VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Aurimas ANSKAITIS

KODUOTO BALSO KOKYBĖS TYRIMAS

DAKTARO DISERTACIJA

TECHNOLOGIJOS MOKSLAI,
ELEKTROS IR ELEKTRONIKOS INŽINERIJA (01T)

VGTU
LEIDYKLA
Vilnius TECHNIKA 2009

Disertacija rengta 2005–2009 metais Vilniaus Gedimino technikos universitete.

Mokslinis vadovas

prof. habil. dr. Algimantas KAJACKAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T).

http://leidykla.vgtu.lt
VGTU leidyklos TECHNIKA 1677-M mokslo literatūros knyga

ISBN 978-9955-28-504-5

© Anskaitis, A., 2009
© Vilniaus Gedimino technikos universitetas, 2009
aurimas.anskaitis@el.vgtu.lt

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Aurimas ANSKAITIS

ANALYSIS OF QUALITY OF CODED VOICE SIGNALS

DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES,
ELECTRICAL AND ELECTRONIC ENGINEERING (01T)

VGTU
LEIDYKLA
Vilnius TECHNIKA 2009

Doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2005–2009.

Scientific Supervisor

Prof Dr Habil Algimantas KAJACKAS (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronic Engineering – 01T).

Reziumė

Disertacijoje nagrinėjama koduoto balso kokybės vertinimo problematika. Pagrindinis dėmesys skiriamas balso kokybės tyrimams, kai perduodama koduota šneka ir prarandami balso paketai.

Darbo tikslas yra patobulinti koduoto balso kokybės vertinimo algoritmus.

Darbo uždaviniai yra šie:

- sukurti matavimo priemonę trumpų balso signalo atkarpų kokybei vertinti;
- apibrėžti koduoto balso segmentų vertės sampratą ir parinkti vertės metrikas;
- išmatuoti bendrinės šnekos balso segmentų verčių skirstinius;
- nustatyti skirtingų koderių sukuriamų iškraipymų ribas;
- ištirti paplitusių koderių inertiškumą, nustatyti kiek laiko pastebima prarastų paketų įtaka sekantiems segmentams.

Disertaciją sudaro įvadas, keturi tiriamieji skyriai ir bendrosios išvados.

Įvade pristatomas darbo naujumas, aktualumas, aptariamas autoriaus indėlis, formuluojami darbo tikslai.

Pirmas skyrius yra apžvalginis – analizuojami balso kokybės vertinimo metodai, jų privalumai ir trūkumai. Kaip savarankiška dalis čia pristatyti autoriaus sudaryti sąrašai lietuviškų žodžių, skirtų šnekos suprantamumo tyrimams.

Antrame skyriuje parodoma, kaip galima išplėsti kokybės vertinimo PESQ (angl. *Perceptual Evaluation of Speech Quality*) algoritmo taikymo ribas. Čia įvedama koduoto balso paketo vertės sąvoka, nustatomi statistiniai paketų vertės skirstiniai.

Trečiame skyriuje nagrinėjami specifiniai koduotos šnekos iškraipymai ir kodavimo parametrų įtaka balso kokybei. Parodoma, kad kodavimo iškraipymų dydis priklauso nuo šnekos signalo ir kinta plačiose ribose. Skyriuje visapusiškai ištirta prarastų paketų įtaka.

Ketvirtame skyriuje nagrinėjamas paketo vertės nustatymo realiu laiku uždavinys. Pasiūlyti metodai viršutiniams paketo vertės nustatymo metodų tikslumo rėžiams nustatyti, sudaryti skaičiavimų atžvilgiu efektyvūs algoritmai, paketo vertei prognozuoti.

Disertacijos tema perskaityti 8 pranešimai mokslinėse konferencijose ir paskelbti 7 straipsniai: 4 – recenzuojamuose žurnaluose, 3 – kituose leidiniuose.

Abstract

The dissertation investigates the problem of quality of coded voice. The main attention is paid to voice quality evaluation under packet loss conditions.

The aim of the work is to improve voice quality evaluation algorithms.

The tasks of the work are:

- construction of the means for measurement of voice quality of short voice signals;
- to define the concept of value of coded voice segment and to choose corresponding value metrics;
- to measure distributions of frame values in standard voice;
- to establish limits of distortions created by different codecs;
- to investigate inertia of wide spread codecs and establish the length of impact of one lost frame.

The dissertation consists of the introduction, 4 chapters, conclusions, list of literature.

Introduction presents the novelty and topicality of the work, tasks and aims of the work are formulated.

The first chapter is overview of voice quality evaluation methods, pros and cons of these methods are analyzed. PESQ algorithm and limits of its applicability are introduced in this chapter too. The lists of Lithuanian words for word intelligibility testing are created.

Chapter two presents the method of signal construction that allows to extend PESQ applicability to short signals. This chapter introduces the concept of frame value. Distributions of frame values are calculated.

Third chapter analyses distortions created by coding. It is shown that coding distortions depends highly on the signal used and limits of distortion variability are calculated. The impact of lost frames is analyzed in this chapter too.

Fourth chapter analyses the problem of frame value calculation in real time. Effective algorithms for this purpose are presented. Approximate upper limits of frame value prognosis precision are calculated.

The results of the thesis were presented in 8 conferences, 7 scientific articles were published on the subject.

Žymėjimai

Simboliai

Q_{PESQ} – balso kokybė PESQ skalėje;
V_{PESQ} – balso kokybės pablogėjimas PESQ skalėje;
D_{sn} – simetriniai iškraipymai n-ojoje PESQ fonemoje;
D_{an} – asimetriniai iškraipymai n-ojoje PESQ fonemoje;
L_{DS} – agreguoti simetriniai iškraipymai PESQ skiemenyje;
L_{DA} – agreguoti asimetriniai iškraipymai PESQ skiemenyje;
d_{sym} – pilnasis simetrinis PESQ iškraipymas;
d_{asym} – pilnasis asimetrinis PESQ iškraipymas;
$x(i)$ – signalo i-asis elementas;
x_i – signalo i-tasis elementas;
FFT – greitoji Furjė transformacija;
N – paketų skaičius eksperimente, signalo elementų skaičius, kadrų skaičius signale;
j – požymių vektoriaus elemento indeksas;
n – paketo eilės numeris;
Θ – kokybės matavimo operatorius;
K – k-artimiausių kaimynų algoritmas;

E – signalo energija;

$X(k)$ – signalo $x(i)$ spektras;

d – atstumas tarp vektorių.

Santrumpos

3SQM – neinvazinis balso kokybės vertinimo algoritmas (angl. *Single Side Speech Quality Measure*);
AMR – adaptyvus daugelio spartų koderis (angl. *Adaptive Multi-Rate codec*);
BER – bitų klaidų tankis (angl. *Bit Error Rate*);
CELP – kodu žadinama tiesinė prognozė (angl. *Code Excited Linear Prediction*);
DB – duomenų bazė (angl. *Data Base*);
dMOS – diferencinis vidutinis nuomonių balas (angl. *Differential Mean Opinion Score*);
DRT – diagnostinis rimavimo testas (angl. *Diagnostic Rhyme Test*);
FER – paketų klaidų tankis (angl. *Frame Error Rate*);
FFT – greitoji Furjė transformacija (angl. *Fast Fourier Transform*);
GSM – pasaulinė mobiliojo ryšio sistema (angl. *Global System for Mobile*);
iQoS – individuali paslaugos kokybė (angl. *individual QoS*);
ITU – Tarptautinė telekomunikacijų sąjunga (angl. *International Telecommunication Union*);
LPC – tiesinės prognozės kodavimas (angl. *Linear Predictive Coding*);
MOS – vidutinis nuomonių balas (angl. *Mean Opinion Score*);
MRT – modifikuotas rimavimo testas (angl. *Modified Rhyme Test*);
PCM – impulsinė kodinė moduliacija (angl. *Pulse Code Modulation*);
PSQM – suvokiamasis balso kokybės matas (angl. *Perceptual Speech Quality Meaure*);
PESQ – suvokiamos balso kokybės vertinimo algoritmas (angl. *Perceptual Evaluation of Speech Quality*);
$PESQ_{mod}$ – algoritmas, aprašantis, kaip taikyti PESQ trumpiems signalams;
dPESQ – diferencinis suvokiamos balso kokybės vertinimo algoritmas (angl. *differential Perceptual Evaluation of Speech Quality*);
QoE – patirties kokybė (angl. *Quality of Experience*);
QoS – paslaugos kokybė (angl. *Quality of Service*).

Turinys

ĮVADAS 1

Problemos formulavimas 1
Darbo aktualumas 2
Tyrimų objektas 3
Darbo tikslas 4
Darbo uždaviniai 4
Tyrimų metodika 4
Darbo mokslinis naujumas ir jo reikšmė 4
Darbo rezultatų praktinė reikšmė 5
Ginamieji teiginiai 5
Darbo rezultatų aprobavimas 5
Disertacijos struktūra 6

1. BALSO KOKYBĖS VERTINIMO UŽDAVINIAI IR METODAI 7

1.1. Balso kokybė telekomunikacijose 7
1.2. Balso kokybės vertinimo metodai 9
1.3. MOS vertinimo metodas 10
1.4. Šnekos suprantamumo vertinimai 11
1.5. Objektyvieji vertinimo metodai 17
1.5.1. Objektyviųjų invazinių vertinimo metodų apžvalga 17

1.5.2. Objektyviųjų neinvazinių vertinimo metodų apžvalga....18
1.6. PESQ algoritmas....19
1.7. Spektrinis šnekos iškraipymų nustatymo metodas....20
1.8. PESQ ypatumai matuojant trumpų signalų kokybę....24
1.9. Pirmo skyriaus išvados ir disertacijos uždavinių formulavimas....28
2. BALSO SEGMENTO VERTĖ....29
2.1. Vertės analizės metrikos ir fundamentalios sąvokos....30
2.2. PESQ algoritmo taikymas trumpiems signalams analizuoti....32
2.3. Paketo vertės nustatymas taikant PESQ algoritmą....36
2.4. Paketų vertingumo tyrimų schema....37
2.4.1. Kai kurie paketų vertingumo dėsningumai....38
2.4.2. Pavienio balso paketo statistinis vertingumas....41
2.4.3. Dviejų paketų statistinis vertingumas....43
2.4.3. Paketo statistinis vertingumas, kai naudojamas balso aktyvumo detektorius....45
2.5. Segmentų vertingumo priklausomybė nuo matavimo intervalo ilgio....47
2.6. Dviejų paketų verčių sumavimas....48
2.7. Antro skyriaus išvados....50
3. BALSO MIKROIŠKRAIPYMŲ ĮTAKOS ANALIZĖ....51
3.1. Tiriamieji signalai ir jų savybės....52
3.2. Skirtingų AMR koderių sukuriami balso iškraipymai....54
3.2.1. Spektro iškraipymai....55
3.2.2. LPC iškraipymai....56
3.2.3. Girdimieji iškraipymai....58
3.3. AMR koderių inertiškumo tyrimas, kai prarandami paketai....60
3.3.1. Girdimieji iškraipymai....62
3.3.2. LPC iškraipymai....66
3.4. Trečio skyriaus išvados....71
4. KITŲ PAKETO VERTĖS NUSTATYMO PRIEMONIŲ PAIEŠKA....73
4.1. Paketo vertės priklausomybė nuo matavimo lango padėties....73
4.2. Paketo vertės nustatymas pagal tiesinės prognozės koeficientus....78
4.3. Netiesinis paketo vertės prognozavimas....81
4.4. Ketvirto skyriaus išvados....84
BENDROSIOS IŠVADOS....85
LITERATŪROS SĄRAŠAS....87
AUTORIAUS MOKSLINIŲ PUBLIKACIJŲ DISERTACIJOS TEMA SĄRAŠAS....93

Contents

INTRODUCTION 1

Problem Formulation 1
Importance of the Thesis 2
The Object of Research 3
The Goal of the Thesis 4
The Task of the Thesis 4
Research Methodology 4
Importance of the Thesis and Scientific Novelty 4
Practical Significance of Achieved Results 5
The Defended Statements 5
Approval of Results 5
Dissertation Structure 6

1. VOICE QUALITY EVALUATION TASKS AND METHODS 7

1.1. Voice Quality in Telecommunications 7
1.2. Voice Quality Evaluation Methods 9
1.3. MOS Method 10
1.4. Speech Intelligibility Evaluation Methods 11
1.5. Objective Methods of Evaluation 17
1.5.1. Overview of Objective Invasive Methods 17
1.5.2. Overview of Objective non-Invasive Methods 18
1.6. PESQ Algorithm 19
1.7. Spectral Method for Voice Quality Distortions Evaluation 20
1.8. PESQ Properties when Using with Short Signals 24

1.9. Conclusions for Chapter 1 and Dissertation Tasks Formulation 28

2. VALUE OF VOICE SEGMENT 29

2.1. Metrics of Value Analysis and Fundamental Concepts 30
2.2. Application of PESQ for Short Signal Analysis 32
2.3. Frame Value Determination using PESQ Algorithm 36
2.4. Diagram of Frame Value Investigation 37
2.4.1. Some Laws of Frame Values 38
2.4.2. Statistical Value of Single Frame 41
2.4.3. Statistical Value of two Frames 43
2.4.3. Statistical Frame Value when Using Voice Activity Detection 45
2.5. Frame Value Dependence on Measurement Segment Length 47
2.6. Sumation of two Frame Values 48
2.7. Conclusions for Chapter 2 50

3. ANALYSIS OF IMPACT OF VOICE MICRODISTORTIONS 51

3.1. Signals under Investigation and their Properties 52
3.2. Voice Distortions Created by Different AMR Codecs 54
3.2.1. Spectral Distortions 55
3.2.2. LPC Distortions 56
3.2.3. Perceptual Distortions 58
3.3. Investigation of AMR Codecs Inertia when Frames are Lost 60
3.3.1. Perceptual Distortions 62
3.3.2. LPC Distortions 66
3.4. Conclusions for Chapter 3 71

4. SEARCH OF OTHER MEANS FOR FRAME VALUE ESTIMATION 73

4.1. The Dependence of Frame Value on Measurement Window Placement 73
4.2. Frame Value Estimation using LPC Coefficients 78
4.3. Non-Linear Frame Value Prognosis 81
4.4. Conclusions for Chapter 4 84

GENERAL CONCLUSIONS 85

REFERENCES 87

LIST OF THE AUTHOR'S SCIENTIFIC PUBLICATIONS ON THE TOPIC OF DISSERTATION 93

Įvadas

Problemos formulavimas

Bet kokioje modernioje telekomunikacinėje sistemoje balsas koduojamas. Atskiri kodai perduodami duomenų paketais. Paketų praradimai perdavimo trakte yra pagrindinis veiksnys, įtakojantis perduoto balso kokybę.

Pagal ITU rekomendaciją P.800 telekomunikacinių sistemų perduoto balso kokybę turi nustatyti specialiai apmokyti ekspertai. Šis būdas sutrumpintai vadinamas MOS (angl. *Mean Opinion Score*). Šnekos kokybė pagal MOS skalę išreiškiama ekspertų nuomonių vidurkiu.

Akivaizdu, kad MOS metodas jį taikant yra brangus, subjektyvus, nes skirtingų ekspertų išvados būna skirtingos. Yra sukurta keletas aparatinių – objektyvių balso kokybės vertinimo būdų. Vienas iš šiuo metu plačiausia taikomų yra PESQ (angl. *Perceptual Evaluation of Speech Quality*) algoritmas, aprašytas ITU rekomendacijoje P.862 (ITU-T Rec. P.862. 2001) ir pripažintas oficialiu standartu.

PESQ algoritmas palygina signalą prieš tiriamąją sistemą su signalu po sistemos ir, naudodamas skaidymo į kadrus apdorojimo principą, formuoja įvertį, kuris, kaip parodė daugkartiniai tyrimai, yra stipriai koreliuotas su daugelio testų MOS vertinimais. Todėl gana dažnai PESQ įverčiai vadinami objektyviais MOS.

Pagal MOS bei PESQ metodikas balso kokybė nustatoma apdorojant gana didelės trukmės balso signalo atkarpas. Konkrečiai PESQ aprašantis standartas rekomenduoja naudoti 8–20 s ilgio signalo atkarpas (ITU-T Rec. P.862. 2001). Perduodant koduotą balsą paketais, atskirų paketų dingimai nepataisomai iškraipo pavienius žodžius – trumpas balso atkarpas. Mūsų atlikti tyrimai parodė, kad taikant PESQ algoritmą trumpų (trumpesnių už 1 s) balso segmentų kokybei vertinti, gaunama didelė neapibrėžtis, kartais siekianti iki 0,55 MOS vertės (Kajackas *et al*. 2008). Todėl pirmieji disertacinio darbo etapai buvo skirti galimybių ir būdų paieškai, kaip pritaikyti PESQ algoritmą trumpų signalų kokybei vertinti. Vienas iš esminių reikalavimų buvo suformuluotas reikalavimas ieškoti būdų, kaip pritaikyti PESQ algoritmą jo nemodifikuojant. Pritaikius oficialiai pripažintą matavimo priemonę išvengiama būtinybės vykdyti naujos priemonės metrologinę patikrą.

Elementarus įvykis, darantis įtaką balso kokybei, yra vieno paketo praradimas. Aišku, kad skirtingi paketai nėra vienodai vertingi, jų praradimai ne vienodai pablogina balso kokybę. Iš to kyla esminis uždavinys, kaip nustatyti vieno paketo informacinę vertę.

Suformulavus paketo informacinės vertės bazinę sampratą kyla ir kiti uždaviniai: kaip kokybės pablogėjimas, atsirandantis dėl kelių prarastų paketų, susijęs su tų paketų informacinėmis vertėmis, kaip paketų informacinė vertė pasiskirsto pavieniuose žodžiuose.

Didesnės spartos koderiai užtikrina vidutiniškai geresnę balso kokybę. Tačiau klausimas, ar ta kokybė visada vienoda, ar ji kinta priklausomai nuo kokybei vertinti naudojamo įrašo? Į šį klausimą ir buvo bandoma atsakyti. Tam reikėjo ištirti, kokie kokybės skirstiniai gaunami koduojant skirtingus įrašus.

Yra darbų, kuriuose siūloma skirtingo vertingumo paketus perduoti skirtingais kodais, svarbesnius geriau apsaugant (Prastholm 2006). Esminė neišspręsta problema, kaip nustatyti siunčiamo paketo vertę. Paketo vertės nustatymas naudojant PESQ algoritmą yra labai imlus skaičiavimams ir sukuria nemažą vėlinimą. Todėl šiame darbe bandoma sukurti mažesnio kompleksiškumo algoritmus, kurie galėtų pakankamai tiksliai realiu laiku prognozuoti būsimo paketo vertę.

Darbo aktualumas

Balso perdavimo sistemos nuolat aktyviai tobulinamos, diegiamos ir plačiai naudojamos. Daugelyje šiuolaikinių telekomunikacinių sistemų balso signalas skaidomas į vienodos trukmės atkarpas, tos atkarpos koduojamos, suformuojami duomenų paketai ir perduodami. Ryšių kanaluose pasitaikantys trukdžiai sukuria klaidas, dėl kurių kai kurie paketai yra prarandami. Bevielėje terpėje paketo

praradimu vadiname procesą, kai cikliniai kodai aptinka, kad pakete yra klaidingų bitų ir toks paketas pažymimas, kaip sugadintas. Tokie iškraipymai, kai prarandami paketai ypatingai būdingi mobiliojo ryšio sistemoms. IP telefonijos sistemose pasitaikantys dideli paketų vėlinimai taip pat transformuojami į paketų praradimus. Jei paketas vėluoja daugiau, negu slenkstinė vertė, tai dekoderis elgiasi taip, lyg to paketo negavo.

Kai perduodant duomenis paketai prarandami, visų sistemų perduodamo balso kokybė pablogėja.

Kuriant šiuolaikines telekomunikacijų sistemas prisilaikoma 1994 m. Tarptautinės telekomunikacijų sąjungos (ITU) E.800 rekomendacijoje (ITU-T Rec. E.800, 1996) suformuluotos paslaugų kokybės (QoS) sampratos, apibūdinamos, kaip vartotojų pasitenkinimą atspindintis paslaugos savybių bendras efektas. Prisilaikant šios sampratos, kuriamos telekomunikacijų sistemos, metodikos ir priemonės, kurios palaiko nustatytą paslaugų kokybės lygį, vertinant visų vartotojų paslaugų kokybės vidurkį. Reikia atkreipti dėmesį, kad paplitusios paslaugų kokybės vidutinių rodiklių vertinimo metodikos sukurtos fiksuotojo telefoninio ryšio sistemoms, kuriose visų vartotojų sąlygos yra panašios, visiems vartotojams teikiamų paslaugų kokybės rodikliai laikytini vienodais.

Mobiliojo ryšio tinkluose, realiai vartotojui suteikiamų paslaugų kokybė, nėra vienoda. Dėl vartotojo judrumo ir kintančių tinklo apkrovų, ji nuolatos kinta ir ne visada, ir ne visur būna pakankamai aukšta.

Balso signalų savybių tyrimas seniai yra plėtojamas. Labai dažnai balso signalų savybės analizuojamos siedami jas su šnekos atpažinimo uždaviniais (Lipeika 1998; Rudžionis 2000). Tačiau šiuolaikinėse sistemose, kai balso kodavimas taip paplitęs, iškyla keletas specifinių uždavinių – šnekos savybių aprašymas tandeme su koderiu ir koderio įnešamų efektų balso signalams analizė. Siedami šias tyrimo kryptis su moderniomis telekomunikacijomis, gauname svarbų vartotojams ir operatoriams uždavinį – koduoto balso kokybės nustatymą.

Balso kokybės vertinimo problematika taip pat domisi ir telekomunikacijų pramonės atstovai. Didieji telekomunikacinių sistemų kūrėjai taip pat sprendžia uždavinius, susijusius su paslaugų kokybe – „Nokia“ (Qulity of Experience 2004), „Nortel Networks“ (Bharrathsingh 2005) ir kiti.

Iš pateiktų faktų matome, kad balso kokybės tyrimai yra plačiai vykdomi, jais vadovaujamasi kuriant telekomunikacijų sistemas, jas prižiūrint. Įvertinus koduoto balso perdavimų telekomunikacijose paplitimą, manome, kad darbas yra aktualus ir atliekamas laiku.

Tyrimų objektas

Darbo tyrimų objektas – koduoto balso kokybė.

Darbo tikslas

Darbo tikslas – patobulinti koduoto balso kokybės vertinimo algoritmus.

Darbo uždaviniai

Suformuluotiems tikslams pasiekti darbe sprendžiami uždaviniai:

1. Sukurti matavimo priemonę trumpų balso signalo atkarpų kokybei vertinti.
2. Apibrėžti koduoto balso segmentų vertės sampratą ir parinkti vertės metrikas.
3. Išmatuoti bendrinės šnekos balso segmentų verčių skirstinius.
4. Nustatyti skirtingų koderių sukuriamų iškraipymų ribas.
5. Ištirti paplitusių koderių inertiškumą, nustatyti kiek laiko pastebima prarastų paketų įtaka tolimesniems segmentams.
6. Ištirti galimybes balso paketo informacinę vertę prognozuoti realiu laiku.

Tyrimų metodika

Darbe taikomi statistikos, tiesinės algebros, signalų analizės metodai. Hipotezės tikrinamos sudarant reiškinių modelius ir atliekant simuliacijas. Algoritmų veikimo tikslumo įvertinimui naudojama koreliacinė analizė.

Darbo mokslinis naujumas ir jo reikšmė

Rengiant disertaciją buvo gauti šie elektros ir elektronikos inžinerijos mokslui nauji rezultatai:

1. Sukurtas balso testams skirtas signalų sintezės būdas, išplečiantis galimybes taikyti PESQ algoritmą trumpų signalo atkarpų kokybei vertinti.

2. Suformuluota balso paketo informacinės vertės samprata. Pasiūlyta metodika ir sudaryti algoritmai balso paketo informacinei vertei skaičiuoti.
3. Nustatytos statistinės charakteristikos, parodančios, kaip balso paketo praradimas veikia signalą laike.
4. Sudaryti algoritmai perduodamo paketo vertei prognozuoti realiu laiku.

Darbo rezultatų praktinė reikšmė

Tyrimų rezultatai gali būti panaudojami kuriant ir optimizuojant naujų kartų balso perdavimo telekomunikacines sistemas. Rezultatai taip pat gali būti panaudojami galinio vartotojo įrangoje kuriant kokybės apskaitos modulius. Tai leistų įgyvendinti apmokėjimo už paslaugas modelį, kai atsižvelgiama į realiai vartotojo gutą kokybės lygį. Šiuo metu tokia apmokėjimo už paslaugas praktika nėra paplitusi.

Ginamieji teiginiai

1. ITU-P. 862 rekomendacija aprašytu PESQ algoritmu vertinant trumpus balso signalus, gaunami rezultatai yra nestabilūs. Trumpus balso signalus galima taip išplėsti, kad standartinis PESQ algoritmas tiktų jų kokybei vertinti.
2. Balso paketo informacinė vertė yra skirtumas tarp balso kokybių, kai paketas priimtas sėkmingai ir kai priimtas su klaida.
3. Balso paketo informacinės vertės skirstinys yra nesimetriško varpo formos.
4. Praradus balso paketą, didžiausi iškraipymai stebimi balso pakete, nutolusiame nuo prarastojo per 20 ms.
5. Balso paketo vertę galima prognozuoti realiame laike. Didžiausias gaunamas koreliacijos koeficientas su tikrąja paketo verte yra 0,6.

Darbo rezultatų aprobavimas

Disertacijos tematika atspausdinti 7 moksliniai straipsniai: keturi – recenzuojamuose žurnaluose, trys – kituose leidiniuose.

Darbo tema buvo perskaityti 8 pranešimai mokslinėse konferencijose:

- 2 pranešimai COST 290 valdymo komiteto posėdžiuose
- 6 pranešimai respublikinėje konferencijoje „*Elektronika*“ 2005, 2006, 2007, 2008 ir 2009 m., Vilniuje.

Disertacijos struktūra

Disertaciją sudaro įvadas, keturi skyriai ir rezultatų apibendrinimas.

Darbo apimtis yra 105 puslapiai, neskaitant priedų, tekste panaudotos 31 numeruotos formulės, 45 paveikslai ir 13 lentelių. Rašant disertaciją buvo panaudota 76 literatūros šaltiniai.

1

Balso kokybės vertinimo uždaviniai ir metodai

Šiame skyriuje aptariami balso kokybės vertinimo uždaviniai telekomunikacijose. Analizuojami veiksniai, galintys turėti įtakos perduodamo balso kokybei. Pagrindinis dėmesys skiriamas ištrintiems paketams – svarbiausiam mobiliojo ryšio sistemų sukuriamam balso iškraipymui. Apžvelgiami ir analizuojami žinomi balso kokybės vertinimo būdai ir priemonės.

1.1. Balso kokybė telekomunikacijose

Visais telekomunikacijų vystymosi laikotarpiais buvo aktualus balso perdavimo kokybės klausimas. Kai telekomunikacinės priemonės buvo analoginės, tai ir vertinimo kriterijai buvo atitinkami. Buvo matuojama signalo triukšmo santykis, interferencinių triukšmų lygis ir kiti. Kokybė buvo nusakoma išmatavus šiuos fizinius parametrus. Kuo jie artimesni idealaus perdavimo sąlygoms, tuo kokybė geresnė.

Skaitmeniniai duomenų perdavimo metodai, kai balsas perduodamas kodais, gerokai pakeitė kokybės vertinimo bendrą vaizdą. Dabartinėse sistemose balso kokybę įtakoja ne tik ryšio kanalo signalo ir triukšmo santykis, bet ir balso

kodavimo algoritmai, perdavime naudojama moduliacijos rūšis, apsaugos nuo klaidų algoritmai, paketinių klaidų maskavimo algoritmai ir kiti veiksniai. Taigi bendras kokybės įvertinimo uždavinys tampa ganėtinai sudėtingas.

ITU paslaugų kokybę (angl. *Quality of Service – QoS*) – apibūdino kaip paslaugos savybių bendrą efektą, atspindintį vartotojo pasitenkinimo lygį (ITU Rec. E.800 1996). Kaip matyti pagal ITU nuostatas QoS yra susitarimo dalykas (Kajackas *et al*. 2004; Batkauskas 2006). Nors šioje disertacijoje analizuojama tik balso paslauga ir jos kokybė, bet QoS vertinimo bendrieji principai galioja ir šiuo atveju.

Taip suformuluota QoS koncepcija pagal OSI modelį orientuota į taikymų lygmenį, nustatoma iš vartotojo pozicijų. Akivaizdu, kad vartotojo pasitenkinimo laipsnis priklauso nuo daugelio techninių faktorių, nuo vartotojo patirties ir savaime aišku – nuo paslaugos kainos. Kaip žinoma, paslaugų kaina yra ir svarbus telekomunikacijų paslaugų vartojimo lygmens valdymo faktorius. Dinamiškai keičiant tarifus vienų laikotarpių apkrova mažinama, kitų laikotarpių – padidinama. Pagal ITU koncepciją QoS neatsiejama nuo paslaugų kainos, nuo tarifų.

Pagal ITU suformuluotą kokybės sampratą balso perdavimo paslaugos kokybė yra subjektyvi. Vertinimo subjektyvumas yra nulemtas tuo, kad vartotojo pasitenkinimo lygis priklauso nuo daugelio dalykų – nuo balso signalo iškraipymų dydžio, nuo paslaugų kainos, jų prieinamumo, alternatyvių paslaugų buvimo ar prieinamumo, nuo vartotojo patirties ir aišku nuo paslaugos kainos.

Šiame darbe kokybės vertinimo tyrimai sukoncentruoti į išskirtines mobiliojo ryšio sąlygas, kai perduodant balso duomenų paketus dalis iš jų būna nepataisomai iškraipomi ir todėl ištrinami. Ištrynus balso duomenų paketą ištrinama ir tam tikra balso signalo atkarpa.

Pastebėkime, kad telekomunikacijų paslaugos kokybė yra ir viso tinklo kokybės bei atskirų jo dalių kokybės indikatorius (Kajackas *et al.* 2000). Kitaip tariant, jei tinklo operatorius turi susidaręs paslaugų kokybės žemėlapį, tai iš jo gali matyti, kuriose tinklo vietose reikia atlikti pakeitimus, įrengti papildomą įrangą. Tačiau toks kokybės nagrinėjimas nieko nepasako apie konkretaus vartotojo konkretaus pokalbio metu patirtą kokybę. Galima išskirti kelias sąvokas, kurios labiau tinkamos vartotojo patirtai kokybei vertinti:

QoE (angl. *Quality of Experience*) sąvoką ITU įvedė 2007 m. pradžioje (Hestnes *et al*. 2003; Alben 1996; Corrie *et al*. 2003; Jain 2004). QoE – subjektyviai suvoktas, bendras paslaugos priimtinumas vartotojui.

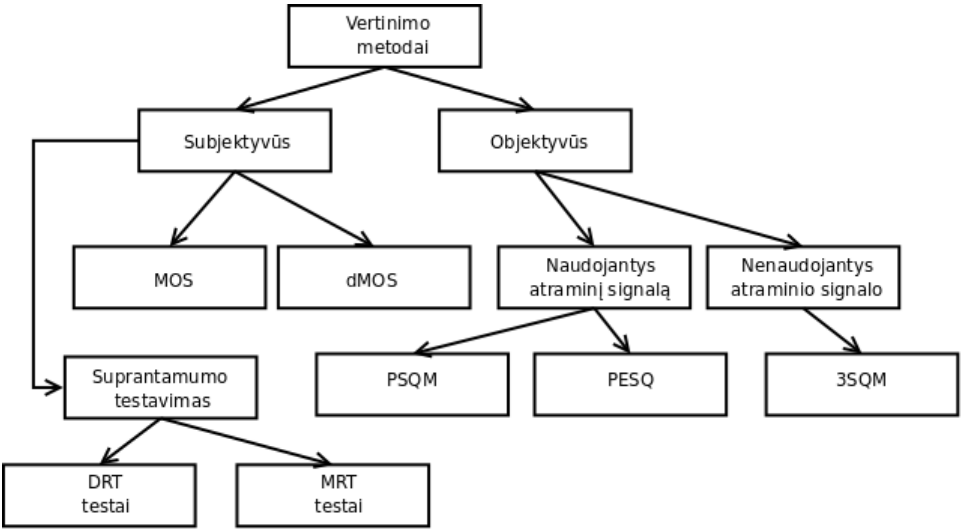
Individuali kokybė – iQoS, apibrėžta (Kajackas *et al*. 2004), akcentuoja individo – konkretaus vartotojo patirtą kokybę, atskiria ją nuo statistiškai matuojamų tinklo charakteristikų.

Taikant individualios kokybės sąvoką galima kalbėti ir apie iQoE. Tai yra individualaus vartotojo gautų paslaugų suvokimo metrika. Panašiai kaip ir iQoS

apibrėžime, esminis dalykas čia yra vartotojo individualumas, jo gautų paslaugų kokybės atskyrimas nuo tinklo vidutinės suteikiamos kokybės. Naudojant iQoE sąvoką balso perdavimams, jau reikia pereiti prie atskirų vartotojų gauto balso kokybės analizės. Toliau apžvelgsime metodus, kurie tinkami tokiai analizei atlikti.

1.2. Balso kokybės vertinimo metodai

Šiuolaikinių balso kokybės vertinimo metodų klasifikacija pateikta žemiau 1.1 paveiksle**.**



1.1 pav. Balso kokybės vertinimo metodų klasifikacija
Fig. 1.1. Classification of voice quality evaluation methods

Kaip matyti, vertinimo metodus galima skirti į dvi dideles grupes – objektyviuosius ir subjektyviuosius. Subjektyvius vertinimus atlieka žmonės – ekspertai. Subjektyvūs vertinimai gali būti gauti, pateikiant vertintojams tik iškraipytą balso signalą. Toks yra MOS (ITU-T Rec. P.800. 1996) metodas. Kitas būdas, kai vertintojai gali klausyti iškraipytą ir neiškraipytą signalus ir juos tarpusavyje palyginti. Toks yra diferencinis MOS – dMOS (angl.D*ifferential MOS*) – metodas.

Objektyvieji vertinimo metodai yra algoritminiai – kompiuteriniai metodai. Jie vertinant naudoja determinuotą algoritmą, kuris visada duoda tokius pačius

rezultatus, jei analizuojami tie patys signalai. Objektyviuosius vertinimo metodus taip pat galime skirti į dvi dideles grupes. Pirmajai priklauso metodai, naudojantys atraminį (etaloninį) signalą (ITU-T Rec. P.862 2001; ITU-R Rec. BS-1387 2001). Jie palygina atraminį signalą su iškraipytuoju ir priskiria signalui įvertį, kuris koreliuotas su MOS įverčiais.

