

**KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIOS FAKULTETAS
KOMPIUTERIŲ KATEDRA**

Donatas Duchovskis

**Aukštesnių eilių statistika grįsto balso
detektavimo algoritmo sudarymas ir tyrimas**

Magistro darbas

**Vadovas
prof. E. Kazanavičius**

KAUNAS, 2006

**KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIOS FAKULTETAS
KOMPIUTERIŲ KATEDRA**

**TVIRTINU
Katedros vedėjas
prof. E. Kazanavičius
2006-05-26**

**Aukštesnių eilių statistika grįsto balso
detektavimo algoritmo sudarymas ir tyrimas**

Informatikos magistro baigiamasis darbas

**Kalbos konsultantė
Lietuvių kalbos katedros lektorė
Dr. I. Mickienė
2006-05-12**

**Vadovas
Prof. E. Kazanavičius
2006-05-26**

**Recenzentas
Doc. P. Kanapeckas
2006-05-26**

**Atliko
D. Duchovskis
2006-05-26**

KAUNAS, 2006

KVALIFIKACINĖ KOMISIJA

Pirmininkas: Laimutis Telksnys, akademikas

Sekretorius: Stasys Maciulevičius, docentas

Nariai: Rimantas Barauskas, profesorius

Raimundas Jasinevičius, profesorius

Jonas Mockus, akademikas

Rimantas Plėštys, docentas

Henrikas Pranevičius, profesorius

Turinys

| | |
|---|-----------|
| PAVEIKSLŲ SĄRAŠAS..... | 6 |
| LENTELIŲ SĄRAŠAS..... | 7 |
| PRATARMĖ | 8 |
| SUMMARY | 9 |
| 1. ĮVADAS | 10 |
| 2. BALSŲ DETEKCIJOS UŽDAVINIO ANALIZĖ..... | 12 |
| 2.1. BALSŲ DETEKCIJA IR TAIKYMAS | 12 |
| 2.2. BALSŲ DETEKTORIAUS ĮVERTINIMAS | 13 |
| 2.3. TRIUKŠMŲ ANALIZĖ | 16 |
| 2.3.1. <i>Automobilio skleidžiamas triukšmas.....</i> | <i>16</i> |
| 2.3.2. <i>Traukinio triukšmas.....</i> | <i>16</i> |
| 2.3.3. <i>Gatvės triukšmas.....</i> | <i>17</i> |
| 2.3.4. <i>Garažo triukšmas.....</i> | <i>17</i> |
| 2.3.5. <i>Triukšmo lygis aplinkoje.....</i> | <i>17</i> |
| 2.4. FUNKCINIAI IR SISTEMINIAI REIKALAVIMAI | 18 |
| 2.5. BALSŲ DETEKCIJOS ALGORITMAI | 19 |
| 2.5.1. <i>Standartizuoti balsų detekcijos algoritmai</i> | <i>19</i> |
| 2.5.2. <i>ITU-T G.729B/G.723.1A.....</i> | <i>20</i> |
| 2.5.3. <i>ETSI GSM-FR/HR/EFR.....</i> | <i>21</i> |
| 2.5.4. <i>ETSI AMR.....</i> | <i>22</i> |
| 2.5.5. <i>TIA/EIA IS-127/733</i> | <i>23</i> |
| 2.6. AUKŠTESNIŲ EILIŲ STATISTIKA GRĮŠTAS ALGORITMAS | 24 |
| 2.7. DIRBTINIAI NEURONINIAI TINKLAI BALSŲ DETEKCIJOJE..... | 26 |
| 3. BALSŲ DETEKCIJOS ALGORITMO TEORINIAI ASPEKTAI | 28 |
| 3.1. TIESINIS PROGNOZUOJANTIS KODAVIMAS | 28 |
| 3.2. ALGORITMO MODELIS | 30 |
| 3.2.1. <i>Tiesinio prognozavimo posistemė.....</i> | <i>31</i> |
| 3.2.2. <i>Aukštesniųjų eilių statistikos skaičiavimo posistemė.....</i> | <i>32</i> |
| 3.2.3. <i>Signalas – triukšmas santykio skaičiavimo posistemė.....</i> | <i>33</i> |
| 3.2.4. <i>Triukšmo tikimybės skaičiavimo posistemė</i> | <i>33</i> |
| 3.2.5. <i>Baigtinio automato posistemė.....</i> | <i>34</i> |
| 3.3. ALGORITMO PATOBULINIMAI | 35 |
| 3.3.1. <i>Momentinis SNR nustatymo algoritmas.....</i> | <i>35</i> |
| 3.3.2. <i>Adaptyvios slenkstinės reikšmės</i> | <i>37</i> |
| 3.3.3. <i>Kadrų klasifikavimo tiesinimas.....</i> | <i>38</i> |
| 3.3.4. <i>Neuroninis tinklas.....</i> | <i>40</i> |

| | |
|---|-----------|
| 4. EKSPERIMENTINIS TYRIMAS..... | 43 |
| 4.1. ALGORITMO REALIZACIJOS METODIKA | 43 |
| 4.2. KALBOS SIGNALO CHARAKTERISTIKŲ TYRIMAS..... | 44 |
| 4.3. MODIFIKACIJŲ REZULTATAI..... | 48 |
| 4.4. ALGORITMŲ Palyginimas | 52 |
| 4.5. ALGORITMO ĮVERTINIMAS PAGAL TIA/EIA-136-250 REKOMENDACIJAS | 54 |
| 5. IŠVADOS..... | 56 |
| 6. LITERATŪRA | 57 |
| SANTRUMPŲ IR TERMINŲ ŽODYNĖLIS..... | 58 |
| PRIEDAS A. ŽMOGAUS KALBOS APARATAS, BALSO SINTEZAVIMAS | 59 |
| PRIEDAS B. PRIDEDAMO CD TURINYS | 61 |

Paveikslų sąrašas

| | |
|--|----|
| 1 pav. Kalbos periodų klasifikavimas | 12 |
| 2 pav. Koduoto kalbos signalo struktūra..... | 13 |
| 4 pav. ITU-T G.729B/G.723.1A algoritmo blokinė diagrama..... | 20 |
| 5 pav. ETSI GSM-FR/HR/EFR algoritmo blokinė diagrama | 21 |
| 6 pav. ETSI AMR1 algoritmo blokinė diagrama | 22 |
| 7 pav. ETSI AMR2 algoritmo blokinė diagrama | 23 |
| 8 pav. TIA/EIA IS-127/733 algoritmų blokinė diagrama | 24 |
| 9 pav. Aukštesnių eilių statistika grįstas balso detektavimo algoritmas | 25 |
| 10 pav. TKP liekanos spektras | 25 |
| 11 pav. Aukščiausio hierarchinio lygmens simulink modelio blokinė schema | 31 |
| 12 pav. Tiesinio prognozavimo posistemė..... | 32 |
| 13 pav. Aukštesnių eilių statistikos skaičiavimo posistemė..... | 32 |
| 14 pav. Signalas – triukšmas santykio skaičiavimo posistemė | 33 |
| 15 pav. Triukšmo tikimybės skaičiavimo posistemė | 34 |
| 16 pav. Baigtinio automato posistemė | 34 |
| 17 pav. Momentinio SNR nustatymo algoritmo blokinė diagrama | 36 |
| 18 pav. Būsenų nestabilumo pavyzdys | 38 |
| 19 pav. Kadru klasifikavimo tiesinimo algoritmas | 39 |
| 20 pav.. Klasifikavimo lyginimo pavyzdys..... | 40 |
| 21 pav. DNT reakcija į signalo parametrus..... | 41 |
| 22 pav. DNT struktūra..... | 41 |
| 23 pav. DNT apmokymo klaidos grafikas | 42 |
| 24 pav. Asimetrijos koeficiento priklausomybė nuo balso signalo..... | 44 |
| 25 pav. Eksceso koeficiento priklausomybė nuo balso signalo | 45 |
| 26 pav. Kalbos ir triukšmo signalų eksceso koeficientų histogramos. | 45 |
| 27 pav. SKR priklausomybė nuo balso signalo. | 46 |
| 28 pav. Gausinio triukšmo tikimybės signale grafikas. | 46 |
| 29 pav. TPK paklaidos grafikas | 47 |
| 30 pav. SNR grafikai | 48 |
| 31 pav. Balso kadru atpažinimo grafikas | 49 |
| 32 pav. Triukšmo kadru atpažinimo grafikas..... | 50 |
| 33 pav. Triukšmo kadru atpažinimo grafikas..... | 51 |
| 34 pav. Balso kadru detektavimo palyginimas | 53 |
| 35 pav. Triukšmo kadru detektavimo palyginimas | 54 |
| 3 pav. Žmogaus vokalinė sistema. [5]..... | 59 |

Lentelių sąrašas

| | |
|---|----|
| 1 lentelė. TIA/EIA-136-250 Testinės situacijos..... | 14 |
| 2 lentelė. Skirtingų šaltinių keliami triukšmo lygiai | 18 |
| 3 lentelė. Kadru ir adityvios konstantos reikšmės..... | 37 |
| 4 lentelė. Algoritmo tyrimo pagal TIA/EIA-136-250 rekomendacijas rezultatai | 54 |

Pratarmė

Šiame darbe aprašomas balso detektavimo algoritmo sudarymas ir tyrimas remiantis E. Nemer, R. Goubran ir S. Mahmoud paskelbta publikacija [1]. Darbas pradėtas Danijos Aalborg universitete studijuojant „Applied Signal Processing and Implementation (ASPI)“ specializaciją ir atliekant semestro projektinį darbą, vadovaujant prof. Per Rubak.

Darbe analizuojamas aukštesnių eilių statistika grįstas balso detektavimo algoritmas, pristatomi algoritmo patobulinimai, gerinantys balso detektavimo rezultatus. Eksperimentiniai algoritmo realizacijos DSP procesoriuje balso detekcijos rezultatai lyginami su šiuo metu standartizuotų balso detektavimo algoritmų rezultatais.

Design and analysis of voice activity detector based on higher order statistics

Summary

This report covers a robust voice activity detection (VAD) algorithm presented in [1]. The algorithm uses higher order statistics (HOS) metrics of speech signal in linear prediction coding (LPC) residual domain to classify noise and speech frames of a signal. Chapters in this report present voice activity detection problem and analysis of environment issues for VAD, deep HOS based and standard algorithms analysis and a real time HOS based voice activity detector model. New improvements (instantaneous SNR estimation, decision smoothing, adaptive thresholds, artificial neural network) to the proposed algorithm are introduced and performance results of the improved algorithm compared to standard VAD algorithms are presented.

1. ĮVADAS

Šiuo metu informacinėms technologijoms sparčiai skverbiantis į žmogaus buitį, intensyviai kompiuterizuojami įvairūs buities įrenginiai ir valdymo prietaisai. Tačiau dažnai susiduriama su komfortabilia įrenginių valdymo problema, ypač kai į įrenginį reikia įvesti tekstą. Dažniausiai tokiu atveju naudojama žodžių įvedimo paraidžiu sistema, tačiau sparčiai kylant vartotojiško komforto reikalavimams, tokios sistemos tampa nepatogios. Valdymo sprendimai, naudojant balso komandas nėra naujiena, tačiau iki šiol jie nėra populiarūs dėl nepatikimo veikimo. O įvedant tekstą jie kol kas sukelia daugiau problemų, nei pageidaujamos naudos. To priežastis – sudėtinga žmogaus balso matematinė analizė ir atpažinimas, bei kritiškas adityvaus triukšmo poveikis analizės metoduose. Norint gauti tikslesnius balso atpažinimo rezultatus, būtina signalo apdorojimui naudoti tik balso signalą. Idealiu atveju, kai apdorojamam signalui nėra adityvaus triukšmo, balso periodą nesunku detektuoti, nustatant atitinkamas slenkstines reikšmes signalo amplitudei. Tačiau realiame pasaulyje balso signalas visada yra iškraipomas įvairiais foniniais ir atsitiktiniais triukšmais, kurių amplitudinės ir dažninės savybės neretai sutampa su parametrais, kurie charakteringi žmogaus balsui.

Viena iš iki šiol neišspręstų problemų analizuojant žmogaus balsą yra ta, kad naudojami skaičiavimo metodai nesugeba optimaliai atskirti balso nuo į jį panašaus triukšmo. Dažnai tokios balso detekcijos problemos sukelia nekorektiškus skaičiavimo rezultatus ir gali sukelti nepageidautinų problemų taikomuosiuose uždaviniuose. Kalbos signalo pradžios ir pabaigos paieška duotame garse yra vienas pagrindinių kalbos signalų atpažinimo pradinio apdorojimo proceso etapų, nuo kuriuo tiesiogiai priklauso kalbos atpažinimo efektyvumas.

Balso detekcijos problema pastaruoju metu dažnai nagrinėjama. Tokį mokslininkų susidomėjimą šiuo uždaviniu paskatino sparti balso komunikacijų plėtra. Tiek GSM, tiek sparčiai plintančios VoIP technologijos specifikacijose numatomas balso detekcijos mechanizmas, kuris padeda mažinti komunikaciniais tinklais siunčiamų duomenų kiekius.

Šiame darbe tiriama balso detekcijos technologija. Apžvelgiami standartiniai, šiuo metu plačiausiai naudojami, balso detekcijos algoritmai, įvertinami jų rezultatai. Tačiau esminis šio darbo objektas yra aukštesniųjų eilių statistikos pritaikymas kalbos detekcijai.

Balso detektavimo aktualumą gali patvirtinti mokslinių publikacijų šia tema įvairovė ir gausa. Iki šiol nėra dominuojančios metodikos ar technologijų, kurios pagelbėtų sudarant balso detekcijos algoritmus. Balso perdavimo technologijų plėtra tik spartina šios temos eskalavimą moksliniuose ir industriniuose tyrimuose. Siekiama pasiūlyti ir sukurti korektiškiau kalbą atpažįstančius algoritmus.

Šio darbo tikslas išanalizuoti ir įvertinti standartizuotus balso detektavimo algoritmus, sukurti ir realizuoti balso detekcijos mechanizmą, taikant mokslinius tyrimus, naudojant aukštesniųjų eilių statistika grįstą algoritmą [1], momentinį signalas-triukšmas santykio nustatymo algoritmą [2] ir dirbtinių neuroninių tinklų technologiją.

Darbe pateikiamas išsamus aukštesniųjų eilių statistika grįsto algoritmo[1] tyrimas, analizuojami ir pirmą kartą siūlomi algoritmo tobulinimo sprendimai. Rezultatai lyginami su standartizuotų balso detekcijos algoritmų rezultatais.

Balso atpažinimo algoritmų tyrimas atliekamas dviem būdais:

- Naudojant aukštesniųjų eilių statistika grįsto bei modifikuoto algoritmo modelį.
- Naudojant algoritmo realizaciją, DSP platformoje.
- Remiantis jau paskelbtomis mokslinėmis publikacijomis.

Atliekant eksperimentus buvo naudojama standartinė balso signalų duomenų bazė[3].

Darbo eigoje sukurtas aukštesniųjų eilių statistika grįsto algoritmo modelis bei realizacija C++ kalboje, DSP platformoje, pateikiami algoritmų įvertinimai. Šie rezultatai gali būti panaudoti tolimesniems tyrimams ar pritaikyti taikomuosiuose uždaviniuose, sprendžiant įrenginių valdymo balsu, balso perdavimo ir atpažinimo problemas.

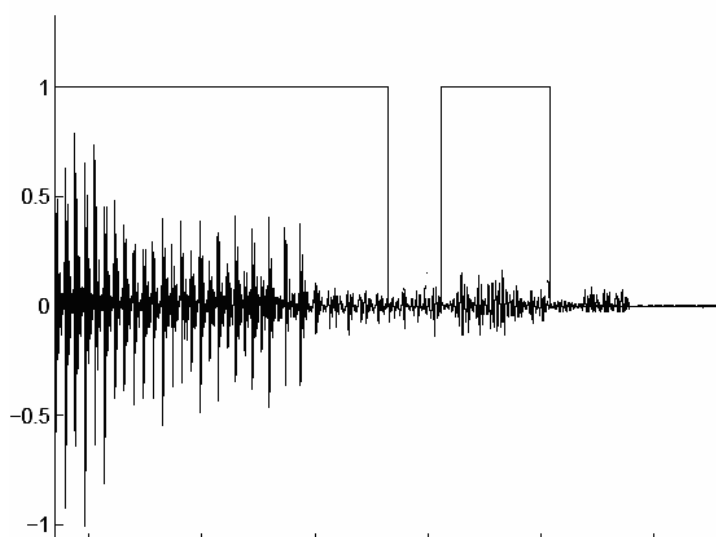
Atliekamas darbas buvo suskirstytas į tokius etapus:

- Balso signalo analizė.
- Kalbos detekcijos proceso bei metodologijų analizė.
- Esamo aukštesniųjų eilių statistika grįsto algoritmo analizavimas ir tyrimas.
- Momentinio signalas-triukšmas nustatymo algoritmo analizė.
- Dirbtinių neuroninių tinklų analizė, balso detekcijos tematikoje.
- Algoritmo modelio sudarymas ir modeliavimas.
- Programinė algoritmo realizacija.
- Kalbos signalo detekcijos tyrimų bei išvadų pateikimas, naudojant sukurtą modelį ir programinę realizaciją.

2. BALSŲ DETEKCIJOS UŽDAVINIO ANALIZĖ

2.1. Balsų detekcija ir taikymas

Balsų aktyvumo detektoriais vadinama balsų specifika pagrįsta priemonė, gebanti išskirti žmogaus balsą iš tylos ar atsitiktinio triukšmo. Vienintelė tokių priemonių funkcija yra klasifikuoti kalbos ir ne kalbos periodus. Paprastai kalbos signalas suskirstomas į kadrus, o balsų detektorius, naudodamasis kadro signalo charakteristikomis ir prieš tai buvusių kadro signalo savybėmis, nustato ar duotasis kadras priklauso kalbos ar tylos laikotarpiui. Yra priimta, kalbos buvimą signalė kuoduoti loginiu „1“, triukšmą, ar tylą – loginiu „0“ (1 pav.)



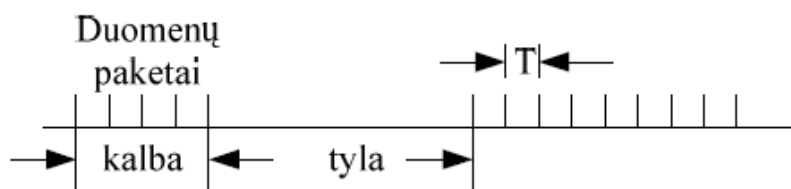
1 pav. Kalbos periodų klasifikavimas

Balsų detektorius panaudojimo spektras yra pakankamai platus. Efektyviai veikiantis balsų detektorius gali būti naudojamas kalbos signalo suspaudimui, balsu valdomiems įrenginiams ir t.t. Tačiau didžiausias poreikis balsų detekcijai yra balsų telekomunikacijų sferoje ir ypač mobilių komunikacijų srityje. Čia yra ypač aktualus perduodamų duomenų kiekis, todėl naudojant balsų detektorius atsiranda galimybė komunikaciniais kanalais perduoti tik naudingą informaciją – kalbą, o tylą arba foninį triukšmą tiesiog pašalinti arba perduoti atskirai naudojant skirtingą suspaudimo technologiją. Kadangi tylai ar foniniam triukšmui užkoduoti ir perduoti reikia daug mažiau informacijos nei kalbai, todėl galima mažinti spartą, skirtą vienam komunikaciniam kanalui. Šiuo metu sistemingai ieškoma būdų, kaip sumažinti balsui perduoti reikalingų duomenų kiekį. Nustatyta, kad kalbant pakaitomis kartojasi trumpi kalbos ir tylos intervalai. Tokia kalbos signalo struktūrą lemia žmogaus kalbos aparato savybės. Žmogus paprastai kalba iškvėpdamas orą ir tyli iškvėpdamas. Kita priežastis, dėl kurios susidaro tylos

intervalai yra tai, kad kalbant telefonu pašnekovai dažniausiai kalba pakaitomis – vienas kalba, kitas klauso. Visų kalbos ir tylos intervalų trukmė yra atsitiktinės. Paprastai, telefoniniuose pokalbiuose, tyla užima nuo 30 iki 50 procentų viso pokalbio [4]. Pritaikius tylos suspaudimo mechanizmą, būdu akivaizdžiai pasiekiamas efektyvesnis tinklo išnaudojimas, vienu metu būtų galima multipleksuoti daugiau balso kanalų tuo pačiu tinklu.

Balso kompresija, balso aktyvumo detektoriais, telefono ryšiui pirmą kartą panaudota TASI (ang.k. *Time Assignment Speech Interpolation*) sistemoje (TASI-A sukurta 1959 m., TASI-B – 1976 m.) Ši balso perdavimo sistema buvo sukurta specialiai tolimajam ryšiui ir taikyta perduodant telefonų pokalbių signalus tarpkontinentiniais kabeliais.

Balso aktyvumo detektoriai taikomi ir moderniose paketinio perdavimo priemonėse (2 pav.). Neperduodant balso duomenų kalbos laikotarpiais lieka laisva ryšio kanalo laidumo dalis. Šią laidumo dalį galima panaudoti kitų pokalbių signalams ar su balso signalu nesusijusiems duomenims perduoti. Mobiliojo ryšio sistemoje tylos laiku galima išjungti siųstuvus. Taip taupoma energija ir mažinami trukdžiai kitoms ryšio priemonėms.



2 pav. Koduoto kalbos signalo struktūra.

Balso detekcija naudinga ir balso komunikacijų vartotojams. Įdiegus efektyvų algoritimą mobiliuosiuose telefonuose, mobiliųjų telefonų siųstuvai galėtų dirbti pertraukiamu režimu, taip taupydami mobiliojo telefono akumulatoriaus tarnavimo laiką.

Atskyrus kalbos periodus iš nepageidaujamo ir dažnai nemalonaus foninio triukšmo ir pakeičiant jį tyla arba maloniu triukšmu, generuojamu nutolusiame įrenginyje, vartotojų pokalbiai būtų daugiau suprantami ir malonesni.

