

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Mohamad ITANI

# ĮVAIRIŲ KALBŲ KALBOS SIGNALŲ KODAVIMO ANALIZĖ IR GERINIMAS

DAKTARO DISERTACIJOS SANTRAUKA

TECHNOLOGIJOS MOKSLAI,  
ELEKTROS IR ELEKTRONIKOS INŽINERIJA (01T)



Vilnius LEIDYKLA TECHNICA 2011

Disertacija rengta 2006–2011 metais Vilniaus Gedimino technikos universitete.  
Mokslinis vadovas

**doc. dr. Šarūnas PAULIKAS** (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T).

**Disertacija ginama Vilniaus Gedimino technikos universiteto Elektros ir elektronikos inžinerijos mokslo krypties taryboje:**

Pirmininkas

**prof. dr. Dalius NAVAKAUSKAS** (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T).

Nariai:

**doc. dr. Jonas RIMAS** (Kauno technologijos universitetas, fiziniai mokslai, informatika – 09P),

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

**prof. habil. dr. Rimvydas SIMUTIS** (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija – 07T),

**doc. dr. Vytautas URBANAVIČIUS** (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T).

Oponentai:

**prof. habil. dr. Gintautas DZEMYDA** (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija – 07T),

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

Disertacija bus ginama viešame Elektros ir elektronikos inžinerijos mokslo krypties tarybos posėdyje 2011 m. kovo 15 d. 10 val. Vilniaus Gedimino technikos universiteto senato posėdžių salėje.

Adresas: Saulėtekio al. 11, LT-10223 Vilnius, Lietuva.

Tel.: (8 5) 274 4952, (8 5) 274 4956; faksas (8 5) 270 0112;

el. paštas: doktor@vgtu.lt

Disertacijos santrauka išsiuntinėta 2011 m. vasario 14 d.

Disertaciją galima peržiūrėti Vilniaus Gedimino technikos universiteto bibliotekoje (Saulėtekio al. 14, LT-10223 Vilnius, Lietuva).

VGTU leidyklos „Technika“ 1859-M mokslo literatūros knyga.

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Mohamad ITANI

# ANALYSIS AND IMPROVEMENT OF MULTILINGUAL SPEECH CODING

SUMMARY OF DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES,  
ELECTRICAL AND ELECTRONIC ENGINEERING (01T)



Vilnius LEIDYKLA  
TECHNIKA 2011

Doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2006–2011.

Scientific Supervisor

**Assoc Prof Dr Šarūnas PAULIKAS** (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronic Engineering – 01T).

**The dissertation is being defended at the Council of Scientific Field of Electrical and Electronic Engineering at Vilnius Gediminas Technical University:**

Chairman

**Prof Dr Dalius NAVAKAUSKAS** (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronic Engineering – 01T).

Members:

**Assoc Prof Dr Jonas RIMAS** (Kaunas University of Technology, Physical Sciences, Informatics – 09P),

**Prof Dr Habil Roma RINKEVIČIENĖ** (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronic Engineering – 01T),

**Prof Dr Habil Rimvydas SIMUTIS** (Kaunas University of Technology, Technological Sciences, Informatics Engineering – 07T),

**Assoc Prof Dr Vytautas URBANAVIČIUS** (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronic Engineering – 01T).

Opponents:

**Prof Dr Habil Gintautas DZEMYDA** (Vilnius University, Technological Sciences, Informatics Engineering – 07T),

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

The dissertation will be defended at the public meeting of the Council of Scientific Field of Electrical and Electronic Engineering in the Senate Hall of Vilnius Gediminas Technical University at 10 a.m. on 15 March 2011.

Address: Saulėtekio al. 11, LT-10223 Vilnius, Lithuania.

Tel.: +370 5 274 4952, +370 5 274 4956; fax +370 5 270 0112;

e-mail: doktor@vgtu.lt

The summary of the doctoral dissertation was distributed on 14 February 2011.

A copy of the doctoral dissertation is available for review at the Library of Vilnius Gediminas Technical University (Saulėtekio al. 14, LT-10223 Vilnius, Lithuania).