Kita grupė – metodai, nenaudojantys etaloninio signalo (ITU-T Rec. P.563 2004). Šie metodai nagrinėdami statistines iškraipyto signalo savybes skaičiuoja kokybės įvertį, kuris taip pat būtų koreliuotas su MOS vertinimais. Aišku, kad pasiekiama koreliacija nėra labai didelė.

Aišku, kad tiksliausiai iQoS sampratas atspindi subjektyvaus vertinimo būdas – MOS, tačiau jis nėra tinkamas realių pokalbių kokybei vertinti. Kompiuteriniai metodai yra spartesni ir kai kurie iš jų gana patikimi, jų vertinimai artimi MOS įverčiams.

Yra dar viena specifinė ribotų taikymų šnekos kokybės vertinimo metodų grupė – suprantamumo testai (angl. *intelligibility testing*). Pastarieji testai nagrinėja panašių žodžių grupes, šių žodžių perdavimą ir žmonių ar automatinio atpažinimo sistemos gebėjimą atskirti panašius žodžius vieną nuo kito.

1.3. MOS vertinimo metodas

Šiuo metu gerai žinomas subjektyvus balso kokybės vertinimo būdas – MOS. Tai intuityviai akivaizdus būdas. Jo esmė – balso įrašų kokybę nustato ekspertai.

MOS įverčius nustato specialiai parinkti ir apmokyti (informuoti) žmonės – ekspertai. Kiekvienas iš jų vertina pateiktus balso įrašus tam tikru balu pagal savo subjektyvų kokybės suvokimą. Bendras MOS įvertis gaunamas suvidurkinus ekspertų nuomones. Yra keletas MOS testų rūšių, tačiau vertinant perduoto balso kokybę svarbiausieji yra klausymosi testai.

Dvi skirtingos MOS versijos gaunamos, kai vertinimo metu vertintojai turi atraminį neiškraipytą signalą palyginimui arba jo neturi. Populiaresnis ir įprastas variantas yra, kai naudojama tik iškraipyta signalo versija. Dažniausiai įverčiams naudojama 5 balų skalė. Vienas balas reiškia, kad iškraipyto signalo kokybė nepriimtina, prasta, sunkiai suvokiama pati šneka, 5 balai reiškia, kad signalo kokybė puiki, vertintojas nepastebi jokių trūkumų.

Pagal MOS standartizuotą versiją (ITU-T Rec. P.800 1996), klausymui turi būti naudojami 2–3 sekundžių trukmės rišlūs sakiniai. Svarbu pastebėti, kad į daugelį šnekos iškraipymų vertintojas gali nekreipti dėmesio, nes sakinio prasmę supranta pagal kontekstą. Iš čia aišku, kad tam tikrais atvejais pasikliauti MOS kokybės testais negalima, nes kai kuriose situacijose svarbu atskirų žodžių

suprantamumas. Tikslesni vertinimai tokiu atveju gali būti gaunami tiriant šnekos suprantamumą. Tam taip pat yra žinomi testavimo metodai.

MOS vertinimai yra sudėtingi, brangūs ir daug laiko reikalaujantys. Kita problema, kad skirtingose laboratorijose atliekant šiuos testus, rezultatai gaunami skirtingi. Dėl šių priežasčių vis labiau patrauklūs tampa objektyvūs balso kokybės vertinimo algoritmai. Šiais laikais MOS testai atliekami retai ir tik tais atvejais, kai reikia gauti esminių išvadų apie balso kokybę.

1.4. Šnekos suprantamumo vertinimai

Yra specifinių su balso kokybe sietinų uždavinių, kai klausantysis turi išgirsti ir suprasti pavienius žodžius ar komandas. Tokiems uždaviniams MOS testai nėra pritaikyti. Kai norima nustatyti kaip sėkmingai suprantami pavieniai žodžiai naudojami taip vadinami suprantamumo rimavimo testai (Delogu 1991).

Diagnostiniame rimavimo teste – DRT – parenkami vienskiemeniai angliški žodžiai, kurie sudaryti iš garsų sekos priebalsė–balsė–priebalsė. Angliškame DRT yra 192 žodžiai, kurie sugrupuoti poromis. Viso yra 96 žodžių poros. Poroje būtinai skiriasi pirmosios priebalsės (1.1 lentelė) (ANSI S3.2-1989 (R1999). 1989). Vykdant DRT testus klausytojui pateikiama rašytine forma žodžių pora ir prašoma pažymėti, kurį žodį pasakė kalbėtojas.

DRT testuose klausytojui pateikiami pavieniai žodžiai, nenaudojami „nešantieji“ sakiniai. Nešantysis sakinys – tai sakinys, kuris naudojamas pateikti klausytojams testinį žodį statistiniame šnekos suprantamumo teste, pavyzdžiui, „Would you write <testo žodis> now“. Testo žodis ištariamas jo nepabrėžiant. Tas pats sakinys naudojamas kiekvienam testo žodžiui. „Nešantieji“ sakiniai garantuoja, kad anksčiau ištartas žodis sukelia aidą ir taip į testo rezultatą įnešamas to garso tikrasis poveikis.

Diagnostinis rimavimo testas paremtas daugybe šnekos ypatybių ir šio testo rezultatai atskleidžia klaidas pirmuose priebalsiuose. Testas gali būti atliktas greitai ir rezultatai pateikti keliais skirtingais būdais. Šį, kaip ir kitus suprantamumo testus, lengva modifikuoti kai atpažinimui naudojama automatinė šnekos atpažinimo sistema.

Modifikuotame rimavimo teste (MRT) (Syrdal *et al*. 1995) naudojami 50 žodžių, sugrupuotų po šešis panašiai skambančius vienskiemenius angliškus žodžius, sąrašas statistiniam suprantamumo tyrimui atlikti (1.2 lentelė).

Kiekvienas žodis sudarytas iš garsų sekos priebalsė–balsė–priebalsė. Šeši žodžiai kiekvienoje eilutėje skiriasi tik pirmos arba paskutinės priebalsės garsu. Klausytojams parodomas šešių žodžių sąrašas ir paprašoma identifikuoti kurį iš tų šešių žodžių pasakė kalbėtojas. Dažniausiai MRT teste naudojamas anksčiau aptartas „nešantysis“ sakinys.

1.1 lentelė. Diagnostiniame rimavimo teste naudojamų žodžių sąrašas
Table 1.1. Word list for diagnostic rhyme test

Garso skambėjimas *(*angl. *Voicing)*		Garso nosingumas *(*angl. *Nasality)*		Garso išlaikymas *(*angl. *Sustenation)*	
veal	feal	meat	beat	vee	bee
bean	peen	need	deed	sheet	cheet
gin	chin	mitt	bit	vill	bill
dint	tint	nip	dip	thick	tick
zoo	sue	moot	boot	foo	pooh
dune	tune	news	dues	shoes	choose
vole	foal	moan	bone	those	doze
goat	coat	note	dote	though	dough
zed	said	mend	bend	then	den
dense	tense	neck	deck	fence	pence
vast	fast	mad	bad	than	dan
gaff	calf	nab	dab	shad	chad
vault	fault	moss	boss	thong	tong
daunt	taunt	gnaw	daw	shaw	chaw
jock	chock	mom	bomb	von	bon
bond	pond	knock	dock	vox	box
zee	thee	weed	reed	yield	wield
cheep	keep	peak	teak	key	tea
jilt	gilt	bid	did	hit	fit
sing	thing	fin	thin	gill	dill
juice	goose	moon	noon	coop	poop
chew	coo	pool	tool	you	rue
joe	go	bowl	dole	ghost	boast
sole	thole	fore	thor	show	so
jest	guest	met	net	keg	peg
chair	care	pent	tent	yen	wren
jab	gab	bank	dank	gat	bat
sank	thank	fad	thad	shag	sag
jaws	gauze	fought	thought	yawl	wall
saw	thaw	bong	dong	caught	thought
jot	got	wad	rod	hop	fop
chop	cop	pot	tot	got	dot

1.2 lentelė. Modifikuotame rimavimo teste naudojamų žodžių sąrašas
Table 1.2. Word list for modified rhyme test

Garso skambėjimas *(*angl. *Voicing)*		Garso nosingumas *(*angl. *Nasality)*		Garso išlaikymas *(*angl. *Sustenation)*	
veal	feal	meat	beat	vee	bee
bean	peen	need	deed	sheet	cheet
gin	chin	mitt	bit	vill	bill
dint	tint	nip	dip	thick	tick
zoo	sue	moot	boot	foo	pooh
dune	tune	news	dues	shoes	choose
vole	foal	moan	bone	those	doze
goat	coat	note	dote	though	dough
zed	said	mend	bend	then	den
dense	tense	neck	deck	fence	pence
vast	fast	mad	bad	than	dan
gaff	calf	nab	dab	shad	chad
vault	fault	moss	boss	thong	tong
daunt	taunt	gnaw	daw	shaw	chaw
jock	chock	mom	bomb	von	bon
bond	pond	knock	dock	vox	box
zee	thee	weed	reed	yield	wield
cheep	keep	peak	teak	key	tea
jilt	gilt	bid	did	hit	fit
sing	thing	fin	thin	gill	dill
juice	goose	moon	noon	coop	poop
chew	coo	pool	tool	you	rue
joe	go	bowl	dole	ghost	boast
sole	thole	fore	thor	show	so
jest	guest	met	net	keg	peg
chair	care	pent	tent	yen	wren
jab	gab	bank	dank	gat	bat
sank	thank	fad	thad	shag	sag
jaws	gauze	fought	thought	yawl	wall
saw	thaw	bong	dong	caught	thought
jot	got	wad	rod	hop	fop
chop	cop	pot	tot	got	dot

MRT rezultatai parodo pirmos ir antros priebalsės garsų skirtumų klaidas. Klausytojo atsakymas gali būti skaičiuojamas kaip:

- teisingai išgirstų žodžių skaičius;

- neteisingai išgirstų žodžių skaičius;
- priebalsių garsų supainiojimo dažnis.

Mokslinėje spaudoje yra paskelbta darbų, kuriuose nagrinėjami įvairių kalbų balso ar šnekos kodavimo, kokybės bei suprantamumo vertinimai. Pastebėta, kad įvairių kalbų kokybės vertinimo skalės gali pastebimai skirtis. Pavyzdžiui CELP šeimos koderiais koduotos skirtingos kalbos (arabų, anglų ir lietuvių) vertinama skirtingais MOS balais (Paulikas 2008). Lietuvių kalba balso kodavimo ypatumų tyrimų nebuvo daryta. Todėl siekiant ištirti balso kodavimo įtaką lietuvių šnekos suprantamumui sudarėme lietuviškų žodžių duomenų bazę (DB). Sudarant DB buvo kombinuoti anksčiau aptarti DRT ir MRT metodai.

Kadangi visi anksčiau minėti suprantamumo testai yra paremti iš anglų kalbos žodžių sudarytomis bibliotekomis ir yra skirti tirti angliškų žodžių suprantamumui, tai norint ištirti lietuvių kalbos žodžių suprantamumą, pirmiausia reikia sukurti lietuviškų žodžių biblioteką, skirtą šiems testams atlikti. Remiantis DRT ir MRT sudarymo taisyklėmis galima sukurti žodžių sąrašą, kur žodžiai skirtųsi priebalsėmis. Kadangi lietuviškų vienskiemenių panašių žodžių, kurie skiriasi viena raide, nėra daug, todėl teko nukrypti nuo anglų kalbos DRT ir MRT ir biblioteką sudarėme iš dviskiemenių žodžių.

Sugrupuotų žodžių, kurie skiriasi viena raide, sąrašai pateikti 1.3 – 1.5 lentelėse. Šiose lentelėse pateikti sugrupuoti žodžiai skiriasi pirmąja arba antrąja priebalse. Šių sąrašų skirtumas tas, kad 1.3 lentelėje yra 66 lietuviškų žodžių, sugrupuotų į 45 žodžių, kurie skiriasi viena priebalse, poras, 1.4 lentelėje yra 48 žodžiai, sugrupuoti po tris panašiai skambančius dviskiemenius lietuviškus žodžius, o 1.5 lentelėje yra 50 žodžių, sugrupuotų po penkis panašiai skambančius dviskiemenius lietuviškus žodžius, kurie skiriasi viena priebalse.

Yra pastebėta, kad šnekos suprantamumui priebalsės yra svarbesnės nei balsės (Barnett 1999; Cutler 1987). Priebalsės tariamos minkščiau nei balsės ir jos lengviau „paskęsta“ fono triukšmuose. Svarbu pastebėti, kad tariant priebalses sukuriamas aukštesnių dažnių garsas nei tariant balses. Ribojant aukštesnių dažnių juostą blogėja šnekos suprantamumas.

Siekiant ištirti balsių iškraipymų reikšmingumą šnekos suprantamumui sukūrėme lietuviškų žodžių biblioteką. Šioje bibliotekoje sugrupuoti žodžiai, kurie skiriasi viena balse. 1.6 ir 1.7 lentelėse yra pateikti žodžių sąrašai, skirti suprantamumo tyrimui, kai prarandama balsė. 1.6 lentelėje yra pateikti 96 lietuviški žodžiai, sugrupuoti į 48 žodžių, kurie skiriasi viena balse, poras, o 1.7 lentelėje pateikti 24 žodžiai, kurie yra sugrupuoti po tris panašiai skambančius dviskiemenius lietuviškus žodžius, kurie skiriasi viena balse.

1.3 lentelė. Žodžių, kurie skiriasi viena priebalse, poros
Table 1.3. Pairs of words with one different consonant

Aĩnis	Aĩris	Gumà	Sumà	Mėsà	Vėsà
Balà	Basà	Fė́ja	Sė́ja	Odà	Olà
Bejė̀	Benė̀	Ýda	Ýla	Parà	Žarà
Bė̀t	Nė̀t	Jòdas	Jògas	Põnas	Póras
Bėdà	Pėdà	Jogà	Togà	Porà	Pozà
Bi̇̀sas	Bi̇̀tas	Kãla	Mãla	Pū̃kas	Rū̃kas
Bū̃vis	Šū̃vis	Kavà	Lavà	Rãgas	Rãtas
Datà	Natà	Kovà	Kopà	Ri̇̀kis	Vi̇̀kis
Dė̃da	Dė̃ga	Kovà	Movà	Sulà	Sumà
Dėžė̃	Vėžė̃	Kóvas	Póvas	Ši̇̀las	Ši̇̀tas
Dorà	Porà	Lãbas	Žãbas	Togà	Tonà

1.4 lentelė. Po tris sugrupuotų žodžių, kurie skiriasi viena priebalse, sąrašas
Table 1.4. Triples of words with one different consonant

Ãmas	Ãras	Ãsas	Laĩdas	Laĩkas	Laĩvas
Aki̇̀s	Ašį̀s	Avi̇̀s	Mãgas	Nãgas	Rãgas
Basà	Kasà	Rasà	Nãgas	Nãmas	Nãras
Bylà	Pylà	Tylà	Sagà	Salà	Savà
Chãmas	Nãmas	Šãmas	Sodà	Sofà	Sojà
Dėlė̃	Dėmė̃	Dėžė̃	Sū̃kis	Šū̃kis	Pū̃kis
Kadà	Kasà	Kavà	Pumà	Pupà	Putà
Ki̇̀nas	Ki̇̀ras	Ki̇̀tas	Vagà	Vatà	Vazà

1.5 lentelė. Po penkis sugrupuotų žodžių, kurie skiriasi viena priebalse, sąrašas
Table 1.5. Groups of five words with one different consonant

Bãdas	Bãkas	Bãlas	Bãras	Bãtas
Bãtas	Mãtas	Pãtas	Rãtas	Vãtas
Bãkas	Jãkas	Lãkas	Pãkas	Tãkas
Bi̇̀tas	Ki̇̀tas	Li̇̀tas	Mi̇̀tas	Ši̇̀tas
Cãras	Gãras	Kãras	Mãras	Nãras
Dėlė̃	Gėlė̃	Lėlė̃	Vėlė̃	Žėlė̃
Dỹla	Kỹla	Šỹla	Tỹla	Žỹla
Lãbas	Lãkas	Lãpas	Lãšas	Lãtas
Mãčas	Mãgas	Mãras	Mãtas	Mãžas
Pãdas	Pãkas	Pãlas	Pãsas	Pãtas

1.6 lentelė. Žodžių, kurie skiriasi viena balse, poros
Table 1.6. Pairs of words with one different vowel

Áibė	Éibė	Vėsà	Visà	Kãlis	Kẽlis
Balà	Bylà	Vazà	Vizà	Kìras	Kùras
Bėdà	Būdà	Želẽ	Žolẽ	Lęšis	Lūšis
Dãnė	Dỹnė	Ãmas	Òmas	Medùs	Midùs
Erà	Yrà	Ãras	Óras	Mólis	Mòlas
Fìga	Fùga	Ãsas	Ũsas	Miẽlas	Miẽlės
Gamà	Gumà	Bãdas	Bũdas	Nãras	Nóras
Ýda	Óda	Bãkas	Bùkas	Pãkas	Pũkas
Jėgà	Jogà	Bãras	Bãris	Pìkas	Pìkis
Kapà	Kopà	Bãras	Bòras	Rìmas	Ròmas
Kavà	Kovà	Bãsas	Bõsas	Rẽžis	Rỹžis
Natà	Notà	Čèkas	Čèkis	Sỹkis	Sũkis
Parà	Porà	Gẽras	Gãras	Šõkis	Šũkis
Sãgė	Sẽgė	Gètas	Gòtas	Tònas	Tùnas
Salà	Sulà	Gìdas	Gùdas	Vãlas	Võlas
Šãkė	Šùkė	Indas	Indis	Žãdas	Žỹdas

1.7 lentelė. Po tris sugrupuotų žodžių, kurie skiriasi viena balse, sąrašas
Table 1.7. Triples of words with one different vowel

Bãtas	Bìtas	Bùtas
Ìlgas	Ilgis	Ilgùs
Lamà	Lemà	Lomà
Lãtas	Lìtas	Lẽtas
Mãras	Mū́ras	Mèras
Mãtas	Mẽtas	Mìtas
Mẽlas	Mòlas	Mùlas
Rãtas	Rẽtas	Rýtas

Sudarytuose lietuviškų žodžių sąrašuose žodžiai sugrupuoti neatsižvelgiant į šnekos savybes, tokias kaip garso skambėjimas, nosingumas, išlaikymas, šnypštimas, žemumas ar glaustumas, kaip į tai atsižvelgta DRT. Mūsų sudaryta biblioteka turi MRT ir DRT savybių.

Vertinant šnekos suprantamumą galima atsižvelgti, pavyzdžiui, į balsių ilgumą ar priebalsių buvimo vietą žodyje (priebalsė yra pirma žodžio raidė ar ne).

1.5. Objektyvieji vertinimo metodai

Šiame poskyryje apibendrintai apžvelgsime objektyvių balso kokybės algoritmų veikimo schemas. Nagrinėsime tik tuos algoritmus, kurie veikia naudodami tik balso signalus, tačiau neatsižvelgia į telekomunikacinės sistemos matuojamus parametrus, tokius, kaip paketų praradimai, BER (angl. *Bit Error Rate*), FER (angl. *Frame Error Rate*), signalo lygis ir kiti.

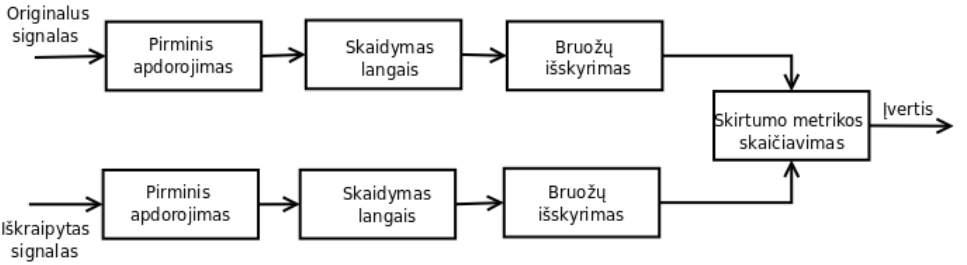
1.5.1. Objektyviųjų invazinių vertinimo metodų apžvalga

Invaziniai balso kokybės vertinimo metodai veikia signalų palyginimo būdu. Vienas signalas laikomas idealiu ir telekomunikacijose tai dažniausiai garso signalas iš karto po mikrofono. Antrasis signalas yra tiriamasis – iškraipytas. Jis gaunamas, kai signalas pereina įvairias transformacijas telekomunikacinėje sistemoje – kvantavimą pagal reikšmes ir kodavimą – dekodavimą.

Visi iki šiol naudojami invaziniai algoritmai turi panašią bazinę struktūrą ir operacijos, atliekamos su signalais taip pat labai bendros. Ši struktūra parodyta 1.2 paveiksle. Pirminio apdorojimo blokas dažniausiai atlieka tokias operacijas, kaip vidurkio iš signalo pašalinimas ir spektro išlyginimas. Po pirminio apdorojimo signalai (originalus ir iškraipytas) skaidomi kadrais, kurių ilgis dažniausiai būna 10–30 ms. Tada iš signalo kiekviename kadre išskiriami balso bruožų vektoriai. Tai gali būti spektriniai, LPC (angl. *Linear Predictive Coding*) ar kitokie požymių vektoriai (Itakura 1970; Chu 2003; Lipeika 1999), kurie turi savyje tam tikrą parametrinę informaciją apie balso signalą atitinkamame kadre. Esminis faktorius čia yra tai, kad kuo daugiau skiriasi atitinkamai originalaus ir iškraipyto signalų požymių vektoriai, tuo labiau tai turės įtakos ir klausomam skirtumo tarp signalų suvokimui. Skirtumo metrika skaitiškai įvertina, kiek tarpusavyje skiriasi atitinkami bruožų vektoriai. Taip pat ši metrika gali agreguoti laike skirtumus tarp skirtingų kadrų vektorių. Galutinis įvertis dažniausiai ir yra tiesiog skirtumo metrika. Pastarąją stengiamasi padaryti tokia, kad gaunami įverčiai būtų kuo labiau koreliuoti su atitinkamais MOS įverčiais.

Paminėsime plačiausiai paplitusius ir žinomiausius invazinius garso kokybės vertinimo algoritmus. Konkrečiai balso kokybei vertinti pirmas plačiai žinomas algoritmas buvo PSQM (angl. *Perceptual Speech Quality Measure*), jis yra apibrėžtas ITU-T P.861 rekomendacijoje. Šis algoritmas dabar praktiškai visuose taikymuose pakeistas PESQ algoritmu. PESQ apibrėžtas ITU-T P.862 rekomendacijose (ITU-T Rec. P.862 2001).

Bendrai garso kokybei įvertinti populiariausias yra PEAQ (angl. *Perceptual Evaluation of Audio Quality*) algoritmas, kuris yra standartizuotas kaip ITU-R BS.1387 rekomendacija (ITU-R Rec. BS.1387 1999).



1.2 pav. Invazinių algoritmų apibendrinta schema
Fig. 1.2. General diagram of invasive algorithms

1.5.2. Objektyviųjų neinvazinių vertinimo metodų apžvalga

Neinvaziniai balso kokybės vertinimo algoritmai veikdami naudoja tik vieną signalą – iškraipytąjį. Akivaizdu, kad tokio pobūdžio algoritmai nėra tokie tikslūs (koreliacijos su MOS vertinimais prasme), kaip invaziniai. Tačiau šiuos algoritmus galima naudoti realioje galinėje vartotojo įrangoje. Tėra tik vienas standartizuotas neinvazinis balso kokybės vertinimo algoritmas – 3SQM (ITU-T Rec. P.563 2004). Čia jį trumpai ir apžvelgsime.

Algoritmas, kaip ir daugelis balso signalų apdorojimo algoritmų, startuoja atlikdamas pirminį apdorojimą, kurį sudaro trys dalys:

- filtravimas imituojant standartinės garso atgaminimo įrangos dažninę charakteristiką;
- šnekos lygio pareguliavimas;
- šnekos ir tylos atskyrimas naudojant balso aktyvumo detektorių;
- pagrindinių šnekos kokybės blogėjimo priežasčių išskyrimas ir jų kiekybinis įvertinimas;
- šnekos trakto analizė ir šnekos nenatūralumas – nustatoma, ar kalbėtojas vyras, ar moteris, nustatoma „robotiško balso“ intervalai ir pasireiškimo stiprumas;
- stipraus pridėtinio triukšmo analizė – nustatoma statinio ir segmentinio triukšmų signalo trukdžio santykis;
- signalo trūkiai, nutilimai ir nykimai įvertinami kaip paskutinis faktorius;
- paskutiniame žingsnyje nustatomas dominuojantis iškraipymas ir pagal tai formuojamas galutinis įvertis. Pats įverčio formavimas atliekamas, kaip visada, tokiu būdu, kad būtų gautas kuo didesnis koreliacijos koeficientas su MOS įverčiais.

Yra ir kitų neinvazinių balso kokybės vertinimo algoritmų. Daugelis jų remiasi statistiniais „švarios“ šnekos modeliais, o patys algoritmai skaičiuoja, kiek nukrypsta matuojamam signalui apskaičiuotas modelis nuo iš anksto sudarytų statistinių modelių.

Pastebėsime, kad neinvaziniai algoritmai visgi neduoda pakankamo tikslumo, kai iškraipymų pobūdis yra sudėtingas, pavyzdžiui, paketų praradimo atveju koduotoje šnekoje. Pagal 3SQM algoritmo kūrėjų duomenis (Opticom, 2004), koreliacija tarp MOS ir 3SQM įverčių visuose testuose yra didesnė, negu 0,8, tačiau ji gerokai mažesnė, negu MOS koreliacija su PESQ.

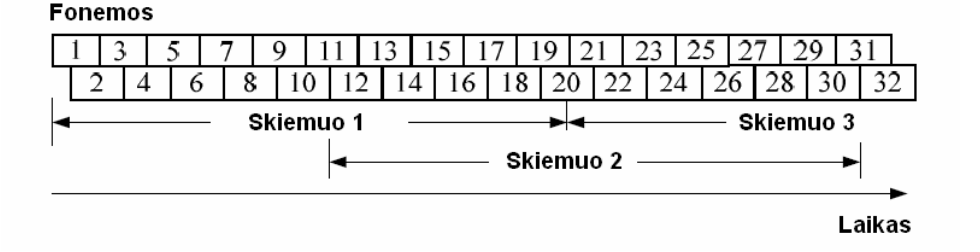
1.6. PESQ algoritmas

Apžvelgsime PESQ algoritmą, kuris sudaro pagrindą daugeliui tolimesnių mūsų tyrimų, nes PESQ algoritmo tikslumas yra bene geriausias iš žinomų standartizuotų algoritmų (Pennock 2002; Rix *et al*. 2001). Tikslumas, vėlgi, suprantamas koreliacijos su MOS prasme.

PESQ algoritmo bendra struktūra iš esmės tokia pati, kaip aukščiau pateikta bendra invazinių algoritmų schema.

Pirminio apdorojimo blokas sudarytas iš lygio ir laiko išlyginimo posistemių. Laiko išlyginimo posistemė suveikia tada, kai iškraipytas signalas yra suvėlintas originalaus signalo atžvilgiu. Mūsų tyrimams tai neaktualu, nes vėlinimo efektų nenagrinėsime. Po pirminio apdorojimo signalai pereina psichoakustinio apdorojimo bloką – čia vyksta signalų skaidymas į kadrus, požymių išskyrimas ir skirtumo metrikos skaičiavimas. Skirtumo tarp signalų metrika parinkta taip, kad ji atspindėtų žmogaus klausos savybes. Kiek detaliau aptarsime skirtumo tarp signalų metrikos skaičiavimą.

Tiek originalus, tiek iškraipytas signalai yra apdorojami naudojant segmentavimą kadrais, tokiais, kaip parodyta 1.3 paveiksle. Pagrindinis apdorojimo vienetas yra fonema, kurios trukmė yra 32 ms. Persidengimas tarp gretimų fonemų yra 50 %. 20 persidengiančių fonemų atitinka „skiemenį“ (angl. *Syllable*), kurio trukmė yra 320 ms. Pastarieji taip pat persidengia 50 %.



1.3 pav. PESQ laiko segmentavimas
Fig. 1.3. PESQ time segmentation

Pirmu psichoakustinio apdorojimo žingsniu skaičiuojami simetriniai D_{sn} ir asimetriniai D_{an} iškraipymai kiekvienai fonemai. Kitu žingsniu signalas yra

atvaizduojamas psichoakustine forma – fonemų iškraipymų vertės D_{sn} ir D_{an} yra agreguojamos kiekviename skiemenyje naudojant tokias formules:

$$L_{\text{DS}} = \sqrt[6]{\frac{1}{20}\sum_{n=1}^{20} D_{sn}^{6}}\ ,\ L_{\text{DA}} = \sqrt[6]{\frac{1}{20}\sum_{n=1}^{20} D_{an}^{6}}\ . \tag{1.1}$$

Galiausiai algoritmas atlieka integravimą per visą balso signalų ilgį. Agregavimas atliekamas naudojant skiemenų iškraipymus ir klasikinį kvadratinį vidurkinimą:

$$d_{\text{sym}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} L_{\text{DS}}^{2}(i)}\ ,\ d_{\text{asym}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} L_{\text{DA}}^{2}(i)}\ , \tag{1.2}$$

čia N – skiemenų skaičius visame signale. Galutinis PESQ įvertis yra paskaičiuojamas kaip

$$Q_{\text{PESQ}} = 4{,}549 - 0{,}1d_{\text{sym}} - 0{,}0309d_{\text{asym}}. \tag{1.3}$$

Toliau dažnai vartosime diferencinį PESQ įvertį, kurį apibrėžkime taip:

$$V_{\text{PESQ}} = 4{,}549 - Q_{\text{PESQ}} = 0{,}1d_{\text{sym}} + 0{,}0309d_{\text{asym}}. \tag{1.4}$$

Toks įvertis tampa lygiu nuliui, jei signalai PESQ algoritmo įėjime yra tapačiai lygūs. Kuo didesnė diferencinio PESQ įverčio reikšmė, tuo didesnis girdimasis (psichoakustinis) skirtumas tarp originalaus ir iškraipytojo signalų.

1.7. Spektrinis šnekos iškraipymų nustatymo metodas

PESQ algoritmas yra gana sudėtingas ir ne visada tinkamas, jei norime panagrinėti skirtumą tarp dviejų labai trumpų signalo atkarpų, pvz. 20 ms trukmės. Čia 20 ms pasirinkta neatsitiktinai, tokia yra daugelyje telekomunikacinių sistemų vienu kodu koduojamo balso segmento trukmė. Tokios trukmės balso segmentus koduoja GSM ir AMR koderiai.

Apskritai, norėdami nustatyti, kaip yra iškraipomi koduoti signalai kai prarandami paketai, naudosime dvi metrikas: PESQ arba dPESQ metriką, skaičiuojamą pagal (1.3) ir (1.4) išraiškas ir spektrinio skirtumo metriką.

Šios abi metrikos skaičiuojamos pagal bendrą invazinių algoritmų schemą, parodytą 1.3 paveiksle. Nuo konkrečios metrikos priklauso tik skirtumo metrikos skaičiavimai.

Patikslinsime algoritmą, kaip skaičiuojama spektrinio skirtumo metrika:

- Nuolatinės dedamosios kompensavimas atliekamas ir etaloniniam, ir iškraipytam signalui naudojant formulę:

$$x_{\text{of}}(i) = x_{\text{inp}}(i) - x_{\text{inp}}(i-1) + 0{,}999 x_{\text{of}}(i-1), \tag{1.5}$$

čia x_{of} – išėjimo signalas iš nuolatinės dedamosios kompensavimo filtro; x_{inp} – įėjimo signalas į pastarąjį filtrą.

- Spektrinio išlyginimo filtravimas atliekamas etaloniniam ir iškraipytam signalams naudojant tokį filtrą

$$x_{\text{pre}}(i) = x_{\text{of}}(i) - 0{,}97 x_{\text{of}}(i-1), \tag{1.6}$$

čia x_{pre} – išėjimo signalas iš spektrinio išlyginimo filtro; x_{of} – įėjimo signalas į pastarąjį filtrą. Kadangi paprastai šnekos amplitudžių spektras linkęs mažėti, kai dažnis didėja, tai šis filtras kompensuoja tokį efektą ir pagerina statistines požymių vektorių charakteristikas.

- Signalas, gautas po paskutinio punkto transformacijos skaidomas į 20 ms trukmės kadrus. Kadangi nagrinėjami šnekos signalai diskretizuoti 8000 Hz dažniu, tai reiškia, kad kiekviename kadre signalą sudaro vektorius, kurio ilgis yra 160 elementų. Kadrus imame nepersidengiančius.
- Skaičiuojam diskrečiąją Furjė transformaciją kiekvienam kadrui atskirai.
- Formuojam bruožų vektorių, kuriame yra 11 elementų, kiekvienam kadrui pagal tolesnį aprašymą. Pirmasis bruožų vektoriaus elementas $F(1)$ imamas lygiu signalo kadre energijai. Bet kuris kitas požymių vektoriaus elementas skaičiuojamas kaip energija 400 Hz pločio dažnių juostoje juostos , t. y.,

$$F(i) = \sum_{j=1+8\cdot(i-1)}^{1+8\cdot(i-2)} X(j), \tag{1.7}$$

čia $i = 2,...,11$; X – signalo x_{pre} diskrečioji Furjė transformacija, skaičiuojama kaip $X = FFT(x_{\text{pre}})$; x_{pre} – signalas nagrinėjamame kadre; j – Furjė transformacijos elemento numeris. Visi bruožų vektoriaus elementai $F(i)$, kai $i \neq 1$ dar normuojami taip, kad jų suma būtų lygi vienetui, t.y., atliekama tokia transformacija:

$$F(i) = \frac{F(i)}{\sum_{i=2}^{11} F(i)}, \text{ kai } i = 2,...,11. \tag{1.8}$$

- Kiekvienam kadrui paskaičiuojamas iškraipymo dydis tarp signalų. Tam naudojama klasikinė Euklido metrika:

$$d(n) = \sqrt{\sum_{i=1}^{11} \left(F_{\text{orig},n}(i) - F_{\text{iškr},n}(i)\right)^2}\,, \tag{1.9}$$

čia $F_{\text{orig,n}}$ ir $F_{\text{iškr,n}}$ – atitinkamai pradinio ir iškraipytų signalų požymių vektoriai n-ajame kadre; i – požymių vektoriaus i-asis elementas.