2.2. Balso detektoriaus įvertinimas

Šiuo metu pagrindinė balso detektorių taikymo sritis yra mobilios komunikacijos. Kadangi nemažai algoritmų yra pasiūlyta ir realizuojama įvairiuose duomenų perdavimo ir suspaudimo technologijose, atsirado būtinybė objektyviai įvertinti balso detektorių veikimo efektyvumą bei jų atitikimą balso komunikacijų reikalavimams.

TIA/EIA-136-250 standarte yra specifikuojama balso detektoriaus veikimo efektyvumo matavimo metodika ir minimalus reikalavimai šią technologiją naudojančioms mobilioms stotims, kurios dirba pertraukiamo siuntimo režimu (ang.k. *Discontinuous transmission (DTX) mode*).

1 lentelė. TIA/EIA-136-250 Testinės situacijos

| Nr. | Balso tipas | Norma- lizacija | Triukšmo tipas | SNR | Nr. | Balso tipas | Norma- lizacija | Triukšmo tipas | SNR |
|-----|-------------|--------------------|-------------------|-----|-----|--------------|--------------------|-------------------|-----|
| 1. | Vyriškas 1 | -36 | Traukinys | 6 | 41. | Moteriškas 1 | -36 | Gatvė | 6 |
| 2. | Vyriškas 1 | -26 | Garažas | 12 | 42. | Moteriškas 1 | -26 | Automobilis | 12 |
| 3. | Vyriškas 1 | -16 | Gatvė | 18 | 43. | Moteriškas 1 | -16 | Traukinys | 18 |
| 4. | Vyriškas 1 | -26 | -- | ∞ | 44. | Moteriškas 1 | -26 | -- | ∞ |
| 5. | Vyriškas 1 | -16 | -- | ∞ | 45. | Moteriškas 1 | -16 | -- | ∞ |
| 6. | Vyriškas 1 | -36 | Automobilis | 18 | 46. | Moteriškas 1 | -36 | Garažas | 18 |
| 7. | Vyriškas 1 | -26 | Traukinys | 12 | 47. | Moteriškas 1 | -26 | Gatvė | 12 |
| 8. | Vyriškas 1 | -16 | Garažas | 6 | 48. | Moteriškas 1 | -16 | Automobilis | 6 |
| 9. | Vyriškas 2 | -26 | Garažas | 6 | 49. | Moteriškas 2 | -26 | Automobilis | 6 |
| 10. | Vyriškas 2 | -26 | Gatvė | 12 | 50. | Moteriškas 2 | -26 | Traukinys | 12 |
| 11. | Vyriškas 2 | -16 | Automobilis | 18 | 51. | Moteriškas 2 | -16 | Garažas | 18 |
| 12. | Vyriškas 2 | -36 | -- | ∞ | 52. | Moteriškas 2 | -36 | -- | ∞ |
| 13. | Vyriškas 2 | -26 | -- | ∞ | 53. | Moteriškas 2 | -26 | -- | ∞ |
| 14. | Vyriškas 2 | -36 | Traukinys | 18 | 54. | Moteriškas 2 | -36 | Gatvė | 18 |
| 15. | Vyriškas 2 | -36 | Garažas | 12 | 55. | Moteriškas 2 | -36 | Automobilis | 12 |
| 16. | Vyriškas 2 | -16 | Gatvė | 6 | 56. | Moteriškas 2 | -16 | Traukinys | 6 |
| 17. | Vyriškas 3 | -36 | Gatvė | 6 | 57. | Moteriškas 3 | -36 | Traukinys | 6 |
| 18. | Vyriškas 3 | -26 | Automobilis | 12 | 58. | Moteriškas 3 | -26 | Garažas | 12 |
| 19. | Vyriškas 3 | -16 | Traukinys | 18 | 59. | Moteriškas 3 | -16 | Gatvė | 18 |
| 20. | Vyriškas 3 | -16 | -- | ∞ | 60. | Moteriškas 3 | -16 | -- | ∞ |
| 21. | Vyriškas 3 | -36 | -- | ∞ | 61. | Moteriškas 3 | -36 | -- | ∞ |
| 22. | Vyriškas 3 | -26 | Garažas | 18 | 62. | Moteriškas 3 | -26 | Automobilis | 18 |
| 23. | Vyriškas 3 | -36 | Gatvė | 12 | 63. | Moteriškas 3 | -36 | Traukinys | 12 |
| 24. | Vyriškas 3 | -16 | Automobilis | 6 | 64. | Moteriškas 3 | -16 | Garažas | 6 |
| 25. | Vyriškas 4 | -26 | Automobilis | 6 | 65. | Moteriškas 4 | -26 | Garažas | 6 |
| 26. | Vyriškas 4 | -36 | Traukinys | 12 | 66. | Moteriškas 4 | -36 | Gatvė | 12 |
| 27. | Vyriškas 4 | -36 | Garažas | 18 | 67. | Moteriškas 4 | -36 | Automobilis | 18 |
| 28. | Vyriškas 4 | -26 | -- | ∞ | 68. | Moteriškas 4 | -26 | -- | ∞ |
| 29. | Vyriškas 4 | -16 | -- | ∞ | 69. | Moteriškas 4 | -16 | -- | ∞ |
| 30. | Vyriškas 4 | -26 | Gatvė | 18 | 70. | Moteriškas 4 | -26 | Traukinys | 18 |
| 31. | Vyriškas 4 | -16 | Automobilis | 12 | 71. | Moteriškas 4 | -16 | Garažas | 12 |
| 32. | Vyriškas 4 | -16 | Traukinys | 6 | 72. | Moteriškas 4 | -16 | Gatvė | 6 |
| 33. | Vyriškas 5 | -26 | Traukinys | 6 | 73. | Moteriškas 5 | -26 | Gatvė | 6 |
| 34. | Vyriškas 5 | -16 | Garažas | 12 | 74. | Moteriškas 5 | -16 | Automobilis | 12 |
| 35. | Vyriškas 5 | -36 | Gatvė | 18 | 75. | Moteriškas 5 | -36 | Traukinys | 18 |
| 36. | Vyriškas 5 | -36 | -- | ∞ | 76. | Moteriškas 5 | -36 | -- | ∞ |
| 37. | Vyriškas 5 | -26 | -- | ∞ | 77. | Moteriškas 5 | -26 | -- | ∞ |
| 38. | Vyriškas 5 | -26 | Automobilis | 18 | 78. | Moteriškas 5 | -26 | Garažas | 18 |
| 39. | Vyriškas 5 | -16 | Traukinys | 12 | 79. | Moteriškas 5 | -16 | Gatvė | 12 |
| 40. | Vyriškas 5 | -36 | Garažas | 6 | 80. | Moteriškas 5 | -36 | Automobilis | 6 |

Testavimui ir vertinimui reikalingi kalbos ir triukšmo įrašai yra pateikiami kartu su standartu. Iš šių įrašų specialiai standartui sukurta programinė įranga sugeneruoja 80 testinių situacijų (1 lentelė.), tam kad būtų galima išbandyti įvertinti balso detektorių įvairiais režimais.

Kalbos medžiaga susideda iš 10 telefoninių pokalbių: 5 yra moteriško ir 5 vyriško balso įrašai. Kalbos įrašai įvertinimams pateikiami žemu, vidutiniu ir aukštu normalizacijos lygiais, su skirtingais triukšmų šaltiniais ir skirtingu signalas – triukšmas santykio lygiu. Balso buvimas signale yra pažymėtas atskiruose failuose, kurie gauti klausant ir rankiniu būdu žymint kalbos sekas signale. Kalbos frazės šiuose failuose yra skirstomos į tris kategorijas: kalbos pradžią, pastovią kalbą ir kalbos pabaigą. Kalbos frazės pradžia laikomi pirmieji trys kalbos kadrai, kalbos pabaiga – paskutiniai trys kadrai, o pastovi kalba yra ta, kuri yra tarp kalbos pradžios ir pabaigos kadru. Standartiškai kalbos kadro ilgis yra 20ms. Toks skirstymas reikalingas, norint tiksliau nustatyti balso detekcijos inertiškumą.

Vertinant balso detektorių, speciali programinė įranga skaičiuoja kiek kartų testuojamas balso detektorius korektiškai sužymėjo kalbos buvimą signale, lyginant su tam tikrame faile teisingai sužymėta kalba. Ši procedūra yra atliekama 80 kartų, pagal 1 lentelėje parodytas situacijas. Iš visų testavimo situacijų gautais duomenimis formuojama rezultatų lentelė susidedanti iš 12 kategorijų, kurių kiekviena atitinka skirtingos normalizacijos ir signalas – triukšmas santykio lygio kombinaciją. Taip įvertinamas balso detektoriaus gebėjimas teisingai nustatyti balso buvimą signale, kalbos pradžioje, pastovioje kalboje, ir kalbos pabaigoje. Be to skaičiuojamas santykinis balso detekcijos faktorius (*dVAF – delta voice-activity factor*) [3]:

$$dVAF = \frac{VAD - T_VAD}{T_VAD}$$

Čia, VAD – testuojamo balso detektoriaus teisingų kalbos balso atpažinimų pataikymo rodiklis – teisingai detektuotų kalbos kadru skaičius iš visų kalbos kadru.

T_VAD – standartus tenkinančio balso detektoriaus rodiklis.

Nepaisant standarte specifiukuotos balso detektorių įvertinimo procedūros, dažnai moksliniuose straipsniuose pateikiami rezultatai būna supaprastinti. Tokiais atvejais balso detektoriumi įvertinti naudojami trys parametrai:

- Teisingai atpažįstamų kalbos kadru tikimybė. Skaičiuojama, koks procentinis santykis teisingai atpažintų kalbos kadru su visais kalbos kadrais.
- Teisingai atpažįstamų triukšmo kadru tikimybė. Skaičiuojama, koks procentinis santykis teisingai atpažintų triukšmo kadru su visais triukšmo kadrais.
- Klaidingų atpažinimų tikimybė. Skaičiuojama, koks procentinis santykis neteisingai klasifikuotų triukšmo ar kalbos kadru su visais signalo kadrais.

Šiai rodikliams visapusiškai perteikia balso detektoriaus gebėjimą klasifikuoti balso ir triukšmo kadrus, tačiau iš jų neįmanoma nustatyti balso detektoriaus savybių, teisingai klasifikuoti kalbos pradžią ir pabaigą.

2.3. Triukšmų analizė

Balso aktyvumo detektoriai dažniausiai taikomi kalbos suspaudimo ir balso perdavimo technologijose. Kadangi telefoniniai pokalbiai vyksta skirtingose aplinkose, tai kalbos įrašai dažnai turi savyje triukšmų iš įvairių triukšmo šaltinių, todėl neįmanoma nuspėti kokio tipo triukšmą reikės klasifikuoti balso detektoriumi. To pasekoje neįmanoma išanalizuoti visų triukšmo šaltinių ir jų generuojamo triukšmo tipų, savybių ir charakteristikų.

TIA/EIA–136–250 standartas specifikuojantis balso aktyvumo detektorių minimalius efektyvumo reikalavimus ir standartinę jų įvertinimo procedūrą, rekomenduoja įvertinimui naudoti ir pateikia keturis skirtingus triukšmų įrašus. Šie įrašai atspindi dažniausiai kalbos signaluose aptinkamus triukšmų tipus: Automobilio skleidžiamas triukšmas, traukinio triukšmas, gatvėje įrašytas triukšmas, garažo triukšmas. Sekančiuose poskyriuose apžvelgiama pagrindinė šių triukšmų specifiška.

2.3.1. Automobilio skleidžiamas triukšmas

Akustinės savybės automobilio salone esančio triukšmo skiriasi priklausomai nuo automobilio tipo ir automobilio savybių, tokių kaip greitis, variklio tipas, kelio dangos tipas, padangų savybės bei vėjo greitis ir kryptis. Didesni ir sunkesni automobiliai skleidžia daugiau triukšmo nei lengvieji. Akustinis variklio triukšmas susideda iš periodinių garso komponentų, kurių dažniai paprastai būna mažesni nei 1000 Hz. Kelio ir padangų trinties keliamas garsas susideda iš atsitiktinių garso komponentų, dažniu mažesniu nei 1000 Hz. Vėjo keliamo triukšmo garso komponentės taip pat atsitiktinės, jų dažnis vyruoja apie 500 Hz. [6].

2.3.2. Traukinio triukšmas

Važiuojančio traukinio keliamas triukšmas labiausiai priklauso nuo traukinio greičio, tačiau papildomų triukšmo variacijų gali atsirasti dėl traukinio variklio tipo, vagonų, ratų, bėgių ir jų polių savybių. Staigesni posūkiai, kurie dažnai aptinkami metro gali priversti traukinius skleisti stiprų aukštų dažnių garsą. Be to požeminiuose traukiniuose ar tuneliuose traukinio keliamas garsas iškraipomas pridėdam jam specifinių akustinių savybių. Dažnai traukinio keliamam triukšme galima aptikti stiprų žemą dažnį (iki 100 Hz) turinčių periodinių komponentų.

2.3.3. Gatvės triukšmas

Gatvės triukšmas pagrįdė susideda iš į žmogaus kalbą panašių triukšmų. (ang.k. human speech-like noises (HSLN)). Tai triukšmas gaunamas sudedant keletą nepriklausomų kalbos signalų. Svarbi savybė tokio triukšmo, yra ta, kad kuo daugiau nepriklausomu kalbos signalų pridedama, tuo triukšmas labiau praranda balso signalui tinkančias savybes ir tampa panašesnis į pastovų triukšmą.

Be to gatvės triukšmą papildo eilė vadinamų buitės garsų: pravažiuojančių automobilių skleidžiamas triukšmas, buitinių ir industrinių variklių keliamas periodinis triukšmas, atsitiktiniai smūginiai garsai. Šiuo triukšmo tipo analizės atskleidžia jo panašumą į Gausinį triukšmą.

2.3.4. Garažo triukšmas

Garažo aplinka turi savo išskirtinį triukšmą, pagrįdė dėl aido kuris daug kartų atsimuša į sienas ir įvairius kitus paviršius, dėl ko gaunama akustinė reverberacija. Spektrinė šiuo triukšmo tipo analizė atskleidžia jo panašumą į gatvės triukšmą, tačiau dėl papildomų atsitiktinių triukšmų ir jų akustinės reverberacijos, šio tipo triukšme aptinkama nemažai žemo dažnio dedamųjų. Šiuo triukšmo savybės panašios į Gausinio triukšmo.

2.3.5. Triukšmo lygis aplinkoje

TIA/EIA–136–250 standartas balso detektoriaus įvertinimui rekomenduoja naudoti 4 skirtingus signalas-triukšmas lygius. Tačiau dažnai aplinkoje pasitaiko kurkas daugiau triukšmo skleidžiančių šaltinių, prie kurių balso detektoriaus veikimo efektyvumas turėtų būti numatytas. Remiantis įvairioje literatūroje surastais duomenimis, 2-oje lentelėje pateikiamas atsitiktinai pasirinktų triukšmo šaltinių keliamas tipinis triukšmo lygis.

2 lentelė. Skirtingų šaltinių keliami triukšmo lygiai

| Triukšmo šaltinis | Triukšmo Lygis dB(A) |
|---|----------------------|
| Pneumatinis kūjis 1 m atstumu | 115 |
| Rankinis pjūklas 1 m atstumu | 115 |
| Tekstilės fabrikas | 103 |
| Laikraščių spaustuvė | 95 |
| Žolepjūvė 1 m atstumu | 92 |
| Sunkvežimis važiojantis 50 km/h greičiu, 20 atstumu | 85 |
| Lengvasis automobilis važiojantis 60 km/h greičiu, 20 atstumu | 65 |
| Pokalbis 1 m atstumu | 55 |
| Tylus kambarys | 40 |

2.4. Funkciniai ir sisteminiai reikalavimai

Atsižvelgiant į balso aktyvumo detektorių taikymo sritis, jiems formuluojami toki reikalavimai:

- Balso komunikacijoms reikalingos balso aktyvumo detektoriaus realizacijos turi veikti realiaame laike. Kalbos suspaudimo technologijoms šis reikalavimas nėra būtinas.
- Kalbos vėlinimas dėl detekcijai reikalingų skaičiavimų neturi būti didesnis nei 30 ms. Tai yra didžiausias leistinas vėlinimas balso komunikacijų įrangai.
- Jeigu naudojama tik kalbos perdavimo technologija, svarbu, kad visos kalbos frazės būtų detektuotos teisingai. Priešingu atveju vartotojams gali kilti sunkumų suprasti apdorotą kalbos signalą. Šis reikalavimas atitinka teisingai atpažįstamų kalbos kadru tikimybę.
- Jeigu algoritmo realizacija yra naudojama vartotojo balso komunikacijų įrenginyje (Pvz.: mobilusis telefonas). Balso detekcijai reikalingu operacijų kiekis turi būti nedidelis, kad tiktų mobilių telefonų procesoriaus pajėgumams ir taupyto akumuliatoriaus energiją.
- Realizacijai mobiliuose įrenginiuose, svarbų vaidmenį turi ir balso detektoriaus naudojamos atminties kiekis, kadangi mobilių įrenginių atmintis yra pakankamai ribota.

Į šiuo reikalavimus vertėtų atsižvelgti parenkant optimalų balso detekcijos algoritmą ar jo realizaciją atskirai taikymo sričiai.

2.5. Balso detekcijos algoritmai

Vienas iš paprasčiausių būdų, leidžiančių nustatyti balso buvimą signale, yra signalo energijos pokyčių įvertinimas. Dažnai balso pradžia signale sutampa su žymiu signalo energijos padidėjimu, todėl nustačius pastovią ar adaptyvią signalo energijos slenkstinę ribą, galima nesunkiai nustatyti kalbos frazės signale. Tačiau dažnai triukšmo amplitudė taip pat būna nepastovi. Atsitiktiniam triukšmo pliūpsniui viršijus nustatytą signalo energijos slenkstinę ribą, šis bus detektuojamas kaip balso signalas. Dar vienas tokio balso nustatymo algoritmo trūkumas tai, kad kadrai, turintys nebalsinius garsus, gali būti klasifikuojami kaip triukšmas, dėl palyginti mažos šių garsų energijos.

Senesniuose balso detekcijos algoritmuose nulio kirtimų skaičius (ang.k. zero-crossing rate) buvo pagrindinis kalbos nustatymo signale kriterijus. Nulio kirtimų skaičius signalui apibrėžiamas kaip parametras, nusakantis kiek kartų, signalas kerta abscisių ašį. Nulio kirtimų skaičius signalui, turinčiam balsių garsus, yra fiksuotas. Tokio signalo 10 ms. kadras turi nuo 5 iki 15 nulio kirtimų, kai tuo tarpu triukšmo signalui šis parametras yra atsitiktinis ir neapibrėžtas. Ši savybė leidžia formuluoti balso nustatymo signale taisyklę, kuri yra nepriklausoma nuo signalo energijos ir daugeliu atvejų leidžia nustatyti mažą energiją turinčius fonemus triukšmingame signale. Tačiau nulio kirtimų skaičium pagrįsti balso detekcijos algoritmai nėra efektyvus, nes klasifikuoja triukšmo kadrus kaip balso, kai triukšmo nulių kirtimų skaičius patenka į kalbos signalui būdingas ribas. Todėl šis algoritmas gali visiškai nefunkcionuoti esant tam tikram triukšmui.

Šiuo metu standartizuoti ir labiausiai paplitę yra du balso detekcijos algoritmai: G.729B ir 3GPP AMR.

2.5.1. Standartizuoti balso detekcijos algoritmai

Pastebėjus balso detekcijos algoritmų praktinę vertę ir taikymo galimybes buvo pasiūlyta daugybė skirtingais principais veikiančių balso detektorių, kai kurie iš jų buvo pasirinkti standartizacijos organizacijų: ITU-T, ETSI ir TAI/EIA. ITU-T išleido G.729 standarto priedą B (G.729B) ir G723.1 priedą A (G723.1A) kaip papildymus 8 kb/s G729 ir 5.3 / 6.3kb/s G723.1 standartams kalbos kodavimui pertraukiamo siuntimo režimu (ang.k. *Discontinuous transmission (DTX) mode*). Europos mobiliųjų telefonų tinklams ETSI rekomendavo GSM-FR, -HR ir -EFR balso detekcijos metodus. Neseniai ETSI pasiūlė dar du naujus balso detekcijos metodus: AMR1 ir AMR2. Pastarieji rekomenduojami naudoti UMTS (trečios kartos mobiliuose tinkluose). Šiaurės Amerikos standartizacijos organizacija TAI/EIA pasiūlė du struktūriškai panašius IS – 127 ir IS – 723 balso detekcijos metodus.

Tradiciskai ETSI siūlomi balso detekcijos metodai buvo pagrįsti tikslesne balso signalo charakterizavimo savybe – spektrine signalo forma. Tiesinio prognozuojančio kodavimo paklaida padidėja kalbos periodais, kai spektrinės kalbos ir triukšmo signalo formos nesutampa. Tačiau paskutinės AMR rekomendacijose, spektrinės signalo formos charakteristikos atsisakyta.

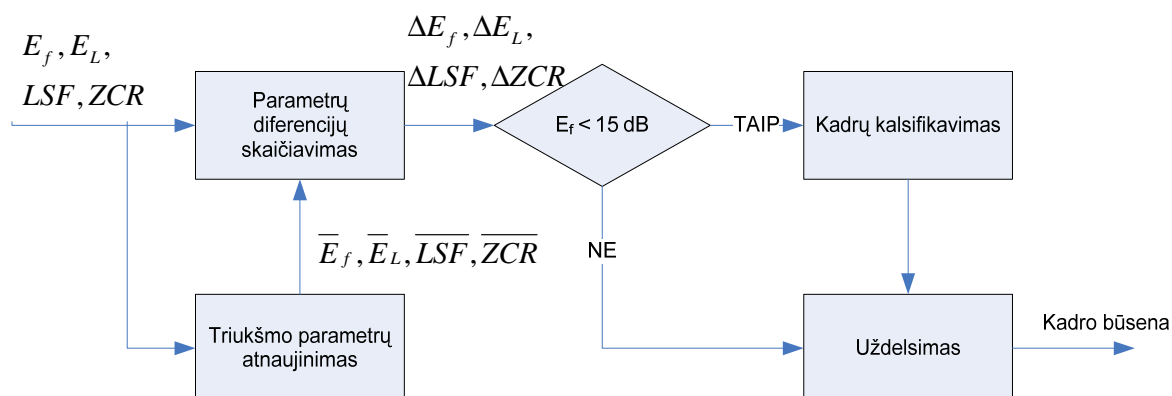
Kituose poskyriuose trumpai apžvelgiami standartizuoti balso detekcijos algoritmai.