© Mohamad Itani, 2011

## Įvadas

**Mokslų problemos aktualumas.** Pastaruoju metu skaitmeninio signalų apdorojimo metodai ir priemonės yra plačiai taikomos visose žmogaus veiklos sferose. Ypatingai reikšmingi adaptyvieji skaitmeninio signalų apdorojimo metodai ir priemonės, nes jie įvertina ir tinkamai reaguoja į aplinkos, sistemos būsenos, signalų charakteristikų kitimą. Vienas iš tokių nestacionarių signalų pavyzdžių yra kalbinis signalas.

Kalba nuo seno yra viena pagrindinių žmonių bendravimo, tautiško išlaikymo bei informacijos apsaugos priemonė, kuri neprarado savo svarbos ir šiuolaikiniame informaciniame pasaulyje. Moderniose elektroninėse telekomunikacinėse sistemose iki šiol viena pagrindinė perduodama informacija yra kalba. Siekiant taupyti kanalo plotį, kalbos signalas yra apdorojamas ir koduojamas balso kodekais. Pastarųjų veikimas yra pagrįstas žmogaus kalbos trakto ir garso susidarymo juose modeliais. Tačiau žmonės pasaulyje šneka įvairiomis kalbomis, kurios skiriasi viena nuo kitos tariamais garsais.

Elektroninių telekomunikacinių sistemų raidoje didžiausią įtaką turėjo anglų kalba, nes pagrindiniai kalbos signalų apdorojimo algoritmai bei jų patobulinimai daugiausiai buvo sukurti angliškai kalbančiose šalyse. Tokių patobulinimų lingvistinis patikimumas kelia tam tikras abejones: ar visi sukurti patobulinimai vienodai gerai tinka įvairioms kalboms? Yra tiesiog nuostabu, kad iki šiol nebuvo atlikta nei viena išsami studija apie kalbos vaidmenį kalbos signalo apdorojime.

Jau yra pastebėta, kad prie skirtingų kalbų kalbos signalų balso kodekai pasižymi skirtingais kalbos kokybės įverčiais. Pavyzdžiui, japonų kalbai kodekų kokybės įverčiai yra apie pusę balo žemesni negu anglų kalbai. Tačiau išsamesni kodekų vertinimo tyrimai tėra atlikti tik su keliomis kalbomis: anglų, vokiečių, prancūzų, suomių, švedų ir japonų. Tuo tarpu arba bei kitų mažesnių tautų, pvz. lietuvių, kalbos kodavimo balso kodekais ypatumai nėra tyrinėti.

Lietuvoje tyrimai susieti su lietuvių kalbos apdorojimu vyksta gana seniai. Nemažą įdirbį šioje srityje turi sukaukę mokslininkai iš Matematikos informatikos instituto, Vytauto didžiojo universiteto bei Kauno technologijos universiteto. Tačiau pastarųjų institucijų mokslininkų darbai šioje tematikoje daugiausiai yra siejami su lietuvių kalbos atpažinimu. Tačiau šiam procesui reikalingi kalbos signalo požymiai yra randami naudojant tradicinius kalbos signalo apdorojimo metodus, kurie buvo sukurti ir ištirti su anglų kalba.

Tuo tarpu pasaulyje, ypač elektroninių telekomunikacijų srityje, stengiamasi įvertinti skirtingų kalbų aspektus. Atliekant balso kodekų testavimus yra rekomenduojama naudotis garsynu iš ITU-T P.50

rekomendacijos priedo. Šiame priede yra surinkti 20 kalbų (dauguma Europos šalių) garso įrašai. Deja lietuvių kalbos garso įrašų čia nėra. Taip pat, net nuo 1995 metų yra dvi Europos sąjungos programos, skirtos kurti telefoniniams pokalbiams orientuotus garsynus: CEC-DGXII programa LRE-RELATOR ir LE-MLAP programa, kurios pagrindu vykdomas SPEECHDAT garsynų projektas. Be to nuo 1995 metų nepelno organizacija ELRA (European Language Resources Association), kuri rūpinasi Europos kalbų išteklių ir įrankių standartizavimu, taip pat vykdo garsyno AURORA kūrimo projektus. Deja, Lietuva nėra dalyvavusi nė viename iš šių projektų.

