „Spartan-6“ lustas.

Prieš paveikiant garso įrašus triukšmu, juose buvo matuojama žodžių trukmė. Tuo tikslu naudotas programiniu paketu „Audacity“. Bandomųjų žodžių trukmė išmatuota 1 ms tikslumu. Tai atlikta norint objektyviai nustatyti san-

tykinį sistemos tikslumą, esant įvairiems signalo ir triukšmo santykiams (S/T), kai lyginama sistemos nustatyta žodžio trukmė su išmatuota naudojant programinį paketą.

Rezultatų lentelėje (1 lentelė) pateikti aritmetiniai santykiųjų tikslumų vidurkiai. Vidutinės vertės išvestos iš visų gautų rezultatų, kurie gauti apdorojant visus įrašus ties atitinkamu S/T.

1 lentelė. Žodžių ribų prie skirtingų S/T santykiųjų santykinis tikslumas

Table 1. Word boundary detection results using different S/N ratios

S/T santykis	Santykinis žodžio ribų nustatymo tikslumas
20 dB	100 %
15 dB	100 %
10 dB	95 %
5 dB	93 %
0 dB	71 %
-5 dB	65 %
-10 dB	58 %

Esant 20 dB ir 15 dB S/T santykiams, žodžio ribų nustatymo tikslumas siekė 100 %. Šiais atvejais buvo tiksliai nustatyta tiek žodžio pradžia, tiek pabaiga.

Esant S/T santykiams 10 dB ir 5 dB, buvo pasiektas atitinkamai 95 % ir 93 % žodžio ribų nustatymo tikslumas. Šio tikslumo pakanka pavienių žodžių atpažintuvui. Beveik visuose įrašuose detektorius neteisingai nustatė žodžio pradžią. Tai nulemta to, kad žodžio pradžioje yra mažesnės amplitudės komponentės, kurios užslopinamos padidėjusio triukšmo signalo.

Prastas žodžio ribų nustatymo tikslumas pasiektas esant S/T santykiui 0 dB ir -5 dB. Abejais atvejais blogai nustatyta tiek žodžio pradžia, tiek pabaiga, nes šiose srityse dažniausiai vyrauja mažesnės amplitudės kalbos komponentės. Esant -5 dB S/T santykiui, detektorius dažnai nustatydavo žodžio pabaigą tyliuose kalbos intervaluose, kurių problema aptarta anksčiau.

Itin prasti rezultatai gauti, kai S/T santykis siekė -10 dB. Pagrindinė prasto tikslumo priežastis – beveik visos kalbos signalo komponentės užgožiamos triukšmo. Šiuo atveju buvo blogai nustatyta tiek žodžio pradžia, tiek pabaiga. Tyliuose intervaluose klaidingai nustatyta žodžio pabaiga.

## Išvados

LPLM modulyje galima įgyvendinti žodžio ribų nustatymo modulį, gebantį:

- 100 % tikslumu nustatyti žodžio ribas, kai S/T santykis siekia 20 dB ir 15 dB;
- 95 % ir 93 % tikslumu nustatyti žodžio ribas, kai S/T santykis siekia 10 dB ir 5 dB.

Šis žodžio ribų nustatymo modulis gali būti įgyvendintas LPLM naudojant 54 konfigūruojamus loginius elementus (0,7 % „Spartan-6 LX45“ lusto išteklių).

Iš gautų rezultatų matyti, kad šį detektorių tikslinga naudoti tada, kai S/T santykis nėra žemesnis nei 5 dB.

## Padėka

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

## Literatūra

- Brotherton, P. 1979. Speaking and not speaking: process for translating ideas into speech, in *Conference on Time and Speech*: 179–209.
- Gevaert, W. 2010. Neural networks used for speech recognition, *Journal of Automatic Control* 20: 1–7. <http://dx.doi.org/10.2298/JAC1001001G>
- Ramirez, J. 2004. Improved voice activity detection combining noise reduction and subband divergence measures, in *International Conference on Spoken Language Processing*.
- Ramirez, J. 2005. Statistical voice activity detection using a multiple observation likelihood ratio test, *IEEE Signal Processing Letters* 12(10): 689–692. <http://dx.doi.org/10.1109/LSP.2005.855551>
- Sakhnov, K., et al. 2009. Approach for energy-based voice detector with adaptive scaling factor, *International Journal of Computer Science* 36: 4–10.
- Stejskal, V., et al. 2010. Voice activity detection based on discriminative weight training incorporating a spectral flatness measure, *Circuits Syst Signal Process*: 183–194.

## ENERGY DETECTOR IMPLEMENTATION IN FPGA FOR ESTIMATION OF WORD BOUNDARIES

L. Stašionis, T. Sledevič

### Abstract

This paper describes implementation of the word boundary estimation module in FPGA. The boundary estimation module is based on energy detector. This module is optimized for implementation in FPGA. It occupies 54 logical elements “Slice” and uses only 0.7% of “Spartan-6 LX45” resources. Experiments with this module were performed at different signal/noise (S/N) ratio. For S/N of 20 dB and 15 dB word boundaries were estimated with 100% accuracy. Acceptable results were also achieved, for S/N ratio of 10 dB and 5 dB, as the estimation accuracy was 95% and 93%, respectively.

**Keywords:** field programmable gate array, word boundary determination, energy detector, silent interval, signal noise ratio.