2.5.2. ITU-T G.729B/G.723.1A

Kaip priedą prie G.729 balso kodavimo metodo, ITU-T išleido G.729 priedą B, kad kodavimas galėtų būti panaudojamas pertraukiamo siuntimo režimui, naudojant balso detekciją, komfortinio triukšmo įterpimą ir sintezaciją. G729B algoritmas kiekvienam 10 ms. kadrai klasifikuoja kaip balsą ar triukšmą naudodamas keturis parametrus:

- Visos dažnių juostos signalo energijos skirtumą, $\Delta E_f = \overline{E_f} - E_f$
- Apatinės dažnių juostos signalo energijos skirtumą, $\Delta E_L = \overline{E_L} - E_L$
- Spektrinius iškreipimus, $\Delta LSF = \sum_{i=0}^9 (\overline{LSF}_i - LSF_i)^2$
- Nulio kirtimų pokytis, $\Delta ZC = \overline{ZC} - ZC$

Kur, E_f , E_L , LSF_i ir ZC visos dažnių juostos signalo energija, apatinės dažnių juostos signalo energija, i-tosios linijos spektrinis dažnis, ir nulio kirtimų skaičius. $\overline{E_f}$, $\overline{E_L}$, \overline{LSF}_i ir \overline{ZC} yra triukšmui charakteringi parametrai, atnaujinami esant foniniam triukšmui. Blokinė algoritmo diagrama pateikta 4 paveikslėlyje.

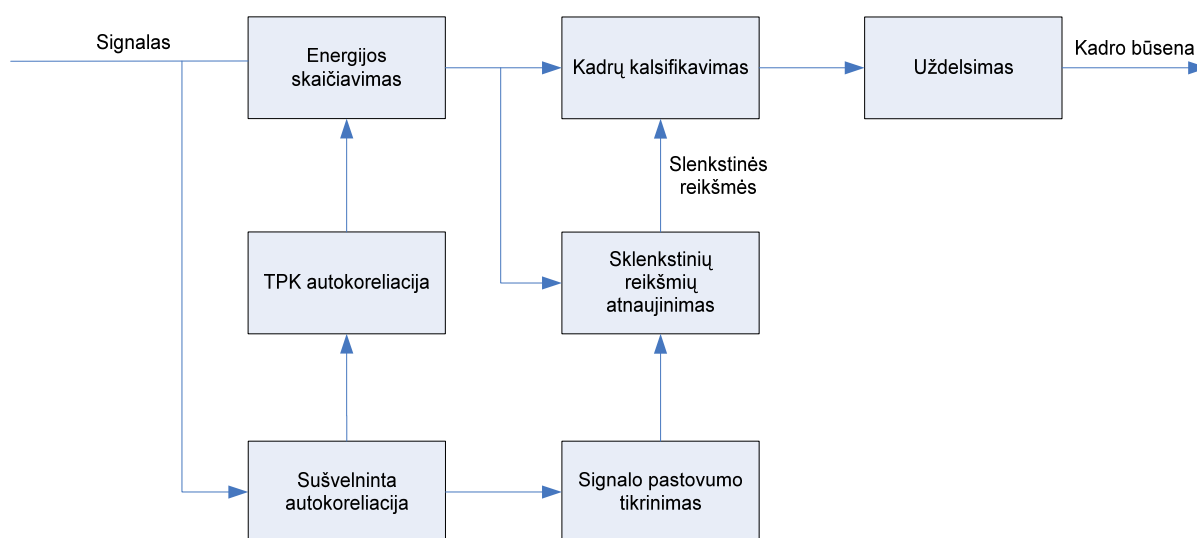


3 pav. ITU-T G.729B/G.723.1A algoritmo blokinė diagrama

Įėjimo parametrai balso detektoriumi skaičiuojami kiekvienam kadre iš įėjimo signalo. Triukšmo parametrai atnaujinami naudojant pirmos eilės autoregresinę schemą ir tik tada, kai visos dažnių juostos signalo energijos skirtumas mažesnis už nustatytą slenkstinę ribą. Delta parametrai ($\Delta E_f, \Delta E_L, \Delta LSF, \Delta ZC$) skaičiuojami iš įėjimo ir triukšmo parametrų. Kadro klasifikavimo sprendimas apskaičiuojamas remiantis besiribojančių regionų klasifikavimo technika bei pridodant papildomą uždelsimą.

2.5.3. ETSI GSM-FR/HR/EFR

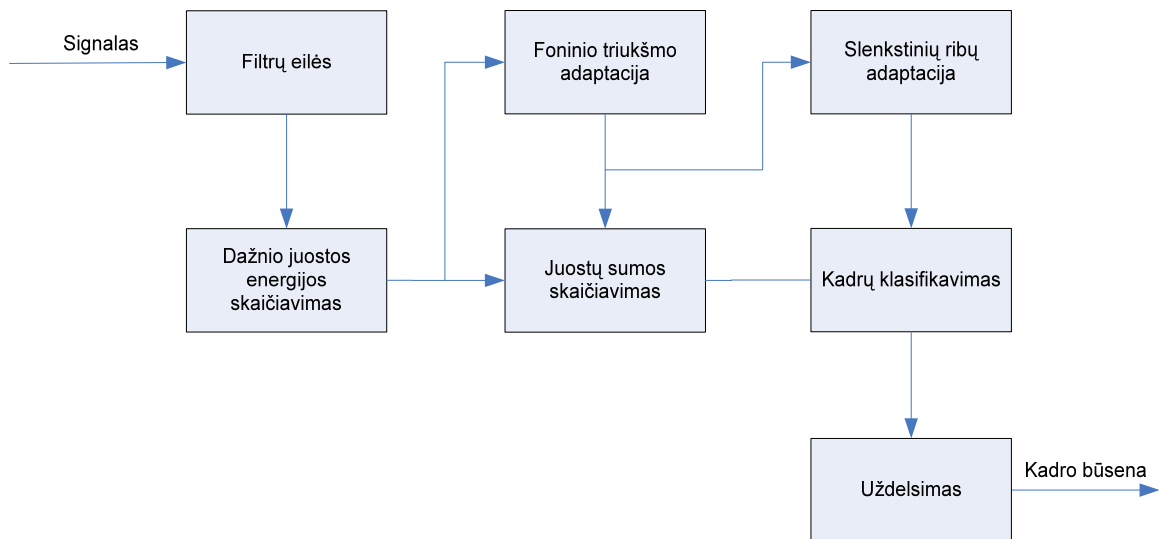
ETSI GSM-FR, -HR ir -EFR balso detektavimo algoritmai struktūriškai panašūs. Visi jie naudoja tiesinio prognozuojamo kodavimo liekanos energiją, kuri lyginama su adaptyvia slenkstine riba. TPK liekanos energija skaičiuojama naudojant esamas ir sulygintas autokoreliacijos reikšmes, kurios nusako spektrines signalo charakteristikas. Yra daroma prielaida, kad jeigu signale yra tik foninis triukšmas, kuris paprastai būna pastovus, tai vidutinė spektrinė signalo forma bus panaši esamo kadro spektrinę formą ir tai sumažins TKP liekanos energiją. Kadro klasifikavimui naudojamos slenkstinės ribos atnaujinamos tik esant foniniam triukšmui, kad parametrai kuo tiksliau atitiktų triukšmo signalo charakteristikoms. Blokinė algoritmo diagrama pateikta 5 paveikslėlyje



4 pav. ETSI GSM-FR/HR/EFR algoritmo blokinė diagrama

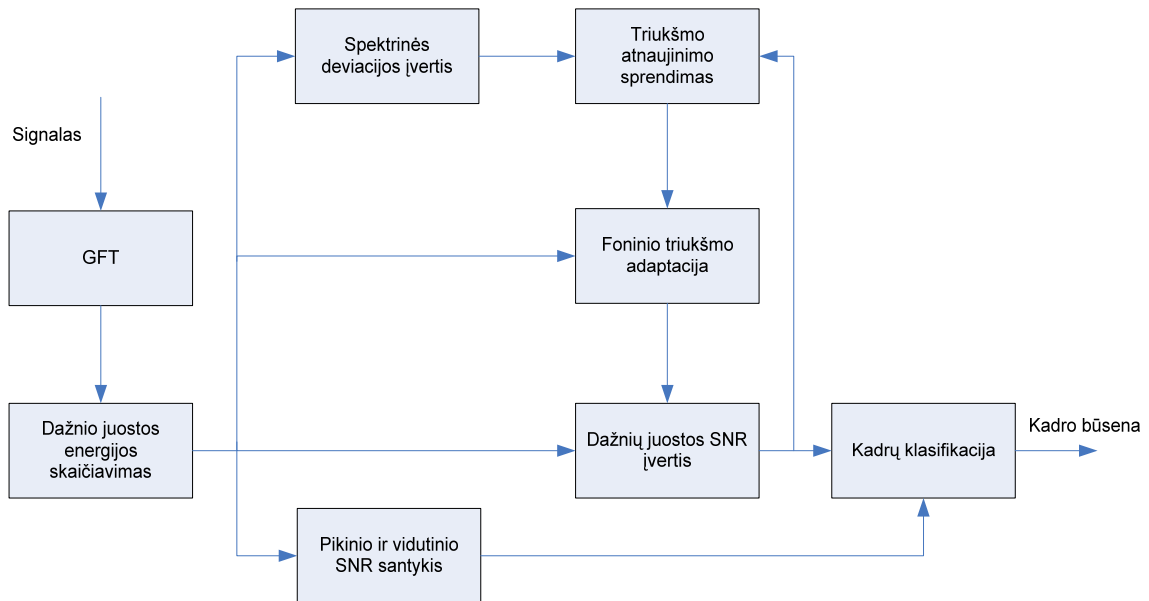
2.5.4. ETSI AMR

AMR1 balso detekcijos algoritme įėjimo signalas suskaidomas į devynias skirtingo pločio dažnio juostas, naudojant filtrų eilės, kur žemesnių dažnių filtrai turi siauresnę dažnių praleidimo juostą o aukštesnių dažnių – platesnę. Tuomet yra skaičiuojama kiekvienos juostos signalo energija bei signalas-triukšmas santykio apytikrė reikšmė. SNR skaičiavimui reikalinga foninio triukšmo energija, skaičiuojama adaptyviu metodu, kuris pagrįstas pirmos eilės autoregresiniu modeliu, bei naudoja vidinę balso detekcijos algoritmo logiką. Sprendimas apie kadro klasifikavimą apsprendžiamas suskaičiavus visų dažnių juostų SNR sumą ir lyginant ją su adaptyvia slenkstine riba. Taip pat naudojama balso segmento užlaikymo technika. Blokinė algoritmo diagrama pateikta 6 paveikslėlyje.



5 pav. ETSI AMR1 algoritmo blokinė diagrama

AMR2 algoritme panašiai, kaip ir AMR1 algoritme signalas sudalinamas į skirtingo pločio dažnio juostas, tačiau AMR2 algoritme signalas dalinama į 16 dažnio juostų, transformavus į dažninę sritį, greitos Furje transformacijos pagalba. SNR kiekvienai dažnio juostai skaičiuojamas naudojant signalo ir foninio triukšmo spektrą. Kiekvienai juostai triukšmo energija yra skaičiuojama, esant foninio triukšmo kadrams, naudojant pirmos eilės autoregresinę schemą. Tam kad išvengti algoritmo jautrumo nepastoviam triukšmui, AMR2 algoritme, kai signalas yra fluktuojantis, padidinamos slenkstinės reikšmės galutiniai kadro klasifikacijai. Tam yra matuojama išankstinė SNR reikšmė kiekvienam kadrai. AMR2 algoritmo blokinė diagrama pateikta 7 paveikslėlyje.

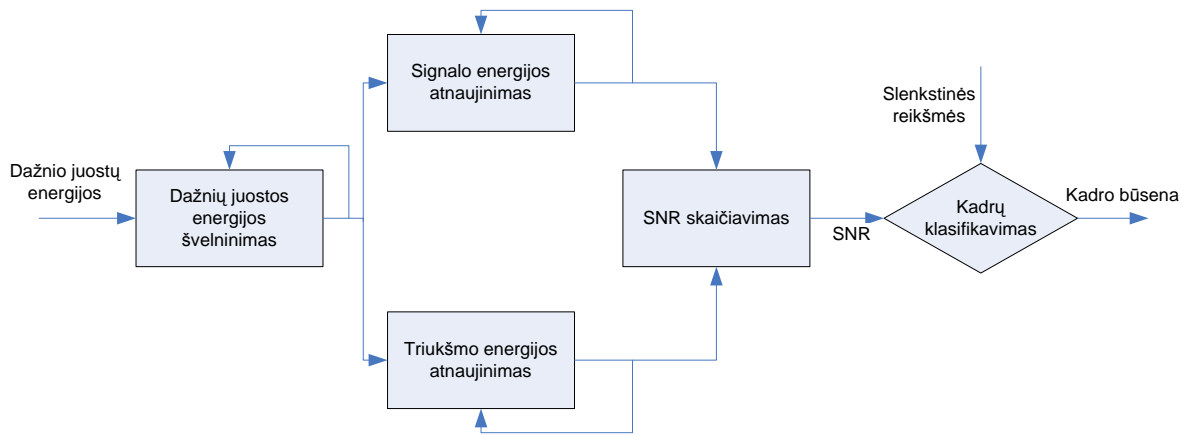


6 pav. ETSI AMR2 algoritmo blokinė diagrama

2.5.5. TIA/EIA IS-127/733

CDMA pagrįstos mobiliųjų komunikacijų sistemos naudoja balso detekcijos funkciją, reikalingos duomenų greitaveikos nustatymo algoritme. Šiaurės Amerikoje naudojami IS-127 ir IS-723 balso detekcijos algoritmai, jie skiriasi tik pokalbio duomenų perdavimo greitaveika. Naudojant balso detekciją, kalbos signalas perduodamas 1 arba $\frac{1}{2}$ perdavimo sparta, o triukšmas $\frac{1}{8}$.

IS-127/733 naudoja dviejų įėjimo signalo dažnio juostų energijas. Triukšmo energija skaičiuojama, kai kadras laikomas triukšmu. Tiek signalo tiek triukšmo energijos atnaujinamos pirmos eilės autoregresine schema Iš šių parametrų skaičiuojamas signalas-triukšmas santykis kiekvienai dažnių juostai. Galutinis sprendimas kadrų klasifikavime skaičiuojamas lyginant SNR su adaptyviom slenkstinėm reikšmėm. IS-127/733 algoritmų blokinė diagrama pateikta 8 paveikslėlyje.



7 pav. TIA/EIA IS-127/733 algoritmo blokinė diagrama

2.6. Aukštesnių eilių statistika grįstas algoritmas

Eksperimentiniai rezultatai įrodo, kad balso aktyvumo detektoriai pagrįsti aukštesnių eilių statistika veikia efektyviau Gausinio tipo triukšmuose, tokiuose kaip gatvės ar garažo triukšmas, nei standartizuoti ir šiuo metu labiausiai paplitę algoritmai.

Kalbos signalo statistinės savybės gali būti efektyviai naudojamos klasifikuoti balso ir Gausinio triukšmo kadrus. Aukštesnių eilių statistiniai parametrai Gausiniam triukšmui lygūs nuliui, kai tuo tarpu balso signalo nelygūs nuliui.

Aukštesnių eilių statistika pagyastas algoritmas naudoja asimetrijos ir eksceso statistinius matavimus, jų santykį, bei eilę kitų signalo parametrų: tiesinio prognozuojančio kodavimo prognozės paklaidą bei viso signalo ir žemesnių dažnių juostos signalas-triukšmas santykį.

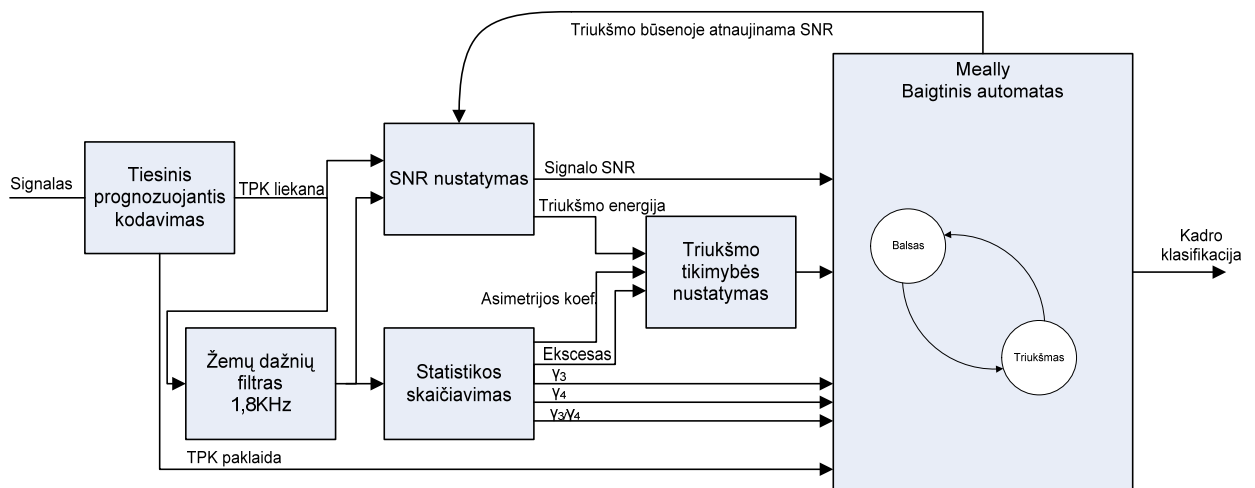
Algoritmas pagrįstas sinusoidiniu kalbos modeliui, vadinamu nulinės fazės harmonine reprezentacija (ang.k. *zero-phase harmonic representation*) [7] Šis modelis pasirinktas dėl jo paprastumo modeliuojant balsinius ir ne balsinius balso garsus. Modeliavimui naudojama dirbtinė fazės funkcija, kuri balso signale esant balsiniui garsui, modeliuoja signalą tos pačios fazės sinusoidėm, o ne balsinius garsus leidžia modeliuoti, kaip skirtingų fazių sinusoides. Šio modelio pagalba yra apibrėžiamos kalbos signalo aukštesnių eilių statistinės savybės [1]. Tokiu atveju, pagal kalbos sinusoidinį modelį, TKP liekanos signalas turi tokias savybes:

- **Balsių garsas** gali būti modeliuojamas kaip deterministinis signalas susidedantis iš sinusoidžių su vienodomis amplitudėmis.

- **Ne balsinis garsas** TKP liekanos atžvilgiu gali būti modeliuojamas kaip harmoninis procesas susidedantis iš sinusoidžių su atsitiktinėmis fazėmis. Bendru atveju ne balsinis garsas modeliuojamas kaip ne Gausinio tipo baltas procesas.

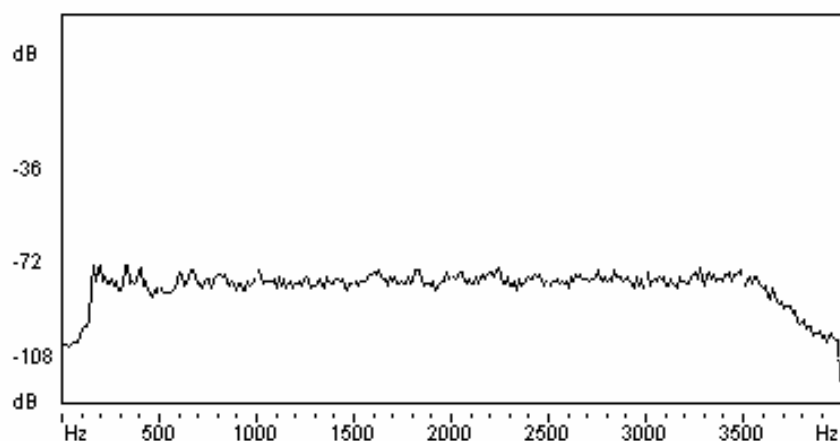
- **Gausinis triukšmas** TKP liekanoje taip pat išlieka nepakitęs.

Algoritmo blokinė diagrama pateikta 9 paveikslėlyje.



8 pav. Aukštesnių eilių statistika grįstas balso detektavimo algoritmas

Iš jos matyti, kad pirmiausia signalo kadras apdorojamas tiesinio prognozuojančio kodavimo. Kadro trukmė rekomenduojama 20 ms. Šiuo atveju iš 8 kHz kalbos signalo įrašo gaunama 80 signalo reikšmių viename kadre. Tiesinio prognozuojančio kodavimui (TPK) opservaciniu metodu nustatyta 10 eilė. TKP liekanos signalas gaunamas filtruojant TKP filtru. Tolesni skaičiavimai atliekami su 20 % persidengiančia TKP liekana (80 naujų signalo kadro taškų apjungiami su 20 taškų iš senesnio kadro). Norint pritaikyti sinusoidinio kalbos modelio savybes, TKP liekanos spektras privalo turėti plokščią spektrinę charakteristiką. Atlikus TKP liekanos signalo spektrinę transformaciją, matoma (10 pav.), kad TKP liekanos spektras yra pakankamai plokščias.



9 pav. TKP liekanos spektras

Iš TKP liekanos, nufiltravus aukštų dažnių dedamąsias, ši naudojama skaičiuoti kiekvieno kadro antros, trečios ir ketvirtos eilės momentam. Momentų reikšmių pokyčiai kadruose sušvelninami autoregresiniu vidurkinimu. Iš šių skaičiuojami asimetrijos ir eksceso koeficientai naudojant tokias formules [1]:

$$a_x = m_3 = E\{x^3\}$$

$$e_x = m_4 - 3m_2^2 = E\{x^4\} - 3E\{x^2\}^2$$

Čia A_x – kadro imties asimetrijos koeficientas; E_x – eksceso koeficientas;

m_2, m_3, m_4 - antros, trečios ir ketvirtos eilės momentai; E – signalo energija.