Taigi yra poreikis sukurti lietuvių kalbos įrašų duomenų bazę tinkančią balso telekomunikacinių sistemų našumo vertinimui. Taip pat yra svarbu ištirti balso kodekų tinkamumą kuoduoti ne anglų kalbos signalams tuo siekiant išsiaiškinti šių kalbų įtaką balso kodekų atkurtos kalbos kokybei. Čia iškyla problema surasti būdus, kurie pagerintų balso kodekų kokybę koduojant skirtingų kalbų signalus. Tai įgalintų balso telekomunikacinių sistemų vartotojams gauti geresnės kokybės paslaugas.

**Tyrimų objektas.** Daktaro disertacijoje tyrimų objektas yra CELP kodekų ir jais dekuoduotos kalbos kokybė.

**Darbo tikslas ir uždaviniai.** Darbo tikslas – ištirti ir pasiūlyti būdus CELP kodekų dekuoduotų lietuviškos ir angliškos kalbų kalbos signalų kokybei pagerinti.

Šiam tikslui pasiekti reikia išspręsti tokius uždavinius:

1. Atlikti analitinę apžvalgą apie įvairių kalbų kalbos signalų kodavimo CELP kodekais aspektus.
2. Sukurti lietuvių kalbos balso įrašų duomenų bazę atitinkančią ITU-T P.50 rekomendaciją ir tinkančią balso kodekų kokybės vertinimui.
3. Ištirti AMR ir Speex kodekų kokybinius rodiklius koduojant lietuvių ir arabų kalbų signalus.
4. Pasiūlyti ir ištirti būdus įvairių kalbų kalbos signalų kodavimui pagerinti.

**Tyrimų metodika.** Rengiant disertaciją buvo taikoma statistinė analizė, kompiuterinis modeliavimas, eksperimentiniai bandymai.

**Mokslinis naujumas.** Rengiant disertaciją buvo gauti šie nauji rezultatai:

1. Eksperimentiniai AMR ir Speex balso kodekų kokybės įverčiai kuoduojant anglų, lietuvių ir arabų kalbų kalbos signalus.

2. Eksperimentiniai PESQ ir 3SQM kalbos kokybės vertinimo algoritmų tinkamumo kalbos signalų kokybės tyrimams įverčiai.
3. Adaptuotos Speex kodeko kodų lentelės leidžiančios pagerinti dekodautos kalbos kokybę.

***Darbo rezultatų praktinė reikšmė.*** Sukurta lietuvių kalbos balso įrašų duomenų bazė atitinkanti ITU-T P.50 rekomendaciją ir tinkanti balso kodekų kokybės vertinimui.

Eksperimentiniai AMR ir Speex balso kodekų kokybės tyrimų kuoduojant anglų, lietuvių ir arabų kalbų kalbos signalus rezultatai.

Adaptuotų lietuviškai ir arabiškai kalboms Speex kodeko kodų lentelių taikymas VoIP balso ryšyje pagerina dekodautos lietuviškos ir arabiškos kalbos kokybę.

Atliktų tyrimų rezultatai buvo panaudoti atliekant vieną mokslinį tiriamąjį projektą remtą Lietuvos valstybinio mokslo ir studijų fondo. Lietuviškos kalbos ypatumų kalbinio signalo kodavime tyrimas. Nr. T–89/09, (Reg. Nr. T–09022).

### ***Ginamieji teiginiai***

1. CELP kodekais dekodautos lietuviškos ir arabiškos kalbos kokybė priklauso nuo panaudotos kalbos.
2. AMR ir Speex kodekai geriau koduoja angliškos kalbos signalus tiek objektyvių, tiek subjektyvių kokybės įvarečių prasme.
3. Fiksuotos Speex kodeko kodų lentelės adaptavimas lietuviškai ir arabiškai kalboms pagerina Speex kodeku dekodautos kalbos kokybę ir jos stabilumą.

***Darbo rezultatų aprobavimas.*** Darbo rezultatai paskelbti trijose mokslinėse publikacijose recenzuojamuose mokslo leidiniuose:

- 2 straipsniuose Mokslinės informacijos instituto duomenų bazės „ISI Web of Science“ leidiniuose;
- 1 tarptautinės konferencijos medžiagoje.