- Paskutiniame žingsnyje agreguojami iškraipymai $d(j)$ naudojant aritmetinį vidurkinimą

$$D = \frac{1}{N}\sum_{i=1}^{N} d(i)\,, \tag{1.10}$$

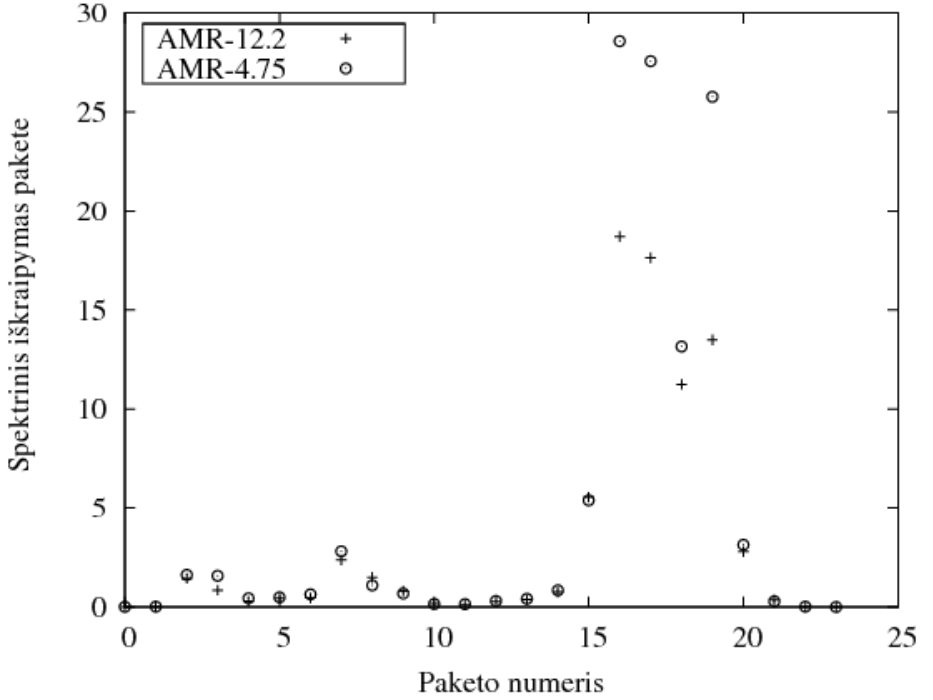
čia N – kadrų skaičius signale.

Aprašytoji signalo iškraipymo vertinimo metrika yra monotoniška iškraipymų didėjimo prasme – ji didesnė, kai iškraipymai didesni. Dar pastebėkime, kad pateiktoji metrika yra praktiškai identiška PESQ simetrinių iškraipymų skaičiavimui, tačiau PESQ algoritmas mūsų sistemoje veikia kaip „juoda dėžė“, todėl tikslinga panagrinėti ir tarpinius rezultatus. Pateiksime vieną pavyzdį. Tarkime, vienas trumpas žodis yra koduojamas AMR 12,2 kb/s ir AMR 4,75 kb/s koderiais (toliau koderius dažnai įvardysime naudodami žymėjimą AMR-n, kur n yra koderio angliškas žymuo, atitinkantis spartą kilobitais per sekundę). Intuityviai aišku, kad kai kodavimas vykdomas 4,75 kb/s koderiu, tai iškraipymai turėtų būti didesni. Grafiškai parodysime, kaip pasiskirstę laike spektriniai iškraipymai $d(i)$. Kadangi nagrinėjamas įrašas yra 20 paketų ilgio (400 ms), tai ir $d_{4,75}(i)$ ir $d_{12,2}(i)$ turės po 20 elementų. Šie vektoriai pavaizduoti 1.4 paveiksle. Kaip ir buvo galima tikėtis, vektoriaus $d_{4,75}(i)$ elementų vertės yra vidutiniškai didesnės, nes šiuo atveju kodavimo iškraipymai didesni. Taigi galime daryti išvadą, kad nagrinėjamas metodas yra pakankamai racionalus būdas sužinoti, kaip stipriai skiriasi du balso signalai. Šiame eksperimente gautas pilnas spektrinis signalo iškraipymas D, kai naudojamas AMR 12,2 kb/s koderiais, yra 0,33. AMR 4,75 kb/s koderiui tas pats pilnasis spektrinio iškraipymo dydis yra 0,46.

Panagrinėkime situaciją, kaip veikia spektrinių iškraipymų skaičiavimo algoritmas, kai skirtinga matavimo lango pradžia. Tolimesnė analizė teisinga daugeliui algoritmų, kurie naudoja signalo dalijimą kadrais. Matavimo langų schema pateikta 1.5 paveiksle. Laikykime, kad pirmame ir 6-ame languose signalas yra nulinis. Žmogus tokį signalą suvoktų kaip tylą. Aktyvus signalas (tiek etaloninis, tiek iškraipytas) yra tik languose $2-5$. Panagrinėkime du pavaizduotus paveiksle skirtingus matavimo langus. Abu langai apima aktyviąją

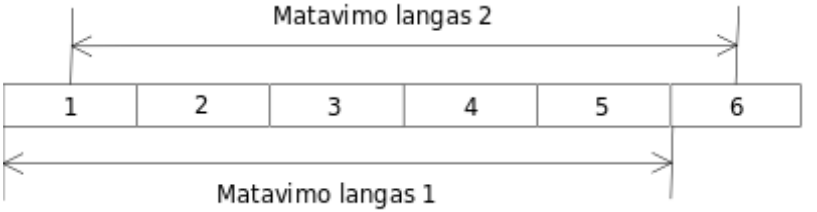
šnekos dalį. Todėl atrodytų, kad iškraipymai ar balso kokybė, matuojami abiejuose languose turėtų būti vienodi. Atlikus eksperimentą pasirodo, kad taip nėra. Pilnas spektrinis iškraipymas, gaunamas matuojant lange 1 yra 0,24, o lange 2 – 0,2. Eksperimente buvo naudota to paties žodžio „yla“ signalo pradinė dalis iš keturių paketų originaliam signalui, o iškraipytas signalas buvo gautas perkoduojant originalųjį ir imituojant paketo praradimą pirmame pakete. Kodavimui dekodavimui buvo naudotas AMR 12,2 kb/s koderis. Matavimo nesutapimų priežastis gana akivaizdi – kai paslenkame matavimo langą, tai pasislenka ir kadrai, kuriuose skaičiuojami bruožų vektoriai, pastarieji šiek tiek pasikeičia – nevienodai originaliame ir iškraipytame signaluose. Dėl to gauname skirtingą galutinį rezultatą.

Panašią problemą turi ir PESQ algoritmas, kuriame taip pat naudojamas signalo dalijimas kadrais. Kitame skyriuje pristatysime metodą (Gursnys 2008, Kajackas *et al*. 2008), kaip galima pakoreguoti PESQ algoritmo taikymą taip, kad jo duodami rezultatai taptų tikslesniais. Pastebėsime, kad aptariama problema aktuali tik tuomet, kai matuojami signalai yra trumpi, o iškraipymai juose impulsinio pobūdžio, pavyzdžiui, paketų praradimai.



1.4 pav. Spektrinio iškraipymo komponentės, kai balsas koduotas skirtingais koderiais

Fig. 1.4. Components of spectral distortion for different codecs



1.5 pav. Matavimo schema kadrų efektams parodyti
Fig. 1.5. Measurement scheme for showing of windowing effects

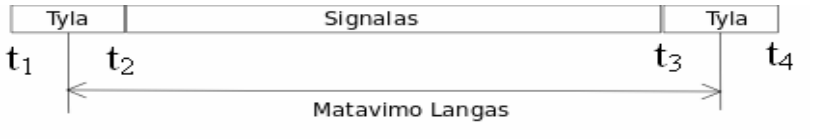
1.8. PESQ ypatumai matuojant trumpų signalų kokybę

PESQ algoritmas pagal savo pobūdį taip pat yra slenkančiais kadrais operuojantis algoritmas. Jau anksčiau parodėme, kad tokie algoritmai gali turėti tam tikrų fundamentalių trūkumų. Čia skaitiškai parodysime, kokie tie trūkumai ir kokiose situacijose jie pasireiškia.

Naudojami eksperimente signalai parodyti 1.6 paveiksle. Analizuojamas signalas yra nuo laiko momento t_2 iki laiko momento t_3. Kaip žinoma PESQ, kaip ir kiti invaziniai algoritmai, naudoja du signalus. Antrasis signalas 1.6 paveikslo schemoje neparodytas.

Laikykime, kad antrasis signalas gautas iš pirmojo perkodavimo būdu ir įvedant vieną klaidą. Operacijos, atliekamos su originaliu ir su perkoduotu signalais, yra visiškai identiškos (dėl to čia sąlyginai rodomas tik vienas signalas). 1.6 paveiksle parodyti technologiniai tylos intervalai laikotarpiuose nuo t_1 iki t_2 ir nuo t_3 iki t_4. Laiko intervalai $t_2 - t_1$ ir $t_4 - t_3$ turi būti lygūs. Matavimo lango ilgis yra $t_3 - t_1$. Balso kokybės matavimo lango pradžia gali būti bet kur nuo laiko momento t_1 iki laiko momento t_2.

Atrodytų, kad nėra jokio skirtumo, kur tarp t_1 ir t_2 prasideda matavimo langas. Išmatuotas kokybės rodiklis turėtų būti gautas toks pat. Tačiau realiai, kai signalas yra trumpas, ir signale įvesta lokalizuota paketo klaida, tai gaunami kintantys matavimo rezultatai. Kitimas priklauso nuo to, kokioje vietoje yra matavimo lango pradžia (Kajackas *et al*. 2008).



1.6 pav. Matavimo lango slinkimo schema analizuojant PESQ algoritmą
Fig. 1.6. Scheme of measurement window displacement for PESQ analysis

Toliau keliais eksperimentais parodysime, kaip matavimo rezultatai priklauso nuo matavimo lango pradžios, kai PESQ algoritmas naudojamas kaip matavimo priemonė.

Pirmam eksperimentui naudosime 1.6 paveiksle parodytą schemą. Patikslinsime kai kurias detales. Tylos intarpai pradinių signalų pradžioje ir gale yra 20 ms ilgio. Patys signalai, naudojami kokybės matavime, yra jau anksčiau naudotas žodis „yla“ ir žodis „dama“. Signalo ilgis yra 400 ms. Iškraipytoji signalo versija yra gaunama perkoduojant originalų signalą AMR-12.2 koderiu ir įvedant paketo klaidą 15-ojo paketo vietoje – ties 300 ms žyma.

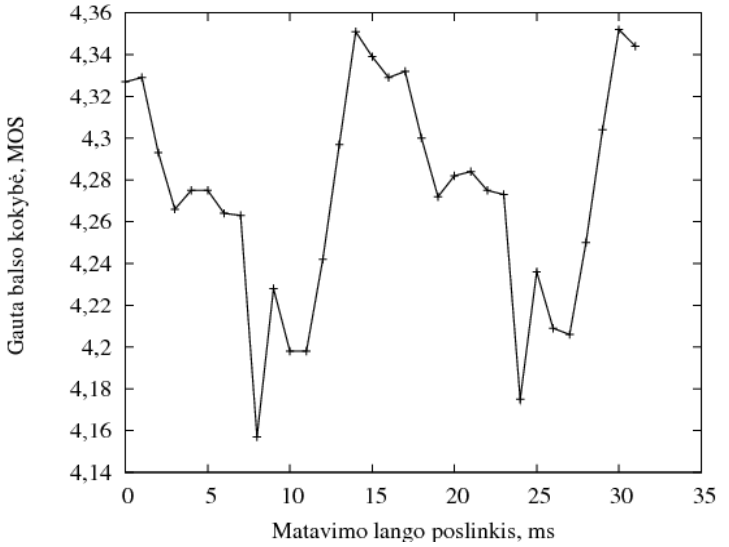
Paaiškinsime, kas atsitinka, kai yra prarandamas paketas AMR ir PCM koderių atveju. PCM koderių atveju naudojome pačią elementariausią galimą strategiją – prarasto paketo vietoje buvo įterpiamas nulinis signalas. AMR veikimas žymiai sudėtingesnis (3GPP TS 26.091 2007). Jei prarandamas pirmas paketas, tai dekoderis pakartoja prieš tai buvusio paketo parametrus ir dekodavimą vykdo, naudodamas šiuos parametrus. Jei iš eilės prarandamas nebe pirmas paketas, tai taip pat naudojami paskutinio teisingai priimto paketo parametrai, tačiau su kiekvienu paeiliui prarastu paketu dar mažinama signalo energija. Praktikoje, kai prarandami 7 paketai paeiliui, tai išėjimo signalas yra tyla.

Matavimo langas eksperimento metu buvo slenkamas 20 ms intervale (nuo t_1 iki t_2) vienos milisekundės žingsniu. Rezultatai, gaunami taikant PESQ algoritmą ir naudojant žodį „yla“, yra pateikti 1.7 paveiksle. Matome, kad matavimo rezultatai yra gan smarkiai išsibarstę ir aiškiai priklauso nuo matavimo lango padėties. Taip pat galime pastebėti pseudoperiodišką rezultatų kitimą su 16 ms periodu – tai patvirtina kadrų įtakos matavimams hipotezę. To paties eksperimento rezultatai, kai signalas yra žodis „dama“, pateikti 1.8 paveiksle. Matome, kad kreivės pobūdis skiriasi, tačiau dėsningumai išlieka tie patys abiems signalams.

Taigi vien toks kokybinis eksperimentas, naudojant vieną signalą, dalinai patvirtina, kad PESQ algoritmas veikia nestabiliai, kai naudojami trumpi signalai su lokalizuotais skirtumais tarp signalų. Jei matuojami signalai ilgi, o trukdžiai iškraipytajame signale stacionarūs ar pseudostacionarūs, tai PESQ algoritmas yra visiškai tinkama balso kokybės matavimo priemonė.

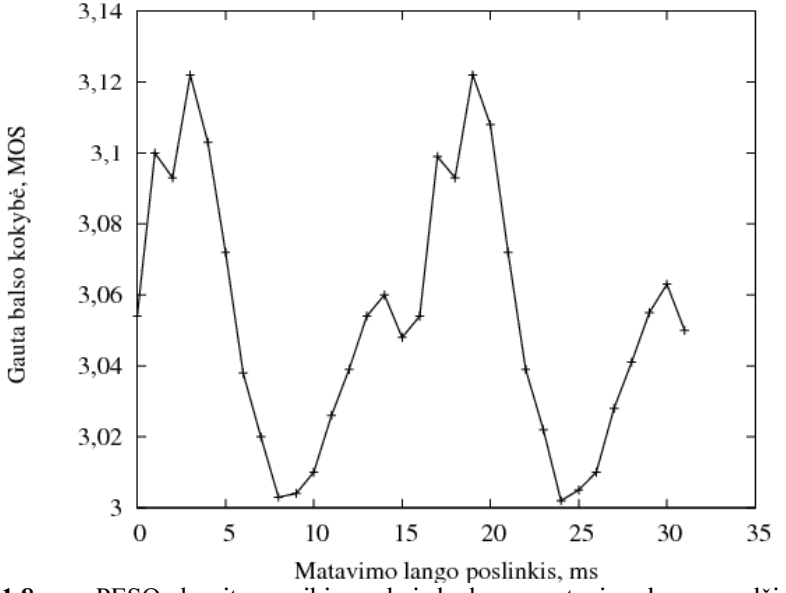
Pateiksime dar vieną statistinį eksperimentą, kuris patvirtina faktą, kad PESQ algoritmas jautrus matavimo lango poslinkiui.

Ir šiame eksperimente naudojamų signalų struktūra bus tokia, kaip 1.6 paveiksle. Šį kartą bus naudojama 100 skirtingų signalų, kurių kiekvieno ilgis 400 ms. Kaip ir anksčiau, iškraipytas signalas yra gaunamas imituojant paketo praradimą penkioliktame pakete. Balso kokybės matavimui bus naudojami du langai – pirmasis nuo t_1 iki t_3, o antrasis – nuo t_2 iki t_4.



1.7 pav. PESQ algoritmo veikimas, kai slenkama matavimo langos pradžia žodyje „yla“

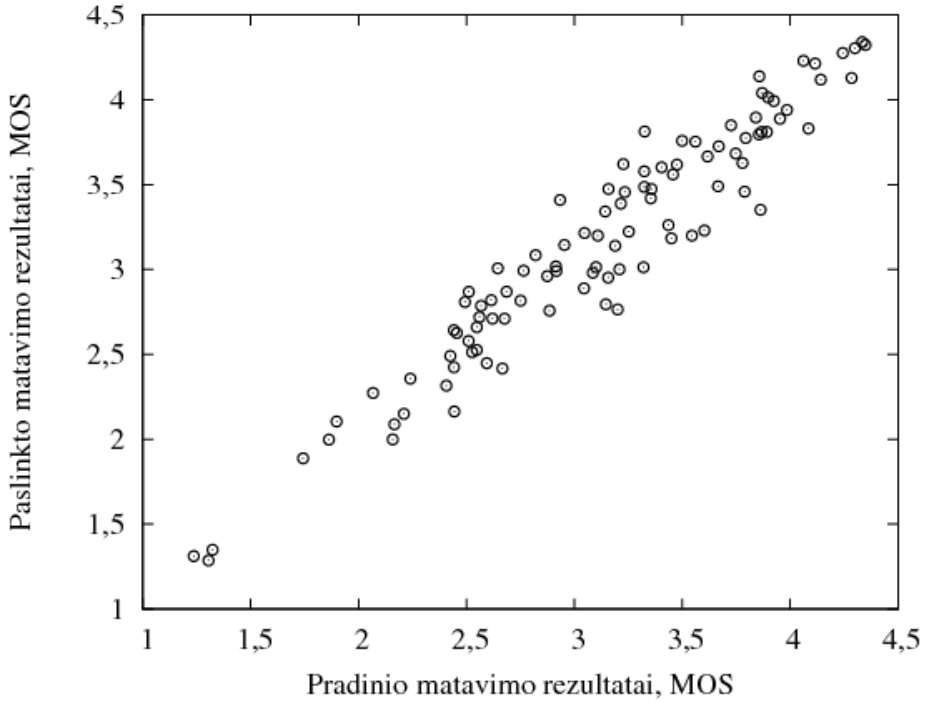
Fig. 1.7. PESQ measurements with different windows with word „yla“



1.8 pav. PESQ algoritmo veikimas, kai slenkama matavimo langos pradžia žodyje „dama“

Fig. 1.8. PESQ measurements with different windows with word „dama“

Jei algoritmas būtų idealus ta prasme, kad jo duodami rezultatai sutaptų su rezultatais, kurių tikisi žmogus, tai aišku, kad matavimo rezultatai naudojant pirmąjį langą, sutaptų su rezultatais, gautais naudojant antrąjį langą. Kitaip sakant, naudojant skirtingus langus pamatuotos kokybės Q_1 ir Q_2, būtų lygios. Jei tokius matavimo rezultatus atidėtume grafike, tai visi taškai būtų ant tiesės, pasvirusios 45 ° kampu. Visų 100 matavimų rezultatai parodyti 1.9 paveiksle. Matavimų rezultatai pradiniam matavimo langui (tarp t_1 ir t_3) ir paslinktam matavimo langui (tarp t_2 ir t_4) atidėti išsibarstymo grafike (angl. *scatter plot*). Minėjom, kad idealaus algoritmo panaudojimo atveju gautumėm visus taškus ant tiesės, pasvirusios 45 ° kampu, t. y., koreliacijos koeficientas būtų lygus vienetui. Nagrinėjamu atveju, kai naudojamas PESQ algoritmas, gaunamas koreliacijos koeficientas yra 0,96. Tai atrodo gana artima vienetui, tačiau visgi išsibarstymas gana didelis, ypač žinant, kad PESQ ir MOS koreliacijos koeficientas yra vos mažesnis – 0,93.



1.9 pav. Balso kokybės matavimas naudojant PESQ algoritmą ir dvi skirtingas matavimo lango padėtis

Fig. 1.9. Voice quality measurement using PESQ algorithm with two different measurement window positions

Aprašytas eksperimentas nėra tik disertacijos autoriaus noras parodyti, kad PESQ algoritmui reikia kažkokio modifikavimo. Tolesniuose skyriuose, nagrinėjant šnekos paketo vertę, panašaus tipo matavimai bus atliekami su daugeliu signalų, todėl toliau pristatysime signalų konstravimo metodą, kuris leis PESQ algoritmą taikyti ir trumpiems signalams, o matavimo rezultatai daug mažiau priklausys nuo matavimo lango padėties.

Pastebėkime, kad nors PESQ yra geriausias šiuo metu standartizuotas invazinis algoritmas, bet jis kritikuojamas kaip netinkamas kai kurių koderių veikimo įvertinimui (Qualcomm Incorporated 2008).

1.9. Pirmo skyriaus išvados ir disertacijos uždavinių formulavimas

- Yra keletas metodų, kurie sukurti matuoti balso kokybę testavimo tikslais. Pripažįstama, kad tiksliausias invazinis kokybės nustatymo algoritmas yra PESQ.
- MOS kokybės vertinimo metodas puikiai tinka stacionariems iškraipymams – triukšmams, filtravimui – tirti, tačiau ne visiems taikymams MOS yra tinkamas pasirinkimas. Jis netinka, kai paketų praradimų yra daug, kai žodžių bei sakinių prasmė tampa neaiški. Tokiu atveju geriau tinkami suprantamumo testai.
- Sudarytos lietuviškų žodžių suprantamumo tyrimo lentelės gali būti naudojamos taip pat, kaip plačiai paplitusios angliškos.
- PESQ algoritmas skirtas sąlyginai ilgų balso atkarpų kokybei vertinti. Algoritmas nėra invariantiškas matavimo lango padėties atžvilgiu. Jis netinka, kai reikia ištirti pavienių žodžių su lokalizuotais iškraipymais kokybę. Tiesiogiai taikant PESQ gaunamos neapibrėžtys, kurios neleidžia algoritmo taikyti trumpų signalų kokybei matuoti. Todėl reikia rasti metodą, kaip PESQ taikymo ribas išplėsti.
- Kadangi nagrinėjama koduotos kalbos kokybė, tai svarbu žinoti, kokios yra atskirų koduotų balso paketų informacinės vertės. Čia kyla uždavinys, kaip tas vertes nustatyti ir kokias reikšmes jos įgyja bei kokias savybes turi.
- Balso paketų informacinių verčių nustatymas yra skaičiavimams imlus uždavinys, todėl tikslinga ieškoti alternatyvių, mažiau skaičiavimų reikalaujančių, paketo vertės nustatymo metodų.

2

Balso segmento vertė

Žmogaus šneka yra analoginis fizinis signalas $x(t)$, kuris sąlyginai lėtai kinta laike. Šiuolaikinės telekomunikacijų ir balso apdorojimo sistemos skaitmenizuoja signalą ir dalija jį į T_f ilgio intervalus. Pastarieji yra koduojami. *i*-asis koduotas segmentas yra vaizduojamas kaip kodas V_i ir yra perduodamas kaip *i*-asis duomenų paketas C_i.

Šiuo metu yra sukurta daug balso koderių (Chu, 2003; ITU-T Rec. G.711. 1993; ITU-T Rec. G.728. 1992; 3GPP TS 26.190). Skirtingi koderiai naudoja skirtingus apdorojimo algoritmus, koduojamus parametrus pasirenka naudodami skirtingus šnekos signalo bruožus. Pastebėkime, kad koduoto balso intervalo ilgis T_f gali būti skirtingas skirtingiems koderiams. Taip pat aišku skiriasi ir kodai V_i.

Apskritai, koderiai kuriami taip, kad dekoduotas signalas būtų kuo mažiau iškraipomas – pagal kokią nors metriką, būtų kuo panašesnis į originalų signalą. Tačiau, kai kodai C_i perduodami ryšio kanalu, atsiranda tikimybė, (kuri dažniausiai gana reali, visai neartima nulinei) kad duomenų paketas C_i, bus prarastas. Šis praradimas gali turėti įvairią fizinę prigimtį. Tai gali būti pernelyg didelis vėlinimas IP tinkle, o gali būti ir fizinis bitų sugadinimas, kai jie priėmimo pusėje neteisingai dekoduojami. Nepriklausomai nuo C_i sugadinimo kilmės, duomenų paketą vadiname prarastu. Tokie prarasti paketai aišku sukelia šnekos signalo iškraipymus, o kartu ir girdimus šnekos kokybės pablogėjimus.

Šiame darbe svarbu tai, kad ne visi perduodami šnekos paketai yra vienodai vertingi šnekos suvokimui ir kokybei. Dažnai šnekos segmentai suvokimo

prasme yra visiškai nevertingi. Pavyzdžiui, tyla, oro įkvėpimas prieš pradedant sakinį ir pan. dažniausiai neturi jokios įtakos šnekos suvokimui. Taip pat yra aišku, kad kai kurie segmentai turi tam tikrą juntamą įtaką girdimumui, bet nėra kritiniai. Galima numanyti, kad yra ir labai svarbių balso segmentų, kurių praradimas gerokai apsunkina šnekos prasmės suvokimą ar patį suvokimo procesą daro nekomfortišką, kai klausytojas yra žmogus. Nepaisant šių faktų, informacija apie balso segmentų skirtingą vertę praktiškai nėra naudojama šiuolaikinėse telekomunikacijose. Bene vienintelis procesas, kur naudojama segmento vertės samprata, yra balso aktyvumo detektavimas. Kaip žinia, moderniose ryšio sistemose tylos atveju perduodami specialūs duomenų paketai. Tylos aktyvumo detektavimas naudojamas GSM, 3G sistemose, internetinės telefonijos tinkluose. Tačiau, jei nustatoma, kad paketas yra „aktyvus“, tai jis perduodamas vienodai apsaugant nuo klaidų, nepriklausomai nuo to, kiek svarbi jame esanti kalbinė informacija. To pasekmė yra tai, kad paketų praradimai įvyksta nepriklausomai nuo to, kiek reikšmingas yra paketas.

Yra darbų, kuriuose siūloma apsaugoti svarbesniuosius balso paketus specialiais klaidas taisančiais kodais (Paraestholm *et al*. 2004) arba sužymint svarbius paketus kaip prioritetinius (De Martin 2001), (Qiao *et al*. 2004). Nepaisant to, yra fundamentalių problemų pačiame paketų vertės nustatyme.

Šiame skyriuje pasiūlysime metodą, kaip apibrėžti paketo informacinę vertę. Siūlomas metodas operuoja koduotos šnekos paketais. Kitaip tariant, paketo vertė priklauso ne tik nuo signalo, esančio tame pakete, bet ir nuo to, kokiu koderiu koduotą signalą analizuojame.

Po to, kai sukursime algoritmą, nustatyti koduoto paketo vertei, tirsime, kaip šnekoje pasiskirstę paketų vertės. Analizuosime sąsajas tarp kelių atskirų paketų verčių ir bendros tų paketų vertės. Parodysime, kaip paketo vertė skiriasi AMR ir PCM (G.711) koderiams.

2.1. Vertės analizės metrikos ir fundamentalios sąvokos

Vertės samprata yra plati kategorija. Šią sampratą yra tyrę filosofai, ekonomistai, psichologai ir kitų sričių mokslininkai. Vertės teorijoje turime sampratą apie vertę kaip tokią ir apie instrumentinę vertę. Pirmoji samprata daugiau filosofinė, o antroji gan konkreti ir reiškia objektų vertę kaip priemonę pasiekti kažkam kitam.

Yra daug darbų, kurie analizuoja segmentinę balso ar šnekos struktūrą. Taip pat dažnai analizuojami atskiri žodžiai ar frazės ir fonologiniai vienetai – fonemos, balsiai, priebalsiai ir pan. Pavyzdžiui, yra žinoma, kad priebalsės yra

labai svarbios signalų suprantamumui, nors jos sudaro tik nedidelė signalų energijos dalį (Kazlauskas 1999).

Šiuolaikiniai balso koderiai balso signalą dalija į atkarpas nepriklausomai nuo fonologinės balso struktūros. Taigi, nėra galimybių panaudoti sukauptų fonologinių žinių, kai siekiame analizuoti atskirų koduotų paketų vertę.

Pagrindinė taikomoji problema yra vertės apibrėžimas ir kiekybinis įvertinimas atskiram balso segmentui.

Balso segmento vertę galima apibrėžti naudojant keletą metodų. Paprasčiausias ir akivaizdžiausias kriterijus būtų signalo energija (Kazlauskas 1999). Tačiau net ir balso aktyvumo detektoriai šiais laikais naudoja modernesnius algoritmus (Prasad *et al*. 2006). Tai tik rodo, kad segmento energija nėra pakankamas rodiklis nustatant segmentinį vertingumą. Kitas galimas požiūris yra klasifikuoti kadrą kaip skardų arba duslų (Kulesza *et al*. 2006). Plati sąvokų grupė bando aprašyti balso segmentą naudojant psichoakustinius modelius. Šie modeliai bando įvertinti žmogaus girdimumo savybes ir jas suformuluoti algoritmiškai. Įvairios metrikos iš šnekos atpažinimo taip pat gali būti naudojamos segmento vertingumui charakterizuoti.

Pabandykime apibendrinti ir formalizuoti balso segmento vertės nustatymo metodiką. Techniniuose vertinimuose svarbu, kad balso segmento vertingumas būtų nustatomas naudojant objektyvų algoritmą. Tarkim, kad turim kriterijų Q šnekos suprantamumo arba kokybės nustatymui. Šis kriterijus apibrėžiamas kaip funkcionalas operuojantis su balso signalu:

$$Q = \Theta[x_1(t), x_2(t)], \tag{2.1}$$

čia $x_1(t)$ yra pradinis balso signalas; $x_2(t)$ yra iškraipytas balso signalas. Iškraipymas dažniausiai įvyksta, kai pradinis signalas pereina per telekomunikacinę sistemą.

Balso segmento vertę apibrėžkime kaip iškraipymo dydį, gaunamą, jei nagrinėjamas balso segmentas (ar jį atitinkantis paketas) yra prarandamas. Taigi bendru atveju paketo (balso segmento) vertę apibrėžiame taip:

$$V_i = \Delta Q = Q_0\text{-}Q_i \tag{2.2}$$

čia Q_0 yra pradinė tam tikros balso atkarpos kokybė pagal naudojamą kriterijų, o Q_i – yra kokybės vertė, gaunama, kai prarastas i-asis paketas. Šis trumpas aptarimas nedetalizuoja, koks yra funkcionalas Θ, tačiau mes nagrinėsime atvejį, kai šis funkcionalas yra PESQ_{mod} algoritmas. Tokiu būdu ankstesnę formulę galima kiek labiau patikslinti ir užrašyti taip:

$$V_i = \Delta\text{Q} = \Theta[x_0(t), x(t)] - \Theta[x_0(t), x_i(t)], \tag{2.3}$$

čia $x_0(t)$ yra pradinis signalas, kurio i-ojo segmento vertingumą norime įvertinti; $x(t)$ yra signalas, gautas perkodavus pradinį nagrinėjamą signalą; $x_i(t)$ yra

signalas, gautas iš pradinio signalo jį perkodavus ir įvedus paketo praradimą *i*-ajame (nagrinėjamame balso segmente). Atkreipsime dėmesį, kad čia atsiranda kodavimo – dekodavimo (perkodavimo) operacija. Tokiu būdu segmento vertė priklauso ir nuo to, kokio koderio atžvilgiu tas segmentas nagrinėjamas. Mūsų tolimesniuose tyrimuose bus naudojami du koderiai – AMR koderių šeima ir taip vadinamas PCM koderis. AMR koderiai ir jų elgesys, kai prarandamas paketas, yra aprašyti čia (3GPP TS 26.091 2007), o PCM koderiu vadinsime koderią, kuris nekeičia signalo kodavimo – dekodavimo metu, o praradimo atveju vietoj 20 ms nagrinėjamo segmento įstato nulius.

Kitas galimas būdas paketo vertei įvertinti yra naudoti tokią metriką: $V_i = \max(\Theta) - \Theta[x(t), x_i(t)]$, čia $\max(\Theta)$ yra maksimali operatoriaus Θ reikšmė, įgaunama visų galimų signalų aibėje. PESQ atveju ši reikšmė yra 4,549. Tokia vertės metrika taip pat didesnę vertę priskiria tiems paketams, kurių praradimas sukelia didesnius nuostolius (signalo iškraipymą).

2.2. PESQ algoritmo taikymas trumpiems signalams analizuoti

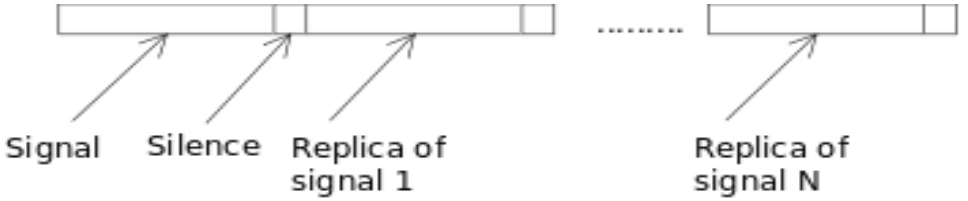
Kadangi PESQ algoritmas bus naudojamas paketų vertei nustatyti, o ankstesniame skyriuje matėme, kad šis algoritmas turi tam tikrų problemų, kai operuoja su trumpais signalais, tai čia parodysime, kaip galima organizuoti matavimo procedūrą naudojant standartinį PESQ algoritmą, kad tų problemų išvengti. PESQ standarto apibrėžtomis sąlygomis veikiantis PESQ algoritmas yra pakankamai tikslus (Rix 2003) ir problemų kyla tik tada, kai norime praplėsti algoritmo taikymo ribas.

Naudojama eksperimento schema parodyta 1.6 paveiksle. (ta pati, kuri buvo naudojama PESQ trūkumams parodyti). Analizuojamas signalas yra nuo laiko momento t_2 iki laiko momento t_3. Tai gali būti bet koks signalas, kuriuo mes domimės. Pastebėkime, kad PESQ, kaip ir kiti invaziniai algoritmai, dirba naudodamas du signalus, bet antrasis signalas šioje schemoje neparodytas. Tačiau laikykime, kad antrasis signalas gautas iš pirmojo perkodavimo būdu (naudojant AMR koderį) ir įvedant vieną klaidą. Operacijos, atliekamos su originaliu ir su perkoduotu signalu, yra visiškai identiškos – dar ir dėl to sąlyginai rodomas tik vienas signalas. Paveiksle taip pat matome, kad laikotarpiuose nuo t_1 iki t_2 ir nuo t_3 iki t_4 yra pridėti technologiniai tylos intervalai. Skirtumai $t_2 - t_1$ ir $t_4 - t_3$ turi būti lygūs. Matavimo lango ilgis yra $t_3 - t_1$. Balso kokybės matavimo lango pradžia gali būti bet kur nuo laiko momento t_1 iki laiko momento t_2.

Intuityviai žmogiškai mąstant, nėra jokio skirtumo, kur tarp t_1 ir t_2 prasideda matavimo langas. Akivaizdu, kad išmatuota kokybė turėtų būti vienoda. Tačiau naudojant originalų PESQ algoritmą taip nėra. Šią problemą toliau ir spręsime specialaus signalo konstravimo pagalba.