Normalizuoti asimetrijos (γ_3) ir eksceso (γ_4) koeficientai skaičiuojami pagal šias formules [1]:

$$g_3 = \frac{a_x}{m_2^{1.5}}$$

$$g_4 = \frac{e_x}{m_2^2}$$

Normalizuoti asimetrijos ir eksceso matai reikalingi, norint išvengti fluktuacijų, kai kinta signalo energija, kas nebūtinai reiškia kalbos periodo pradžią.

Kai signalo-triukšmo santykio lygis būna nedidelis (daug triukšmo) γ_3 ir γ_4 tikslumas silpnėja, todėl įvedamas papildomas parametras asimetrijos ir eksceso koeficientų santykis (*SKR*) [1]:

$$SKR = \frac{a_x^2}{e_x^{1.5}}$$

Skirtingi santykio laipsniai atsiranda dėl nevienodo signalo energijos poveikio asimetrijos ir eksceso koeficientam. *SKR* santykiu eliminuojamas signalo energijos poveikis.

2.7. Dirbtiniai neuroniniai tinklai balso detekcijoje

Dirbtinis neuroninis tinklas (DNT), tai matematinių ir algoritminių metodų rinkinys, skirtas įvairių uždavinių sprendimui. DNT atsirado siekiant sukurti dirbtines sistemas sugebančias vykdyti „protingus“ skaičiavimus panašius į tuos, kuriuos kasdien vykdo žmogaus smegenys. neuroninis tinklas – tai sistema, susidaranti iš daugelio paprastų, dirbančių lygiagrečiai skaičiavimo elementų, kurių funkcija apsprendžiama tinko struktūros, o skaičiavimai vyksta pačiuose elementuose ar mazguose. DNT naudojama kaip klasifikavimo pagal daug požymių priemonė, leidžianti suskaidyti į sritis įėjimo erdvę. Ši savybė leido

pritaikyti DNT balso signalų analizės uždaviniuose. Ypač plačiai DNT naudojamas balso fonemų atpažinimo uždaviniuose.

Balso detekcijos uždaviniuose DNT nėra naujiena. Nors industriniuose sprendimuose DNT pagrindu grįsti balso detekcijos algoritmai nėra paplitę, tačiau mokslinėse publikacijos galima sutikti keletą DNT grįstų algoritmų. Ypač plačiai cituojamas Jotaro Ikedo 1998m. pasiūlytas algoritmas [10]. Publikuojamuose balso detekcijos algoritmuose, DNT naudojamas klasifikuoti balso ir triukšmo kadrams pagal signalo analizės metu apskaičiuotus parametrus. Dažniausiai kadrų klasifikavimui naudojama daugiasluoksnio perceptrono DNT struktūra, tačiau galimos ir kitos DNT struktūros (Pvz. radialinių bazinių funkcijų DNT [11]).

3. BALSŲ DETEKCIJOS ALGORITMO TEORINIAI ASPEKTAI

3.1. Tiesinis prognozuojantis kodavimas

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

TPK metode daroma prielaida, kad kalbos signalas yra inicijuojamas garso šaltinio kintamo skerspjūvio vamzdelio viename iš galų. Žmogaus garso šaltinis yra virpančios *balso stygos*. Šis šaltinis charakterizuojamas intensyvumu (garsumu) ir dažniu (pagrindiniu tonu). Gerklės ir burnos kalbos traktas formuoja kintamo skersmens vamzdelį. Kalbos traktas charakterizuojamas rezonansiniais dažniais, vadinamais *formantėmis*.

Tiesinė prognozės metodu yra įvertinamos formantės, atskiriant jas nuo kalbą generuojančio šaltinio, lemiančio kalbos garsumą ir toną. Formančių eliminavimas iš kalbos signalo kartais vadinamas *atvirkštiniu filtravimu*, ir likęs signalas vadinamas *liekamuoju (liekana)*.

Kadangi kalbą generuojantis traktas kinta laike, tiesinė prognozė atliekama trumpais laiko tarpais, vadinamais *kadrais, langais* (frames). Kaip taisyklė, vienos sekundės trukmės kalbos signalas skaidomas į 30-50 kadru.

Pagrindinė TKP metodo problema yra įvertinti formantes remiantis įrašytu kalbos signalu. Metodu tikslas yra gauti tam tikrą diferencialinę lygtį, kurios esmė yra išreikšti kuo tiksliau eilinę garso slėgio imtį, remiantis keliomis prieš tai žinomoms kalbos signalo reikšmėmis. Kadangi naudojama tiesinė aproksimacija, tai gaunama skirtuminė diferencialinė lygtis yra vadinama tiesine prognoze, o pats signalo kodavimo metodas vadinamas *tiesinės prognozės* metodu.

Diferencialinės išraiškos koeficientai vadinami *prognozės (tiesinės) koeficientais*. Šie koeficientai aprašo formantes, taigi metodo esmė yra apskaičiuoti tiesinės prognozės koeficientus. Formulės apskaičiuoti LPC koeficientus yra gaunamos minimizuojant liekamojo signalo vidutinę kvadratinę paklaidą.

Rezultate yra gaunama tiesinė algebrinių lygčių sistema prognozės koeficientų atžvilgiu. Praktiškai tenka spręsti tokias problemas:

- Apskaičiuoti tiesinių lygčių sistemos matricos koeficientus ir laisvuosius narius pagal kalbos signalo fragmento duomenis.
- Išspręsti gautą tiesinę lygčių sistemą.

Naudojami keli matricos koeficientų gavimo metodai (autokoreliacijos, kovariacijos, rekursyvinės gardelės), kurie garantuoja sprendinio vienatį ir skaičiavimų efektyvumą.

Tarkime $\{x_n\}_{n=0,\dots,N-1}$ yra signalo fragmento duomenys. Pažymėjus $\{a_m\}_{1 \leq m \leq M}$ tiesinės prognozės koeficientus (Čia M žymi ieškomų prognozės koeficientų skaičių). Pagal tiesinės prognozės metodą eilinė duomenų reikšmė x_n yra aproksimuojama keletu prieš tai žinomų duomenų reikšmių tiesine išraiška. Tiesinės išraiškos koeficientai pasirinktam duomenų fragmentui nekeičiami (jie nepriklauso nuo indekso n). Pažymėję tiesinės prognozės paklaidą ε_n , gaunama tokia išraiška:

$$x_n = \sum_{m=1}^M a_m x_{n-m} + e_n$$

Lygtis tiesinės prognozės koeficientai $\{a_m\}_{1 \leq m \leq M}$ gaunami minimizuojant vidutinę kvadratinę paklaidą, atžvilgiu prognozės koeficientų:

$$err = \left(\sum_n x_n - \sum_{m=1}^M a_m x_{n-m} \right)^2 = \sum_n (e_n)^2$$

Minimizuojant funkciją err reikia apskaičiuoti jos išvestines kintamųjų a_m atžvilgiu ir prilyginti jas 0. Taip gaunama tokia tiesinių lygčių sistema:

$$\sum_n x_{n-k} \left(x_n - \sum_{m=1}^M a_m x_{n-m} \right) = 0, \quad k = 1, 2, \dots, M$$

Tuo atveju, kai $x_n=0$ visiems $n < 0$, pastarąją lygčių sistemą galima perrašyti taip:

$$\sum_{m=1}^M a_m R_{k-m} = R_k, \quad k = 1, 2, \dots, M$$

Čia R_n žymi $\{x_n\}$ duomenų autokoreliacijos seką:

$$R_n = \sum_k x_{n+k} x_k$$

Nežinomų prognozės koeficientų atžvilgiu gaunama algebrinė tiesinių lygčių sistema. Išsprendus šią sistemą randami prognozės koeficientai a_m , $m=1, 2, \dots, M$. Tačiau tiesioginis gautos algebrinių tiesinių lygčių sprendimo metodas nėra efektyvus. Dažniausiai naudojamas Levinsono-Durbin algoritmas, kurio efektyvumas $O(M^2)$ yra žymiai geresnis už tiesioginio sprendimo veiksmų skaičių $O(M^3)$.

Levinsono-Durbin algoritmo tinkamumas atsiskleidžia užrašius gautąją lygčių sistemą matricinine forma.

$$\mathbf{Xa} \approx \mathbf{b}$$

$$X = \begin{bmatrix} x(1) & 0 & \dots & 0 \\ x(2) & x(1) & \ddots & \vdots \\ \vdots & x(2) & \ddots & 0 \\ x(m) & \vdots & \ddots & x(1) \\ 0 & x(m) & \ddots & x(2) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & x(m) \end{bmatrix}, \quad a = \begin{bmatrix} 1 \\ a(2) \\ \vdots \\ a(p+1) \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Analizuojant gautą lygčių sistemą atžvilgiu nežinomųjų, kurie ir yra tiesinės prognozės koeficientai, matyti kad lygčių sistemos matrica yra Tioplico, t.y. matricos gretimos eilutės ir stulpeliai yra panašūs ir skiriasi tik poslinkiu.

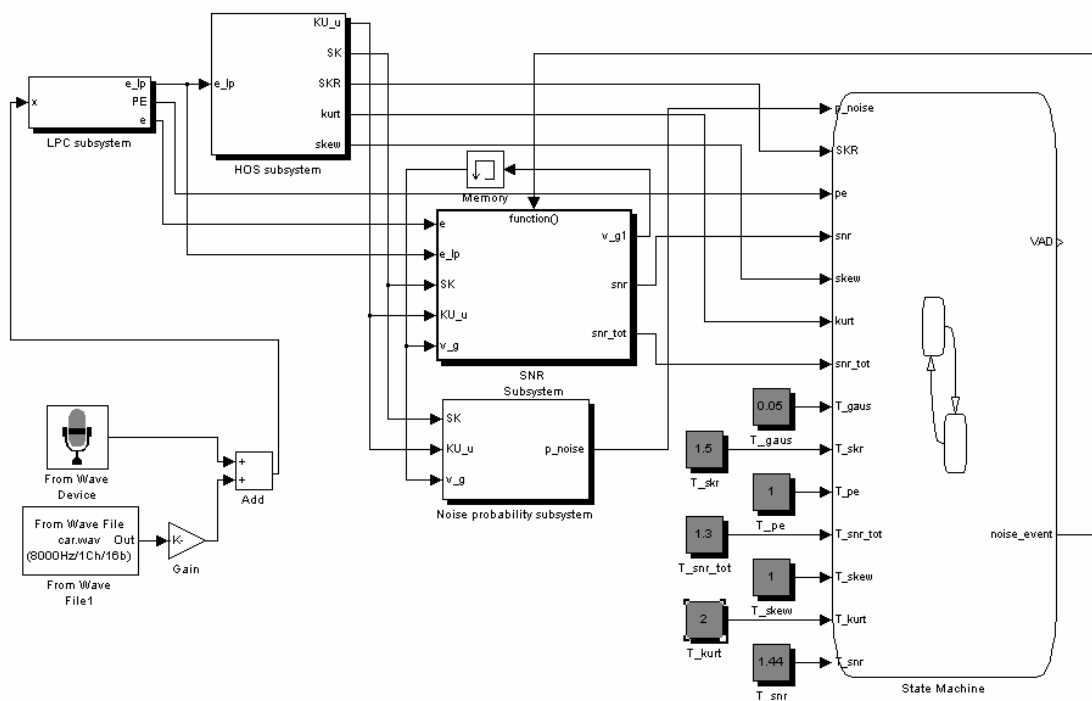
Dešinė lygčių sistemos pusė yra specifinė ir yra poslinkis pirmojo stulpelio.

Lygčių sistemom su tokia specifika spręsti egzistuoja efektyvus metodas, vadinamas autorių vardu Levinson-Durbin algoritmu. Metodas rekursyviai ieško tiesinės prognozės koeficientus, atlikus paskutinę iteraciją gaunamas galutinis sprendinys.

Levisono-Durbin algoritmo sudėtingumas yra $O(M^2)$ eilės. Tai eile mažiau, nei tiesioginio lygčių sistemos sprendimo metodo sudėtingumo, kuris gaunasi lygčių sistemą sprendžiant kintamųjų eliminavimo metodu.

3.2. Algoritmo modelis

Algoritmo modelio sudarymui pasirinkta naudoti grafinė realaus laiko sistemų modeliavimo aplinka „SIMULINK“, kuri yra neatsiejama dalis „MATLAB“ paketo. Ši modeliavimo aplinka pasižymi greitu sistemos modelio kūrimo procesu. Algoritmas suskaidomas į vienareikšmiškas funkcijas atliekančius modulius, kurie hierarchinių „SIMULINK“ blokų pagalba apjungiami. Toks sistemų kūrimo ir modeliavimo būdas leidžia greitai surasti ir ištaisyti modelio ir tuo pačiu algoritmo spragas. 11 paveikslėlyje pateiktas aukštesniųjų eilių statistika grįsto algoritmo aukščiausio hierarchinio lygmens „SIMULINK“ modelio blokinė schema. Ji susideda iš keturių hierarchinių posistemų, baigtinio automato posistemės su slenkstinėmis reikšmėmis ir signalo įvedimo sąsajos. Kiekviena posistemė atspindi dalies algoritmo skaičiavimus. Posistemų viduje yra „SIMULINK“ funkciniai blokai atliekantys tai posistemiai nustatytus skaičiavimus arba posistemė sudaliname hierarchiškai į mažesnes posistemas.

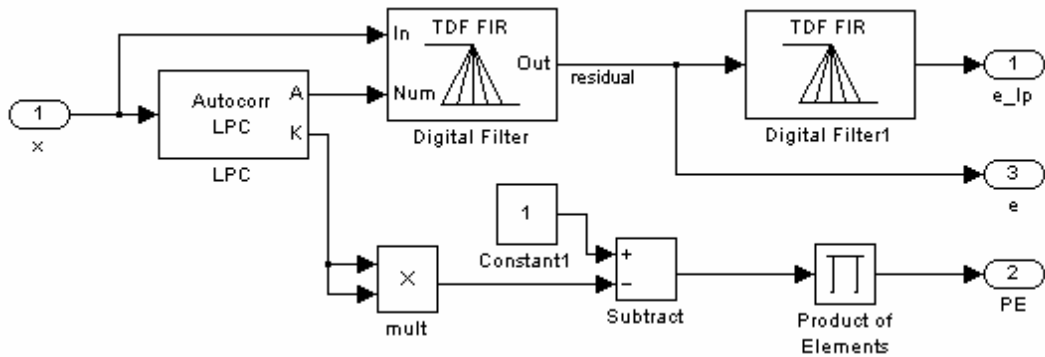


10 pav. Aukščiausio hierarchinio lygmens simulink modelio blokinė schema

Įėjime, algoritmo modeliui pajungta signalo įvedimo sąsaja susidedanti iš mikrofono sąsajos ir garso failo duomenų įvedimo sąsajos. Šie „SIMULINK“ blokai leidžia prie modelio pajungti mikrofoną per kompiuterio garso plokštės sąsają. Tokiu būdu algoritmo modeliavimas atliekamas realiaje laike, realiomis sąlygomis, skirtingais SNR lygiais. Sujungimai tarp posistemių vaizduoja duomenų srautus tarp algoritmo modulių. Sekančiuose poskyriuose pateikiama modelio posistemių struktūra.

3.2.1. Tiesinio prognozavimo posistemė

Signalo įvedimo sąsaja (mikrofonas ir garso failo duomenų įvedimo sąsaja) teikia 160 reikšmių įėjimo signalo duomenų kadrus tiesinio prognozavimo skaičiavimo posistemiai. 12 paveikslėlyje pavaizduota šios posistemės struktūra. Čia įėjimo signalas x yra kalbos arba triukšmo signalas, sujungtas su tiesinės prognozės analizės bloku (*LPC*) parinktu iš standartinių „SIMULINK“ funkcinių blokų bibliotekos.

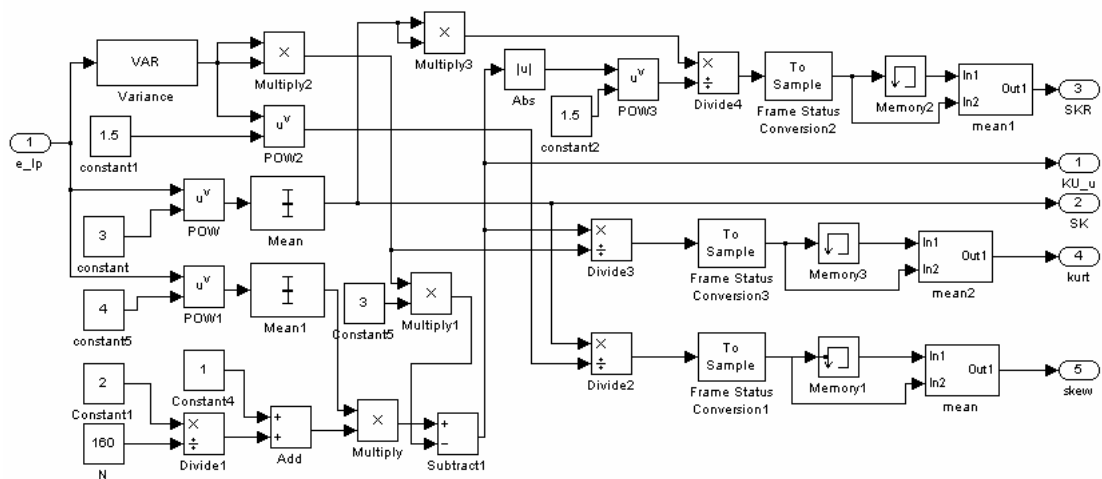


11 pav. Tiesinio prognozavimo posistemė

Bloko veikimas pagrįstas autokoreliaciniu metodu. Tokiu metodu atliekant dešimtos eilės tiesinę prognozę, gaunami atspindžio koeficientai (K) ir tiesinės prognozės koeficientai (A). Pastarieji yra naudojami tiesinės prognozės liekanai gauti, filtruojant įėjimo signalą FIR filtru (*Digital Filter*). Antrasis skaitmeninis filtras (*Digital Filter 1*) naudojamas kaip dažninis filtras, tiesinės prognozės liekanos žemų dažnių juostai (iki 1,8 KHz) gauti. Atspindžio koeficientai naudojami, tiesinės prognozės paklaidai suskaičiuoti.

3.2.2. Aukštesnių eilių statistikos skaičiavimo posistemė

Tiesinės prognozės liekanos žemų dažnių juosta teikiama tolimesniems skaičiavimams į aukštesnių eilių statistikos skaičiavimo posistemę, kurios struktūra pateikta 13 paveikslėlyje.



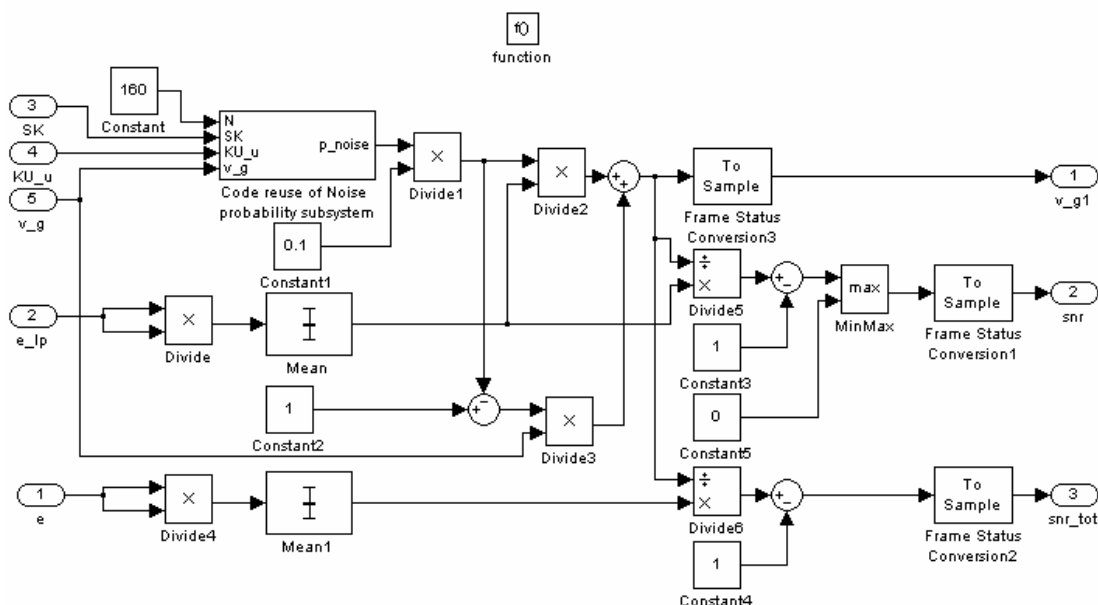
12 pav. Aukštesnių eilių statistikos skaičiavimo posistemė

Šioje posistemėje naudojant 2.7 skyrelyje pateiktomis išraiškomis skaičiuojami sklaidos (*skew*) ir eksceso koeficientai (*kurt*), signalo energija normalizuotos jų versijos (*SK* ir *KU_u*) ir

atitinkamas jų santykis (*SKR*). Šioje sistemoje naudojami tik funkciniai blokai iš standartinės „SIMULINK“ bibliotekos. Gilesnių hierarchinių lygių nėra.

3.2.3. Signalas – triukšmas santykio skaičiavimo posistemė

SNR skaičiavimus atliekanti posistemė skiriasi nuo kitų tuo kad skaičiavimai joje vykdomi priklausomai nuo sužadavimo signalo. Išorinis signalas išeinantis iš baigtinio automato posistemės įjungia skaičiavimus posistemėje tuomet, kai signalo kadras klasifikuojamas kaip triukšmas. Toks veikimas reikalingas norint tiksliau skaičiuoti triukšmo energijos reikšmę. Posistemės blokinė diagrama pateikta 14 paveikslėlyje.



13 pav. Signalas – triukšmas santykio skaičiavimo posistemė

SNR posistemėje apskaičiuojami šie parametrai:

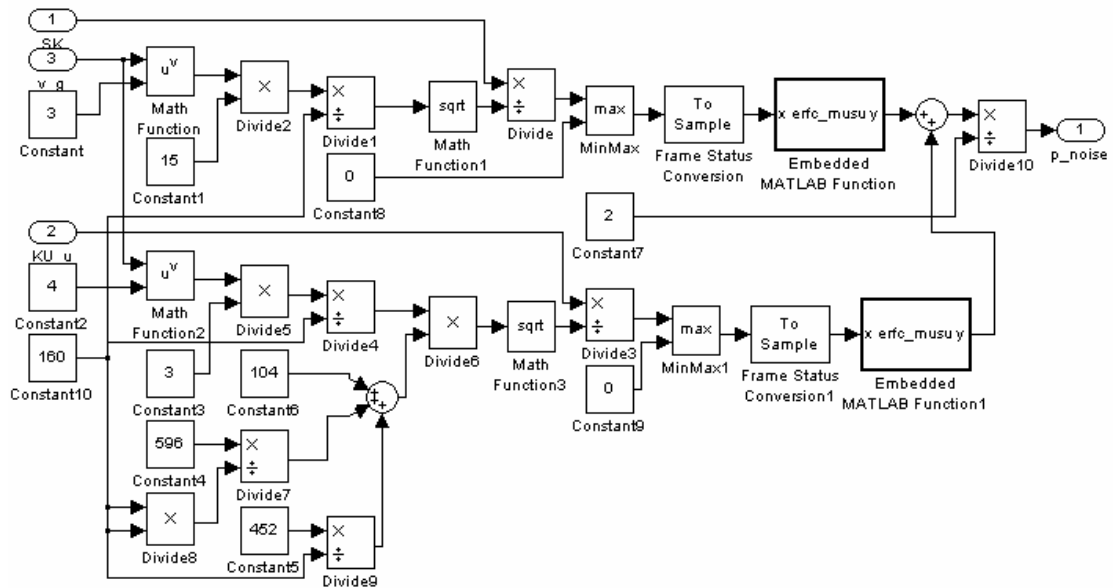
- *Snr* - apatinės dažnių juostos SNR,
- *snr_tot* - visos dažnių juostos SNR,
- *v_g1* - triukšmo energijos reikšmė.

Šioje sistemoje iškviečiama triukšmo tikimybės posistemės funkcijos.

3.2.4. Triukšmo tikimybės skaičiavimo posistemė

Aukštesnių eilių statistiniai matai, apskaičiuoti 3.3.2 skyrelyje pateiktoje posistemėje, naudojami nustatyti tikimybei, kad kadre esantis signalas yra Gausinis triukšmas. Šios

posistemė blokinė schema pateikta 15 paveikslėlyje. Vienintelis posistemės išėjimas yra gausinio triukšmo tikimybė apskaičiuojama pagal 2.7 skyrelyje pateiktas formules. Posistemėje panaudota įterptinė „MATLAB“ paketo funkcija – komplementarinė klaidos funkcija, integruojanti Gausine kreivę nuo nustatytos reikšmės.



14 pav. Triukšmo tikimybės skaičiavimo posistemė

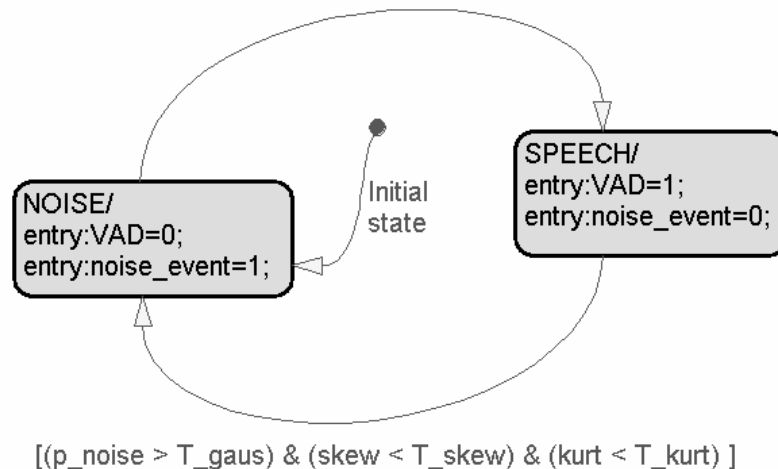
3.2.5. Baigtinio automato posistemė

Šios posistemės paskirtis imituoti dviejų būsenų baigtinio automato darbą. Posistemės (16 Paveikslėlis) kūrimui panaudoti „MATLAB *stateflow*“ įrankiai.

```

[(p_noise < T_gaus) |
 (SKR < T_skr & (snr > T_snr | pe < T_pe)) |
 (snr_tot > T_snr_tot) ]

```



15 pav. Baigtinio automato posistemė

Iėjimai baigtinio automato posistemiai paduodami iš aukščiau skyreliuose pateiktų posistemių ir slenkstinių reikšmių konstantų. Įėjimo parametrai tarnauja kaip būsenų perėjimo sąlygų kintamieji tarp triukšmo ir kalbos būsenų. Pradinė baigtinio automato būsena nurodoma „triukšmas“ specialia anotacija.

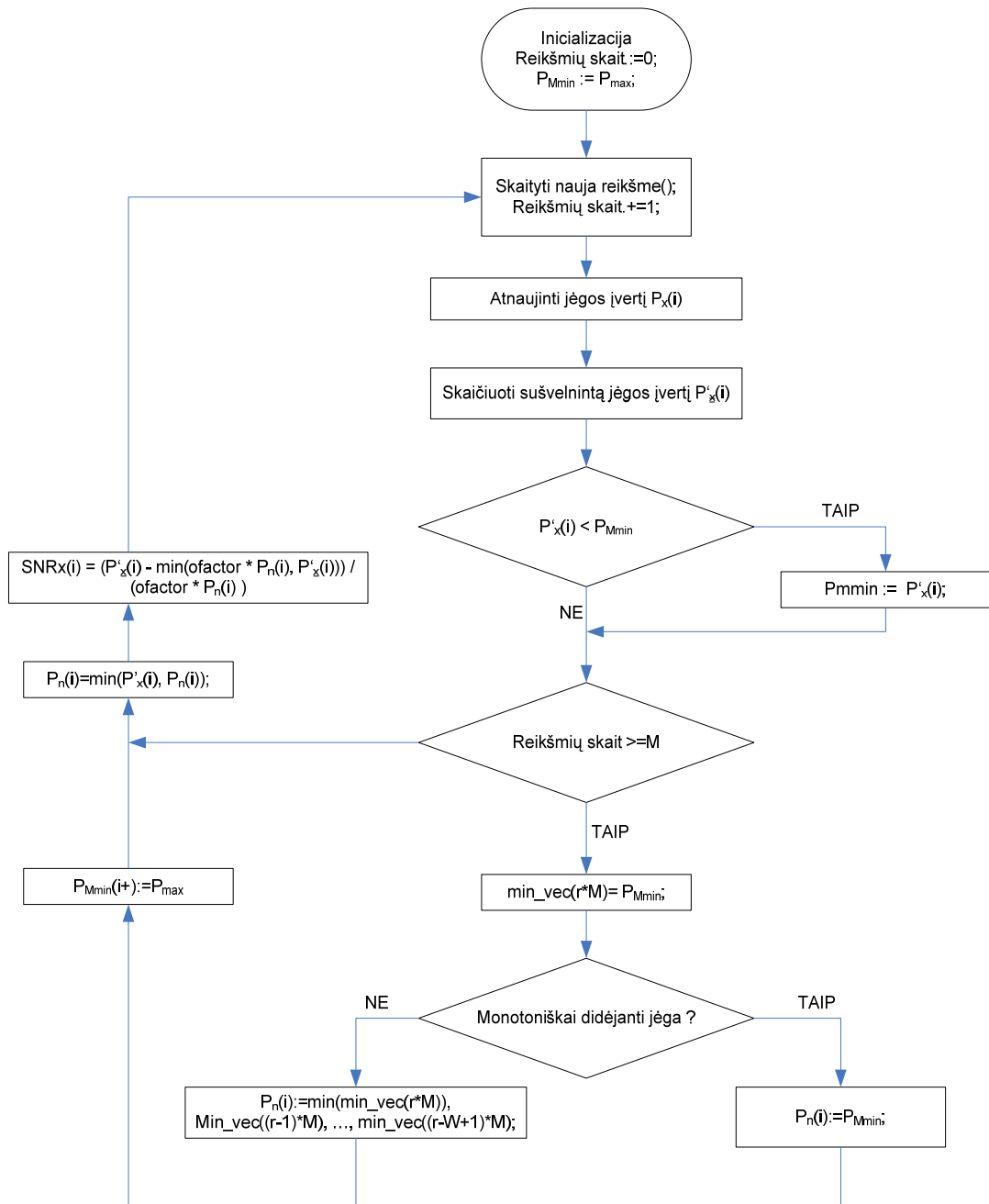
Šioji posistemė turi du dvejetainius išėjimus, kadro klasifikavimo reikšmė ir specialų išėjimą – triggerį, signalas triukšmas santykio skaičiavimo posistemiai.

3.3. Algoritmo patobulinimai

3.3.1. Momentinis SNR nustatymo algoritmas

Pirminiam aukštesnių eilių statistika grįstam algoritme SNR nustatymo proceso veiksmingumas kelia pagrįstų abejonių. Triukšmo energija apskaičiuojama tuomet, kai signalo kadras laikomas triukšmu. Šiuo atveju triukšmo energijos apskaičiavimo teisingumas tampa priklausomas nuo kadro detektavimo korektiškumo. Tačiau pats kadru klasifikavimas vyksta remiantis tuo pačiu SNR, priklausančiu nuo signalo triukšmo energijos. Tokia abipusė priklausomybė gali privesti prie neteisingų SNR reikšmių apskaičiavimo ir tuo pačiu mažina algoritmo galimybes teisingai detektuoti balso kadrus. Eksperimentiniai SNR rezultatų tyrimai patvirtina šio skaičiavimo ydingumą.

Norint pagerinti algoritmo detektavimo rezultatus, reikia pakeisti SNR nustatymo algoritmą efektyvesniu, nuo kadru klasifikacijos nepriklausančiu algoritmu. Vienas iš tinkamų ir kaip mokslinėse publikacijose teigiama, efektyvių algoritmų yra momentinis SNR nustatymo algoritmas [8]. Šis algoritmas tinkamas tik balso signalų SNR nustatymui. Algoritmo blokinė schema, atskleidžianti jo veikimo principą pateikta 17 paveikslėlyje.



16 pav. Momentinio SNR nustatymo algoritmo blokinė diagrama

SNR įvertis skaičiuojamas kiekvienai įėjimo signalo reikšmei, pagal žemiau pateiktą formulę:

$$SNR = \frac{P_x - \min(\text{ofactor} \cdot P_n, P_x)}{\text{ofactor} \cdot P_n}$$

Čia P_x – sušvelnintas jėgos įvertis; P_n – triukšmo įvertis, ofactor – įverčio konstanta.

Papildomai naudojama įverčio konstanta *ofactor*, nes tikroji triukšmo jėgos reikšmė yra didesnė nei minimali jėgos reikšmė. *Ofactor* konstanta parenkama nuo 1.3 iki 2.

3.3.2. Adaptyvios slenkstinės reikšmės

Išanalizavus klasifikavimo baigtiniam automatui teikiamų signalo parametrų reikšmių priklausomybę nuo signale esančio triukšmo lygio pastebėta, kad aukštesnių eilių statistiniai parametrai: asimetrijos ir eksceso koeficientai yra nepriklausomi nuo triukšmo lygio signale. Vieninteliai priklausomi nuo triukšmo lygio yra signalas – triukšmas santykį apibūdinantys parametrai (momentinė SNR reikšmė, pilnos ir žemų dažnių juostos signalo SNR). Kadangi signalo SNR pagrindu priklauso nuo aplinkos, kurioje balso signalas yra formuojamas, SNR parametram reikalingos adaptyvios slenkstinės reikšmės. Slenkstinės reikšmės yra skaičiuojamos pagal šią formulę.

$$T_{SNR} = C + \frac{1}{N} \sum_{i=0}^{N-1} K_i$$

Čia T_{SNR} - slenkstinė reikšmė, C - adityvi konstanta, K_i - i -tojo triukšmo kadro energija, N - kadro kiekis.

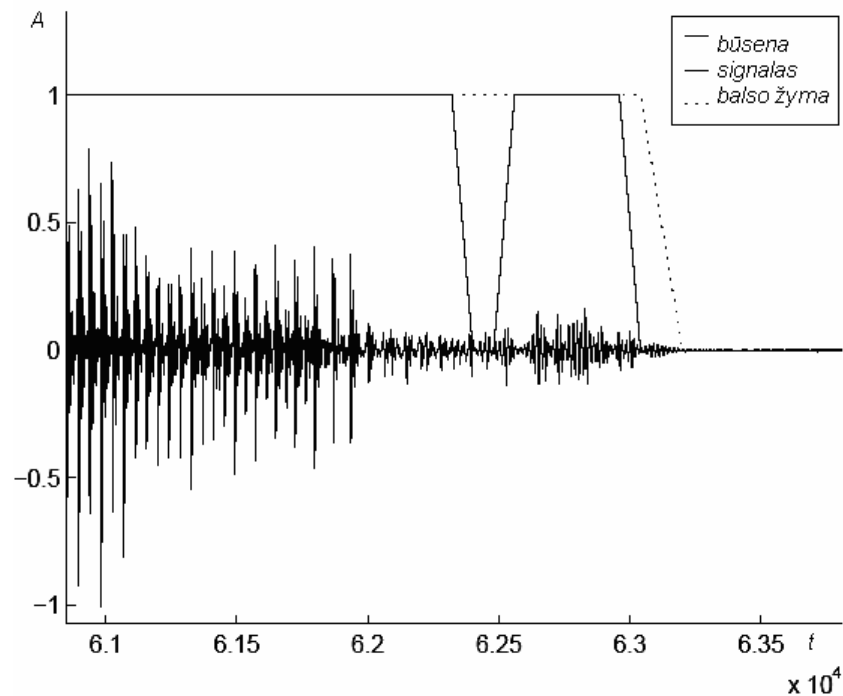
Pagal pateiktą formulę slenkstinė reikšmė apskaičiuojama kaip praėjusių triukšmo kadro energijos vidurkis pridėjus adityvią konstantą. Optimalus kadro kiekis ir adityvios konstantos reikšmės surastos observaciniu metodu ir pateiktos 3 lentelėje.

3 lentelė. Kadro ir adityvios konstantos reikšmės

| Slenkstinė reikšmė | Triukšmo kadro skaičius | Adityvios konstantos reikšmė |
|---------------------------|-------------------------|------------------------------|
| Momento SNR | 5 | 0,5 |
| Pilnos dažnių juostos SNR | 7 | 0,9 |
| Žemų dažnių juostos SNR | 7 | 3 |

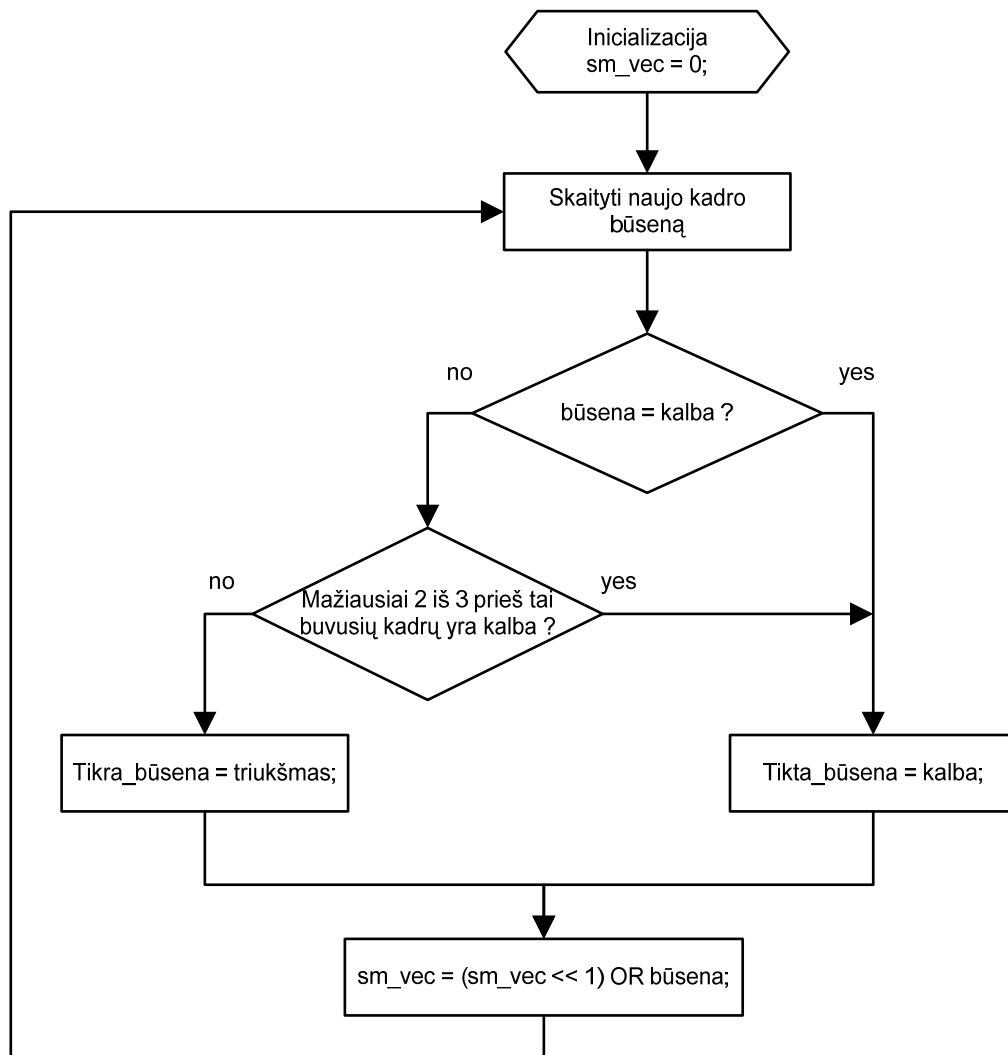
3.3.3. Kadru klasifikavimo tiesinimas

Stebinant aukštesnių eilių statistika grįsto algoritmo modelio rezultatus pastebėta, kad klasifikavimą atliekantis baigtinis automatas turi tendenciją pereiti iš balso į triukšmo ir vėl atgal iš triukšmo į balso būsenas vientisoje kalbos frazėje. Ši tendencija turi neigiamą įtaką kalbos detekcijos rezultatams (18 paveikslėlis).



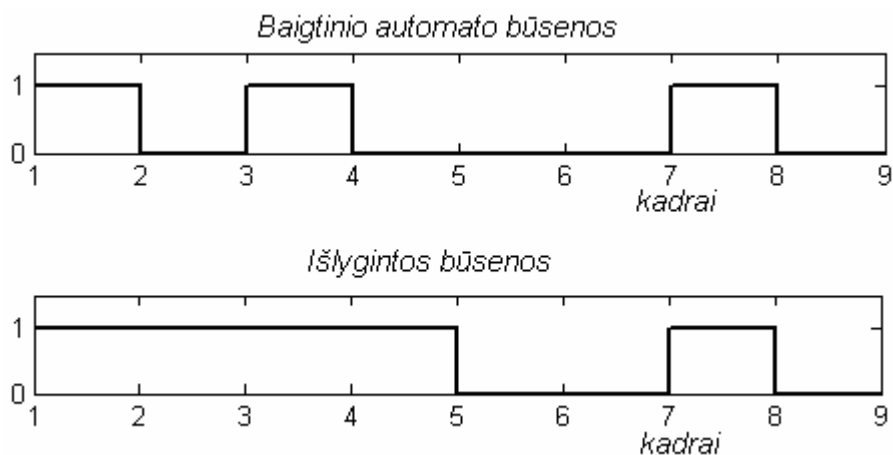
17 pav. Būsenų nestabilumo pavyzdys

Tam kad algoritmas kadrus klasifikuotu stabiliau, be reikalo neįterpdamas triukšmo kadra į vientisą kalbos frazę, reikalingas algoritmo patobulinimas – kadru klasifikavimo tiesinimas. Esamo kadro klasifikavimui įtaką turi daryti ką tik klasifikuotų kadru reikšmės. 19 paveikslėlyje pateikiama kadru klasifikavimo tiesinimo algoritmas, veikiantis realiame laike.