Norėdami išvengti matavimo rezultato priklausomybės nuo matavimo lango pozicijos, turime atlikti tam tikrą skirtingų matavimo rezultatų vidurkinimą, kai matavimo langas slenkamas vieno segmento ribose. PESQ atveju tas matavimų vidurkinimas turi būti atliekamas 16 ms ribose, nes toks yra segmentų persidengimo laikas. Pats matavimų vidurkinimas atliekamas naudojant specialiai sukonstruotą signalą. Esminis dalykas šiame konstravime – nepakeisti pradinio signalo taip, kad tai atrodytų visiškai kitoks matavimas, nei pradinis, o taip pat, sukonstruotame signale tos pačios matuojamo signalo dalys turi atsidurti vis kitoje PESQ matavimo lango fonemų dalyje. Pati signalo konstravimo idėja ta, kad turimą signalą reikia replikuoti tam tikrą skaičių kartų ir replikas sudėti vieną prie kitos. Tam, kad taip pat gautumėme matavimo rezultatų vidurkinimą, į tarpus tarp replikų reikia įterpti tylą – nulinį signalą. Klausimai, į kuriuos reikia atsakyti yra šie: kokio ilgio tylos intarpus naudoti tarp signalo replikų ir kiek signalo replikų reikia tam, kad gaunamas matavimo rezultatas būtų stabilus ir nepriklausytų nuo matavimo lango padėties.

Kokybiškai signalo pratęsimo procedūra pavaizduota 2.1 paveiksle.



2.1 pav. Prailginto signalo konstravimas
Fig. 2.1. Construction of extended signal

Patį signalo konstravimą galima užrašyti tokia formule:

$$v_{\Sigma}(t) = \sum_{i=1}^{N-1} x_{\text{prad}}(t - iT_{\text{prad}}), \tag{2.4}$$

čia x_{prad} yra pradinis signalas su gale pridėtu dT ilgio tylos (nulių) intervalu, dT parenkamas lygiu 16/N ms, kur N, kaip ir 1.6 formulėje, yra pradinių signalų replikavimų skaičius. Tokiu būdu T_{prad} yra lygus originalaus signalo ilgio ir dT sumai. Dar vienas reikalavimas yra, kad originalaus signalo ilgis būtų kartotinis 16 ms. Jei taip nėra, tai šį signalą prieš replikavimo procedūrą reikia prailginti iki tokio ilgio. Mūsų bandymuose (Kajackas *et al*. 2008) buvo nustatyta, kad tinkami

replikavimo skaičiai yra 8, kuris naudotas (Guršnys 2008), ir 16. Šiame darbe bus naudojama 16 signalo replikavimų visuose PESQ algoritmo taikymuose.

Susumuosime PESQ algoritmo taikymą trumpiems signalams tokiu algoritmu. Pradiniai duomenys, tai 2 signalai (originalus ir iš šio gautas iškraipytasis), kuriems norime pamatuoti PESQ metriką. Šie signalai, aišku, yra vienodo ilgio. Toliau nuosekliai:

- Jei signalų ilgis nėra kartotinis 16 ms, tai juos prailginame iki artimiausio kartotinio skaičiaus. Prailginimas atliekamas naudojant atitinkamo ilgio nulinį vektorių.
- Prie gautųjų signalų pridedame nulinį vektorių, kurio ilgis 16/N milisekundžių, čia N – replikavimo kartų skaičius. Konkrečiai mes naudosime vektorių, kurio ilgis 1 ms, nes naudojame 16 replikavimų.
- Gautą vektorių pridedame vienas prie kito (replikuojame) N kartų.
- Gautam ilgam signalui (jo ilgis NT_{prad}) taikome įprastą PESQ algoritmą.

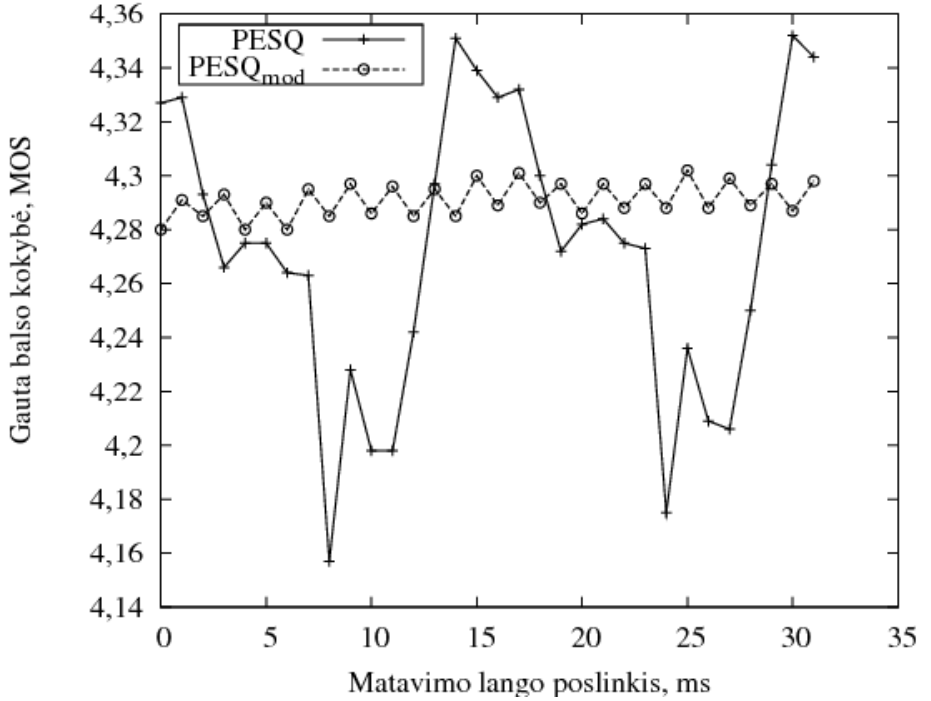
Gautas įvertis ir yra tikslesnis, suvidurkintas kokybės įvertis ir gaunamas rezultatas daug mažiau priklauso nuo matavimo lango pozicijos.

Toliau, kai norėsime pabrėžti, kad signalų kokybei matuoti naudojama aukščiau aprašyta procedūra, tai visą procedūrą sąlyginai žymėsime PESQ_{mod}.

Toliau keliais eksperimentais parodysime, kad PESQ algoritmas, naudojamas su šitaip modifikuotais signalais, yra tikslesnis, negu originalus PESQ metodas.

Pirmam eksperimentui naudosime 1.6 paveiksle parodytą schemą. Patikslinsime kai kurias detales. Tylos intarpai pradinių signalų pradžioje ir gale yra po 20 ms ilgio. Patys signalai, naudojami kokybės matavime, yra jau anksčiau naudotas žodis „yla“. Signalo ilgis yra 400 ms. Iškraipytoji signalo versija yra gaunama perkoduojant originalų signalą AMR-12.2 koderiu ir įvedant paketo klaidą 15-ojo paketo pozicijoje – ties 300 ms žyma.

Matavimo langas eksperimento metu buvo slenkamas dvidešimties milisekundžių intervale (nuo t_1 iki t_2) naudojant vienos milisekundės žingsnį. Rezultatai, gaunami taikant nepakeistą PESQ algoritmą ir taikant mūsų pasiūlytą algoritmą, yra pateikti 2.2 paveiksle. Paveiksle rutuliukais parodyta originalaus PESQ algoritmo gaunami rezultatai. Matom, kad šie rezultatai yra ganėtinai smarkiai išsibarstę ir aiškiai priklauso nuo matavimo lango padėties. Taip pat pastebimas pseudoperiodiškas rezultatų kitimas su 16 ms periodu. PESQ algoritmas, taikomas išplėstiems signalams, duoda žymiai mažiau išsibarsčiusius rezultatus – pastarieji iš esmės yra originalių PESQ matavimų vidurkis. Tačiau išsibarstymas lieka ir modifikuoto algoritmo atveju, tik jis yra žymiai mažesnis.

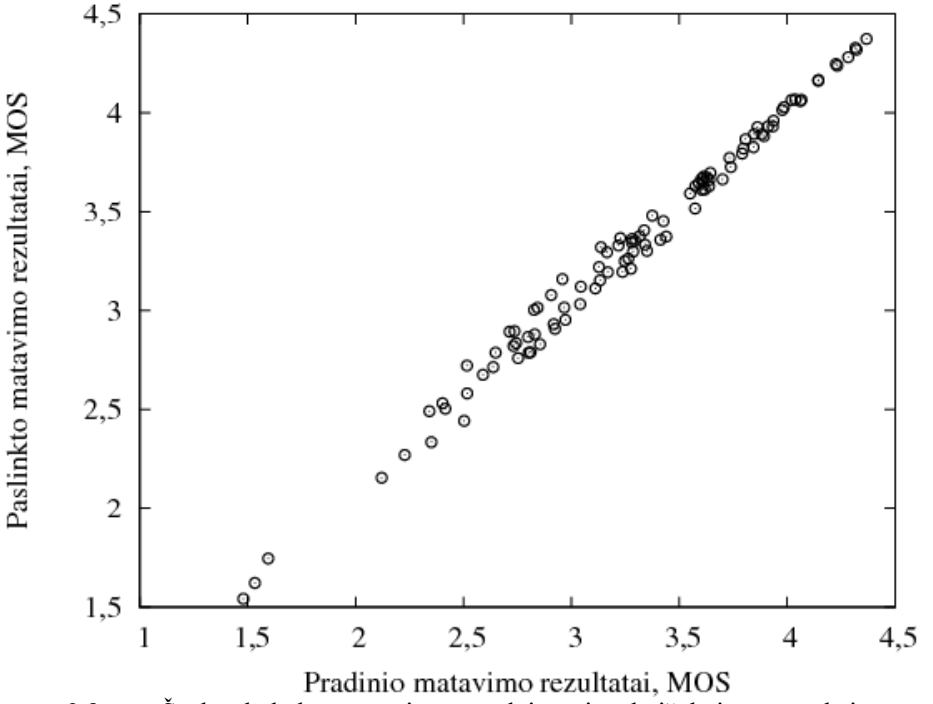


2.2 pav. PESQ ir modifikuoto PESQ algoritmų palyginimas
Fig. 2.2. Comparison of PESQ and modified PESQ algorithms

Taigi vien toks kokybinis eksperimentas, naudojant vieną signalą, dalinai patvirtina, kad modifikuotas PESQ algoritmas veikia stabiliau, negu originalusis. Tiesa, nauda iš algoritmo modifikavimo yra tik tuo atveju, kai naudojami trumpi signalai su lokalizuotais skirtumais tarp signalų. Jei matuojami signalai ilgi, o trukdžiai iškraipytajame signale stacionarūs ar pseudostacionarūs, tai originalusis PESQ algoritmas yra visiškai tinkama balso kokybės matavimo priemonė.

Pateiksime dar vieną statistinį eksperimentą, kuris patvirtina faktą, kad $PESQ_{mod}$ yra naudingas signalų mikrostruktūros tyrime.

Eksperimentas yra analogiškas pirmame skyriuje atliktam, kai buvo naudojami du matavimo langai ir gaunami du kokybės įverčiai kiekvienam signalui. Priminsime, kad gauta koreliacija originaliam PESQ algoritmui tarp šių kokybės įverčių buvo 0,96. Atlikus visiškai analogišką matavimą ir naudojant visiškai tokius pačius signalus, tik naudojant aprašytą signalų išplėtimo metodą ir PESQ algoritmą, gauname išsibarstymo grafiką, kuris parodytas 2.3 paveiksle. Matome, kad gaunami duomenys daug labiau koreliuoti. Koreliacijos koeficientas šiuo atveju yra 0,995. Jei analizuotumėme, kiek koreliacijos



2.3 pav. Šnekos kokybės matavimas naudojant signalų išplėtimo metodą ir PESQ algoritmą

Fig. 2.3. Speech quality measurement using signal extension method and PESQ

koeficientas skiriasi nuo vieneto, tai šiuo požiūriu modifikuotas PESQ algoritmas yra vidutiniškai 8 kartus tikslesnis, negu originalusis.

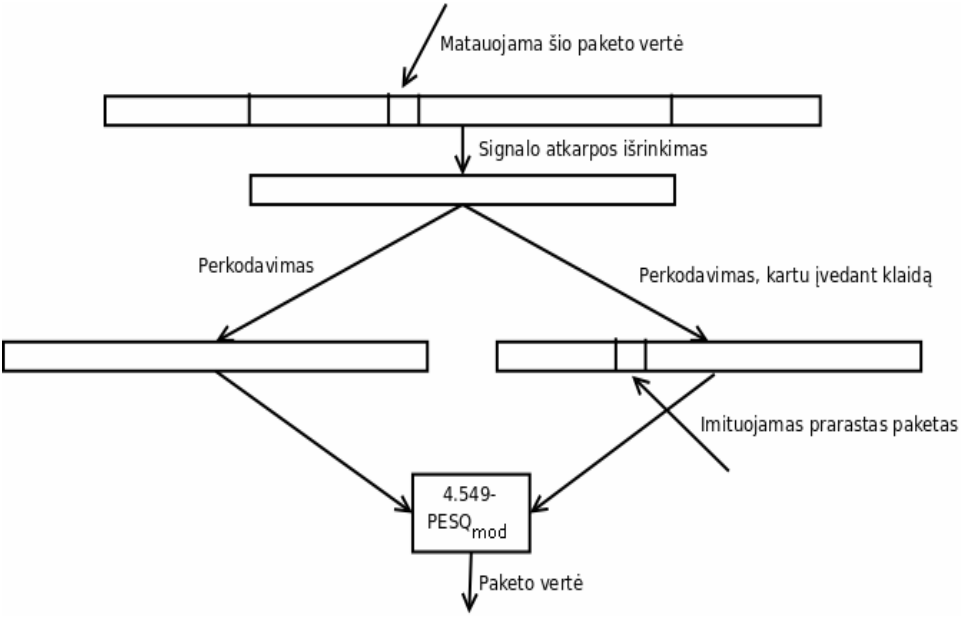
2.3. Paketo vertės nustatymas taikant PESQ algoritmą

Tarkime, kad norime pamatuoti tam tikro šnekos paketo vertę. Aišku, tas paketas nėra atskira signalo atkarpa – jis yra tam tikrame šnekos signale. Pastarasis gali būti bet kokio ilgio.

2.4 paveikslas parodo atskiro paketo vertės matavimo procesą schematiškai.

Paveikslo viršuje sąlyginai parodytas signalas su pažymėtu paketu, kurio vertę matuosime. Pirmu žingsniu išrenkame 0,5 s ilgio signalo atkarpą (toks signalo ilgis atitinka apytiksliai vieną ištartą žodį), kuri turi savyje nagrinėjamą paketą. Signalo atkarpa išrenkama taip, kad nagrinėjamas paketas prasidėtų ties 240-aja šios atkarpos milisekunde. Kitu žingsniu atliekame du perkodavimus.

Pirmas perkodavimas naudodamas tam tikrą koderį, koduoja ir dekoduoja signalą. Antras perkodavimas koduoja signalą, imituoja paketo klaidą ties 240-ąja ms (12 paketu) ir dekoduoja signalą. Tokiu būdu gaunami signalai $x(t)$ ir $x_i(t)$ iš 2.1 poskyrio. Paskutiniu žingsniu šie signalai paduodami į $PESQ_{mod}$ algoritmą. Gautas rezultatas atimamas iš 4,549 ir šis skirtumas yra galutinė reikšmė, kurią norėjome apskaičiuoti.



2.4 pav. Paketo vertės matavimo schema
Fig. 2.4. Diagram of frame value measurement

Aptartame vertės metode yra vienas parametras, kuris paveiksle neparodytas. Tai yra koderis. Gaunama paketo vertė priklauso nuo to, koks koderis naudojamas, nes skirtingi koderiai duos skirtingą rezultatą po paketo praradimo.

2.4. Paketų vertingumo tyrimų schema

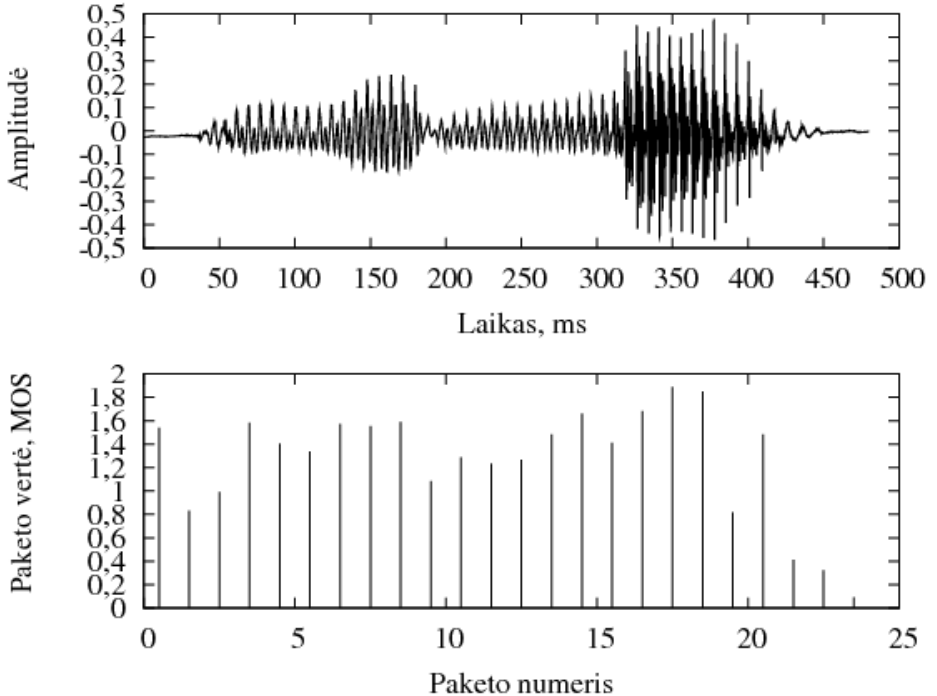
Pateiksime statistinius ir vizualinius duomenis, parodančius, kiek vertingi yra šnekos paketai. Nustatinėjant statistinę vertę naudojama tyrimų schema panaši į 2.4 paveiksle pavaizduotą. Aišku, norint gauti statistinių duomenų, vertės nustatymo bandymai bus kartojami daug kartų (tikslus skaičius bus nurodytas konkrečiuose eksperimentuose).

2.4.1. Kai kurie paketų vertingumo dėsningumai

Čia vizualiai parodysime, kaip atrodo balso signalas, kur yra paketų ribos ir kokios yra tų paketų vertės. Tam paimsime du atskirus balso signalus. Pirmasis atitinka jau nagrinėtą žodį „yla“, o antrasis – žodį „namas“. Taip pat naudosime tris skirtingus balso koderius – PCM, AMR-12.2 ir AMR-4.75 koderius. Skaičiai, einantys po AMR koderio žymens nurodo, kokį duomenų srautą koderis sukuria. Šie AMR koderiai buvo pasirinkti, nes jie yra ribiniai – vieno sparta didžiausia, o kito mažiausia.

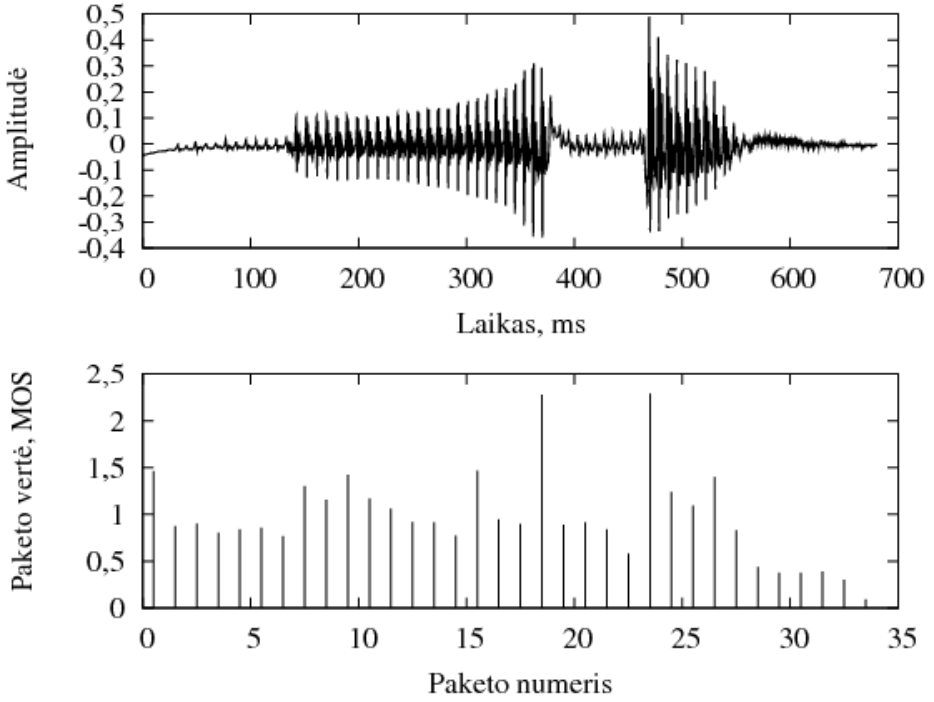
Paketų vertės, kai naudojamas PCM koderis žodžiui „yla“ parodyta 2.5 paveiksle. Tam pačiam PCM koderiui gaunamos paketų vertės žodžiui „namas“ parodytos 2.6 paveiksle.

Priminsime, kad PCM koderis signalą perduoda nepakeistą, o paketų praradimo atveju įterpia nulius prarastame laiko intervale. Pavyzdžiui, jei signale prarandamas antras paketas (paketus indeksuojant nuo nulio), tai signalo perkoduoto signalo vertės nuo 320-os iki 480-os bus prilygintos nuliui, nes paketo ilgis yra 20 ms, o tai atitinka 160 imčių.



2.5 pav. Paketų vertės žodžiui „yla“ (PCM koderis)
Fig. 2.5. Frame values for word „yla“ (PCM codec)

Iš 2.5 ir 2.6 paveikslų matome, kad naudojant PCM koderį, paketo vertę gan sunku nuspėti iš signalo. Ko gero, pagrindinis pastebimas dalykas tas, kad aukštos energijos paketai yra vertinami labiau, negu žemesnės energijos paketai. Bet ši tendencija irgi labai apytikrė. Kai kurie žemos energijos paketai vertinami kaip vertingi, tačiau tokiu vertinimu galima suabejoti. Tarkim, pradžioje signalų praktiškai yra tyla, o atitinkami paketai įvertinami kaip ganėtinai vertingi.



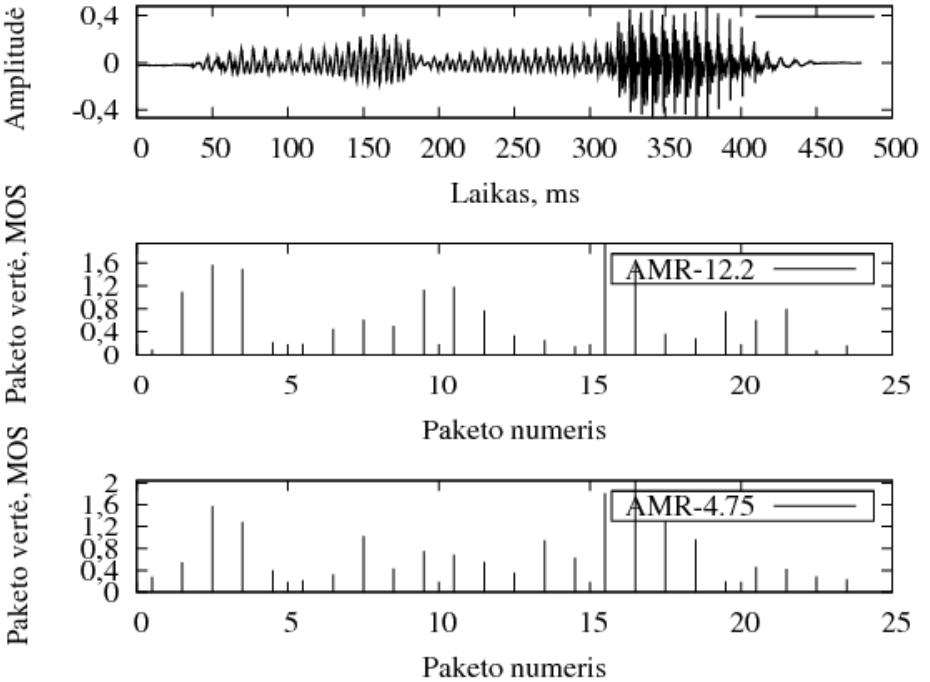
2.6 pav. Paketų vertės žodžiui „namas“ (PCM koderis)
Fig. 2.6. Frame values for word „namas“ (PCM codec)

2.7 ir 2.8 paveiksluose pateikti analogiški grafikai ką tik nagrinėtiems, tik šį kartą naudoti AMR koderiai. Viršutinėje grafikų dalyje parodytos paketų ribos. Antroje grafikų dalyje parodytos paketų vertės, kai vertės matavimui naudojamas AMR-12.2 koderis. Trečioje grafikų dalyje parodytos paketų vertės, kai naudojamas AMR-4.75 koderis.

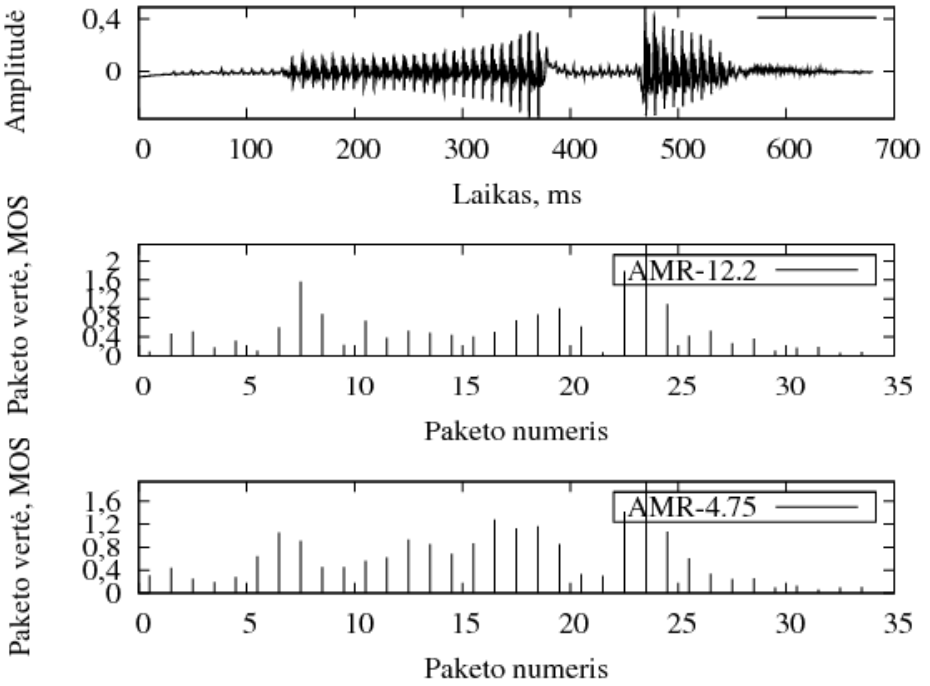
Pastebima gana ryški tendencija, kada paketo vertė yra didelė, kai kodavimui naudojami AMR koderiai – labai vertinami paketai, kurie pradeda šnekos pliūpsnį (Kajackas, Anskaitis 2005; Kajackas *et al*. 2005). Taip yra dėl AMR paketų pakeitimo algoritmo inertiškumo. Kadangi modernūs koderiai yra dinaminės sistemos su tam tikru būviu dekodavimo metu, tai paketo praradimas

tą būvį pakeičia. Paketų pakeitimo algoritmas sukonstruotas taip, kad šis būvio praradimas yra itin svarbus tada, kai signalas nestacionarus, o būtent tokį nestacionarumą ir gauname šnekos pliūpsnių pradžioje.

Apibendrindami pateiktus grafikus pastebime štai ką. PCM kodavimo atveju vertingi paketai turi didelę energiją. AMR kodavimo atveju vertingi yra paketai tuomet, kai koduojamas signalas yra stipriai nestacionarus. Abiem atvejais paketų vertingumas priklauso nuo to, kokia yra koduojamo signalo struktūra (dažninė sudėtis), tačiau kokį nors dėsningumą čia pastebėti yra gana sunku. Penktame skyriuje parodysime keletą tokių dėsningumų skaitiškai. Dar vienas įdomus dalykas tas, kad paketų vertės naudojant AMR-12.2 ir AMR-4.75 koderius skiriasi nežymiai. Taip yra todėl, kad žemesnio duomenų srauto koderiai, kai nėra klaidų, įneša didesnius iškraipymus, tačiau pats klaidų įvedimas šiuos iškraipymus padidina ne ką labiau, nei naudojant aukštesnio duomenų srauto koderį.



2.7 pav. Paketų vertės žodžiui „yla“ (AMR koderiai)
Fig. 2.7. Frame values for word „yla“ (AMR codec)



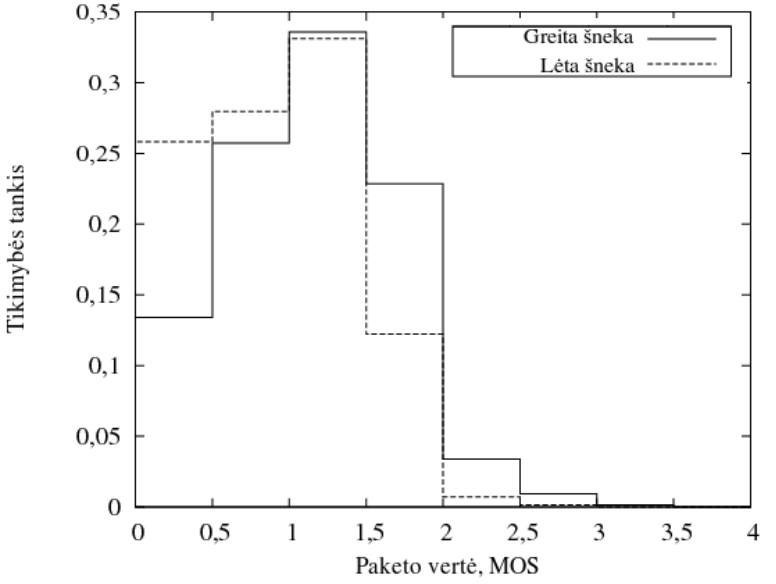
2.8 pav. Paketų vertės žodžiui „namas“ (AMR koderiai)
Fig. 2.8. Frame values for word „namas“ (AMR codec)

2.4.2. Pavienio balso paketo statistinis vertingumas

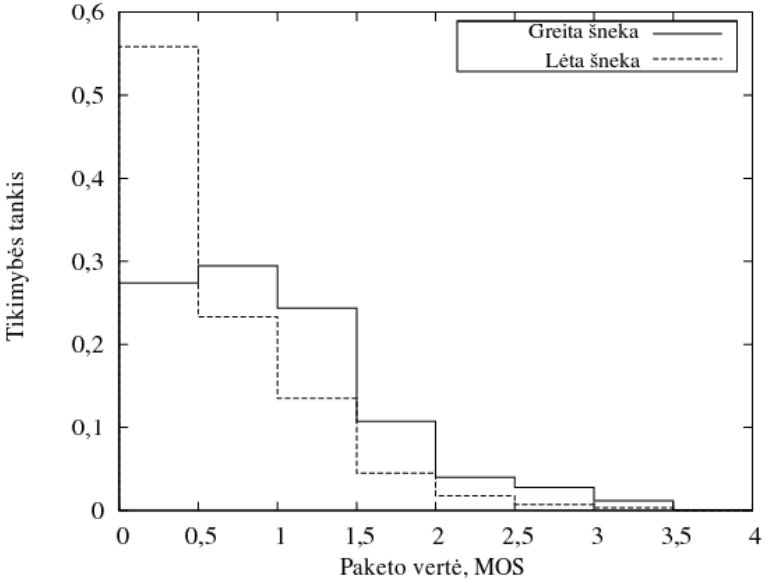
Trumpai apie tai, ką norime pasiekti atlikdami šiuos bandymus. Pagrindinis tikslas yra nustatyti vieno ir keliu paketų vertingumo skirstinį. Tai bus daroma naudojant du šnekos tipus (sąlyginai vadinsime „greita“ ir „lėta“ šneka). Kaip pirmo šnekos tipo atstovas bus naudojamas įrašas iš CNN žinių. Antrojo šnekos tipo atstovas yra interviu sporto tematika. Abu įrašai (ir visi tolimesniuose tyrimuose naudojami) yra diskretizuoti 8 kHz dažniu ir kvantuoti pagal vertes naudojant 16 bitų. Pirmojo įrašo šneka yra daug dinamiškesnė, joje mažiau pauzių, negu antrojo įrašo šnekoje. Taigi parodysime ne tik kaip yra pasiskirstę paketų vertės šnekoje, bet ir kaip pasiskirstymas priklauso nuo šnekos pobūdžio.

Paketų vertės skirstiniams (histogramoms sudaryti) buvo naudojama po 2.000 paketų vertės matavimų kiekvienam šnekos tipui.

2.9 paveiksle pateikta vieno paketo vertės histograma, kai naudojamas PCM koderis. 2.10 paveiksle pateiktas analogiškas paveikslas, kai naudojamas AMR-12.2 koderis.



2.9 pav. Vieno paketo vertės histograma (PCM koderis)
Fig. **2.9**. Histogram of one frame value (PCM codec)



2.10 pav. Vieno paketo vertės histograma (AMR-12.2 koderis)
Fig. **2.10**. Histogram of one frame value (AMR-12.2 codec)

Nesunku pastebėti, kad daugiau vertingų paketų yra „greitoje" šnekoje. Tai gana natūralu, nes šio tipo šnekoje mažiau tylos (analizėje nenaudotas balso aktyvumo detektorius), o šnekos dinamiškumas įtakoja tai, kad daugiau yra nestacionarių intervalų, kurie yra ypač vertingi AMR kodavimo atveju.

Toliau pateiksime keletą statistinių charakteristikų, paskaičiuotų iš šių histogramų. Šias charakteristikas pateiksime tik AMR-12.2 koderiui. Charakteristikos yra tokios – vidurkis, standartinė deviacija, tikimybė, kad paketo vertė bus mažesnė už vienetą, tikimybė, kad paketo vertė bus tarp 1 ir 2, tikimybė, kad paketo vertė bus tarp 2 ir 3 ir tikimybė, kad paketo vertė bus didesnė, nei 3. Visa tai parodyta 2.1 lentelėje. Matome, kad vidutiniškai vienas paketas yra vertingesnis „greitoje“ šnekoje. Esminis skirtumas pastebimas vidutinės vertės zonoje (kai paketo vertė yra tarp vieno ir dviejų). Šioje zonoje net dvigubai daugiau yra paketų iš „greitos“ šnekos. Atkreipkime, dėmesį, kad „lėtoje“ šnekoje sąlyginai daug labai vertingų paketų, kai paketo vertė viršija tris. Tokio paketo praradimas labai smarkiai sugadina signalą, todėl idealioje sistemoje šie paketai turėtų būti apsaugomi nuo praradimo stipriau. Tačiau „lėtoje“ šnekoje daugiau yra mažai vertingų paketų. Tokius paketus idealioje sistemoje būtų galima apsaugoti mažiau, nes jų praradimas nelabai kenkia galutinei šnekos kokybei.