18 pav. Kadru klasifikavimo tiesinimo algoritmas

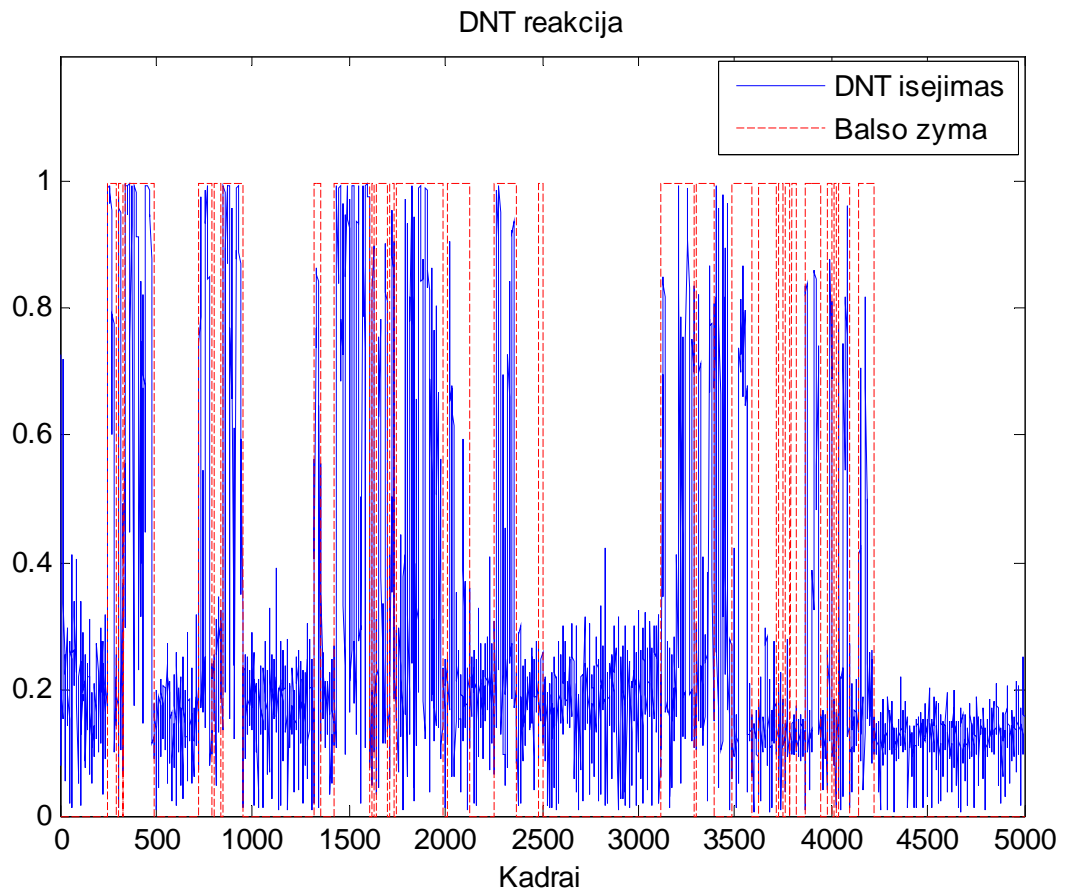
Kai kadru klasifikavimo baigtinis automatas pereina iš kalbos į triukšmo būseną, algoritmas analizuoja tris ankstesnių kadru reikšmes. Jeigu du iš trijų ankstesnių kadru buvo klasifikuojami kaip kalba, tai esamo kadro būseną priskiriama kalbai. Algoritmas nesudėtingas tačiau efektyvus, nes turi savybę po kiekvienos kalbos frazės, kalbai priskirti dar vieną papildomą kadru (20 paveikslėlis). Tokiu būdu yra realizuojamas netik kadru klasifikavimo nestabilumo tiesinimas, bet ir užlaikymo periodas (*ang.k. hangover period*).



19 pav.. Klasifikavimo lyginimo pavyzdys

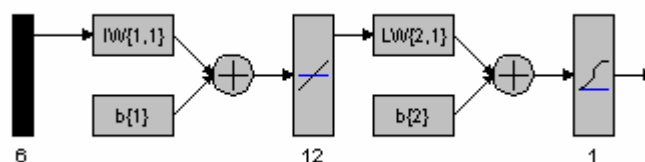
3.3.4. Neuroninis tinklas

Dirbtiniai neuroniniai tinklai gali sėkmingai atlikti klasifikavimo uždavinius. Siekiant pagerinti balso detektavimo algoritmo kadro klasifikavimo rezultatus, buvo numatyta DNT panaudoti vietoj klasifikavimo baigtinio automato. Tačiau modeliuojant tokia algoritmo variaciją buvo pastebėta, kad DNT atpažinimo rezultatus tik pablogina. Balso ir triukšmo kadro klasifikavimo degradaciją lemia balsą ir triukšmą charakterizuojančių parametrų specifika. Signalų aukštesnių eilių statistiniai skaičiavimai netiksliai apibūdina signalo turinį ir dažniausiai nusako tik balsinių garsų buvimą signale. Kalbos frazėse tokie garsai dažniausiai pasitaiko žodžio pradžioje ir lemia baigtinio automato būsenos pakeitimą iš „triukšmas“ į „balsas“. Kalbos pabaigai būdingi parametrai aktyvuoja perėjimą iš „balsas“ į „triukšmas“. Tokios savybės visiškai nebūdingos DNT, todėl jo naudojimas duoda žymiai prastesnius klasifikavimo rezultatus nei baigtinis automatas (21 paveikslėlis).



20 pav. DNT reakcija į signalo parametrus.

Modeliuojant balso detekcijos algoritmą su įvairiomis DNT struktūromis pastebėta, kad vienos krypties atgalinės grįžties (feed-forward backpropagation) DNT išėjimas pakankamai gerai apibūdina balso kadrus. Šią savybę nuspręsta panaudoti kaip išvestinį parametą formuojantį baigtinio automato būsenos kaitą iš „triukšmas“ į „balsas“. Detalesnė tinklo struktūra pateikiama 22 paveikslėlyje.



21 pav. DNT struktūra

Tinklas turi tris dirbtinių neuronų (DN) sluoksnius. Pirmasis vadinamas įėjimo sluoksniu ir turi šešis DN atitinkančius DNT įėjimui paduodamus signalo parametrus. Vidinis sluoksnis vadinamas paslėptuoju ir turi 12 DN. Šis kiekis parinktas observaciniu metodu, kadangi nėra

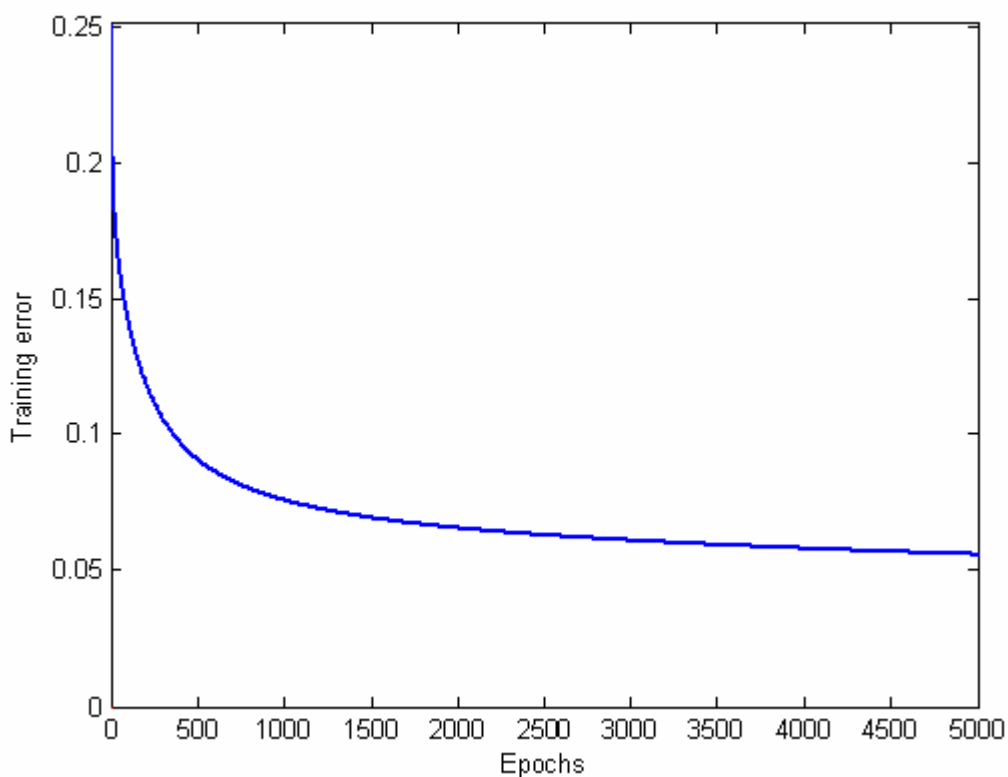
DNT metodikos nurodančios paslėpto sluoksnio DN skaičiaus parinkimą. Išėjimo sluoksnio dirbtiniam neronui parinkta sigmoidinė (loginė) perėjimo funkcija tam, kad išėjimo reikšmės būtų [0;1] ribose.

$$Y = \frac{1}{1 + e^{-a(s-w_0x_0)}}$$

Vidinio sluoksnio perėjimo funkcija yra tiesinė, kadangi įėjimo parametrai teoriškai gali įgyti neriboto dydžio reikšmes.

DNT apmokomas gradientiniu metodu, naudojant 500 kadrų signalo parametrų pavyzdžius ir tiksliniam vektoriui naudojant rankiniu būdu teisingai sužymėtus balso ir triukšmo kadrus.

Tinklas apmokomas per 5000 epochų, pasiekiant 0,056 apmokymo klaidos reikšmę. Apmokymui naudojant daugiau epochų, apmokymo klaida kinta nežymiai. Apmokymo grafikas pateiktas 23 paveikslėlyje.



22 pav. DNT apmokymo klaidos grafikas

4. EKSPERIMENTINIS TYRIMAS

4.1. Algoritmo realizacijos metodika

Algoritmo gebėjimą atpažinti balso signalą, apytiksliai galima įvertinti naudojantis „MATLAB“ ir „SIMULINK“ terpėje sukurtais modeliais. Tačiau modelių simuliacija naudojant ilgus garso signalo įrašus užima daug laiko. Siekiant efektyviai ir kuo tiksliau, įvairiomis sąlygomis eksperimentiškai įvertinti balso detektavimo algoritmą ir pasiūlytus patobulinimus, buvo suprogramuota keleta algoritmo realizacijų.

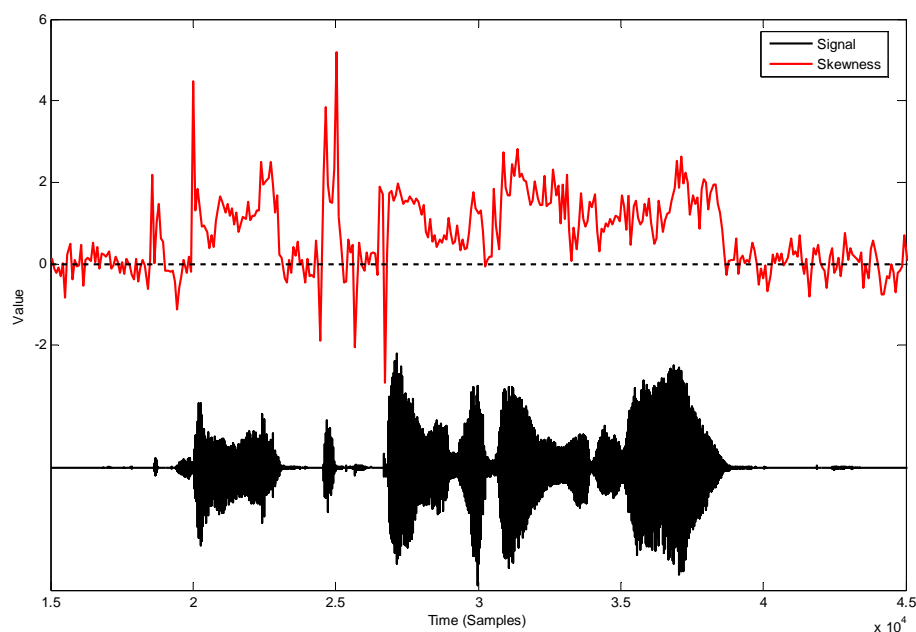
Vienas iš greičiausių būdų modelį paversti pilnaverte programinę realizaciją yra pasinaudoti „SIMULINK Real-time workshop“ priedu. Šis „SIMULINK“ įrankis gali sugeneruoti ir sukompiluoti programinį kodą populiariausiems DSP procesoriams. Tokiu būdu užtenka vieno mygtuko paspaudimo ir modelis pradeda veikti realiame procesoriuje. Tai pirmas būdas įsitikinti algoritmo veikimo teisingumu realiomis sąlygomis. Tačiau šis būdas turi ir trūkumų. Atlikus realizacijų greitaveikos ir užimamos atminties palyginimus pastebėta, kad sugeneruojamas kodas nėra efektyvus. Realizacija veikia palyginti lėtai ir užima daug vietos. Atlikti kodo optimizavimą pakankamai sudėtinga, o kaikuriais atvejais visai neįmanoma, dėl naudojamų specifinių „Simulink“ funkcijų bibliotekų. Be to šis būdas netinkamas gauti ir apibendrinti balso atpažinimo rezultatams.

Naudojantis „Texas Instrument Code Composer Studio“ programine įranga, algoritmas realizuotas TMS320C6713 DSP procesoriuje. Šioje realizacijoje programuojant siekta maksimaliai išnaudoti TI C67XX DSP architektūros privalumus. Išnaudoti DMA ir L2 spartinančiosios atminties privalumai, signalų apdorojimo skaičiavimams panaudotos optimizuotos funkcijų bibliotekos. Pilnai funkcionuojanti programinė realizacija naudoja mažiau procesoriaus ir atminties resursų. Ši realizacija ar jos dalys gali būti tiesiogiai panaudotos balso apdorojimo uždaviniuose.

Eksperimentinio tyrimo tikslui algoritmo realizacija buvo suprogramuota „Borland C++ Builder“ programiniu paketu. Gautas rezultatas – programa veikianti personaliniame kompiuterį, Windows terpėje, skirta greitai optimalių slenkstinių reikšmių paieškai ir balso atpažinimo rezultatams gauti. Programoje pilnai realizuoti visi pasiūlyti patobulinimai.

4.2. Kalbos signalo charakteristikų tyrimas

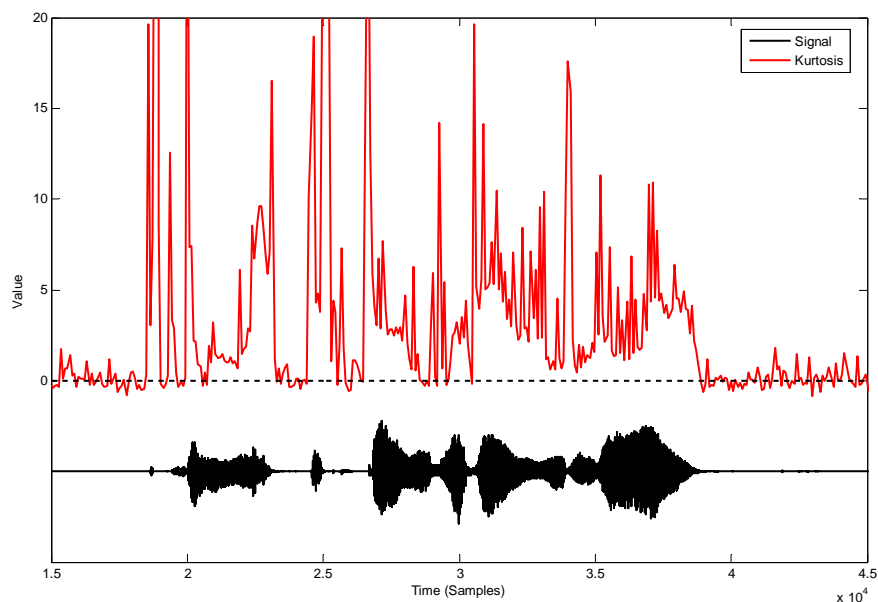
Šiame skyrelyje pateikiamos kalbos signalą charakterizuojančių parametų rezultatai. Nuo šių charakteristikų atitikimo teoriniams pastebėjimams priklauso algoritmo klasifikavimo teisingumas. Iš šių parametų klasifikavimo, baigtinis automatas apsprendžia ar kadre yra triukšmas ar balsas. Žemiau pateikiami rezultatai gauti modeliuojant ir eksperimentuojant su aukštesnių eilių statistika grįsto algoritmo realizacija.



23 pav. Asimetrijos koeficiento priklausomybė nuo balso signalo

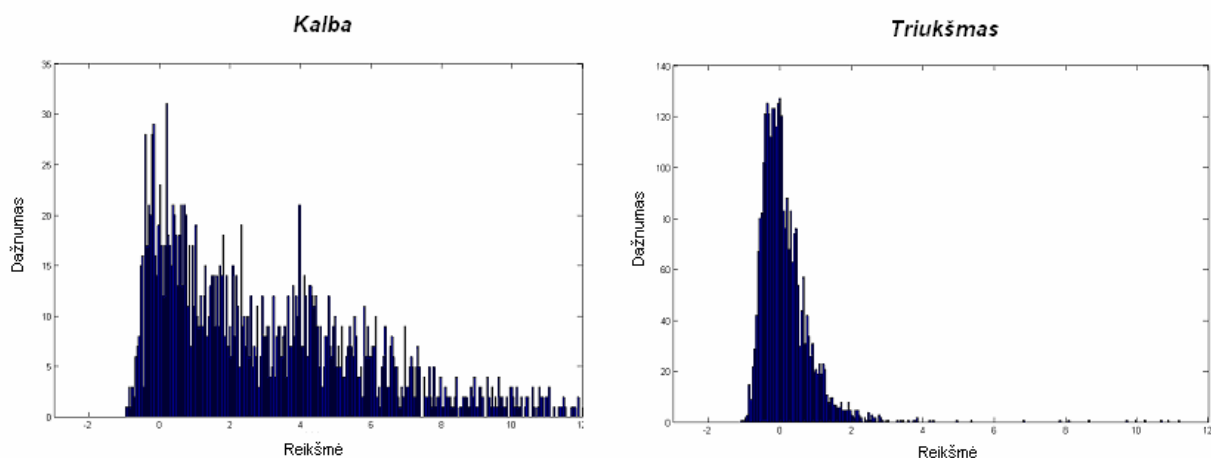
24 paveikslėlyje pateikiama signalo asimetrijos koeficiento kitimo laike charakteristika lyginant su balso signalu. Paveikslėlyje matoma, kad nesant balso signalui asimetrijos koeficientas svyruoja apie nulį, tačiau vos tik atsiradus balsui signale, asimetrijos koeficiento reikšmės ženkliai pakinta. Eksperimentuojant su skirtingomis SNR reikšmėmis nustatyta, kad asimetrijos koeficientas yra nepriklausomas nuo triukšmo kiekio signale.

Analogiška yra ir eksceso koeficiento charakteristika (25 paveikslėlis). Ekscesas svyruoja apie nulį nesant balso signalui, tačiau jam atsiradus tampa ženkliai teigiamas. Šis parametras taip pat nepriklauso nuo SNR.



24 pav. Eksceso koeficiento priklausomybė nuo balso signalo

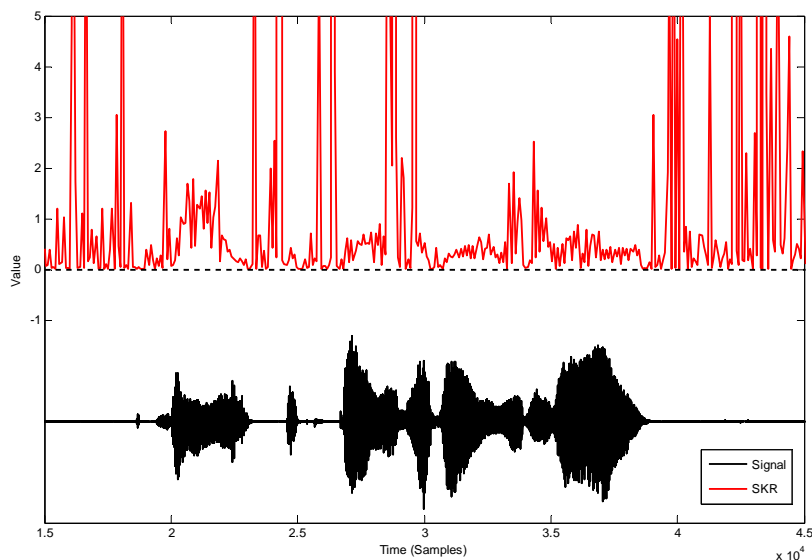
26 paveikslėlyje matomos eksceso reikšmių histogramos vienodo ilgio kalbos ir triukšmo signalui. Triukšmo signalui panaudotas važiuojančio automobilio salone įrašytas triukšmas. Nepaisant to, kad automobilio keliamas triukšmas turi daug periodinių komponentių, histogramoje matoma, kad triukšmo eksceso koeficiento pasiskirstymas panašus į Gausinį. Kalbos signalo eksceso histogramoje matoma, kad nemažai eksceso reikšmių svyruoja apie nulį, tai yra dėl to kad kalbos frazėse yra tarpai.



25 pav. Kalbos ir triukšmo signalų eksceso koeficientų histogramos.

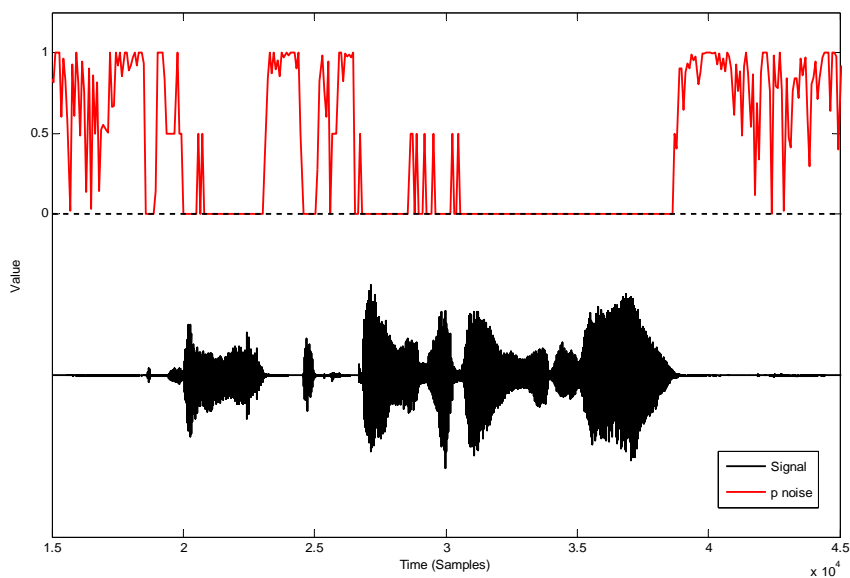
Asimetrijos ir eksceso koeficientų atitinkamų laipsnių santykio (SKR) priklausomybės nuo balso signalo grafikas pateiktas 27 paveikslėlyje. Šis santykis yra visada teigiamas ir teoriškai turėtų būti nedidesnis už vienetą esant balso signalui, ir neapibrėžtai didelis jam

nesant. Ši tendencija pastebima ir eksperimentiniuose rezultatuose, tačiau balse esant nebalsiniam garsui SKR gali viršyti vieneta. Nesant balsui SKR taip pat gali įgyti reikšmes paprastai būdingas balso signalui. Kadangi SKR priklauso nuo asimetrijos ir eksceso koeficientų, todėl šis santykis taip pat nepriklausomas nuo triukšmo kiekio signale.



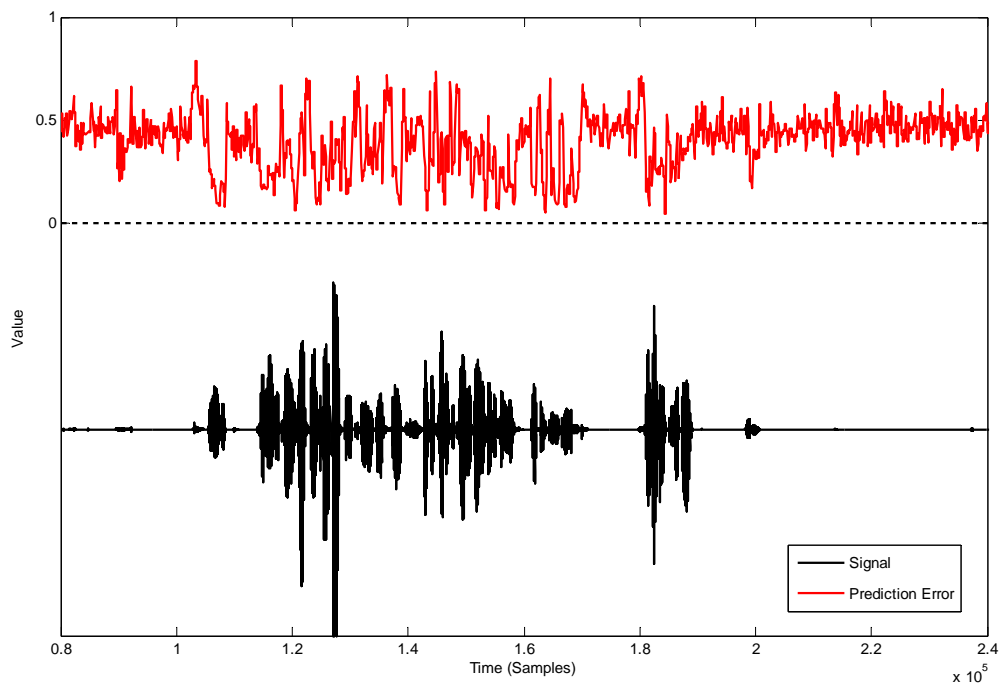
26 pav. SKR priklausomybė nuo balso signalo.

Parametro apibūdinančio gausinio triukšmo tikimybę signale grafikas pateiktas 28 paveikslėlyje. Iš grafiko matoma kad šis parametras pakankamai tiksliai aptinka balso dalis signale. Eksperimentuojant nustatyta, kad šis parametras nepriklauso nuo SNR.



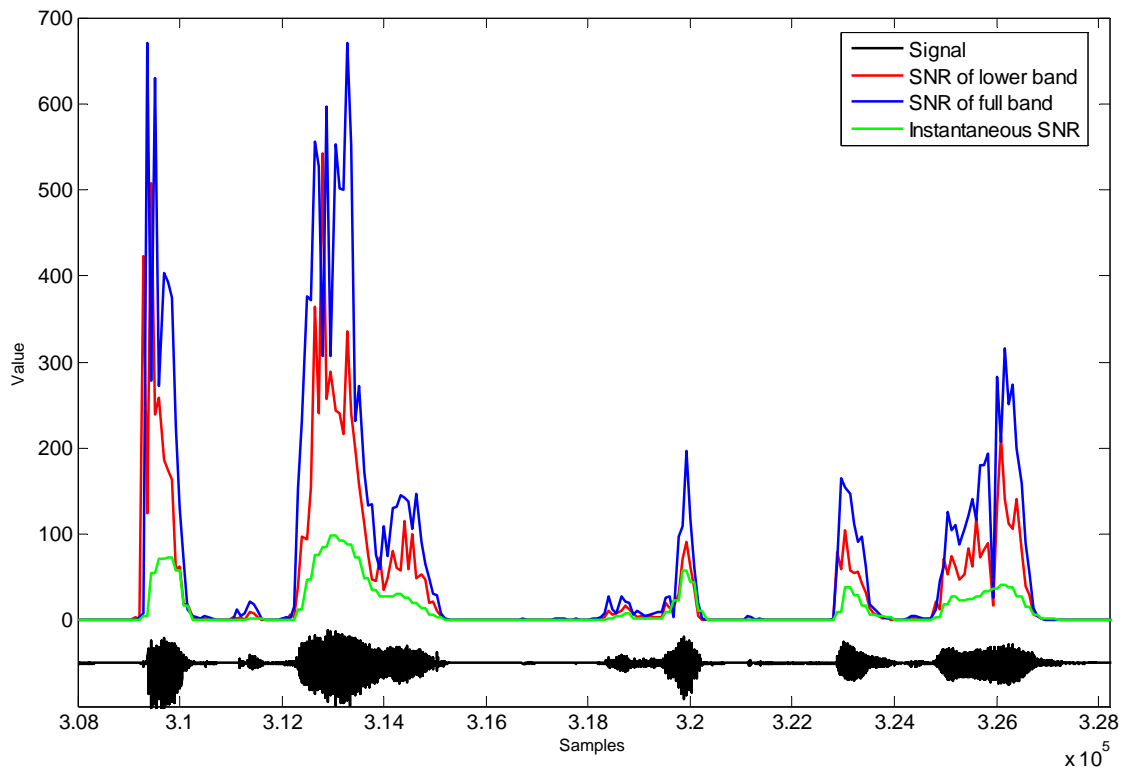
27 pav. Gausinio triukšmo tikimybės signale grafikas.

Grafikas, pateiktas 29 paveikslėlyje, vaizduoja TPK paklaidos reikšmių kitimą. Esant balso signalui, TPK paklaida paprastai turi mažesnes reikšmes nei esant triukšmui. Tačiau esant nebalsiniam garsui, TPK paklaidos reikšmės patenka į triukšmui būdingų TPK paklaidos reikšmių intervalą.



28 pav. TPK paklaidos grafikas

30 paveikslėlyje palyginimui pateikiami visi signalas-triukšmas santykį apibūdinantys įverčiai. Kaip ir galima tikėtis, SNR pilnos dažnių juostos yra nežymiai didesnis nei žemų dažnių juostos SNR. Skirtumas nedidelis, nes paprastai balso signalas turi mažai aukštesnių nei 1.8 KHz dažnio komponentių. Grafike pastebimas ryškus momentiniu SNR nustatymo algoritmo gautų įverčių skirtumas. Tai yra todėl, kad momentinis SNR nustatymo algoritmas skaičiuoja SNR įvertį kiekvienai signalo reikšmei, o ne kadru. Palyginus su tikru, skaičiavimais nustatytu SNR, šis algoritmas tiksliausiai įvertina SNR reikšmę.



29 pav. SNR grafikai

4.3. Modifikacijų rezultatai

Šiame skyrelyje apibendrinami aukštesnių eilių statistika grįšto balso detektavimo algoritmo patobulinimų eksperimentiniai rezultatai ir palyginami su straipsnyje publikuoto algoritmo[1] rezultatais. Kadangi publikuotame straipsnyje[1] algoritmo rezultatai pateikiami ne visiems triukšmo tipams, kurie nurodyti TIA/EIA-136-250 balso detektavimo algoritmo įvertinimo rekomendacijoje, todėl palyginime naudojami publikuoto algoritmo rezultatai gauti eksperimentiškai. Eksperimentiniai rezultatai nevisiškai sutampa su straipsnyje publikuojamais, nes algoritmo realizacijoje naudojamas kitoks tiesinės prognozės analizės metodas, be to straipsnyje nenurodomos tikslios slenkstinės reikšmės, klasifikavimo baigtiniam automatui. Slenkstinės reikšmės tiesiogiai lemia balso detektavimo rezultatus, todėl eksperimentuose panaudotos slenkstinės reikšmės surastos specialiai sukurta programine įranga ir teikiančios įmanomai geriausius rezultatus

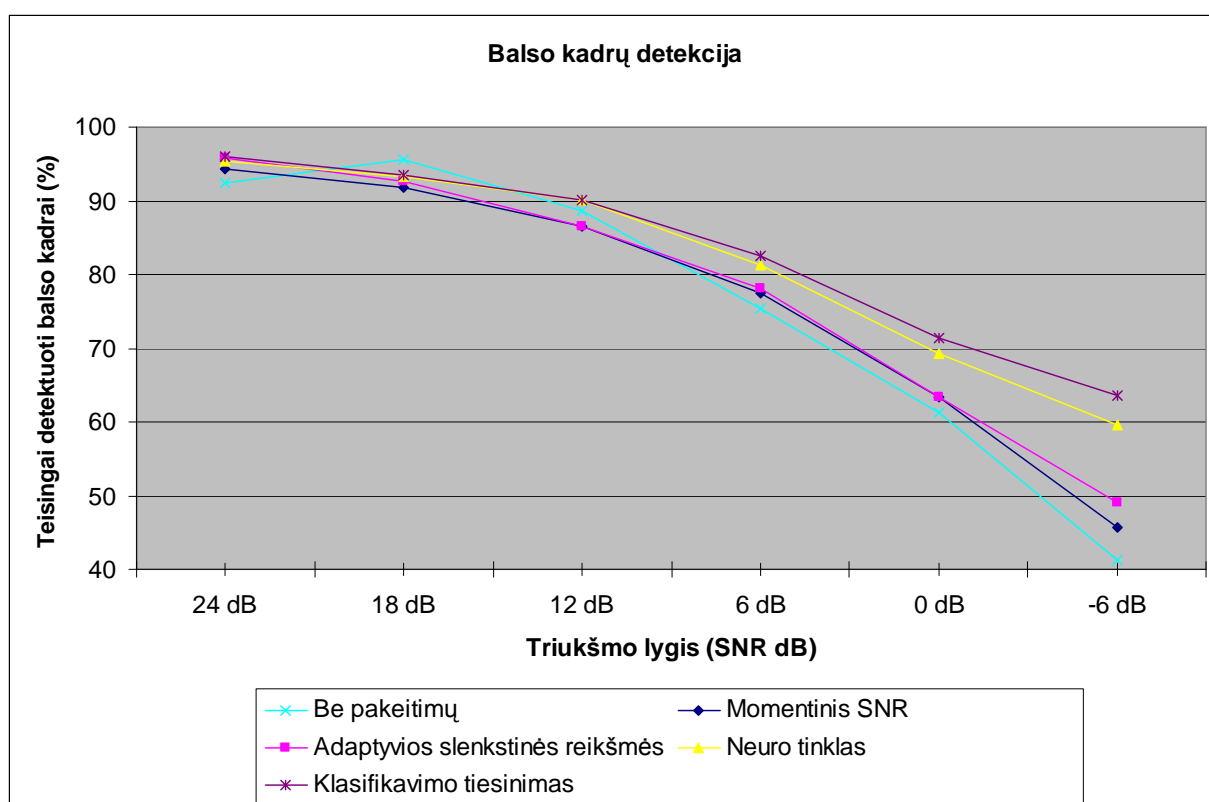
Algoritmo modeliavimo ir realizacijų tyrimo metu naudoti visi keturi 2.4.1 – 2.4.4 skyreliuose išanalizuoti triukšmo tipai. Tyrimas atliktas didesniame nei straipsnyje pateikiama SNR spektre, siekiant detaliau išsiaiškinti algoritmo savybes. Galutinis rezultatas gaunamas išvedant to paties SNR signalo, skirtinguose triukšmuose, rezultatų aritmetinį vidurkį.

Algoritmas ir jo patobulinimai lyginami pagal gebėjimą atpažinti balso (31 paveikslėlis) ir triukšmo (32 paveikslėlis) kadrus. Šie rodikliai skaičiuojami kaip santykis teisingai klasifikuotų balso ar triukšmo kadru su visais signalu esančias to paties tipo kadrais. Bendrą algoritmo gebėjimą klasifikuoti apibūdina klaidingai detektuotų kadru rodiklis, skaičiuojamas padalinus klaidingai detektuotas kadrus iš visų signalo pavyzdyje esančių kadru skaičiaus (33 paveikslėlis).

Atliekant eksperimentinius algoritmo tyrimus, kiekvienas patobulinimas buvo nuosekliai pridėdamas prie pirminės algoritmo realizacijos. Pirmiausia ištirtas pirminis, be patobulinimų, algoritmas visuose 24 dB, 18 dB, 12 dB, 6 dB, 0 dB, -6dB SNR signaluose. Sekančiame etape algoritmui įdiegtas momentinis SNR nustatymas. Taip žemiau pateikta tvarka buvo algoritmui buvo įdiegiamas kiekvienas patobulinimas, paliekant jau esamus. Algoritmo patobulinimų įdiegimo tvarka:

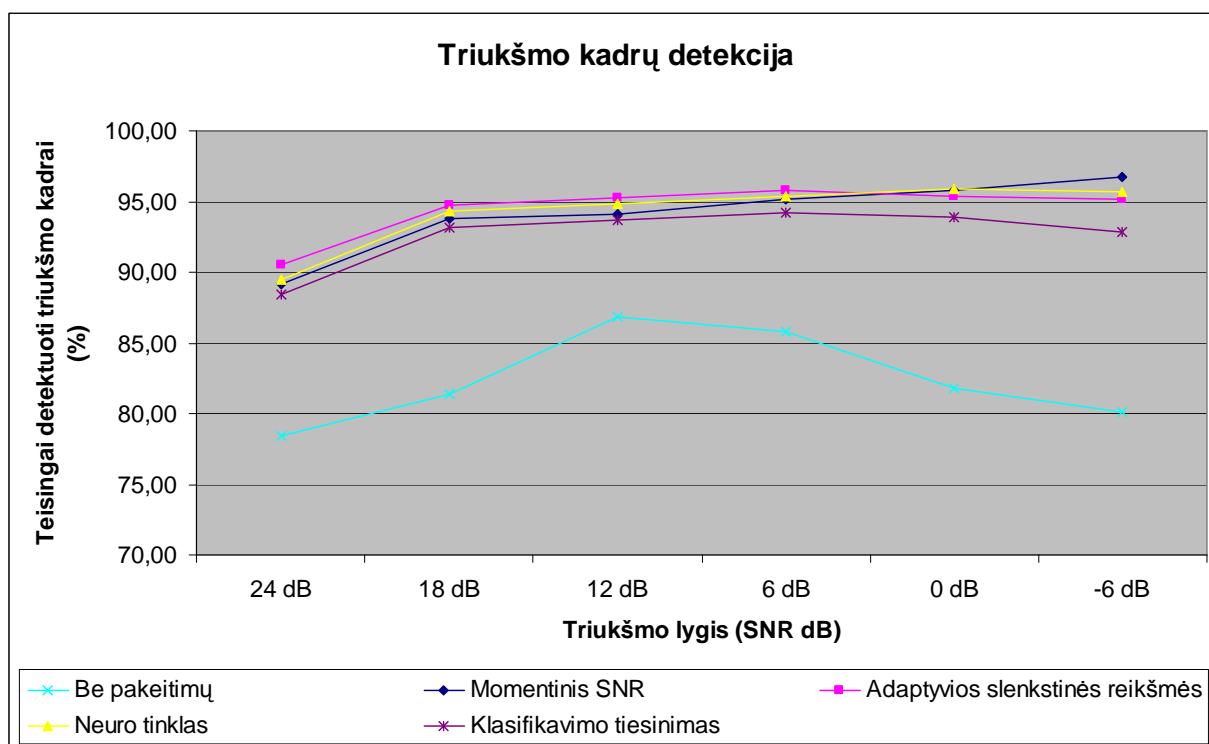
1. Momentinis SNR nustatymas.
2. Adaptyvios slenkstinės reikšmės klasifikavimo baigtiniam automatui.
3. Neuroninis tinklas.
4. Kadru klasifikavimo tiesinimas.

Tokiu būdu žemiau pateiktuose grafikuose galima įvertinti bendrą algoritmo rezultatų pasikeitimą įdiegus patobulinimus.



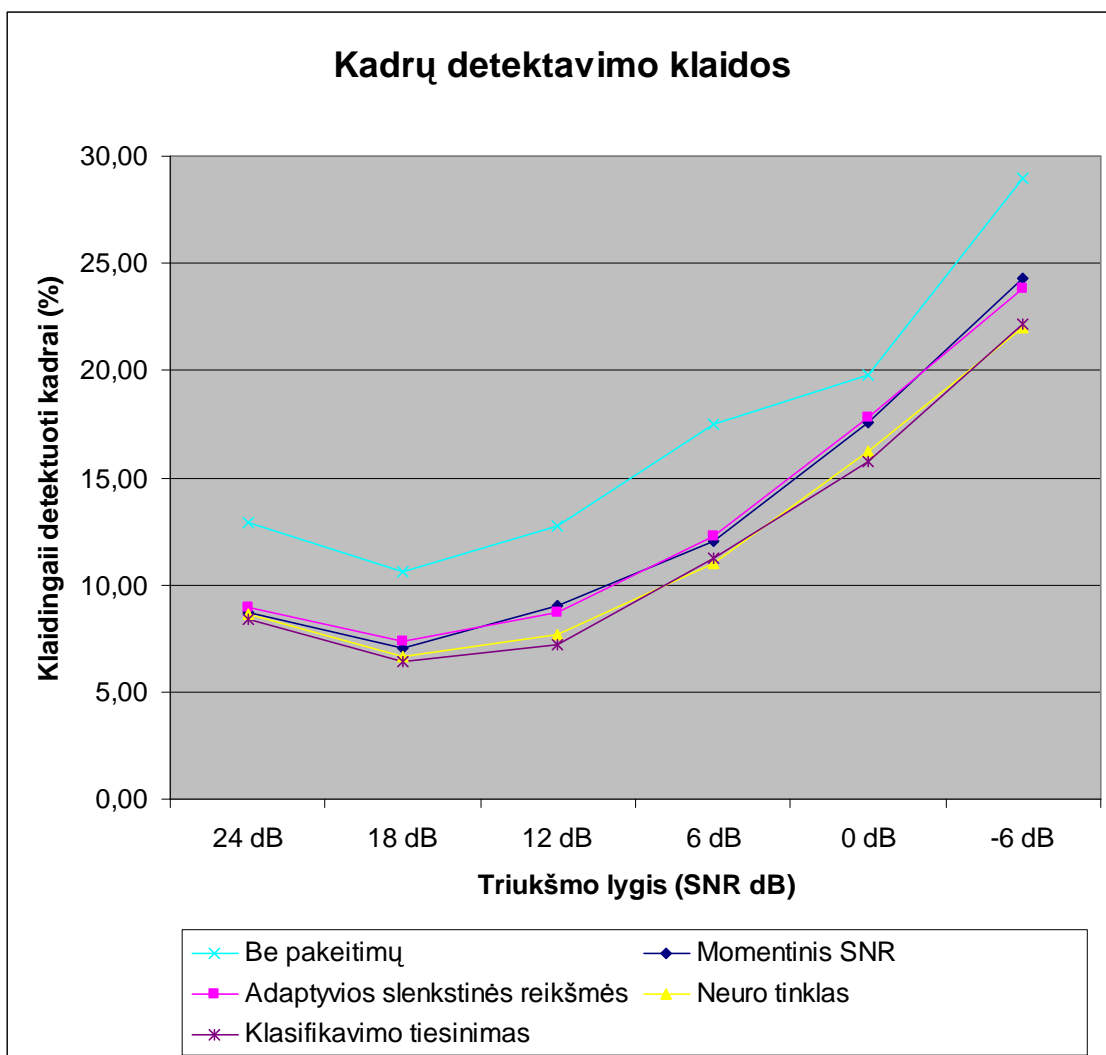
30 pav. Balso kadru atpažinimo grafikas

Iš 31 ir 32 paveikslėliuose pateiktų grafikų matyti kad pirminis, straipsnyje publikuotas, aukštesnių eilių statistika grįstas balso detekcijos algoritmas balso kadrus atpažįsta geriau nei triukšmo. Įvesti patobulinimai balso detektavimą pagerina nežymiai, tačiau triukšmo detekcija pasidaro daug geresnė. Didėjant triukšmo lygiui krinta algoritmo gebėjimas teisingai klasifikuoti balso kadrus.



31 pav. Triukšmo kadų atpažinimo grafikas

Didžiausias įdiegtų patobulinimų efektyvumas pasireiškia triukšmo kadų detekcijoje. Visi patobulinimai išskyrus kadų klasifikavimo tiesinimą pagerina triukšmo detekciją. Kadų klasifikavimo tiesinimas pablogina triukšmo kadų detekciją, tačiau ženkliai pagerina balso kadų atpažinimą. Tokius rezultatus lemia algoritmo savybė po kiekvienos kalbos frazės, kalbai priskirti vieną papildomą kadą.