2.1 lentelė. Vieno paketo vertės statistinės charakteristikos AMR-12.2 koderiui
Table 2.1. Statistical properties of one frame value for AMR-12.2 codec

Charakteristika	„greita“ šneka	„lėta“ šneka
Vidurkis	1,12	0,97
Standartinė deviacija	0,71	1,09
P(vertė < 1)	0,47	0,68
P(1<vertė<2)	0,42	0,21
P(2<vertė<3)	0,09	0,04
P(vertė>3)	0,02	0,06

2.4.3. Dviejų paketų statistinis vertingumas

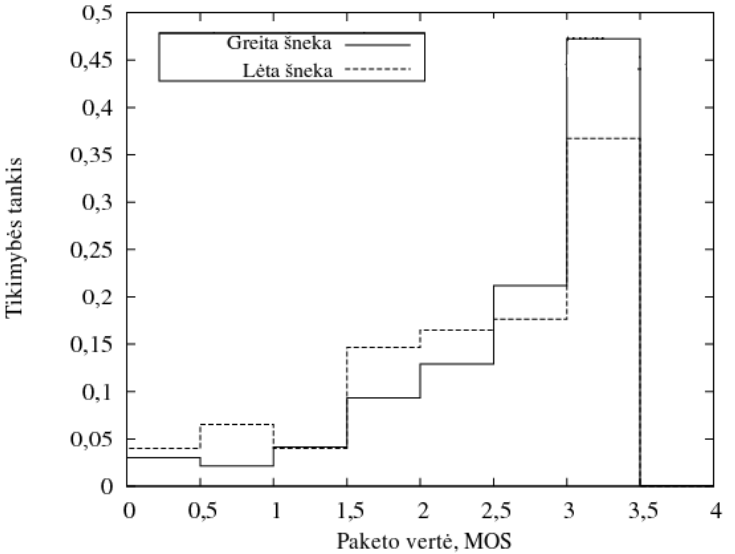
Panašiai kaip prieš tai buvusiame skyrelyje nagrinėjome vieną prarastą paketą, dabar panagrinėkime, kokia yra dviejų paeiliui prarastų paketų statistinė vertė. Eksperimentas visiškai analogiškas, kaip ir anksčiau buvęs, tik dabar imituosime dviejų paeiliui esančių paketų praradimą.

2.11 paveiksle pavaizduota dviejų paketų statistinė vertė, kai naudojamas PCM koderis.

Matome, kad šiuo atveju labai sumažėja tikimybė, kad du paketai bus nevertingi. Tačiau tendencija, kad daugiau nevertingų paketų yra „lėtoje“ šnekoje, išlieka.

2.2 lentelėje pateiksime statistines dviejų paketų vertės charakteristikas AMR-12.2 koderiui. Lentelė savo forma yra analogiška 2.1 lentelei.

Palyginę 2.1 ir 2.2 lenteles matome, kad dviejų iš eilės einančių paketų vertė yra gerokai didesnė, nei vieno paketo. Čia ypač išryškėja, kad „lėtos“ šnekos atveju yra sąlyginai mažai reikšmingų šnekos intervalų, nes tikimybė, kad dviejų paeiliui einančių paketų vertė bus mažesnė už vienetą, yra net 0,42. „Greitai“ šnekai ši tikimybė sudaro vos 0,17.



2.11 pav. Dviejų paketų vertės histograma (PCM koderis)
Fig. 2.11. Histogram of two frames value (PCM codec)

2.2 lentelė. Dviejų paketų vertės statistinės charakteristikos AMR-12.2 koderiui
Table 2.2. Statistical properties of two frames value for AMR-12.2 codec

Charakteristika	**„greita“ šneka**	**„lėta“ šneka**
Vidurkis	1,71	1,40
Standartinė deviacija	0,75	1,08
P(vertė < 1)	0,17	0,42
P(1<vertė<2)	0,50	0,40
P(2<vertė<3)	0,25	0,10
P(vertė>3)	0,07	0,08

Apibendrinant galima pastebėti, kad statistinė vieno ar kelių paketų vertė priklauso nuo šnekos pobūdžio ir kalbėtojo savybių. Akivaizdi tendencija tame, kad kuo daugiau paketų paeiliui prarasta, tuo labiau jie tampa vertingi. Paketo vertė taip pat smarkiai priklauso nuo to, kokį koderį naudojant ji buvo matuota.

2.4.3. Paketo statistinis vertingumas, kai naudojamas balso aktyvumo detektorius

Ankstesniuose poskyriuose atlikti bandymai su balso signalo atkarpa, nekreipiant dėmesio į tai, kad kai kurie paketai gali būti mažai vertingi iš principo – juose gali būti tyla. Panagrinėsime, kokią įtaką turi tokių paketų eliminavimas iš analizės. Tam naudosime šiek tiek modifikuotą bandymo schemą, kurią toliau aprašysime.

Kaip ir anksčiau, analizei imame po 2.000 greitos ir lėtos šnekos atkarpų. Kitu žingsniu matuojame kiekvieno paketo vertę pagal aukščiau aprašytą algoritmą. Jei balso aktyvumo detektorium nustatom, kad pakete yra tyla, tai tokį matavimo rezultatą tiesiog atmetame. Iš gautų matavimo duomenų brėžiame histogramas atskirai greitai ir lėtai šnekai. Tai reiškia, kad matavimus pakartojam atskirai abiems šnekos įrašams.

Trumpai apie naudojamą balso aktyvumo detektorių. Tai energetinis balso aktyvumo detektorius, kurio veikimą galima aprašyti taip:

- Paskaičiuokime kiekvieno (n-ojo) AMR paketo energiją naudodami tokią išraišką:

$$E(n) = \sum_{i=160\cdot n}^{160\cdot n+159} x^2(i), \tag{2.3}$$

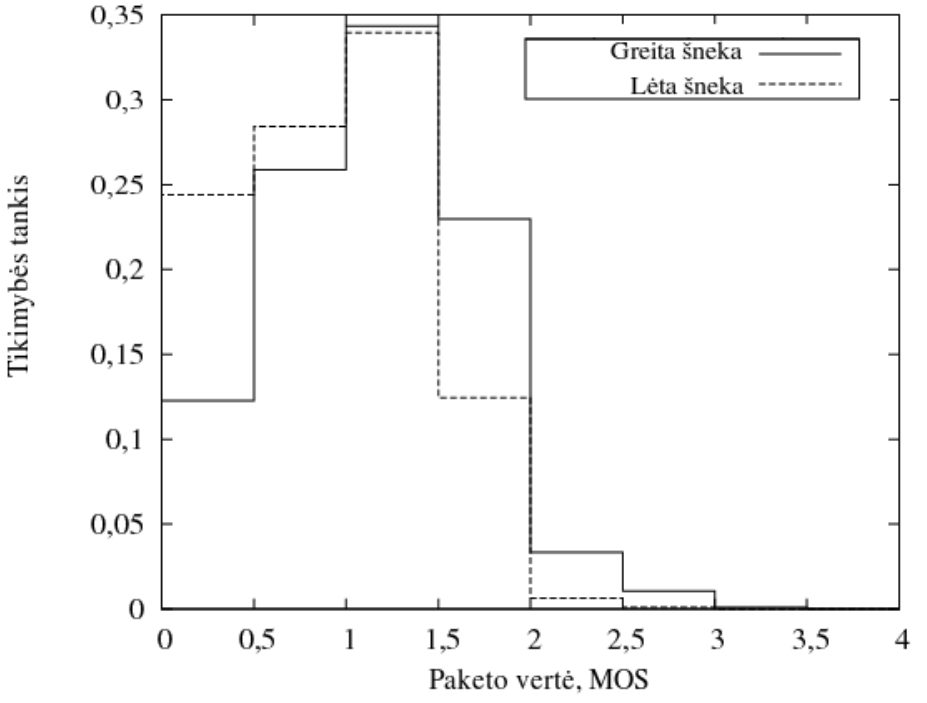
  čia x yra šnekos signalas, o $n = 0,...,1999$ – paketo numeris. Surikiuokime $E(n)$ elementus didėjimo tvarka ir tokiu būdu gautą vektorių pažymėkime $E_1(n)$. Kitaip, $E_1(n) = \mathrm{SORT}(E(n))$, kur „SORT“ yra vektoriaus elementų rikiavimo didėjimo tvarka operatorius.
- Pasirinkime slenkstinę energiją E_{sl}, lygią $E_1(200)$. Kitaip tariant, slenkstinė energija yra didžiausia paketo energija iš 10 % paketų su mažiausia energija.
- Jei eilinio paketo energija mažesnė, negu $1{,}2E_{sl}$, tai laikome, kad šiame pakete yra tyla. Tokiu atveju slenkstinę energiją atnaujiname pagal formulę: $E_{slnew} = 0{,}8\ E_{slold} + 0{,}2\ E(k)$ ir pereinam prie tolimesnio paketo analizės. Čia E_{slnew} yra nauja slenksčiui naudojama energijos reikšmė, o E_{slold} yra sena slenkstinės energijos reikšmė
- Jei paketo energija didesnė, negu E_{sl}, tai pažymime šį paketą kaip aktyvų ir pereinam prie kito paketo analizės.

Algoritmo aprašyme pateikti koeficientai yra standartiniai (Sohn *et al* 1999; El-Maleh *et al* 1997).

Pašalindami paketus, kuriuose yra tyla, tikimės, kad histogramos, atitinkančios greitą ir lėtą šnekas, supanašės. Tačiau gaunamas supanašėjimas nėra didelio mąsto. Kitaip sakant, ganėtinai žymūs skirtumai tarp histogramų išlieka nepaisant tylos pašalinimo.

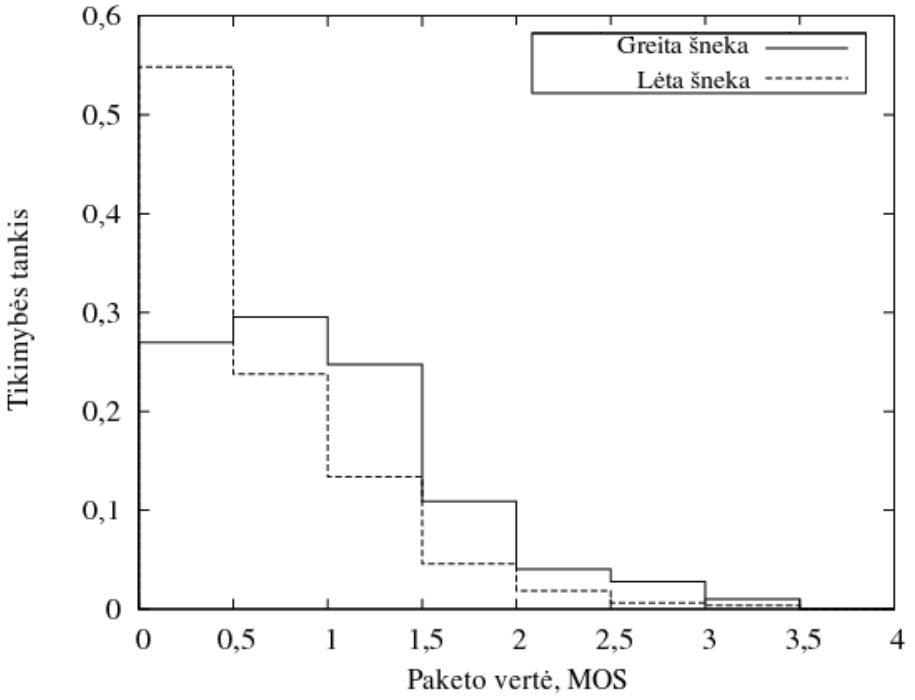
2.12 paveikslas rodo, kaip paketo vertė yra pasiskirsčiusi, kai neanalizuojami paketai, atitinkantys tylą. Matome, kad paketų, kurių vertė yra labai maža (mažesnė, negu 0,5) vistiek daugiau yra lėtos šnekos signale. Nors skirtumas nebėra toks ryškus, kaip nenaudojant balso aktyvumo detektoriaus.

2.13 paveikslas rodo, paketo vertės histogramas, kai naudojamas AMR-12.2 koderis. Tendencijos išlieka tos pačios – skirtumai tarp histogramų yra mažesni, negu tuo atveju, kai nenaudojamas balso aktyvumo detektavimas, tačiau jie išlieka ir gana žymūs.



2.12 pav. Paketo vertės histograma, kai naudojamas balso aktyvumo detektorius (PCM koderis)

Fig. 2.12. Frame value histogram when using VAD and PCM codec



2.13 pav. Paketo vertės histograma, kai naudojamas balso aktyvumo detektorius (AMR-12.2 koderis)
Fig. 2.13. Frame value histogram when using VAD and AMR-12.2 codec

2.5. Segmentų vertingumo priklausomybė nuo matavimo intervalo ilgio

Iki šiol matuodami paketo vertę naudojom fiksuotą signalo ilgį, lygų 0,5 s. Šiame intervale buvo lokalizuotas iškraipymas – vieno ar kelių paketų praradimas. Akivaizdu, kad jei matavimui naudotume kitokį signalo ilgį, tačiau iškraipymą paliktume tokį patį, tai rezultatas būtų gaunamas kitoks.

Prisiminkime PESQ algoritmo galutinio iškraipymo formavimą (1.2) ir tarkime, kad tik dvi $L_{DA}(i)$ reikšmės yra nenulinės, kai $i = j$ ir $i = j+1$. Toks atvejis gaunamas, jei signalo iškraipymas (iškraipyto signalo skirtumas nuo atraminio signalo) yra lokalizuotas pusėje PESQ skiemens. Taip pat nagrinėkime tik simetrinį iškraipymą, nes asimetrinio iškraipymo analizės atvejis yra analogiškas. Tokiu atveju (1.2) tampa:

$$d_{\text{asym}} = \frac{\sqrt{L^2{}_{\text{DA}}(j) + L^2{}_{\text{DA}}(j)}}{\sqrt{N}}. \tag{2.4}$$

Matome, kad iškraipymo reikšmė, o tuo pačiu ir PESQ rezultatas, priklauso nuo PESQ skiemenų skaičiaus, kuris, savo ruožtu, tiesiog proporcingas signalo ilgiui (Kajackas et al 2008). Iš čia gauname išvadą, kad turėdami bet kokius PESQ matavimus, jei jie atlikti su lokalizuotom, pavienėm klaidom, turime nurodyti ir matavimo intervalo ilgį. Aišku, šis ilgis gali būti sąlyginai laisvai pasirenkamas ir priklauso tik nuo sprendžiamo uždavinio. Disertacijose daugelis matavimų buvo atliekama naudojant 0,5 s ilgio signalo atkarpas. Toks signalo ilgis yra apytiksliai lygus vienam ištartam žodžiui, o trumpesnis ilgis kai kurių žodžių nesutalpintų.

2.6. Dviejų paketų verčių sumavimas

Iki šiol nagrinėjome, kokia yra vieno ar kelių paeiliui prarastų paketų vertė. Tačiau nebandėme nustatyti, kaip yra susiję atskirų paketų vertės ir suminė dviejų paketų vertė. Kitaip tariant, tarkim, kad prarandam du paketus, nutolusius vienas nuo kito laike. Nutolimas gali būti nulinis, kas reiškia, kad prarandami paketai yra greta, o gali būti ir taip, kad tarp prarastų paketų yra nesugadintų. Keliam klausimą – ar dviejų prarastų paketų vertę galima gauti kaip atskirų paketų verčių tam tikrą sumą?

Pažymėkime atskirų paketų vertes $Q_1(i)$ ir $Q_2(i)$, kur i reiškia matavimui naudojamo šnekos intervalo numerį. Abiejų paketų suminę vertę pažymėkime $Q_{12}(i)$.

Jei prarasti paketai vienas nuo kito yra nutolę pakankamai toli ir pirmojo paketo praradimas neturi įtakos antram paketui, tai pagal PESQ agregavimo formulę 1.2 gautume, kad:

$$Q_{12}(i) = \sqrt{Q_1^2(i) + Q_2^2(i)}. \tag{2.5}$$

Bendru atveju pabandykime dviejų paketų vertę prognozuoti naudodami tokią išraišką:

$$Q'_{12}(i) = \sqrt{Q_1^2(i) + k \cdot Q_2^2(i)}, \tag{2.6}$$

čia koeficientas k yra randamas minimizuojant išraišką:

$$\sum_{i=0}^{1999} (Q_{12}(i) - Q'_{12}(i))^2. \tag{2.7}$$

Sumavimo rėžiai tokie todėl, kad koeficiento skaičiavimui, kaip ir anksčiau, imta 2000 šnekos intervalų. Šios išraiškos išvestinė gaunama tokia:

$$\sum_{i=0}^{1999} Q_2^2(i)\left(1-\frac{Q_{12}(i)}{\sqrt{Q_1^2(i)+kQ_2^2(i)}}\right). \tag{2.8}$$

Prilyginus šią išraišką nuliui ir suprastinę, gauname tokią lygtį:

$$\sum_{i=0}^{1999}\left(\sqrt{Q_1^2(i)+kQ_2^2(i)}-Q_{12}(i)\right)=0\,. \tag{2.9}$$

Šios lygties uždara forma išspręsti nepavyksta, bet skaitinis sprendimas elementarus, nes funkcija akivaizdžiai didėjanti ir tolydi bei lengva rasti intervalą, kurio galuose ženklai priešingi. Todėl sprendimui naudojome elementarų intervalo dalijimo pusiau metodą.

Dviejų paketų vertės prognozavimo tikslumą nurodys dydžio $Q_{12}(i)-Q'_{12}(i)$ vidutinis kvadratinis nuokrypis, t.y.

$$RMSE=\sqrt{\frac{1}{2000}\sum_{i=0}^{1999}\left(Q_{12}(i)-Q'_{12}(i)\right)^2}\,. \tag{2.10}$$

Eksperimentai buvo atliekami pasirenkant skirtingą atstumą tarp paketų. Kiekvienam skirtingam atstumui tarp prarandamų paketų nuosekliai skaičiuodavome $Q_1(i)$, $Q_2(i)$, $Q_{12}(i)$, k, *RMSE*, kur RMSE yra vidutinis kvadratinis nuokrypis, aprašoma (2.10) formule. Rezultatai, gaunami esant skirtingam atstumui tarp prarandamų paketų, pateikti 2.3 lentelėje. Nulinis atstumas reiškia, kad prarandami gretimi paketai, atstumas „1“ reiškia, kad atstumas tarp prarastų paketų lygus vienetui ir t. t. Duomenys lentelėje pateikti, kai eksperimentui buvo naudojamas „greitos“ šnekos įrašas.

2.3 lentelė. Dviejų paketų vertės prognozavimo duomenys
Table 2.3. Data for prognosis of two frame value

Atstumas tarp prarastų paketų	0	1	2	3
„*k*“ vertė	1,03	1,14	0,92	0,89
RMSE	0,213	0,208	0,095	0,075

Iš lentelės matome, kad sumuojant dviejų atskirų paketų vertes pagal (2.1), gauname gana tikslią šių dviejų paketų vertės prognozę. Jau tuo atveju, kai paketai nutolę per 3 nesugadintus paketus, daugeliui praktinių tikslų prognozės tikslumas tampa pakankamas – kokybės matavimo variacija iki 0,1 dažniausiai yra neesminė.

2.7. Antro skyriaus išvados

- PESQ algoritmas nėra tinkamas trumpų signalo atkarpų kokybei nustatyti, nes gaunama nepageidaujama rezultatų priklausomybė nuo matavimo lango pozicijos signalo atžvilgiu.
- Pasiūlytas signalų formavimo metodas, kuris leidžia naudoti PESQ algoritmą trumpiems signalams. Metodas vidutiniškai sumažina matavimų sklaidą 8 kartus naudojant koreliacijos koeficientą kaip kriterijų.
- Koduoto šnekos paketo informacine verte siūlome laikyti to paketo įtakos dydį šnekos kokybei. Įtaka nustatoma pasirinkto kokybės rodiklio degradavimo dydžiu, susidarančiu, kai tiriamas paketas prarandamas.
- Nustatyta balso duomenų paketų informacinės vertės kokybinė priklausomybė nuo kodavimo būdo. Koduojant PCM būdu, paketų su didesne energija vertė yra didesnė. Kai balsas koduotas AMR kodais vertingesni tie paketai, kurie perteikia greitai kintančias signalo dalis.
- Pasiūlyta formulė (2.1), leidžianti kelių prarastų paketų įtaką kokybei prognozuoti pagal šių paketų vertes.
- Apibendrintai galime teigti, kad „lėtos“ šnekos paketo vertė vidutiniškai yra mažesnė, negu „greitos“ šnekos.

3

Balso mikroiškraipymų įtakos analizė

Šiame skyriuje, siekiant darbe suformuluoto bendro tikslo, tiriama balso kodavimo būdo ir kodavimo parametrų įtaka balso kokybei. Dėmesys sutelktas į du reiškinius:

- Kodavimo skirtingais AMR koderiais įtaka signalui ir jo kokybei;
- Prarasto paketo įtakos išplitimas laike, kai pastebima prarasto paketo įtaka po jo sekančiam signalui.

Mokslinėje literatūroje yra duomenų apie koderių įtaką balso kokybei (3GPP TR 26.975 V5.0.0. 2002; 3GPP TR 26.976 V5.1.0 (2003 09); Holma 2003; Ding 2003; Sun 2001). Yra žinoma, kad skirtingų kalbų (pvz. anglų, japonų) gaunama skirtinga kokybė, net jei naudojamas vienas ir tas pats koderis (Paulikas 2008). Tačiau skelbti darbai neapima visų su balso kodavimu sietinų aspektų – nenagrinėja, kaip koduoto balso kokybė statistiškai priklauso nuo koduojamo signalo.

Šiame skyriuje tyrimai vykdomi pagal bendrą schemą:

- Parenkamas balso signalas $v(t)$.
- Signalas $v(t)$ koduojamas ir dekoduojamas naudojant tiriamą koderį, sukuriant modifikuotą kopiją $v^*(t)$.
- Analizuojami modifikuotos kopijos $v^*(t)$ pokyčiai ir bendros savybės.
- Eksperimentai kartojami su daugeliu skirtingų balso atkarpų ir daromos išvados.

3.1. Tiriamieji signalai ir jų savybės

Tyrimams buvo naudojamas CNN žinių įrašas (anglų kalba). Toje signalo atkarpoje, su kuria buvo atliekami tyrimai, tylos intervalai trumpi, kalbėtojas yra vyras, o pačią šneką būtų galima pavadinti greitakalbe – tai tipiška šneka, kokia skaitomos žinios.

Tyrimai buvo vykdomi naudojant AMR koderius. Šie koderiai koduoja 20 ms trukmės balso segmentus (toliau vadinamus baziniais segmentais) ir kiekvienam segmentui suformuoja duomenų paketą.

Nėra standartizuotų metodų, kaip apibūdinti bendras ilgos šnekos signalo atkarpos savybes. Trumpos šnekos signalo atkarpos (10–20 ms ilgio) paprastai apibūdinamos naudojant tam tikrus spektrinius požymius (Lipeika 2002).

Bendroms naudotų signalų savybėms apibūdinti panaudoti suvidurkinti amplitudžių spektrai. Spektrinės signalų charakteristikos buvo apskaičiuojamos taip:

- Kiekvienam balso baziniam segmentui $x(i)$, $i = 0, 1, 2,... 159$ (160 imčių atitinka 20 ms) paskaičiuojamas diskretusis amplitudžių spektras naudojant diskrečiąją Furjė transformaciją:

$$X(k) = \sum_{n=0}^{159} x(i) \cdot e^{-\frac{2 \cdot \pi \cdot j}{160} \cdot k \cdot i}, \tag{3.1}$$

  kur j yra menamas vienetas. Amplitudžių spektras $X_n(k)$, kur $k = 0, 50, 100,... 4000$ Hz yra dažnis, o n – paketo numeris, gaunamas iš $X(k)$ pagal formulę: $X_n(k) = |X_k|$, kai $k = 0, 1, 2,... 80$. Kitaip sakant, pirma kiekvienam baziniam segmentui paskaičiuojame diskrečiąją Furjė transformaciją, tada atmetame antrąja gauto vektoriaus pusę (nes ji simetriška pirmajai, kadangi signalas realusis) ir paimame likusio vektoriaus modulį. Gautasis vektorius atspindi signalo bazinio segmento amplitudinį spektrą.
- Galutinis rezultatas – spektrų vidurkis skaičiuojamas pagal formulę

$$X(k) = \frac{1}{3000} \cdot \sum_{n=1}^{3000} X_n(k). \tag{3.2}$$

Vykdant eksperimentus buvo pastebėta, kad dauguma statistinių dėsningumų tampa reguliariais (keičiant įrašus pakinta sąlyginai nedaug), jei tyrimams naudojama 3.000 šnekos paketų, todėl (3.2) vidurkinama 3000 spektrų $X_n(k)$.

3.1 paveiksle parodytas naudojamo signalo gaunamas amplitudžių spektro vidurkis. Tai gana būdingas šnekos signalų spektras be matomų nestandartinių

nukrypimų. Tyrimams, aišku, ir buvo siekiama naudoti niekuo neišsiskiriantį signalų rinkinį.

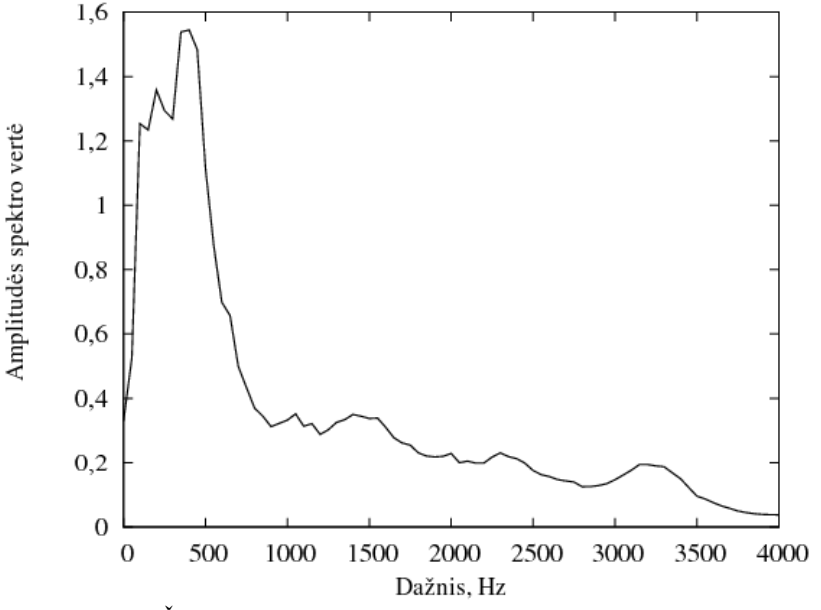
Kita signalų savybėms apibūdinti panaudota charakteristika, yra bazinių šnekos signalo segmentų energijų skirstiniai, kurie parodo, kaip pasiskirsto signalo energija baziniuose segmentuose Viename baziniame segmente (*k*-ajame) energija skaičiuojama taip:

$$E(k) = \sum_{n=160*i}^{160*k+159} x^2(n) . \tag{3.3}$$

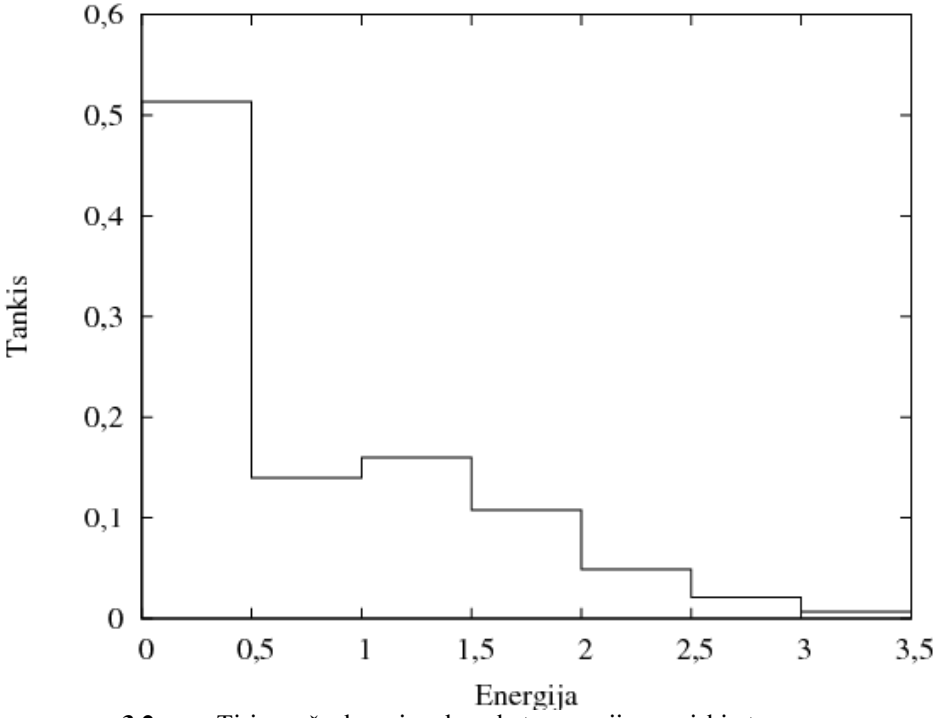
Iš energijų segmentuose skirstinių galima spręsti, pavyzdžiui, kokia nagrinėjamoje atkarpoje yra tylos dalis (energija artima nuliui) ir kiek pasitaiko didžiausia energija pasižyminčių atviros, balsius atitinkančios, šnekos segmentų.

Energijų skirstiniai buvo sudaryti paskaičiuojant energijos vertes kiekvienam iš 3.000 paketų. Energijos verčių histograma pateikta 3.2 paveiksle.

Analizuojant gautas histogramas ir pasirinkus šnekos segmento energiją kaip paprasčiausią kriterijų šnekos aktyvumui detektuoti, gauname, kad nagrinėjamų šnekos įrašų tylos dalis priklausomai nuo tylai priskyrimo slenksčio yra 10–30%. Tokie rezultatai sutampa su literatūros šaltinių duomenimis apie tylos intervalų procentą standartinėje šnekoje (Brady 1965; Brady 1968; Brady 1968; Kajackas, Šaltis 2003; Šaltis 2004).



3.1 pav. Šnekos signalo įrašo amplitudžių spektro vidurkis
Fig. 3.1. Average of amplitude spectrums of speech record



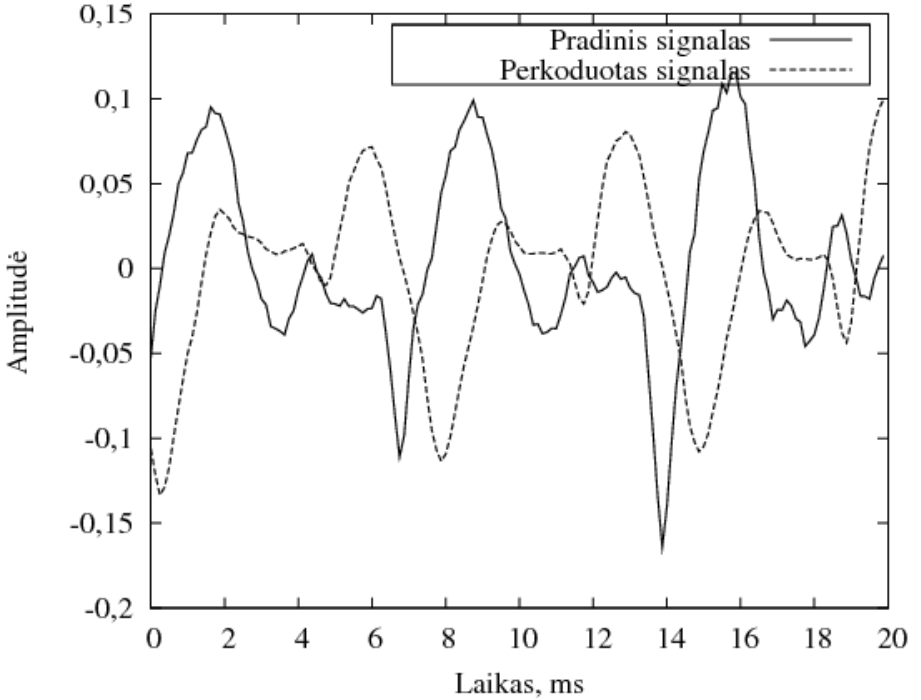
3.2 pav. Tiriamo šnekos signalo paketų energijos pasiskirstymas
Fig. 3.2. Energy distribution of speech sample under investigation

3.2. Skirtingų AMR koderių sukuriami balso iškraipymai

Šiame poskyryje nagrinėsime, kaip skirtingi AMR koderiai iškraipo balso signalus. AMR koderiai ir jų specifikacijos detaliai aprašyti (3GPP TS 26.071 2004). Šis nagrinėjimas disertacijos kontekste specifinis tuo, kad nagrinėsime tik kodavimo (o ne prarastų paketų) įtaką šnekos kokybei. Bazinis nagrinėjimo modelis toks:

- Parenkamas balso signalas $v(t)$.
- Signalas $v(t)$ koduojamas – dekoduojamas AMR-n koderiu.
- Nagrinėjamos originalaus bei perkoduotų signalų charakteristikos naudojant PESQ ir metriką, pagrįstą spektro iškraipymais. Taikant PESQ įvertinami girdimieji iškraipymai. Spektriniai iškraipymai vertinami pagal LPC (angl. *Linear Predictive Coding*) koeficientų vidurkių pokyčius.

Šio tyrimo objektui iliustruoti 3.3 paveiksle parodyta, kaip atrodo pradinio signalo 20 ms atkarpa (atitinkanti vieną paketą) ir tas pat signalas, bet perkoduotas (koduotas ir dekoduotas) AMR-4.75 koderiu. Iš grafikų matosi, kad signalo perkodavimas gerokai pakeičia patį signalą. Taip pat matyti, kad sukuriamą iškraipymą gana sunku apibūdinti kiekybiškai. Bandysime statistiškai apibūdinti iškraipymo dydį, kai perkoduojami skirtingi signalai.



3.3 pav. Signalo kodavimo AMR-4.75 koderiu rezultatas
Fig. 3.3. Coding of signal using AMR-4.75 codec

3.2.1. Spektro iškraipymai

Formaliai parodysime, kaip kodavimas skirtingais koderiais keičia dažninę signalų sudėtį. Tuo tikslu buvo atliktas eksperimentas, kurio rezultate gavome duomenis, kaip kodavimas keičia signalo amplitudžių charakteristiką. Detaliai eksperimentas atrodo taip:

1. Parinkta pakankamai ilga signalo atkarpa, kuri suskaidoma į 3.000 bazinių segmentų, kurių kiekvienas koduojamas ir dekoduojamas nepriklausomai AMR-4.75 ir AMR-12.2 koderiais.