32 pav. Triukšmo kadrų atpažinimo grafikas

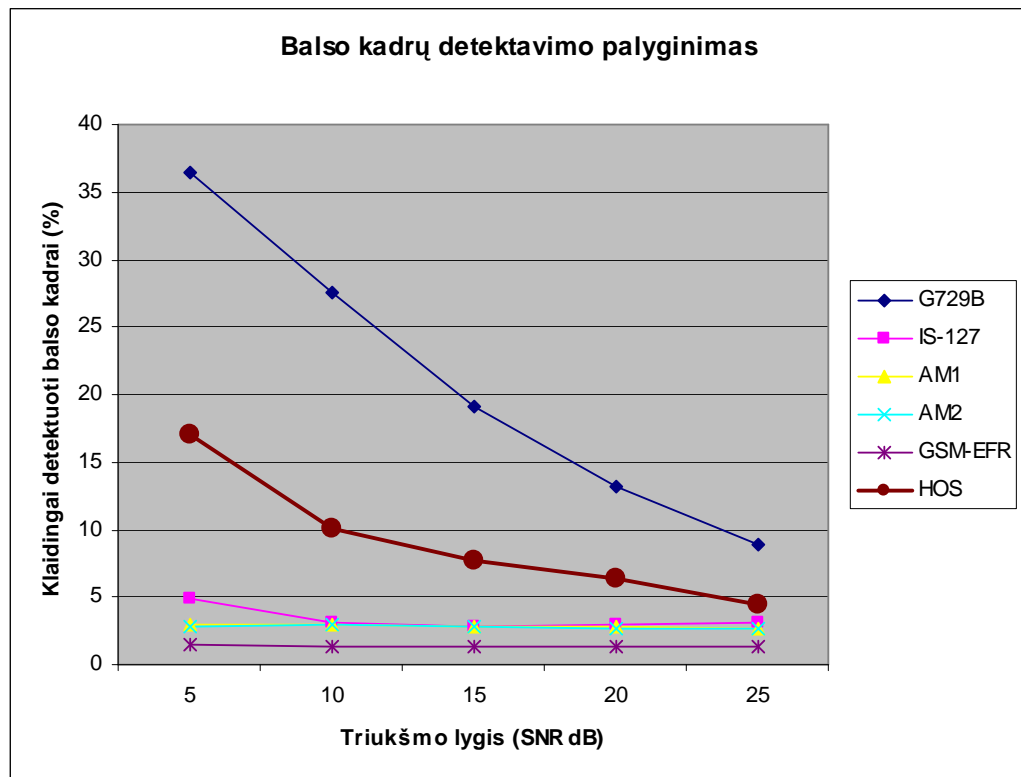
Algoritmo gebėjimą teisingai klasifikuoti tiek balso, tiek triukšmo kadrus įvairiuose triukšmo lygiuose galima pamatyti 33 paveikslėlyje pateiktame grafike. Kiekvienas įdiegtas patobulinimas algoritmo gebėjimą klasifikuoti kadrus pagerina. Grafike matoma, kad aukštesnių eilių statistika grįstas balso detektavimo algoritmas ir modifikuotos algoritmo versijos geriausiai kadrus klasifikuoja esant 18 dB SNR signalui. Kylant triukšmo lygiui algoritmui darosi sunkiau išskirti kadrus. Čia reikia atkreipti dėmesį, kad algoritmo rezultatai pablogėja ir esant labai mažam triukšmui. Tokia algoritmo savybė pastebėta ir modeliuojant algoritmą su kalbos signalu be triukšmo (∞ dB SNR). Viena iš priežasčių algoritmo rezultatų degradacijai nesant triukšmui, yra ta, kad aukštesnių eilių statistiniai parametrai ne visada teisingai atspindi kadre esančio signalo kilmę, o klasifikavimo baigtinio automato būsenų perėjimo taisyklės formuojamos pagrinde pagal šiuos statistinius parametrus.

4.4. Algoritmų palyginimas

Standartizuotų balso detektavimo algoritmų gebėjimą klasifikuoti balso ir triukšmo kadrus galima ištirti tik eksperimentiniu būdu. Nemažai mokslinių publikacijų ir tiriamųjų darbų paskelbta šia tema. Tačiau dažniausiai tyrimams naudojami balso ir triukšmo signalų duomenų pavyzdžiai, skiriasi, todėl rezultatų panaudojimas tolimesniems tyrimams komplikuojasi. Nepaisant skirtingų tiriamojo signalo pavyzdžių, apytikris algoritmų rezultatų palyginimas įmanomas. Eksperimentiniuose tyrimuose naudojami kalbos signalo pavyzdžiai dažniausiai būna iškarpos iš tikro žmonių pokalbio. Tokiuose signalo pavyzdžiuose balso ir tylos kadru proporcija yra labai panaši, todėl tyrimo rezultatų palyginimui skirtingų balso pavyzdžių naudojimas didelės paklaidos nesudaro.

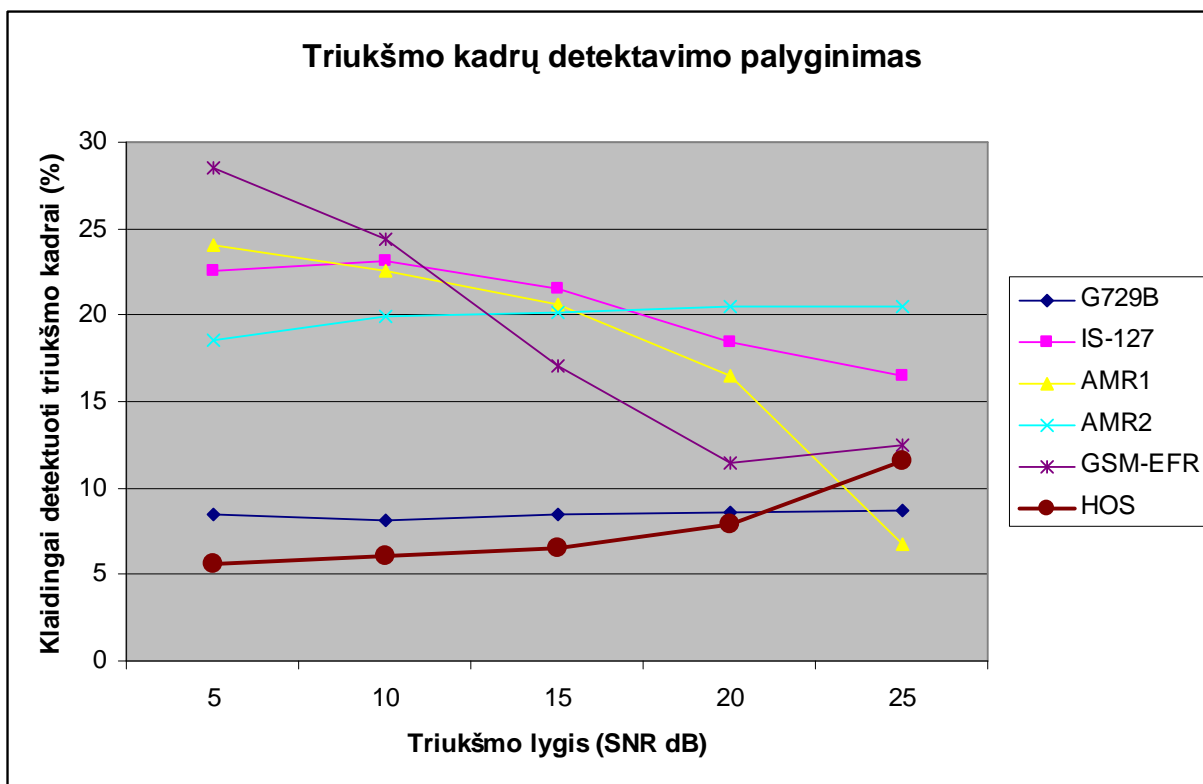
Šiame skyrelyje pateikiami aukštesnių eilių statistika grįsto balso detektavimo algoritmo ir standartizuotų algoritmų gebėjimo klasifikuoti kadrus palyginimai. Aukštesnių eilių statistika grįsto balso detektavimo algoritmo eksperimentinio tyrimo rezultatai naudojami palyginime gauti naudojant tokios pačios trukmės, tačiau kitus, kalbos signalo pavyzdžius, nei naudotus standartizuotų algoritmų rezultatų palyginimui [12]. Siekiant didesnio palyginimo tikslumo kalbos pavyzdžiai parinkti taip, kad kalbos ir tylos kadru proporcija būtų lygi 43/57, tokia kuri naudojama standartizuotų algoritmų palyginime [12].

34 paveikslėlyje parodytas standartizuotų ir aukštesnių eilių statistika grįsto balso detektavimo algoritmo klaidingai detektuotų balso kadru rezultatų grafikas. Aukštesnių eilių statistika grįstas balso detektavimo algoritmas balso periodus detektuoja geriau nei G729B visuose lygintuose triukšmo lygiuose, tačiau kiti standartizuoti algoritmai balsą detektuoja geriau.



33 pav. Balso kadru detektavimo palyginimas

Toks žymus balso detektavimo skirtumas gali būti paaiškintas 35 paveikslėlyje. Čia pateikiami klaidingai detektuotų triukšmo kadru palyginimai. Visi, balso kadru detektavimo prasme, veikiantys geriau algoritmai, sunkiai klasifikuoja triukšmo kadrus. Tokią savybę galima būtų paaiškinti tuo, kad telefoniniuose pokalbiuose (kur ir naudojami šie algoritmai) daug svarbiau yra detektuoti balso kadrus tam, kad kalbos periodai būtų aiškūs ir nepertraukti, nei kartais praleisti triukšmo kadrus. Tuo tarpu aukštesnių eilių statistika grįsto balso detektavimo algoritmas pritaikytas kuo tiksliau klasifikuoti ir triukšmo ir balso kadrus. Triukšmo kadru klasifikavime aukštesnių eilių statistika grįsto balso detektavimo algoritmas visumoje geriau detektuoja triukšmo kadrus.



34 pav. Triukšmo kadru detektavimo palyginimas

4.5. Algoritmo įvertinimas pagal TIA/EIA-136-250 rekomendacijas

Vienas iš metodų įvertinti algoritmo gebėjimą atpažinti balso kadrus yra aprašomas 2.2 poskyryje analizuotoje TIA/EIA-136-250 rekomendacijose. Naudojantis kartu su rekomendacija teikiama specialia programine įranga aukštesnių eilių statistika grįstas balso detektavimo algoritmas buvo tiriamas pagal nurodytas rekomendacijas. Tyrimo rezultatai pateikiami 4 lentelėje. Pateiktoji lentelė automatiškai suformuojama tyrimo programos darbo pabaigoje.

4 lentelė. Algoritmo tyrimo pagal TIA/EIA-136-250 rekomendacijas rezultatai

| | Normalizacija, dBov | SNR, dB | Pc_on | Pd_ss | Pc_off | dVAF |
|----|---------------------|---------|-------|-------|--------|-------|
| 01 | -36 | ∞ | p | p | p | 0.382 |
| 02 | -36 | 18 | 0.171 | 0.139 | 0.301 | p |
| 03 | -36 | 12 | 0.23 | 0.243 | 0.379 | p |
| 04 | -36 | 06 | 0.287 | 0.324 | 0.369 | p |
| 05 | -26 | ∞ | p | p | 0.034 | 0.226 |
| 06 | -26 | 18 | 0.019 | 0.024 | 0.107 | p |
| 07 | -26 | 12 | 0.005 | 0.04 | 0.106 | p |
| 08 | -26 | 06 | 0.037 | 0.048 | 0.119 | p |
| 09 | -16 | ∞ | 0.002 | p | 0.012 | 0.001 |
| 10 | -16 | 18 | p | p | p | p |
| 11 | -16 | 12 | p | p | p | p |
| 12 | -16 | 06 | p | p | 0.017 | p |

Balso detektoriaus gebėjimas atpažinti balso kadrus įvertinami pagal 12 skirtingų normalizacijos lygio ir signalas-triukšmas santykio variacijų. Įvertinami gebėjimai detektuoti kalbos pradžią (*Pc_on*), pastovią kalbą (*Pd_ss*), ir kalbos pabaigą (*Pc_off*). Rezultatai, kurie tenkina rekomendacijose nustatytas normas žymimi raide „p“, priešingu atveju pateikiama santykinė skirtumo nuo normos reikšmė. Lentelėje galima pastebėti, kad aukštesnių eilių statistika grįstas balso detektavimo algoritmas tenkina 23 reikalavimus ir 25 iš 48 netenkina. Didžiausia problema algoritmui detektuoti balso kadrus žemam normalizacijos lygį. Kalbos pabaigą algoritmas atpažįsta sunkiau nei pradžia ir pastovią kalbą. Tai galima paaiškinti tuo. Kad žmogaus balso pradžią dažniausiai sudaro balsiniai garsai, o pabaigoje seka nebalsiniai, savo prigimtimi ir savybėmis labai panašūs į triukšmą.

5. Išvados

- Balso detekcijos algoritmai turi didelę praktinę reikšmę telekomunikacijose ir balso analizės uždaviniuose, kaip signalo apdorojimo priemonė.
- Mokslinėse publikacijose yra pristatoma daug balso detekcijos algoritmų, tačiau nei vieno jų veikimas nėra idealus. Balso detekcija yra plati sfera moksliniams tyrimams.
- Aukštesnių eilių statistiniai parametrai yra gera, tačiau nepakankama priemonė balso kadru detekcijai.
- Aukštesnių eilių statistika grįstas balso detekcijos algoritmas kadrus klasifikuoja geriau nei G.729B tačiau balso kadru detekcijos rezultatai yra prastesni nei kitų standartizuotų algoritmų.
- Aukštesnių eilių statistika grįsto balso detekcijos algoritmo rezultatai yra priklausomi nuo slenkstinių reikšmių, kurios publikacijoje nėra pateikiamos, parinkimo.
- Darbe pristatomi aukštesnių eilių statistika grįsto balso detekcijos algoritmo patobulinai visumoje pagerina algoritmo gebėjimą klasifikuoti kadrus.
- Klasifikavimo baigtinio automato pakeitimas DNT balso detektavimo rezultatus pablogina dėl signalą charakterizuojančių parametrų specifikos.
- Standartizuoti balso detekcijos algoritmai pritaikyti geriau atpažinti balso kadrus nei triukšmo.
- Aukštesnių eilių statistika grįstas balso detekcijos algoritmas ir jo modifikacijos sėkmingai realizuotos TMS320C6713 DSP procesoriaus platformoje.
- Pateikta algoritmo realizacija gali būti panaudota balso signalo apdorojimo taikomuosiuose uždaviniuose.

6. Literatūra

- [1] Goubran, R.; Nemer, E.; ir Mahmoud, S. *Robust Voice Activity Detection Using Higher-Order Statistics in the LPC Residual Domain*. IEEE transactions on Speech and Audio Processing, 2001.
- [2] Martin, R. *An efficient algorithm to estimate the instantaneous SNR of speech signals*. EUROSPEECH, 1993.
- [3] Benyassine, A.; Shlomot, E.; ir Su, H. *ITU-T recommendation G.729, annex B, a silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications*” IEEE Communication, 1997, Nr. 5, p. 64–72.
- [4] Voice over IP provider list [interaktyvus]. – [Žiūrėta 2005 m. balandžio 2 d.]. Prieiga per internetą: <<http://www.voipproviderslist.com/voice-over-ip-calculator/>>
- [5] Lietuvių kalbos literatūros istorija. Kalbos padargai [interaktyvus]. – [Žiūrėta 2006 m. balandžio 17 d.]. Prieiga per internetą: <http://ualgiman.dtiltas.lt/kalbos_padargai.html>
- [6] Kalbos analizės, sintezės, atpažinimo, kaupimo, patikimo atgaminimo principai. Vokalinio trakto ir kalbos sintezės modeliai [interaktyvus]. – [Žiūrėta 2006 m. balandžio 19 d.]. Prieiga per internetą: <<http://www.ik.ku.lt/lessons/konspekt/processing/traktas.htm>>
- [7] McAulay, R.; ir Quatieri, T. *Speech analysis/synthesis based on a sinusoidal representation*. IEEE, 1986.
- [8] E. Dermatas, G. Kokkinakis, *Speech endpointing in noisy environment using neural networks*, International Conference on Digital Signal Processing, Nicosia, Cyprus, 1995.
- [9] *DARPA Neural Network Study*, AFCFA International 1988 m., p.60
- [10] Ikedo, J. *Voice activity detection using neural network*. IEICE Trans Commun., 1998, E81-B, (12), pp. 2509–2513
- [11] Hong-Ik Kim, Sung-Kwon Park. *Voice activity detection algorithm using radial basis function network*. IEICE Transactions on Communications 2005, E88-B(4):1653-1657
- [12] Kondo, A.M. *Digital speech: coding for low bit rate communication systems*. Wiley, 2004.

Santrumpų ir terminų žodynėlis

CDMA (*Code division multiple access*) – duomenų paketų multipleksavimo technologija, mobiliųjų telefonų tinkluose.

DAC (*digital-to-analog converter*) – keitiklis paverčiantis skaitmeninį signalą analoginiu.

DMA (*Direct memory access*) – technologija leidžianti skaityti ir rašyti į operacinę atmintį nepriklausomai nuo centrinio procesoriaus.

DNT – Dirbtinis neuroninis tinklas

DSP (*Digital signal processing*) – Skaitmeninis signalų apdorojimas.

DTX (*Discontinuous transmission*) – metodas perduoti duomenis tik tuomet, kai signale yra balsas.

ETSI (*European Telecommunications Standards Institute*) – Europos telekomunikacinių standartų institutas, nepriklausoma, nepelno organizacija.

GSM (*Global System for Mobile Communications*) – šiuo metu populiariausias standartas mobiliųjų telefonų tinkluose.

L2 (*Level 2*) – antro lygio spartinančioji atmintis (kešas).

LPC (*Linear predictive coding*) – Tiesinis prognozuojantis kodavimas (**TKP**). Kalbos signalų analizės technologija, dažniausiai naudojama kalbos suspaudimui ir parametrų nustatymui.

SNR (*Signal-to-noise ratio*) – Signalas triukšmas santykis. Signalų ir foninio triukšmo galių santykis išreiškiamas logaritminėje skalėje.

TIA/EIA (*Telecommunications Industry Association /Electronic Industries Alliance*)

UMTS (*Universal Mobile Telecommunications System*) – Viena iš trečios kartos (3G) mobiliųjų telefonų technologijų.

VAD (*Voice activity detection*) – Balso aktyvumo detekcija arba detektorius.

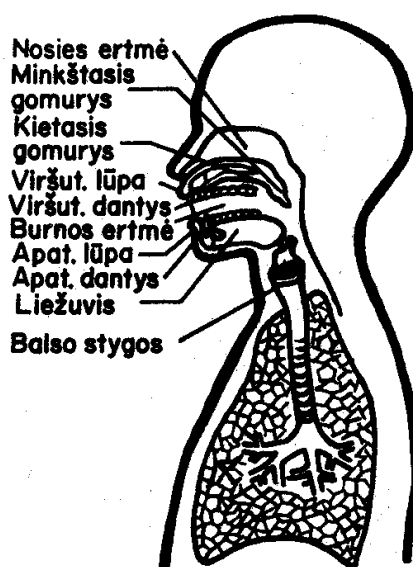
VoIP (*Voice over Internet Protocol*) – balso ryšys perduodamas duomenų perdavimo tinklais naudojant interneto protokolą.

PRIEDAS A. Žmogaus kalbos aparatas, balso sintezavimas

Norint pritaikyti signalų apdorojimo technologijas žmogaus balso detekcijos problemai, būtina išanalizuoti žmogaus balso prigimtį – garsų skleidimo sistemą.

Garsai susidaro veikiant kalbos padargams. Kalbos padargai yra šie:

- lūpos (viršutinė ir apatinė);
- dantys (viršutiniai ir apatiniai);
- liežuvis;
- gomurys, kuris baigiasi liežuvėliu;
- nosies, burnos, ryklės ertmės;
- gerklos su balso stygomis;
- kvėpuojamoji gerklė;
- bronchai;
- plaučiai.



35 pav. Žmogaus vokalinė sistema. [5]

Vienas svarbiausių kalbos padargų yra gerklos su jose esančiomis balso stygomis. Oras eina iš plaučių pro kvėpuojamąją gerklę ir balso stygas. Jeigu balso stygos neįtemptos ir nevirpa, oras išeina be kliūčių. Jeigu oras eidamas virpina įtemptas balso stygas – atsiranda garsas. Dažniausiai oras išeina pro burnos ertmę, kur liežuvis, dantys ir gomurys padeda susidaryti garsui. Tik tariant *m* ir *n*, oras išeina ne pro burnos, o pro nosies ertmę.

Kalbos sintezacija remiasi realia žmogaus vokaline sistema, kurią sudaro oro siurblys (plaučiai), nuosavo žadinimo virpesių generatorius (gerklos), šnypščiantis šaltinis, vokalinės

ertmės - suderinamas perdavimo traktas ir triukšmų bei burnos angos per kurias išeina vokalinis išėjimo signalas. Žmogaus vokalinė sistema savyje jungia sužadavimo šaltinį su vokaliniu traktu-laike kintančiu filtru. Balso garsai - balsės ir priebalsių tono dedamosios generuojamos kaip aukšto dažnio impulsų seka, gaunama iš balso stygų, kurios vibruoja laike kintančiu aukščiau. Nebalsiniai garsai, tokie kaip f, b, s, t,... yra sužadinami iš triukšmų šaltinio, kurį realizuoja lūpų ar vokalinio trakto turbulentiškumas. Panašiai yra sukonstruoti ir akustiniai muzikiniai instrumentai, kuriuose atskirti sužadavimo šaltinis ir rezonansinės ertmės.

Dirbtinis vokalinės sistemos modelis pagrįstas sužadimu, kurį sudaro-komutuojamas balto triukšmo ar periodinių impulsų garso aukščio šaltinis, duodantis balsinius ar nebalsinius signalus, amplitudės daugintuvas, keičiantis signalų stiprumą, vokalinį traktą modeliuojantis filtras, atliekantis kiekvieno impulso klostę su filtro impulsine reakcija ir tuo suformuojantis vokalinio trakto charakteringuosius garsus-formantes ir DAC, išduodantis susintetintą kalbos signalą[6].

Kalbos sintezatorių projektuotojai bando aproksimuoti spektrinius intensyvumus, kurdami vokaderius modeliuoja pačias sistemas, naudodami tiesinį prognozuojantį kodavimą LPC (ang.k. Linear predicting coding) arba suspaudžia pačias kalbos bangas.

PRIEDAS B. Pridedamo CD turinys

| Katalogas | Turinys |
|-----------|---|
| Matlab | Kataloge yra Matlab modelio programiniai išeities tekstai. Dėl autorinių teisių TIA balso ir triukšmo signalų duomenys nėra pateikiami, tačiau gali būti parsiusiti atskirai iš oficialios TIA svetainės internete. |
| Simulink | Kataloge yra Simulink modelis (<i>model.mdl</i>) ir modifikuotas modelis, programinio kodo generavimui TMS320C6713 procesoriniai plokštei. |
| CCS | Kataloge yra Code Composer Studio projektas ir išeities tekstai C kalba. |
| PDF | Kataloge yra magistrinio darbo ataskaitos elektroninė kopija PDF formate. |