2. Kiekvienam iš 3.000 segmentų paskaičiuojami amplitudžių spektrai. Signalo segmentų prieš kodavimą $x(t)$ amplitudžių spektrai pažymėti $X_n(k)$. Dekoduotų signalo segmentų amplitudžių spektrai pažymėti $Y_n(k)$.
Indeksas n abiem atvejais reiškia naudojamo paketo numerį (indeksą), k reiškia dažnį.

3. Suvidurkinta „amplitudės dažninė charakteristika“ koderiui paskaičiuojama naudojant išraišką:

$$H(k) = \frac{1}{3000} \cdot \sum_{n=1}^{3000} \frac{Y_n(k)}{X_n(k)}. \tag{3.4}$$

Šią išraišką sunormuojame taip, kad vidutinė charakteristikos reikšmė būtų lygi vienetui. Normavimas vykdomas pagal tokia formulę:

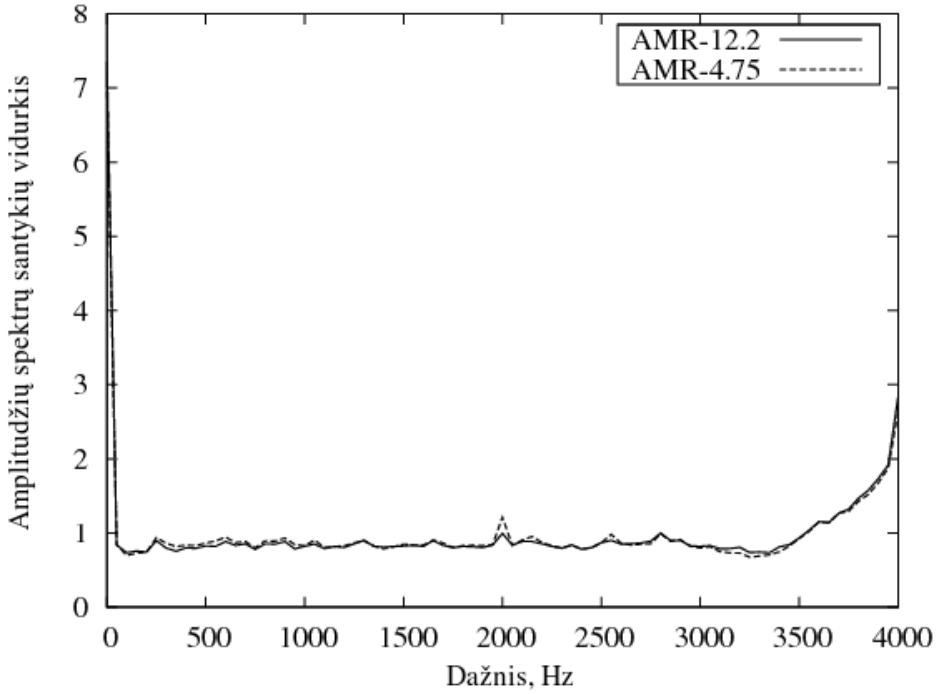
$$H(k) = 81 \cdot \frac{H(k)}{\sum_{k=0}^{80} H(k)}. \tag{3.5}$$

„Amplitudės dažninė charakteristika“ parašyta kabutėse, nes kodavimas – dekodavimas be abejo yra netiesinis operatorius. Formaliai čia nereikėtų taikyti tiesinių sistemų analizės metodų. Tačiau toks pasirinktas analizės būdas sudaro galimybes vaizdžiai parodyti, kaip skirtingi koderiai pakeičia signalų spektrus.

Eksperimento rezultatas – gautos normuotos „perdavimo charakteristikos“ parodyta 3.4 paveiksle. Matome, kad spektriniai amplitudžių iškraipymai nėra dideli. Įdomiau tai, kad nematyti kokio nors žymesnio skirtumo tarp didesnės spartos koderio AMR-12.2 ir mažos spartos koderio AMR-4.75 sukeliamų iškraipymų. Atliktas eksperimentas parodo, kad koderių amplitudiniai spektriniai iškraipymai nėra svarbiausi analizuojant kodavimo – dekodavimo kokybę.

3.2.2. LPC iškraipymai

Panagrinėsime, kaip kodavimas iškraipo balso signalo LPC požymius. Analizei pasirinkti LPC požymiai, nes jų naudojimas gan plačiai paplitęs (Lipeika 1999), jų išskyrimas sąlyginai nesudėtingas (Chu 2003), o savybės gan plačiai tyrinėtos. Nors šnekos atpažinime (šnekos atpažinimas yra ta sritis, kur įvairūs balso signalų požymių vektoriai plačiausiai tyrinėti) yra efektyvesnių požymių (Lipeika 2005; Rudžionis 2005), bet mūsų sprendžiami uždaviniai kiti, todėl standartiniai LPC požymiai yra tinkamas pasirinkimas pradiniams tyrimams.



3.4 pav. „Amplitudinės dažninės charakteristikos“ AMR-12.2 ir AMR-4.75 koderiams

Fig. 3.4. Amplitude frequency response for AMR-12.2 and AMR-4.75 codecs

Trumpai aptarsime naudojamą LPC požymių apibrėžimą. Kiekvienam balso baziniam segmentui paskaičiuojami koeficientai $a_1, a_2, ..., a_{10}$ tokie, kurie minimizuoja funkciją:

$$\sum_{n=0}^{159}\left(x(n)-\sum_{i=1}^{10}a_i\cdot x(n-i)\right)^2, \tag{3.6}$$

čia $x(n)$ – nagrinėjamo šnekos signalo atkarpa; $x(n) = 0$, kai $n < 0$. Konkretus minimizavimo metodas aptartas (Chu 2003).

LPC iškraipymai AMR-4.75 ir AMR-12.2 koderiams buvo tiriami naudojant tokį eksperimentą. Signalą, sudarytą iš 3.000 AMR paketų, perkoduojame ir paskaičiuojame, kaip perkodavimas iškraipo LPC koeficientus. Toks tyrimas leidžia kiekybiškai įvertinti skirtingų koderių sukuriamus iškraipymus, o taip pat tų iškraipymų ribas. Detaliau eksperimentas aprašytas žemiau:

1. Koduokime ir dekoduokime visą 3000 bazinių segmentų ilgio signalą naudojant AMR-12.2 ir AMR-4.75 koderius.
2. Kiekvienam iš 3.000 bazinių segmentų paskaičiuojami LPC koeficientai, kai: signalas neiškraipytas (prieš kodavimą gauti koeficientai pažymėti $A_i(n)$) ir signalas iškraipytas – po dekodavimo (šie koeficientai pažymėti $B_i(n)$). Indeksas i abiem atvejais reiškia naudojamo paketo numerį, o n – LPC koeficiento numeris i-ajame pakete.
3. Kiekvienam iš 3.000 bazinių segmentų paskaičiuojame koderių sukuriamus LPC koeficientų iškraipymus. Naudojama Euklido metrika tarp LPC požymių vektorių:

$$D_i = \sqrt{\sum_{n=1}^{10} (A_i(n) - B_i(n))^2} \ . \tag{3.7}$$

4. Naudodami duomenis D_i, kur $i = 1, 2, \ldots, 3000$ braižome histogramas nagrinėjamiems koderiams ir paskaičiuojame iškraipymų vidurkius bei standartines deviacijas.

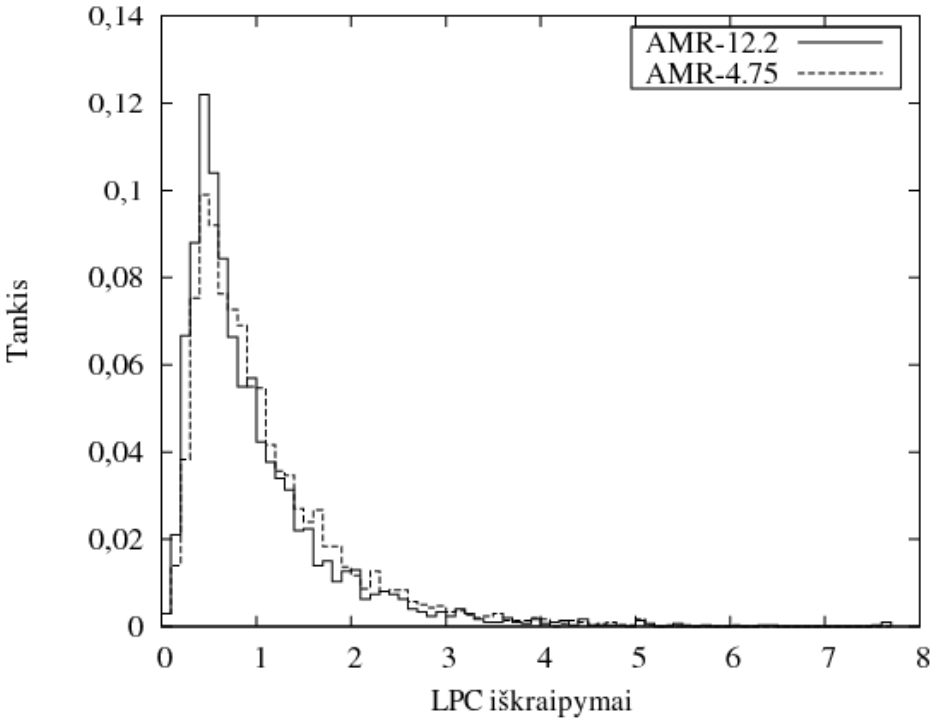
Aprašyto eksperimento sukuriami rezultatai pateikti 3.5 paveiksle. Matome, kad AMR-4.75 koderio sukuriami LPC iškraipymai yra ne daug didesni, nei AMR-12.2 iškraipymai. Iškraipymų vidurkiai yra atitinkamai 0,96 – AMR-12.2 koderiui ir 1,07 – AMR-4.75 koderiui.

Šie rezultatai leidžia pateikti hipotezę, kad girdimoji balso kokybė, kuri matuojama MOS arba PESQ metodais yra menkai koreliuota su spektro bei LPC koeficientų iškraipymais. Šį faktą galima paaiškinti PESQ algoritmo fonemų agregavimo formulės (1.1) specifika. Šeštasis laipsnis formulėje labai išryškina net ir menkus skirtumus spektro iškraipymuose.

3.2.3. Girdimieji iškraipymai

Girdimieji iškraipymai, priešingai nei spektriniai bei LPC, naudoja integralinę metriką. Subjektyvus girdimųjų iškraipymų nustatymo metodas yra MOS testai. Šiame darbe visur MOS testai keičiami objektyviais matavimais taikant PESQ ar PESQ_{mod} algoritmą.

Eksperimentas, kai norima nustatyti skirtingų koderių įtaką šnekos kokybei, yra gana paprastas. Imame 3.000 skirtingų šnekos signalų ir juos perkoduojame AMR-12.2 ir AMR-4.75 koderiais. Kiekvienam iš 3.000 signalų paskaičiuojame dPESQ kokybės degradavimo metriką ir iš gautų duomenų brėžiame histogramas bei skaičiuojame vidurkius.



3.5 pav. LPC koeficientų iškraipymai naudojant AMR-12.2 ir AMR-4.75 koderius

Fig. 3.5. Distortions of LPC coefficients for AMR-12.2 and AMR-4.75 codecs

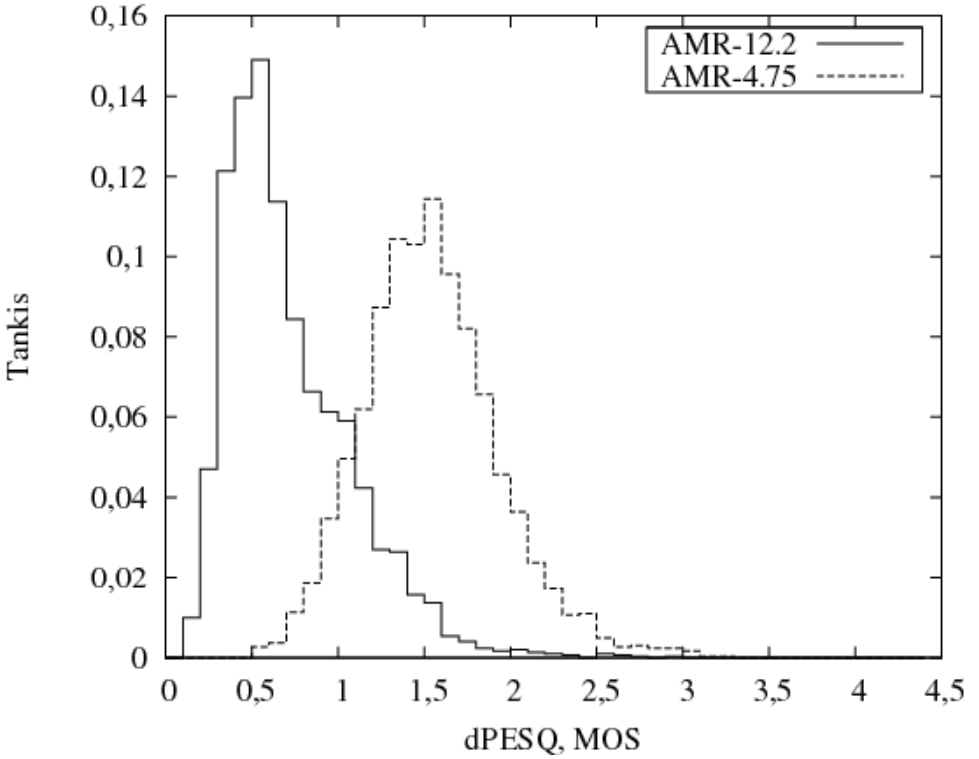
Aprašyto eksperimento rezultatai pateikti 3.6 paveiksle. Matome, kad šiuo atveju AMR-4.75 koderis signalą iškraipo žymiai labiau, negu AMR-12.2 koderis.

Vidutinis AMR-12.2 iškraipymas yra 0,73, o vidutinis AMR-4.75 koderio sukuriamas iškraipymas yra 1,55 MOS vienetai. Taigi MOS skalėje AMR-12.2 sukuria kokybės pablogėjimą dvigubai mažesnį, negu pats žemiausias AMR koderis.

Kitas įdomus pastebėjimas – matome, kad histogramos yra gana plačios – jų standartinės deviacijos sąlyginai didelės. Tai reiškia, kad skirtingi žodžiai, koduoti tuo pačiu koderiu, gali įgauti kokybės pablogėjimą nuo 0,5 iki 3,0 (AMR-4.75 koderiui).

Šie duomenys leidžia suabejoti, kad koderį galima apibūdinti tuo, kaip jis sumažina kokybę. Kaip matome iš atliktų eksperimentų, gaunamas kokybės įvertis labai priklauso nuo naudojamo signalo.

Analizuojant trumpus šnekos signalus – mikrostruktūrą, koderio sukuriamus kokybės degradavimus realiai galima apibūdinti tik statistiškai – tikimybiniais skirstiniais. Vienu skaičiumi išreikšta kokybė šiuo atveju nėra pakankamas rodiklis.



3.6 pav. Girdimieji (dPESQ) iškraipymai naudojant AMR-12.2 ir AMR-4.75 koderius

Fig. 3.6. Perceptual distortions for AMR-12.2 and AMR-4.75 codecs

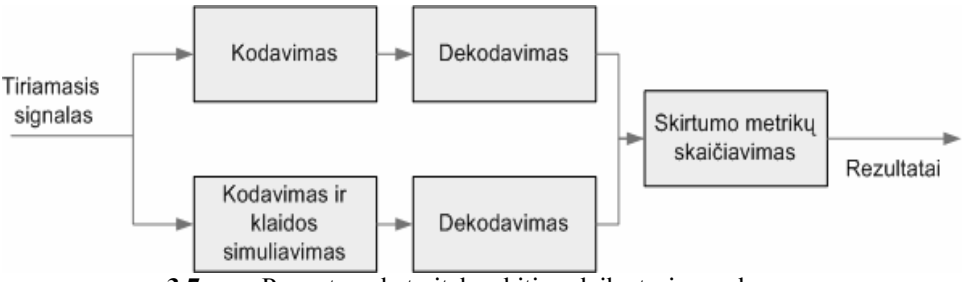
3.3. AMR koderių inertiškumo tyrimas, kai prarandami paketai

Prarastas paketas iškraipo ne tik signalą, atitinkantį šį paketą, bet ir toliau esantį signalą. Tyrimo tikslas yra kiekybiškai įvertinti, kaip visas signalas yra iškraipomas laike.

Dauguma modernių koderių (AMR taip pat) yra inertiški – jie veikia panašiai, kaip rekursinis filtras. Tam, kad dekoduotų *i*-ąjį paketą, dekoderis

naudoja ne tik informaciją, esančią *i*-ajame pakete, bet ir informaciją iš $i - 1$ paketo. Tokiu būdu, jei perdavimo metu yra sugadinamas $i - 1$-asis paketas, tai net ir turint tikslius duomenis *i*-ajame pakete, signalas jame nebus atstatomas tiksliai (lyginant su tuo, koks signalas būtų, jei paketo praradimo nebūtų buvę). Iš šio dėstymo aišku, kad sugadinus *i*-ąjį paketą bus paveikti ir $i + 1$, ir $i + 2$ ir visi paketai iki $i + N$-ojo. Aišku, kad jei daugiau prarastų paketų nebus, tai po kurio laiko koderis vėl grįš į sinchronizuotą būseną, palyginti su tokiu koderiu, kuriame klaida nebuvo įvykusi. Čia ir vėliau, kai sakome „dekoderiai sinchronizuoti“, tai turim mintyje, kad signalai po dekodavimo abiejuose dekoderiuose 3.7 paveiksle yra vienodi. Tokiu būdu kiekvienas paketinis praradimas išsinchronizuoja dekoderius, Po kurio laiko jie sinchronizuojasi, jei prarastų paketų daugiau nėra.

Dėl išdėstytų motyvų tyrimams gerai tinka apibendrinta schema pateikta 3.7 paveiksle. Ši schema leidžia įvertinti tik paketo praradimo įnešamą signalo iškraipymą, o paties koderio įtaka yra eliminuojama. Tai lengva matyti, jei pastebėsim, kad tuo atveju, kai apatinėje šakoje klaida nėra įvedama, tai abu signalai (viršutinėje ir apatinėj šakose) yra lygūs. Ir jie yra lygūs nepriklausomai nuo to, koks koderis naudojamas.

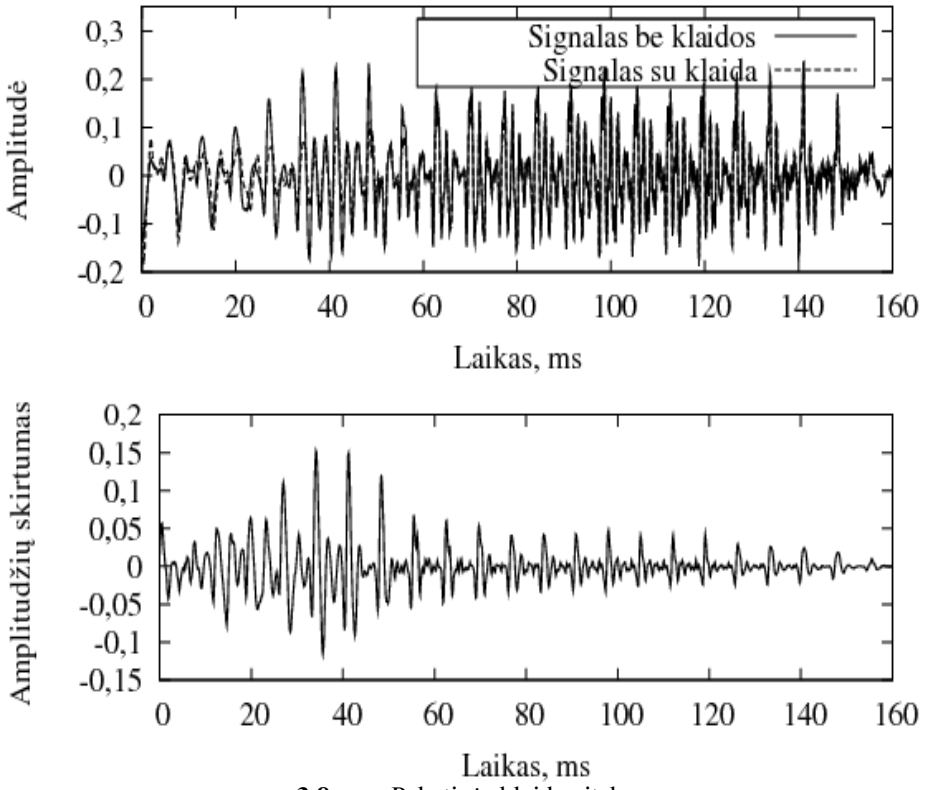


3.7 pav. Prarasto paketo įtakos kitimo laike tyrimo schema
Fig. 3.7. Diagram of investigation of the impact of lost frame

Tyrimo metu lyginami du signalai – pirmas, kuris gautas perkodavus pradinį signalą, bet neįvedus klaidų. Antrasis signalas gaunamas užkodavus, įvedus klaidą ir dekodavus pradinį signalą. Įvairios metrikos (PESQ ir LPC požymiais grįstos) yra naudojamos aprašyti skirtumams tarp šių signalų. Priklausomai nuo to, kaip ir kuriai signalo daliai taikysime metriką, gausime rezultatus, atskleidžiančius vis kitą prarasto paketo pasireiškimo formą.

3.8 paveiksle parodytas pavyzdys, kaip atrodo signalai po dekoderių, naudojant AMR-12.2 koderį. Viršutinėje paveikslo dalyje pateikti jau aptarti signalai, o apatinėje – skirtumas tarp minėtų signalų. Paketo praradimas įvyksta laiko momentu 0, t.y, tarp 0 ir 20-os ms. Matome, kad skirtumas tarp signalų gan ženklus prarasto paketo vietoje, dar didesnis, kai atsiranda vieno paketo

vėlinimas, o toliau mažėja. Po 160 ms po paketo praradimo dekoderis jau praktiškai yra grįžęs į sinchronizuotą būseną.



3.8 pav. Paketinės klaidos įtaka
Fig. 3.8. The impact of lost frame

Toliau, kaip ir ankstesniame poskyryje, nagrinėsime girdimuosius ir spektrinius iškraipymus. Šį kartą papildomai tirsime, kaip iškraipymų dydis kinta laike.

3.3.1. Girdimieji iškraipymai

Kadangi girdimieji iškraipymai pagal savo prigimtį yra integraliniai, tai norėdami nustatyti jų kitimą laike, turime naudoti kokį nors netiesioginį metodą. Paketų praradimo atveju siūlome naudoti tokį metodą. Tarkime, kad buvo prarastas *i*-asis paketas. Norime sužinoti, kokie buvo girdimieji iškraipymai praėjus 0, 20, 40, 60 ms po prarasto paketo pradžios. Tam naudosime daugelį

matavimo langų. Norėdami sužinoti girdimuosius iškraipymus nuliniu laiko momentu naudosime langą nuo 0 iki 0,5 s, norėdami sužinoti iškraipymus, kai vėlinimas nuo prarastojo paketo yra 20 ms naudosime matavimo langą nuo 0,02 s iki 0,52 s. Ir taip toliau. Aišku, kad tokiu būdu gaunama funkcija bus nedidėjanti (praktiškai ji visada bus griežtai mažėjanti). Laiko momentą, kai liekamieji iškraipymai bus mažesni, negu 10 % nuo pradinio iškraipymų dydžio, galima laikyti dekoderio atstatyto normalaus veikimo pradžia.

Eksperimentas buvo atliekamas imant 3.000 signalų, kiekvienam iš jų buvo paskaičiuojama, kaip kinta dPESQ metrika, kai kokybės matavimo langas slenkamas laike. Galutinis rezultatas yra šių dPESQ metrikų kitimo laike vidurkis. Detaliai eksperimentas aprašytas žemiau:

1. Iš viso pradinio signalo, kuris yra 3.050 bazinių segmentų ilgio, pasirenkame signalo dalį tarp i-tojo ir $i + 35$-ojo segmentų. Tai atitinka laiko momentus signale $i \cdot 0{,}02$ s ir $(i + 35) \cdot 0{,}02$ s. Pažymėkime taip pasirinktą signalą z_i.

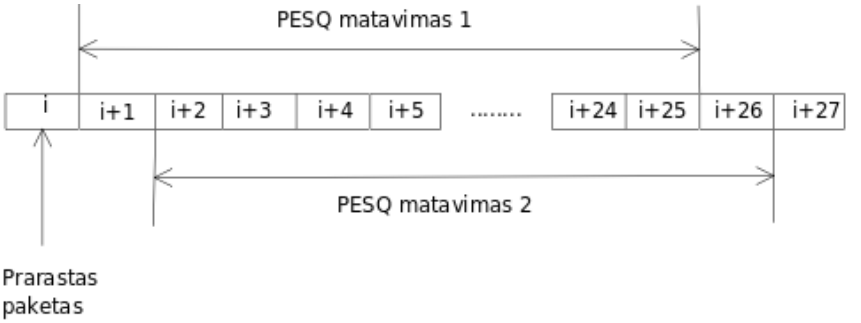
2. Matavimo signalą z_i koduojame ir dekoduojame. Iš dekoderio gautą signalą pavadinkime x_i. Kitas signalas gaunamas prieš dekodavimą imituojant klaidos atsiradimą nuliniame baziniame signalo z_i segmente. Šį signalą pavadinkime y_i.

3. Paskaičiuojame gaunamus iškraipymus, atitinkančius i-tąjį matavimo signalą ir matavimo vėlinimą v (čia v yra vėlinimas paketais ir gali būti 0–8) (vėlinimas šiame kontekste yra PESQ matavimo lango poslinkis prarasto paketo atžvilgiu), naudodami diferencinę PESQ metriką: $D_{iv} = 4{,}549 - PESQ(x_{iv}, y_{iv})$, čia x_{iv} yra signalo x_i dalis nuo $v \cdot 20$ ms iki $v \cdot 20 + 500$ ms. Analogiškai su y_{iv}.

4. Kartokime punktus 1, 2, 3 naudodami $i = 0, 1, 2, \ldots, 2999$. Naudodami gautą vektorių (skaičių rinkinį) D_{iv} brėžkime histogramas skirtingiems v. Taip pat paskaičiuojame girdimųjų iškraipymų vidurkius skirtingiems vėlinimams v.

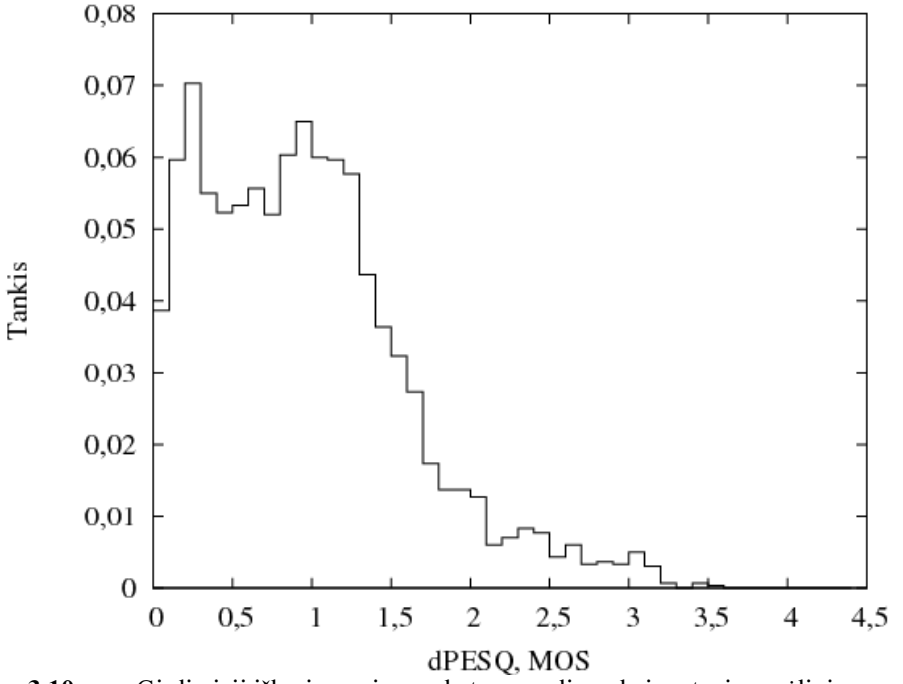
Eksperimento vykdymo algoritmą galėtų padėti paaiškinti 3.9 paveiksle. Čia pavaizduoti du skirtingi matavimo langai, atitinkantys $v = 1$ ir $v = 2$ matavimo lango poslinkius (vėlinimus). Tokiu būdu antrasis pavaizduotas matavimo langas neapima klaidos įtakoto signalo $i + 1$-ajame pakete.

Eksperimento rezultatas parodo, kokie yra statistiškai liekamieji iškraipymai, atsirandantys dėl vieno paketo praradimo. Aišku, kad šie iškraipymai bus tuo mažesni, kuo toliau nuo klaidos juos analizuosime. Tuo atveju, kai matavimo langas yra paslinktas taip toli, kad gaunamas kokybės iškraipymas lygus nuliui, sakome, kad paketo praradimas nebeturi įtakos signalui.

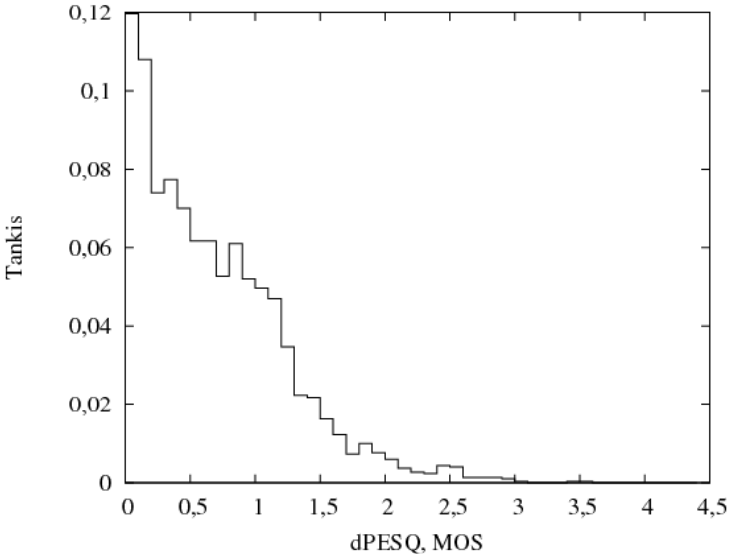


3.9 pav. Girdimųjų iškraipymų matavimo schema esant prarastam paketui
Fig. 3.9. Diagram of perceptual distortions measurement when one frame is lost

3.10–3.13 paveiksluose pateiktos girdimųjų iškraipymų histogramos, kai matavimo vėlinimas *v* po paketo praradimo yra 0, 1, 2, 3 paketai. Iš paveikslų matyti, kad kai praeina 60 ms nuo prarasto paketo pabaigos ($v = 3$), tai liekamieji girdimieji iškraipymai jau yra gerokai sumažėję lyginant su pradine histograma, kai vėlinimo nėra (3.10 paveikslas).

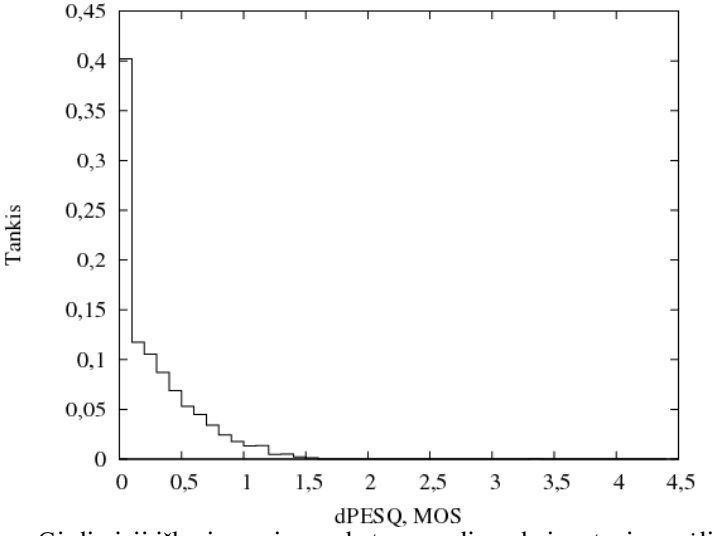


3.10 pav. Girdimieji iškraipymai po paketo praradimo, kai matavimo vėlinimas 0 paketų
Fig. 3.10. Perceptual distortion histogram when measurement delay is 0 frames



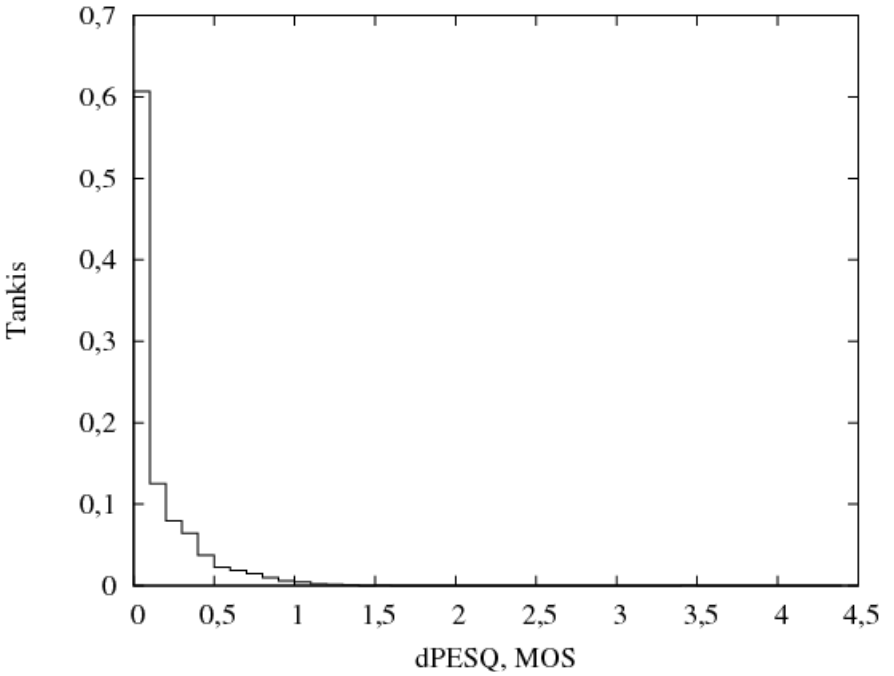
3.11 pav. Girdimieji iškraipymai po paketo praradimo, kai matavimo vėlinimas 1 paketas

Fig. 3.11. Perceptual distortion histogram when measurement delay is 1 frame



3.12 pav. Girdimieji iškraipymai po paketo praradimo, kai matavimo vėlinimas 2 paketai

Fig. 3.12. Perceptual distortion histogram when measurement delay is 2 frames



3.13 pav. Girdimieji iškraipymai po paketo praradimo, kai matavimo vėlinimas 3 paketai
Fig. 3.13. Perceptual distortion histogram when measurement delay is 3 frames

3.1 lentelėje pateiksime girdimųjų iškraipymų vidurkius skirtingiems matavimo vėlinimams. Matome, kad reali vieno prarasto paketo įtaka tęsiasi apie 100 ms, t. y., 5 AMR paketus. Po 100 ms girdimieji iškraipymai siekia mažiau, negu 10 % girdimųjų iškraipymų, kai vėlinimo nėra.

3.1 lentelė. Girdimųjų iškraipymų vidurkiai skirtingiems vėlinimams
Table 3.1. Averages of perceptual distortions for different delays

Vėlinimas	0	1	2	3	4	5	6	7	8
Iškraipymas	0,98	0,7	0,31	0,17	0,1	0,07	0,05	0,04	0,04

3.3.2. LPC iškraipymai

Spektrinių iškraipymų kitimas laike po prarasto paketo bus skaičiuojamas taip. Po to, kai prarandamas *i*-asis paketas, matuojame LPC iškraipymus *i*-ajame, $i+1$ – ajame , $i+2$-ajame ir t.t. paketuose. Taip gauname vieno konkretaus (*i*-

ojo) paketo praradimo sukuriamo iškraipymo kitimą laike. Vidurkindami gautus iškraipymus skirtingiems *i*, gauname šnekos koderio charakteristiką – vidutinį iškraipymų kitimą laike. Ši charakteristika gana aiškiai nurodys, kurioje vietoje didžiausi iškraipymai, kiek laiko jie trunka po paketo praradimo ir kaip gęsta.

Čia panagrinėsime dar viena efektą. Jei prarandamas ne vienas paketas, bet keletas, tai dekoderis smarkiau išsinchronizuojamas. Taigi panagrinėsime, kokie yra ir keleto prarastų paketų sukuriami iškraipymai ir koks jų kitimas laike.

Simuliavimui naudojome 3.000 skirtingų signalų. Kiekviename signale buvo įvedama viena ir dvi paketinės klaidos ir paskaičiuojami LPC iškraipymai pakete su įvesta klaida (bei dviem klaidom) ir tolimesniuose paketuose. Galutinis rezultatas yra šių iškraipymų vidurkis.

1. Iš viso pradinio signalo, kuris yra 3.050 paketų ilgio, pasirenkame signalo dalį tarp *i*-tojo ir *i* + 25-ojo paketų. Tai atitinka laiko momentus signale signale *i*·0,02 s ir (*i* + 25)·0,02 s. Pažymėkime taip pasirinktą signalą z_i.

2. Matavimo signalą z_i koduojame ir dekoduojame – rezultate gaunamą signalą pavadinkime x_i. Kitas signalas gaunamas prieš dekodavimą imituojant klaidos atsiradimą nuliniame baziniame signalo z_i segmente. Šį signalą pavadinkime y_i.

3. Paskaičiuokime gaunamus LPC iškraipymus, atitinkančius *i*-tąjį paketą ir vėlinimą *v* (kur *v* gali būti 0–8), naudodami tokią formulę:

$$D_{iv} = \sqrt{\sum_{j=1}^{10} \left(\left(LPC(x_{iv})\right)[j] - \left(LPC(y_{iv})\right)[j] \right)^2} , \qquad (3.8)$$

čia x_{iv} yra signalo x_i dalis nuo laiko momento *v*·20 ms iki (*v* + 1)·20 ms. LPC – LPC koeficientų skaičiavimo operatorius, *j* formulėje žymi gaunamą *j*-ąjį LPC koeficientą.

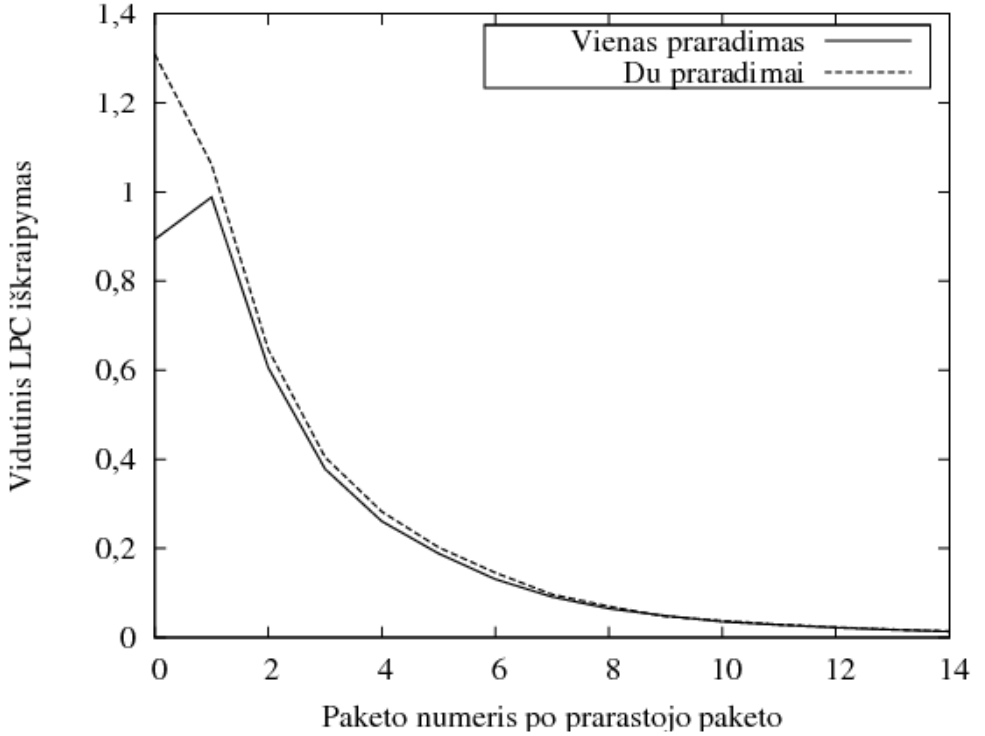
4. Kartokime punktus 1, 2, 3 naudodami *i* = 0, 1, 2,…, 2999. Naudodami gautą skaičių rinkinį D_{iv} paskaičiuokime vidutinį LPC iškraipymo kitimą laike vidurkindami D_{iv} pagal *i*.

3.14 paveikslas rodo, kaip kinta vidurkintas LPC iškraipymas laike, kai prarandamas vienas arba du paketai paeiliui. Pastebėkime, kad vieno paketo praradimo atveju signalas labiausiai sugadinamas ne prarastajame pakete, o viename vėliau. Taip yra todėl, kad prarastasis paketas dažniausiai nemažai koreliuotas su prieš tai buvusiu, kuo pasinaudodamas paketų pakeitimo algoritmas neblogai atstato signalą. Pirmasis gi po prarasto paketas mažai koreliuotas su paskutiniu geru paketu, o dekoderio resinchronizacija dar irgi nespėja gerai suveikti. Dar toliau jau pradeda veikti resinchronizacija ir LPC iškraipymas pradeda greitai mažėti. Nagrinėjant LPC vidutinį iškraipymą sakytume, kad dekoderis resinchronizuojasi po 8–10 paketų nuo prarastojo. Tai atitinka apie 160–200 ms. Atkreipkime dėmesį, kad šis skaičius skiriasi nuo to,

kuris gautas naudojant girdimųjų iškraipymų metriką. Čia ir vėl veikia tas faktorius, kad skaičiuojant girdimuosius iškraipymus, spektrinių iškraipymų vertės yra keliamos šeštuoju laipsniu.

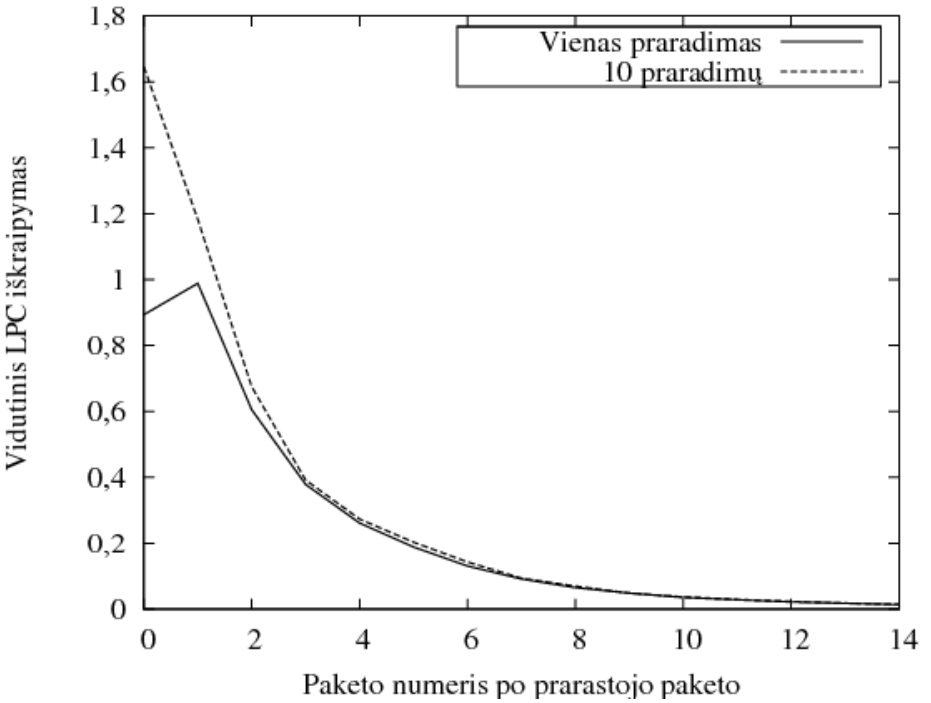
Du paeiliui prarasti paketai aiškų dekoderio sinchronizaciją pablogina labiau, nei vienas. Ir tai atsispindi paveiksle. Tačiau skirtumas nuo pirmojo paketo yra mažesnis, negu 5 %.

3.15 paveikslas rodo ribinį atvejį, kai prarandama 1 ir 10 paketų paeiliui ir tada skaičiuojami vidutiniai LPC iškraipymai. Atkreipkime dėmesį, kad 10 paeiliui prarastų paketų panaikina bet kokią dekoderių sinchronizaciją. Tačiau galutinis rezultatas yra vos daugiau įtakojamas, negu kai prarandami du paketai paeiliui. Iš čia galime daryti išvadą, kad praktiškai net ir vienas prarastas paketas pilnai išsinchronizuoja dekoderius. Sinchronizacijos atsistatymo laikas nepriklauso nuo to, kiek paketų buvo prarasta.



3.14 pav. Vidutinių LPC iškraipymų kitimas laike po vieno ir dviejų paketų praradimo

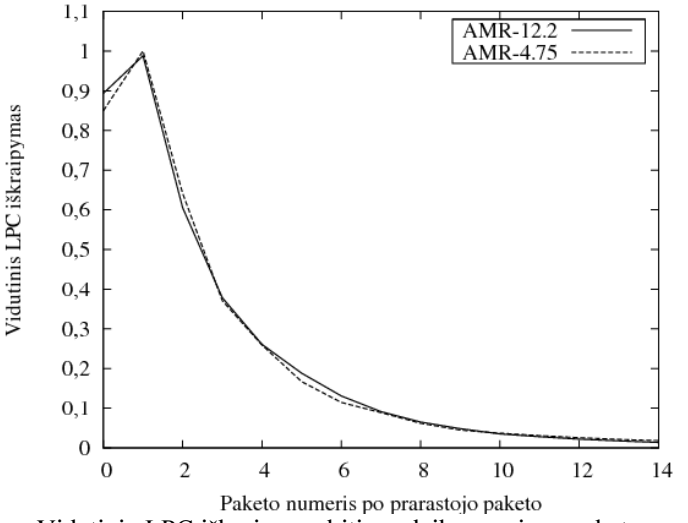
Fig. 3.14. Change of LPC distortion over time after one and two frame losses



3.15 pav. Vidutinių LPC iškraipymų kitimas laike po vieno ir dešimties paketų praradimo

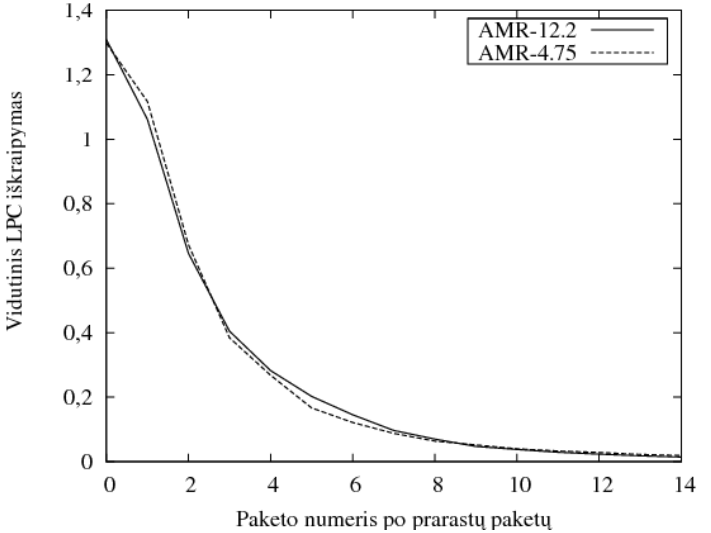
Fig. 3.15. Change of LPC distortion over time after one and ten frame losses

Iki šiol prarasto paketo įtakos kitimo laike tyrimai buvo atlikti naudojant AMR-12.2 koderį. Minėjome, kad prarasto paketo įtaka nėra labai skirtinga, jei naudojami kiti koderiai. Atlikime eksperimentą, kuris pateiktų duomenis apie tai, kaip skirtingi koderiai įtakoja signalo kitimą po paketo praradimo. Eksperimentas bus toks, kaip ir paskutinysis – tirsime, kaip kinta LPC iškraipymai laike, kai naudojami AMR-12.2 ir AMR-4.75 koderiai. Kai imituojamas tik viena paketinė klaida, tai gaunami rezultatai parodyti 3.16 paveiksle. Dviejų paketų praradimo atveju gaunamas rezultatas parodytas 3.17 paveiksle. Iš pateiktų grafikų tegalima daryti vienintelę logišką išvadą – AMR klaidos įtakos kitimas laike praktiškai nepriklauso nuo to, kokios spartos koderis naudojamas. Tai nėra akivaizdus dalykas, bet gan tikėtinas. Reikalas čia tame, kad žemesnės spartos koderiai turi žemesnę bazinę kokybę – tuo atveju, kai klaidų nėra, jie labiau iškraipo signalą. O signalo iškraipymas po klaidos yra panašus bet kurio koderio atveju, nes klaidos taisymo algoritmai naudojami tie patys nepriklausomai nuo koderio.



3.16 pav. Vidutinių LPC iškraipymų kitimas laike po vieno paketo praradimo skirtingiems AMR koderiams

Fig. 3.16. Change of LPC distortion over time after one frame loss for different AMR codecs



3.17 pav. Vidutinių LPC iškraipymų kitimas laike po dviejų paketų praradimo skirtingiems AMR koderiams

Fig. 3.17. Change of LPC distortion over time after two frame loss for different AMR codecs

3.4. Trečio skyriaus išvados

- Matuojant trumpų signalų kokybę negalima formuluoti išvados „koderis X užtikrina kokybę Y“. Taip yra todėl, kad gaunama kokybė priklauso nuo konkretaus signalo, kurį naudojant kokybė vertinama.
- Paketų praradimo atveju realių koderių palaikomą balso kokybę galima charakterizuoti kokybės skirstiniais. Empirinis skirstinys dažnai turi gana platų reikšmių intervalą – AMR-12.2 koderio bazinė kokybė (dPESQ) kinta nuo 0,1 iki 2,5 MOS balų.
- Net nežymūs spektriniai iškraipymai sukuria didelius girdimus iškraipymus. Skirtingos spartos AMR koderiai sukuria mažai besiskiriančius spektrinius iškraipymus, bet girdimųjų (įvertinamų PESQ) iškraipymų skirtumai yra dideli.
- Prarasto paketo įtaka naudojant AMR koderį tęsiasi iki 100 ms. Tai svarbus rezultatas paketo vertės nustatymo algoritmui. Norint nustatyti paketo vertę reikia atsižvelgti į visą paketo praradimo sukeliamą įtaką. Tai reiškia, kad realiu laiku paketo vertės naudojant $PESQ_{mod}$ nustatyti neįmanoma.
- AMR koderių paketų pakeitimo algoritmo ypatumai pasireiškia tuo, kad praradus duomenų paketą, didžiausi iškraipymai stebimi ne prarastajame, bet po jo sekančiame segmente.

4

Kitų paketo vertės nustatymo priemonių paieška

Pagal pradinį apibrėžimą balso paketo vertė nustatoma taikant $PESQ_{mod}$. Šio algoritmo esminis trūkumas jo sudėtingumas skaičiavimų prasme. Šiame skyriuje analizuojama galimybė nustatyti balso paketo vertę, taikant mažesnio kompleksiškumo būdus, pagrįstus LPC ir spektrinių požymių vektoriais bei reikšmių prognozavimo algoritmais. Tiriamas paprastesnių balso paketo vertės nustatymo algoritmų tikslumas**.**

4.1. Paketo vertės priklausomybė nuo matavimo lango padėties

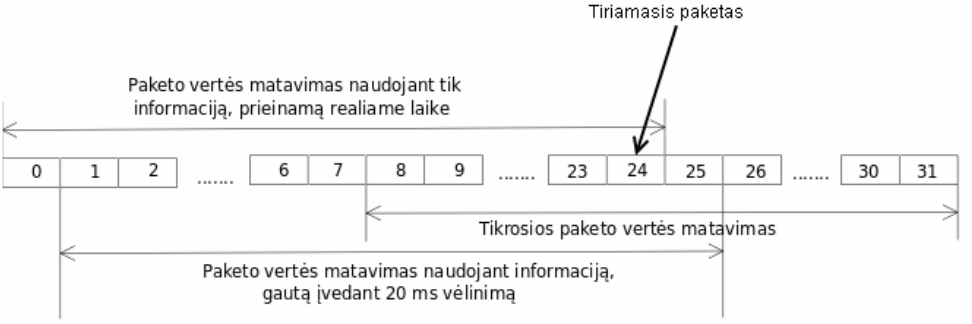
$PESQ_{mod}$ algoritmas paketo vertę nustato analizuodamas palyginti didelės (lyginant su vienu paketu) trukmės balso signalo atkarpą, todėl tikslios paketo vertės gavimas yra neekonomiškas skaičiavimų apimties atžvilgiu.

Paketo vertei nustatyti naudojamas signalas, laike apimantis tiriamąjį paketą. Tai reiškia, kad realiu laiku paketo vertės nustatyti naudojant 2 skyriuje apibrėžtus algoritmus neįmanoma. O toks paketo vertės nustatymas gali būti

naudingas realiose situacijose. Tarkim, pažymint paketų prioritetus prieš juos siunčiant (Prastholm *et al*. 2006; Pavilanskas 2007).

Paketo vertės nustatymo schemą reikia pakeisti taip, kad ji veiktų realiu laiku. Vienas iš būdų – matuoti balso kokybę naudojant $PESQ_{mod}$ algoritmą nuo tam tikro laikotarpio praeityje iki paskutinio koduojamo balso paketo pabaigos. Tokiu būdu būtų gautas paketo vertės įvertis, kuris, aišku, nebus lygus tiksliai paketo vertei, nes prarandama signalo iškraipymo informacija, esanti už prarandamojo paketo. Tačiau žinome, kad AMR koderio prarasto paketo įtaka tęsiasi ne itin ilgai, todėl galime tikėtis, kad realiu laiku gaunamas įvertis bus gera tikrosios paketo vertės prognozė. Atitinkamai, jei leisime paketo išsiuntimą vėlinti vienu paketu (20 ms) tai gausime tikslesnį išsiunčiamo paketo vertės įvertį.

Tikrosios paketo vertės ir paketo vertės alternatyvių nustatymo metodų tyrimų schema pavaizduota 4.1 paveiksle. Schemoje parodyta, kad matuojama 24-o paketo vertė, taip pat skaičiuojamos šios vertės aproksimacijos realiame laike ir įvedus 20 ms vėlinimą. Tikroji paketo vertė matuojama imituojant paketo praradimą ir skaičiuojant $PESQ_{mod}$ įvertį nuo 8-o iki 32-o paketų, t. y., matavimo langas pilnai apima šį paketą ir visus jo praradimo sukeliamus efektus. Realaus laiko paketo vertės įvertis gaunamas matuojant nuo nulinio iki 24-o paketo pabaigos. Aišku, kad tokiu būdu gaunant duomenis neatsižvelgiama į tai, kad paketo praradimas turėjo įtakos ir tolimesniems paketams (signalui). Paketo vertės įvertis gaunamas tuo tikslesnis, kuo didesnis vėlinimas įvedamas. Matysime, kad jau vėlinant 60 ms, koreliacijos koeficientas tarp tikrosios paketo vertės ir jos įverčio, yra didesnis, negu 0,99.



4.1 pav. Paketo vertės įverčių skaičiavimo laikinė diagrama
Fig. 4.2. Time diagram of frame calculation of value estimation

Balso paketo informacinės vertės įvertį, gautą naudojant aptartą schemą, laikysime viršutiniu paketo vertės prognozavimo tikslumo rėžiu. Kitaip sakant, postuluojame, kad įverčio skaičiavimas naudojant $PESQ_{mod}$ matavimą ir signalą

iki esamojo momento, statistiškai yra tikslesnis, negu bet kuris kitas paketo vertės įverčio skaičiavimo metodas.

Toliau aprašysime eksperimentus, kurių pagalba gauti įverčių skaičiavimo tikslumo viršutiniai rėžiai esant įvairiems vėlinimams. Prognozės tikslumą čia ir vėliau vertinsime naudodami koreliacijos koeficientą – po to, kai paketų aibei paskaičiuojamos prognozuojamos tų paketų vertės, skaičiuojame šių reikšmių koreliacijos koeficientą su tikrosiomis paketų vertėmis. Gautasis koreliacijos koeficientas ir nurodys prognozės tikslumą.

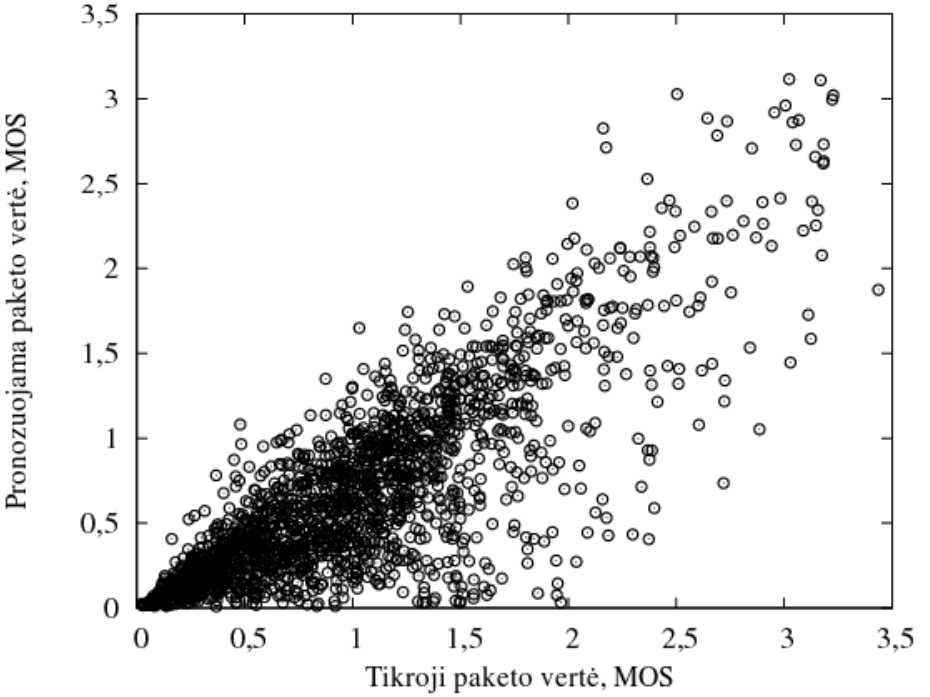
Eksperimentas viršutiniams prognozės tikslumo rėžiams nustatyti gan paprastas. Imame 2.000 signalų, kiekviename jų imituojame paketo klaidą ir skaičiuojame tikrąsias paketų vertes bei apytiksles paketų vertes su vėlinimu, lygiu 0, 20, 40 ms. Koreliacijos koeficientai tarp tikrųjų paketų verčių ir apytikslių, yra viršutiniai prognozės tikslumo rėžiai. Detaliai eksperimento vykdymas aprašytas žemiau:

- Iš ilgos šnekos signalo atkarpos paimama 2.000 nuosekliai einančių šnekos paketų;
- Išmatuojama kiekvieno paketo vertė taikant 2 skyriuje aprašytą algoritmą. Gautą reikšmių vektorių pažymime X, kur vektoriaus elementai žymimi X_i, $i = 1, 2,\ldots, 2000$, o indeksas i atitinkama tiriamo paketo numerį.
- Išmatuokime kiekvieno paketo vertę naudodami PESQ_{mod} algoritmą ir signalą nuo τ ms iki $\tau + 480$ ms (laikome, kad kiekvieno signalo, apimančio tiriamąjį paketą, laiko ašis prasideda nuo nulio). Gautą duomenų vektorių pažymėkime $Y(\tau)$, atskiri vektoriaus $Y(\tau)$elementai žymimi $Y(\tau)_i$ ir reiškia i-tojo paketo vertės prognozę, kai naudojamas vėlinimas τ lyginant su realaus laiko matavimu. Parametras τ reiškia vėlinimą ir gali būti keičiamas. Realiai mus dominančios τ vertės yra 0, 20, 40 ms.
- Paskaičiuokime koreliacijos koeficientus tarp X ir $Y(\tau)$ naudodami standartinę formulę

$$c(\tau) = \frac{\sum_{i=1}^{2000} (X_i - \overline{X}) \cdot (Y(\tau)_i - \overline{Y}(\tau))}{\sqrt{\sum_{i=1}^{2000} (X_i - \overline{X})^2} \cdot \sqrt{\sum_{i=1}^{2000} (Y(\tau)_i - \overline{Y}(\tau))^2}}, \qquad (4.1)$$

čia brūkšnelis virš dydžio žymi aritmetinį vidurkį. Taip pat nubrėžkime gautų duomenų išsibarstymo grafikus. Gaunamas koreliacijos koeficientas rodo, kiek tiksliai paketo vertę galima prognozuoti naudojant vėlinimą τ.

Kai paketo vertę nustatinėjame realiu laiku ($\tau = 0$) naudodami aprašytą metodą, tai gaunami duomenys pavaizduoti 4.2 paveiksle. Gaunamas koreliacijos koeficientas šiuo atveju yra 0,822. Čia jau galime pastebėti, kad tikroji paketo vertė yra statistiškai didesnė, negu prognozuotoji. Šis rezultatas buvo tikėtinas, nes matuojant tikrąją paketo vertę į matavimo langą patenka daugiau iškraipymų, o matuojant vertę realiame laike tam tikra iškraipymų dalis yra atmetama, nes tiriamas paketas yra matavimo lango gale ir jo poveikis tolimesniems paketams nepapuola į langą. Regresijos linija, nubrėžta per šiuos duomenis būtų pasvirusi mažesniu, negu 45 ° kampu. Nepaisant šių trūkumų, gaunamas koreliacijos koeficientas yra gana didelis ir toks vertės nustatymo metodas yra tinkamas kai kuriems uždaviniams.



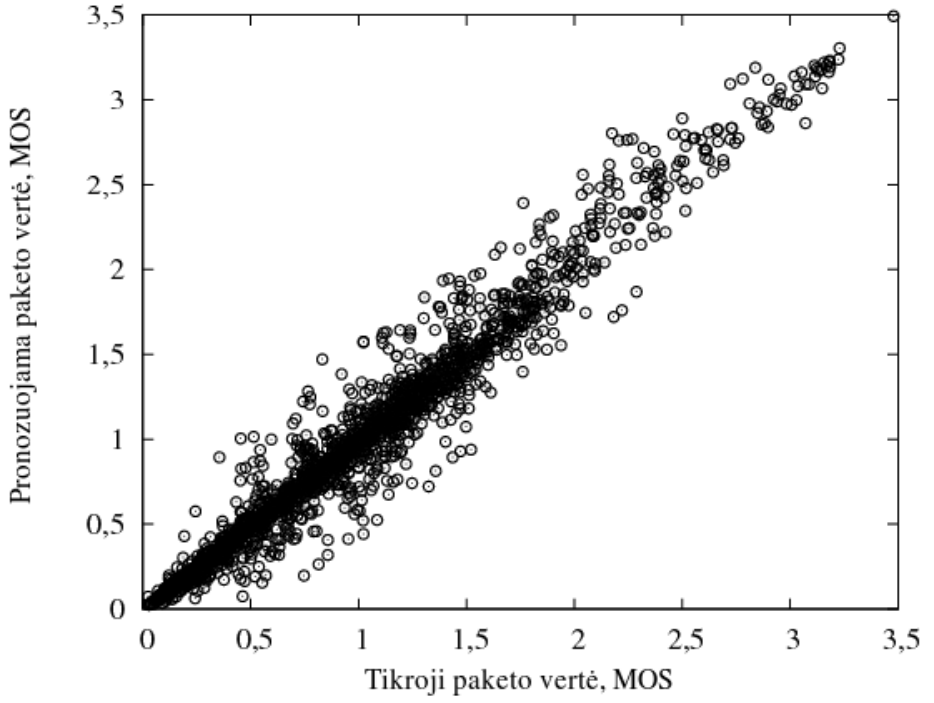
4.2 pav. Paketo vertės „optimalus“ nustatymas realiame laike
Fig. 4.2. Optimal calculation of frame value in real time

Kai paketo vertei apskaičiuoti naudojami 20 ir 40 ms vėlinimai, tai gaunami išsibarstymo grafikai pavaizduoti 4.3 ir 4.4 paveiksluose. Galima sakyti, kad toks vėlinimo įvedimas leidžia tiksliai prognozuoti paketo vertę visiems praktiniams taikymams. Koreliacijos koeficientas tarp tikrosios paketo vertės ir prognozės,

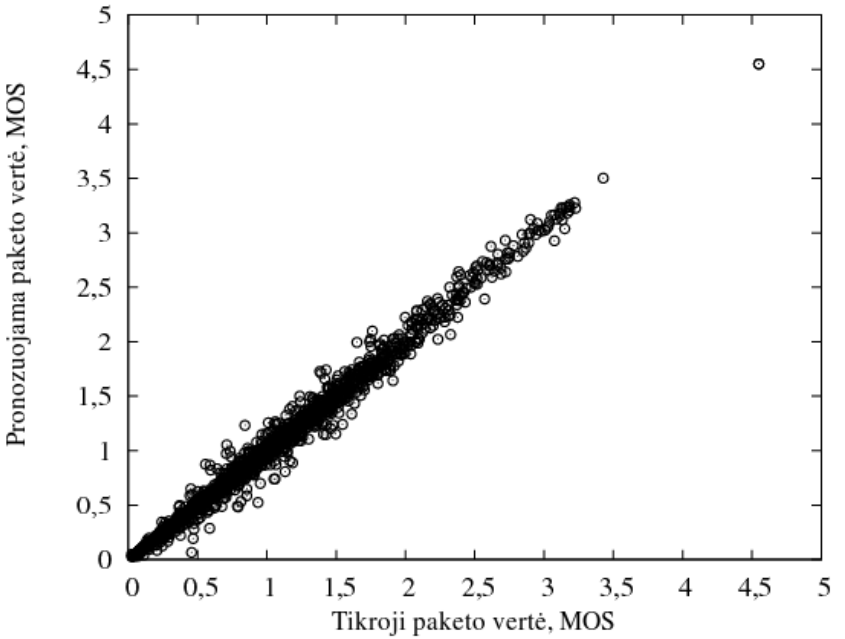
naudojant 20 ms vėlinimą, yra 0,978, o kai vėlinimas 40 ms, tai koreliacijos koeficientas pasiekia 0,994.

Taigi, matome, kad vėlinimas per 40 ms leistų praktiškai tiksliai nustatyti paketo vertę. Tačiau čia yra kelios problemos – toks papildomas vėlinimas (o kur dar vėlinimas dėl kitų skaičiavimų) jau yra didelis realiose sistemose. Kita problema tame, kad skaičiavimai, reikalingi šiam paketo vertės įverčiui gauti, yra daug resursų reikalaujantys ($PESQ_{mod}$ algoritmo taikymas matavimo lange). $PESQ_{mod}$ algoritmas veikia apytiksliai realiu laiku naudojant Pentium IV procesorių.

Tolimesniuose skyriuose paanalizuosime sąlyginai paprastus ir mažai skaičiavimų reikalaujančius algoritmus, kuriuos galime taikyti paketo vertei prognozuoti. Analizuosime, kokį paketo vertės prognozės tikslumą galima pasiekti alternatyviais algoritmais.



4.3 pav. Paketo vertės „optimalus“ nustatymas naudojant 20 ms vėlinimą
Fig. 4.3. Optimal calculation of frame value with 20 ms delay



4.4 pav. Paketo vertės „optimalus“ nustatymas naudojant 40 ms vėlinimą
Fig. 4.4. Optimal calculation of frame value with 40 ms delay

4.2. Paketo vertės nustatymas pagal tiesinės prognozės koeficientus

PESQ algoritmo neinvazinių alternatyvų paieškos uždavinys nėra naujas (Chen 2006), tačiau daugelis kuriamų algoritmų yra labai bendri neinvaziniai algoritmai (Raja 2008; Conway 2004). Mes pasiūlysime specifiškesnius metodus – jais siekiama įvertinti perduodamo balso kokybę su sąlyga, kad konkretus paketas yra prarastas. Kitaip sakant, keliamas klausimas „kokia bus balso kokybė, jei šis paketas bus pamestas“? Mes nebandome sukurti bendro neinvazinio kokybės vertinimo algoritmo. Todėl čia aprašomi algoritmai nėra pertekliniai.

Poskyryje nagrinėsime paketo vertės prognozavimo uždavinį naudojant tiesinės prognozės požymių vektorių kombinaciją. Taip pat čia apibrėšime ir visą tyrimų schemą, kuri teisinga ir vėliau nagrinėjamiems netiesiniams prognozavimo metodams.

Taigi, paketo vertės prognozė realiu laiku bus atliekama naudojant balso paketų požymių vektorius iš dviejų paketų – iš to, kurio vertę bandome nustatyti

ir iš prieš tai buvusio. Antras variantas, kurį nagrinėsime, bus paketo vertės prognozavimas, kai įvedamas vieno paketo (20 ms) vėlinimas. Šiuo atveju naudosime dar ir požymių vektorių iš suvėlintojo paketo (iš viso bus naudojami požymių vektoriai iš trijų paketų). Abiem atvejais požymių vektoriai iš skirtingų paketų bus sudedami į vieną vektorių (konkatenuojami) ir naudojami kaip vienas didelis požymių vektorius.

Eksperimentuose naudosime dviejų tipų požymių vektorius – 1-ame skyriuje aptartus spektrinius ir LPC koeficientus. Spektriniai koeficientai dar yra paelemenčiui padauginti iš tokio vektoriaus $w = [11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1]$. Kitaip tariant, po to kai paskaičiuojame energijas vienodo 400 Hz pločio juostose, tai jas dauginame iš nurodyto vektoriaus w elementų. Šis dauginimas atliekamas tam, kad žemesnio dažnio juostos būtų laikomos vertingesnėmis ir tai atitinka auditorinį klausos modelį (Allen 1995; Paulikas 1999). Tai reiškia, kad suminė paketo energija dauginama iš 11, energija juostoje nuo 0 iki 400 Hz dauginama iš 10, energija juostoje nuo 400 iki 800 Hz dauginama iš 9 ir t. t. Tokiu būdu žemesnio dažnio juostos yra labiau įvertinamos.

Įprasta, kad bandant sudaryti duomenų modelį (o paketo vertės prognozė iš bruožų vektorių ir yra toks modelis) turime 2 duomenų rinkinius – duomenis mokymuisi ir testavimui (Russel 2002). Duomenys mokymuisi yra tie, iš kurių sudaromas modelis, atliekama optimizacija ieškant modelio parametrų ir panašiai. Naudojant testavimo duomenis tikrinamas gaunamo modelio tikslumas apibendrintose sąlygose.

Toliau visur kaip duomenys mokymui ir testavimui bus naudojama informacija apie 2.000 balso paketų iš skirtingų angliškų įrašų (CNN žinios).

Tiesinę paketo vertės prognozę atliksime naudodami tokį eksperimentą (aprašymas tik realaus laiko prognozei, bet analogiškai ir tuo atveju, kai įvedamas vėlinimas, tik papildomai būtų pridedamas vėlinto paketo bruožų vektorius):

- Skaičiuojame tikrąsias paketų vertes mokymo ir testavimo duomenims. Pažymėkime jas $Q_{\text{tren}}(i)$ ir $Q_{\text{test}}(i)$, čia $i = 1,2,\ldots, 2000$. Paskaičiuokime kiekvienam nagrinėjamam paketui bruožų vektorių, kuris susideda iš LPC arba spektrinių koeficientų iš esamo ir prieš tai buvusio paketų. Šiuos „sudėtinius“ požymių vektorius pažymėkime $V_{\text{tren}}(i)$ ir $V_{\text{test}}(i)$. Kiekvienas iš $V(i)$ yra vektorius iš 22 elementų (šalia sudėti LPC ar spektriniai koeficientai, gauti iš esamojo ir prieš tai buvusių paketų).
- Paskaičiuokime tiesinės prognozės koeficientus a_j minimizuodami tokį reiškinį:

$$\sum_{n=1}^{2000}\left(Q_{\text{tren}}(n) - \sum_{j=1}^{22} a_j \cdot V_{\text{tren}}(n)_j\right)^2. \qquad (4.2)$$

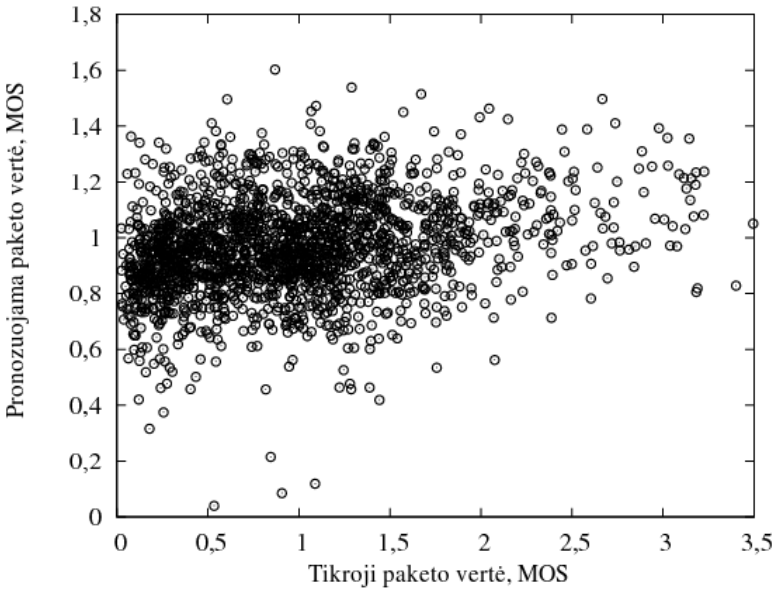
Minimizavimas atliekamas naudojant standartinį mažiausių kvadratų metodą (Russel 2002).

- Paskaičiuokime gaunamas paketų verčių prognozes naudojant koeficientus a_j mokymo ir testavimo duomenims:

$$\hat{Q}_{\text{tren}}(n) = \sum_{j=1}^{22} a_j \cdot V_{\text{tren}}(n)_j, \quad \hat{Q}_{\text{test}}(n) = \sum_{j=1}^{22} a_j \cdot V_{\text{test}}(n)_j . \qquad (4.3)$$

- Paskaičiuokime gaunamus koreliacijos koeficientus ir nubrėžkime išsibarstymo grafikus naudodami tikrąsias ir prognozuotas paketų vertes.

Kai naudojami LPC koeficientai realiame laike ir tiesinė paketo vertės prognozė, tai gauname tokį išsibarstymo grafiką mokymo duomenims, kaip parodyta 4.5 paveiksle. Iš grafiko matome, kad tiesinė prognozė iš nagrinėjamo ir prieš tai buvusio paketų LPC koeficientų yra mažai patikima. Gaunamas koreliacijos koeficientas su tikrąja paketo verte yra 0,27. Akivaizdu, kad tiesinė prognozė „nesuvokia“ duomenų kompleksiškumo ir atlieka iš esmės vidurkinimą. Kitaip sakant, algoritmo išduodama reikšmė yra artima paketo vertės vidurkiui. Tai gana tipiškas tiesinės reikšmių prognozės elgesys, kai įėjimo duomenys turi ganėtinai sudėtingą pobūdį.



4.5 pav. Paketo vertės nustatymas naudojant LPC koeficientus ir tiesinę prognozę realiame laike

Fig. 4.5. Frame value calculation using LPC coefficients and linear prognosis

Koreliacijos koeficientas prognozei iš LPC koeficientų realiame laike mokymo duomenims yra 0,27, testavimo duomenis – 0,24. Taigi, nors pati prognozė nėra tiksli, bet ji gana stabili.

Kai įvedamas vėlinimas, tai gaunami koreliacijos koeficientai 0,31 ir 0,24 atitinkamai mokymo ir testavimo duomenims. Taigi, ryškaus pagerėjimo nėra. Šiuo atveju koreliacijos koeficientas testavimo duomenims pablogėja daugiau. Akivaizdu, čia pradeda reikštis problemos dimensiškumas (kai naudojam vėlinimą, tai yra 33 optimizuojami koeficientai vietoj 22). Apibendrintai galima teigti, kad tiesinė paketo vertės prognozė iš LPC koeficientų duoda prastus rezultatus.

Spektrinių koeficientų naudojimas pasirodė pranašesnis tiesinės prognozes atveju. Kai nėra vėlinimo, tai gaunami koreliacijos koeficientai 0,47 ir 0,46 atitinkamai mokymo ir testavimo duomenims. Prisiminkime, kad naudojant $PESQ_{mod}$ paketo vertės prognozei realiame laike, didžiausias gaunamas koreliacijos koeficientas su tikrosiomis paketų vertėmis yra 0,82. Taigi spektriniai koeficientai žymiai efektyviau leidžia prognozuoti paketo vertę, bet visgi rezultatai dar nėra artimi „optimaliai“ prognozei. Dar vienas teigiamas šių koeficientų bruožas – puikus duomenų apibendrinimas (angl. *generalization*).

Kai įvedamas vieno paketo vėlinimas, paketo vertės nustatymas iš spektrinių koeficientų dar pagerėja. Tokiu atveju koreliacijos koeficientai mokymo ir testavimo duomenims yra 0,53 ir 0,51. Taigi, tai jau visai neblogi rezultatai, tačiau dar gan tolimi idealiems. Pakartotinai matome, kad algoritmo apibendrinimas yra žymiai geresnė naudojant spektrinius koeficientus, negu naudojant LPC.

4.3. Netiesinis paketo vertės prognozavimas

Ankstesniame skyrelyje matėme, kad tiesinis paketo vertės prognozavimas, nors ir duoda pozityvius rezultatus, bet akivaizdūs ir metodo trūkumai. Algoritmas „nesupranta“ duomenų sudėtingumo ir pasirenka vienintelę įmanomą strategiją – vidurkinimą. Todėl akivaizdu, kad siekiant geresnių prognozavimo rezultatų reikia kažkur įvesti netiesiškumą. Tam padaryti yra daug metodų. Vieni iš populiariausių yra dirbtiniai neuronų tinklai (Navakauskas 1999). Tačiau mes pasirinkome neparametrinę prognozę naudojant *k*-artimiausių kaimynų algoritmą (Russel 2002). Bendrinis algoritmo pavadinimas yra „k-artimiausių kaimynų algoritmas“. Trumpai apžvelgsime šį algoritmą.

Bet kokia prognozės technika yra modelis, bandantis nustatyti priklausomybę tarp priklausomų (mūsų atveju tai tikrosios paketų vertės) ir nepriklausomų (požymių vektoriai) kintamųjų. Pažymėkime nepriklausomus kintamuosius X_i. Kiekvienas šis kintamasis yra daugiamatis vektorius.

Priklausomus kintamuosius pažymėkime y_i. Bendru atveju tai taip pat vektoriai, tačiau dažniausiai tai būna skaliarai – taip ir laikykime, nes ir mūsų nagrinėjamu atveju taip yra – paketų vertės aišku yra skaliarai. Bet kuri regresijos technika daro prielaidą, kad yra tam tikra deterministinė ar tikimybinė (kartais priklausomi kintamieji y_i būna su triukšmu) funkcija F tokia, kad $F(X_i) = y_i$ visiems i. Regresijos tikslas yra, gavus bet kokį nepriklausomą kintamąjį X, nustatyti, koks priklausomas kintamasis jį atitinka. Kitaip sakant, iš turimų duomenų X_i ir y_i siekiama sukonstruoti funkcijos F artinį.

- Gavus nepriklausomą kintamąjį X randame k artimiausių nepriklausomų kintamųjų iš aibės X_i, $i = 1, 2,\ldots, N$, kur N yra nepriklausomų kintamųjų skaičius. Artumas nustatomas naudojant Euklido metriką:

$$d_i = \sqrt{\sum_{j=1}^{M} \left(X(j) - X_i(j)\right)^2} \,. \tag{4.4}$$

  Čia M yra nepriklausomų kintamųjų vektoriaus ilgis.
  Gautų k artimiausių nepriklausomų kintamųjų indeksų aibę pažymėkime I.
- Priklausomo kintamojo reikšmė, atitinkanti nepriklausomąjį kintamąjį X nustatoma taip: $y = \frac{1}{k}\sum_{I} y_i$.

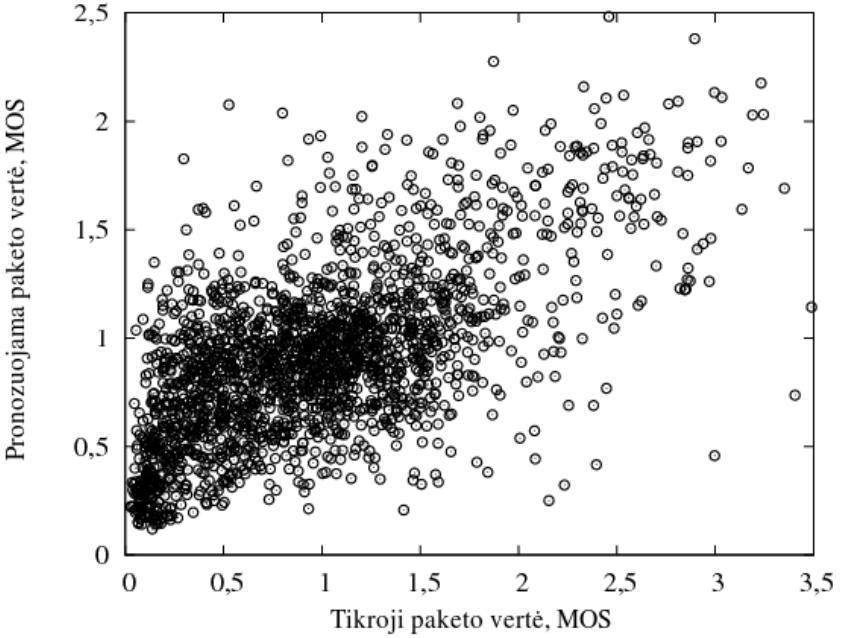
Trumpai tariant, k-artimiausių kaimynų regresija išrenka k arčiausiai esančių (pagal d) nepriklausomų kintamųjų ir grąžina nepriklausomus kintamuosius atitinkančių priklausomų kintamųjų vidurkį. Kadangi algoritmas yra neparametrinis, tai jo veikimą galima aprašyti tokia formule:

$k = K(\boldsymbol{X}, \boldsymbol{Y}, X)$, kur K žymi algoritmo taikymo operatorių; $\boldsymbol{X}$ – sukauptus nepriklausomus kintamuosius (jie gauti iš mokymo duomenų); $\boldsymbol{Y}$ – sukauptus priklausomus kintamuosius (taip pat gauti iš mokymo duomenų);X – analizuojamas nepriklausomas kintamasis, kuriam norime rasti atitinkamą priklausomą kintamąjį.

Aprašysime eksperimentus, naudojant k-artimiausių kaimynų algoritmą paketo vertei prognozuoti. Iš esmės, šie eksperimentai identiški tiesinės prognozės eksperimentui, tik vietoj tiesinės prognozės naudojamas k-artimiausių kaimynų algoritmas. Tyrimams visur naudosime $k = 10$, kadangi empiriškai tai pasirodė geriausias atvejis iš bandytų. Taigi eksperimentas visiškai identiškas tiesinės prognozės eksperimentui, tik paketo vertė prognozuojama naudojant k-artimiausių kaimynų algoritmą vietoje tiesinės kombinacijos.

Eksperimento rezultatai, kaip ir anksčiau, rodo, kad LPC koeficientai yra prastesni paketo vertės prognozei, negu spektriniai. Tačiau nepriklausomai nuo naudojamų koeficientų, išsibarstymo grafikai turi panašų pobūdį. Pateiksime tik

vieną išsibarstymo grafiką – paketo vertės prognozės realiame laike naudojant spektrinius koeficientus testavimo duomenims. Grafikas pateiktas 4.6 paveiksle. Čia akivaizdu, kad vertės prognozavimas gerokai tikslesnis. Pastebimas neigiamas bruožas tas, kad regresijos linija visgi nėra lygi 45 ° ir algoritmas nesugeba prognozuoti ekstremalių (didelių ir mažų, artimų nuliui) reikšmių. Taip yra dėl naudojamo vidurkinimo ir fakto, kad ekstremalios reikšmės duomenyse pasitaiko sąlyginai retai. Naudojant LPC koeficientus gaunami rezultatai pateikti 4.1 lentelėje.



4.6 pav. Paketo vertės nustatymas naudojant spektrinius koeficientus ir *k*-artimiausių kaimynų prognozę realiame laike (testavimo duomenims)
Fig. 4.6. Frame value calculation using spectral coefficients and k-nearest neighbours prognosis in real time (testing data)

4.1 lentelė. Paketo vertės prognozė naudojant LPC koeficientus ir *k*-artimiausių kaimynų algoritmą. Koreliacijos koeficientai su tikrosiomis paketų vertėmis
Table 4.1. Frame value prognosis using LPC coefficients and k-nearest neighbours prognosis. Correlations with true frame value

	Koreliacijos koeficientai, kai naudojami tokie duomenys	
Duomenys	Realaus laiko	Su 20 ms vėlinimu
Mokymo	0,602	0,604
Testavimo	0,412	0,406

Iš lentelės matome, kad LPC koeficientai nagrinėjamam uždaviniui nelabai tinka. Nepaisant to, kad mokymo duomenis gaunamos koreliacijos neblogos, bet algoritmo efektyvumas smarkiai krinta testavimo duomenims. Tai rodo, kad LPC koeficientai prastai atspindi girdimąsias šnekos suvokimo savybes.

Žymiai geriau veikia algoritmas, kai naudojami spektriniai koeficientai. Šiuo atveju gaunami rezultatai parodyti 4.2 lentelėje. Matome, kad rezultatas mokymo duomenims ir realaus laiko prognozei jau gana artimas „idealiai“ prognozei. Aišku, testavimo duomenims pastebimas gan ryškus prognozės kokybės sumažėjimas. Rezultatus, kai naudojamas vėlinimas gana sunku paaiškinti, bet manome, kad taip yra dėl problemos dimensiškumo padidėjimo ir su tuo susijusio didesnio triukšmo rezultatuose atsiradimo.

Šiame skyriuje nagrinėti algoritmai moderniame kompiuteryje veikė bent 10 kartų greičiau, negu pradinis paketo vertės nustatymo metodas.

4.2 lentelė. Paketo vertės prognozė naudojant spektrinius koeficientus ir k-artimiausių kaimynų algoritmą. Koreliacijos koeficientai su tikrosiomis paketų vertėmis
Table 4.2. Frame value prognosis using spectral coefficients and k-nearest neighbours prognosis. Correlations with true frame value

	Koreliacijos koeficientai, kai naudojami tokie duomenys	
Duomenys	Realaus laiko	Su 20 ms vėlinimu
Mokymo	0,764	0,722
Testavimo	0,601	0,655

4.4. Ketvirto skyriaus išvados

1. Kai naudojamas $PESQ_{mod}$ algoritmas, balso paketų vertes galima prognozuoti artimu realiam laiku, jei naudojama labai galinga skaičiavimo technika. Geriausias prognozės koreliacijos koeficientas šiuo atveju pasiekiamas 0,8.
2. Taikant tiesinės prognozės būdus statistiškai nepavyksta tiksliai nustatyti paketo vertės.
3. Netiesinė paketo vertės prognozė pagal spektro koeficientus naudojant *k*-artimiausių kaimynų algoritmą gali patenkinamai pakeisti $PESQ_{mod}$ algoritmo taikymą – gaunamas koreliacijos koeficientas didesnis, negu 0,6.
4. Gauti paketų vertės prognozės algoritmai parodo, kad paketų vertę įmanoma prognozuoti realiame laike su koreliacijos koeficientu, didesniu, nei 0,6.

Bendrosios išvados

1. Pasiūlytas specialus tiriamųjų signalų sudarymo būdas, kuris leidžia trumpų balso atkarpų kokybės PESQ būdu vertinimo rezultatų sklaidą sumažinti vidutiniškai 8 kartus.
2. Sukurta balso segmentų vertės matavimo metodika ir matavimo priemonė, kurie sudaro galimybes įvertinti kiekvieno atskiro paketo potencialų indėlį į šnekos kokybę.
3. Sudaryti pavyzdiniai balso paketų vertės skirstiniai, parodantys kiek šnekoje yra vertingų ir kiek mažai svarbių paketų. Nustatyta, kad paketo vertės skirstinys yra asimetriškos formą.
4. Ištirti įvairiais koderiais koduotos šnekos iškraipymų dydžiai. Pastebėta, kad koderio perteikiamos šnekos kokybė labai priklauso nuo paties koduojamo signalo. Kokybės kitimo ribos, kai šneka koduojama AMR-12.2 koderiu, yra nuo 0,5 iki 3 MOS vienetų. Tai reiškia, kad bazinio kodavimo sukuriamo iškraipymo vienu skaičiumi apibūdinti negalima.
5. Nustatyta, kad girdimieji iškraipymai turi daug didesnes standartines deviacijas, nei spektriniai. Įrodyta, kad AMR koderiai paketo praradimo atveju labiausiai iškraipo po prarandamojo einantį balso segmentą.
6. Sukurti algoritmai paketo vertei prognozuoti pritaikant spektrinius požymius. Pasiektas paketo vertės prognozės koreliacijos koeficientas su tikrąja paketo verte yra 0,6. Toks tikslumas leidžia parinkti skirtingą kodavimo būdą skirtingos informacinės vertės balso paketams.

Literatūros sąrašas

3GPP TR 26.975 V5.0.0. 2002. Performance Characterization of the AMR Speech Codec, Technical Report 3rd Generation Partnership Project, Technical Specification Group Services and System Aspects.

3GPP TR 26.976. 2002. AMR-WB Speech Codec Performance Characterization. 3GPP Technical Specification.

3GPP TS 26.071. 2004. Mandatory Speech CODEC Speech Processing Functions. AMR Speech Codec. General Description V6.0.0.

3GPP TS 26.091. 2007. AMR Speech Codec; Error Concealment of Lost Frames V7.0.0.

3GPP TS 26.192. 2007.AMR Wideband Speech Codec; Comfort Noise Aspects. 3GPP Technical Specification.

3GPP TS 26.193.AMR Wideband Speech Codec; Source Controlled Rate Operation.3GPP Technical Specification.

3GPP TS 26.194.AMR Wideband Speech Codec; Voice Activity Detector (VAD). 3GPP Technical Specification.

Alben, L. 1996. Quality of Experience: Defining the Criteria for Effective Interaction Design, *Interactions* 3(3): 11–15.

Allen, J. B. 1995. *Speech and Hearing in Communication*. ASA edition, Acoustical Society of America, New York. 451 p.

ANSI S3.2-1989 (R1999). 1989. Method for Measuring the Intelligibility of Speech over Communications System.

Barnett, P. W. 1999. Overview of Speech Intelligibility, in *Proceedings of the Institute of Acoustics* 21(5): 1-8.

Batkauskas, V. 2006. *Mobiliojo tinklo paslaugų kokybės vertinimo ir gerinimo būdų tyrimas*. Daktaro disertacija. Vilniaus Gedimino technikos universitetas. Vilnius: Technika.

Bharrathsingh, K. 2005. Quality of Experience as an Integral Part of Network Engineering, *Nortel Technical Journal* 1: 31–36 [žiūrėta 2009 gegužės 20 d.]. Prieiga per internetą: <http://www.nortel.com/corporate/news/collateral/ntj1_quality.pdf>.

Brady P. T. 1965. A Technique for Investigating on-off Patterns of Speech, *Bell System Technical Journal* 44(1): 1–22.

Brady P. T. 1968. A Statistical Analysis of on-off Patterns in 16 Conversations, *Bell System Technical Journal* 47(1): 73–91.

Brady P. T. 1969. A Model for Generating on-off Speech Patterns in two Way Conversation, *Bell System Technical Journal* 48(9): 2445–2472.

Chen, G; Parsa, V. 2006. A Bayesian Estimator for Non-intrusive Speech Quality Evaluation in Psychoacoustic Domain, in *Signal Processing and Information Technology*, 438–441.

Chu,W. C. 2003. *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders*. John Wiley & Sons, Inc. 584 p.

Conway, A. E. 2004. Output-Based Method of Applying PESQ to Measure the Perceptual Quality of Framed Speech Signals, in *Wireless Communications and Networking Conference,* 2521–2526.

Corrie, B. *et al*. 2003. Towards Quality of Experience in Advanced Collaborative Environments, in *Third Annual Workshop on Advanced Collaborative Environments*, 1-8.

Cutler, A.; Carter, D. M. 1987. The Predominance of Strong Initial Syllables in English Vocabulary, *Computer Speech and Language* 3(2): 133–142.

De Martin, J. C. 2001. Source-driven Packet Marking for Speech Transmission over Differentiated Services Networks, in *Proceedings of IEEE ICASSP,* 753–756.

Delogu, C.; Paoloni, A.; Pocci, P. 1991. New Directions in the Evaluation of Voice Input Output Systems, *Selected Areas in Communications* 9(4): 566–573.

Ding, L.; Goubran, R. 2003. Assessment of Effects of Packet Losson Speech Quality in VoIp, in *Proceedings of Haptic, Audio and Visual Environments and Their Applications*, 49–54.

El-Maleh, K; Kabal, P. 1997. Comparison of Voice Activity Detection Algorithms for Wireless Personal Communications Systems, *in Proeedings of. IEEE Canadian Conference on Electrical and Computer Engineering*, 470–473.

Guršnys, D. 2008. *Balso kokybės vertinimo metodai ir priemonės mobiliojo ryšio sistemoms.* Doctoral dissertation. Vilnius Gediminas technical university. Vilnius: Technika.

Hestnes, B.; Brooks, P.; Heiestad, S.; Ulseth, T.; Aaby., C. 2003. Quality of Experience in Real-Time Person-Person Communication – User based QoS Expressed in Technical Network QoS Terms, in *Proceedings of the 19th International Symposium on Human Factors in Telecommunication*, 3–10.

Holma, H.; Melero, J.; Vainio, J.; Halonen, T.; Makinen, J. 2003. Performance of Adaptive Multirate (AMR) Voice in GSM and WCDMA, in *Proceedings of Vehicular Technology Conference*, 2177–2181.

Itakura F; Saito, S. 1970. A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies, *Electronics and Communications in Japan* 53A: 36–43.

ITU-R Rec. BS-1387. 2001. Method for Objective Measurements of Perceived Audio Quality (PEAQ).

ITU-T Rec. E.800. 1996. Terms and Definitions related to Quality of Service and Network Performance including Dependability.

ITU-T Rec. G711. 1992. Coding of Speech at 16 kbit/s using Low-delay Code Excited Linear Prediction.

ITU-T Rec. G711. 1993. Pulse Code Modulation (PCM) of Voice Frequencies.

ITU-T Rec. P.563. 2004. Single ended Method for Objective Speech Quality Assessment in Narrowband Telephony Applications.

ITU-T Rec. P.800. 1996. Methods for Subjective Determination of Transmission Quality.

ITU-T Rec. P.862. 2001. Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for end-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs.

Jain, R. 2004. Quality of Experience, *Multimedia, IEEE* 9(1): 95–96.

Kajackas, A.; Batkauskas, V.; Medeišis, A. 2004. Individual QoS Rating for Voice Services in Cellular Networks, *IEEE Communications Magazine* 42(6): 88–93.

Kajackas, A.; Medeišis, A. 2000. Apie mobiliojo ryšio paslaugų kokybės rodiklius bei jų kontrolę, iš *Elektronika-2000* konferencijos „Lietuvos mokslas ir pramonė“ medžiagos. Kaunas: Technologija, 2000, 135–138.

Kajackas, A.; Šaltis, A. 2003. Balso signalų kalbos ir tylos intervalų modeliai ir jų taikmas, *Elektronika ir elektrotechnika* 47(5): 7–12.

Kazlauskas, K. 1999. Noisy Speech Intelligibility Enhancement, *Informatica* 10(2): 171–188.

Kulesza, M.; Szwoch, G.; Czyżewski, A. 2006. High Quality Speech Coding using Combined Parametric and Perceptual Modules, in *Proceedings of World Academy of Science, Engineering and Technology*, 244–249.

Lipeika, A. 2005. Formantiniai požymiai atpažįstant kalbą, *Information sciences* 34(1): 215–219.

Lipeika, A.; Lipeikienė, J. 1999. Speaker Recognition based on the use of Vocal Tract and Residue Signal LPC Parameters, *Informatica* 10(4): 377–388.

Lipeika, A.; Lipeikienė, J.; Telksnys, L. 2002. Development of Isolated Word Speech Recognition System, *Informatica* 13(1): 37–46.

Lipeikienė, J.; Lipeika, A. 1998. Language Egineering in Lithuania, *Informatica* 9(4): 449–456.

Makinen, J. Ojala, P. Toukomaa, H. 2004. Performance Comparison of Source Controlled GSM AMR and SMV Vocoders, in *Proceedings in Intelligent Signal Processing and Communication Systems,* 151-154.

Navakauskas, D. 1999. *Artificial Neural Networks for the Restoration of Noise Distorted Songs Audio Records*. Doctoral dissertation. Vilnius Gediminas technical university. Vilnius: Technika.

Nokia White Paper. 2004. Quality of Experience (QoE) of Mobile Services: Can It Be Measured and Improved? [interaktyvus]. 2004. [žiūrėta 2009 gegužės 20 d.]. Prieiga per internetą: <http://www.nokia.com/NOKIA_COM_1/Operators/Downloads/Nokia_Services/whitepaper_qoe_net.pdf>.

Opticom [interaktyvus]. 2004. Advanced non-Intrusive Voice Quality Testing,[žiūrėta 2009 05 01]. Prieiga per internetą <http://www.opticom.de/download/3SQM-WP-290604.pdf>.

Paraestholm, S.; Jensen, S. S; Andersen S.V.; Murthi M. N. 2004. On Packet Loss Concealment Artifacts and their Implications for Packet Labeling in Voice over IP, in *IEEE International Conference on Multimedia and Expo (ICME)*, 1667–1670.

Paulikas, Š. 1999. *Restoration of Noise Distorted Accents Audio Records*. Doctoral dissertation. Vilnius Gediminas technical university. Vilnius: Technika.

Paulikas, Š.; Itani, M. 2008. Influence of Languages on CELP Codecs Performance, *Information Technology and Control* 37(2): 141–144.

Pavilanskas, L. 2007. *Adaptation of Wireless Acces MAC Protocol forReal Time Packet Flows*. Doctoral dissertation. Vilnius Gediminas technical university. Vilnius: Technika.

Pennock, S. 2002. Accuracy of the Perceptual Evaluation of Speech Quality (PESQ) Algorithm, in *Measurement of Speech and Audio Quality in Networks Line Workshop*, 1 –19.

Prasad, R. V.; Vijay, S.; Shankar, H. N.; Pawelczak, P.; Niemegeers, I. 2006. Voice Activity Detection for VoIP-An Information Theoretic Approach, in *Global Telecommunications Conference*, 1–6.

Prastholm, S.; Jensen, S.S.; Andersen, S.V.; Murthi, M.N. 2006. Network Resource Allocation for Perceptually Based Unequal Packet Protection in Voice Communication, in *Proceedings of Acoustics, Speech and Signal Processing*, 14–19.

Qiao, Z.; Sun, L.; Heilemann N.; Ifeachor, E. 2004. A New Method for VoIP Quality of Service Control use Combined Adaptive Sender Rate and Priority Marking, in *Proceedings of IEEE International Conference on Communication,* 1473–1477.

QUALCOMM Incorporated. 2008. PESQ Limitations for EVRC Family of Narrowband and Wideband Speech Codecs.

Raja, A.; Azad, R. M.; Flanagan, C.; Picovici, D. and Ryan, C. 2006. Non-Intrusive Quality Evaluation of VoIP Using Genetic Programming, in *Proceedings of the 1st International Conference on Bio inspired Models of Network, Information and Computing Systems, 1-8.*

Raja, A.; Flanagan, C. 2008. *Real-Time, Non-intrusive Speech Quality Estimation: A Signal-Based Model*, in *Proceedings of Genetic Programming, 11th European Conference*, EuroGP 2008, Naples, Italy, March 26-28, 37-48.

Rix, A. W. 2003. Comparison Between Subjective Listening Quality and P.862 PESQ Score, in *Proceedings of Online Workshop Measurement of Speech and Audio Quality in Networks*, 17–25.

Rix, A. W., Beerends J.G., Hollier, M. P. and Hekstra A.P., Perceptual Evaluation of Speech Quality (PESQ) – a New Method for Speech Quality Assessment of Telephone Networks and Codecs, in *IEEE Signal Processing Society International Conference on Acoustics, Speech, and Signal Processing,* 749-752.

Rudžionis, A.; Rudžionis, V. 2000. Noisy Speech Detection and end Pointing, in *Proc of. Workshop Voice Operated Telecom Services: Do they have a Bright Future?*, 79–82

Russell, S.; Norvig, P. 2002. *Artificial Intelligence: A Modern Approach*. ISBN 0-13-790395-2. 1057 p.

Šaltis, A. 2004. *Radijo technologijos vartotojų prieigų tinkluose*. Daktaro disertacija. Vilniaus Gedimino technikos universitetas. Vilnius: Technika.

Sohn, J.; Kim, N. S.; Sung W. 1999. A Statistical Model-Based Voice Activity Detection, *IEEE Signal Processing Letters* 6(1): 1–3.

Sun, L; Wade, G.; Lines, B.; Ifeachor, E. 2001. Impact of Packet Loss Location on Perceived Speech Quality, in *Internet Telephony Workshop 2001*, 114-122.

Syrdal, A. K.; Bennett, W. R.; Greenspan,L. S. 1995. *Applied Speech Technology*. CRC Press. 656 p.

Autoriaus mokslinių publikacijų disertacijos tema sąrašas

Straipsniai recenzuojamuose mokslo žurnaluose

Kajackas, A.; Anskaitis, A. 2005. Investigation the Ability of Objective Measures of the Perceptual Speech Quality in Mobile Networks, *Electronics and Electrical Engineering* 7(63): 10–15.

Kajackas, A.; Anskaitis, A.; Guršnys, D. 2005. Individual Quality of Service Concept in Next Generations Telecommunications Networks, *Electronics and Electrical Engineering* 4(60): 11–16.

Kajackas, A.; Anskaitis, A.; Guršnys, D. 2008. Peculiarities of Testing the Impact of Packet Loss on Voice Quality, in *Electronics and Electrical Engineering* 2(82): 35–40.

Kajackas,A.; Anskaitis, A. 2009. An Investigation of the Perceptual Value of Voice Frames, *Informatica* (priimtas spausdinti).

Straipsniai kituose leidiniuose

Kajackas, A.; Anskaitis, A.; Guršnys, D.; Pavilanskas, L. 2005. Estimation of QoS Dynamics in the Wireless Networks, *The 4th COST 290 meeting. October 13–14*. Prieiga per internetą: <http://www.cost290.org/td2005/tds/td05050.pdf>.

Kajackas, A.; Anskaitis, A.; Guršnys, D. 2009. User Perceived Quality of Service, in *Traffic and QoS Management in Wireless Multimedia Networks: Cost 290 Final Report: Lecture Notes in Electrical Engineering* 31: 125–128. Ed. by E. Koucheryavy *et al*. ISBN 978-0-387-85572-1.

Kajackas, A.; Anskaitis, A.; Guršnys, D. 2009. Estimating Individual QoS, in *Traffic and QoS Management in Wireless Multimedia Networks: Cost 290 Final Report: Lecture Notes in Electrical Engineering* 31: 180–183. Ed. by E. Koucheryavy *et al*. ISBN 978-0-387-85572-1.