Rezultatai viešinti 4 mokslinėse konferencijose Lietuvoje ir užsienyje:

- 9-toje IEEE tarptautinėje konferencijoje „Signal Processing and Information Technology“. 2009 m. gruodžio 14–17, – Ajman, UAE.
- Tarptautinėje konferencijoje „ELECTRONICS“, 2007 ir 2008 metais.
- Lietuvos jaunųjų mokslininkų konferencijoje „Mokslas – Lietuvos ateitis“ 2007.

*Darbo apimtis.* Darbą sudaro bendra darbo charakteristika, 5 skyriai, išvados, literatūros sąrašas, publikacijų sąrašas ir priedai. Bendra disertacijos apimtis – 102 puslapiai, 20 iliustracijų, 9 lentelės ir 1 priedas.

## **1. Literatūros apžvalga apie įvairių kalbų kalbos kodavimą**

Įvairių tautų kalbų garsai skiriasi nuo anglų kalbos garsų savo formančių struktūra. Tai lemia, kad kvantavimas pritaikytas anglų kalbos garsams gali modifikuoti šias struktūras. To pasėkoje kodu žadinama tiesinės prognozės (CELP) kodavimas gali modifikuoti tikruosius ne anglų kalbos garsus. Darbe atlikti ir literatūroje aprašyti balso kodekų patikros testai rodo, kad dekodautos kalbos kokybė priklauso nuo kalbos. Deja, kalbos įtakos balso kodekų kodavimo kokybei tyrimai yra atlikti su didžiosiomis pasaulio kalbomis. Tuo tarpu, mažesnių tautų– lietuvių, arabų – kalbos, kuriant elektronines balsinio ryšio sistemas, paprastai nėra vertinamos.

Darbe siekiama:

1. Ištirti kalbos kokybės vertinimo būdus tinkamus koduoto balso kokybės vertinimui.
2. Sukurti lietuvių kalbos garsų įrašų duomenų bazę atitinkančią ITU-T rekomendacijos P.50 priedą ir tinkančią balso kodekų kokybės tyrimams.
3. Ištirti AMR ir Speex kodekų kokybę kuoduojant ne anglų kalbos garso signalus.
4. Pasiūlyti ir ištirti būdus tinkamus įvairių kalbų garso signalų kodavimui.

## **2. Kalbos kokybės vertinimas**

Šiame skyriuje pateikiamos kalbos signalo kokybės samprata, jos vertinimo būdai. Atliekami pirminiai kodekų kodavimo gerumo koduojant lietuviškos ir arabiškos kalbų signalus tyrimai.

Gauti rezultatai rodo, kad objektyvūs kalbos kokybės įverčiai (MOS) nėra gerai tinkami žmogaus suvokiamai kalbos kokybei vertinti.

Šiame skyriuje atliktų pradinų tyrimų rezultatai parodė, kad CELP kodekai gana gerai koduoja lietuvių ir anglų kalbų signalus, tačiau dekodautos kalbos kokybės svyravimai yra gana dideli, lyginant su kokybės svyravimais kuoduojant anglų kalbos signalus (2.1 ir 2.2 lentelės).

Šie rezultatai parodė, kad kalba įtakoja CELP kodekų veikimą, tačiau siekiant aiškiai atskleisti šią įtaką, reikalingi išsamesni tyrimai su geresnės kokybės kalbos įrašais.



**2.1 lentelė.** 3SQM kokybės vertinimo rezultatai arabų (AR), lietuvių (LT) ir anglų (EN) kalboms

MOS	AMR			Speex		
	AR	LT	EN	AR	LT	EN
Vidurkis	4,18	4,91	3,54	3,45	4,97	3,63
Nuokrypis	0,70	0,14	0,44	1,50	0,07	0,55

**2.2 lentelė.** PESQ kokybės vertinimo rezultatai arabų (AR), lietuvių (LT) ir anglų (EN) kalboms

MOS	AMR			Speex		
	AR	LT	EN	AR	LT	EN
Vidurkis	3,31	2,39	3,33	4,15	2,64	4,08
Nuokrypis	0,26	0,41	0,17	0,11	0,41	0,12

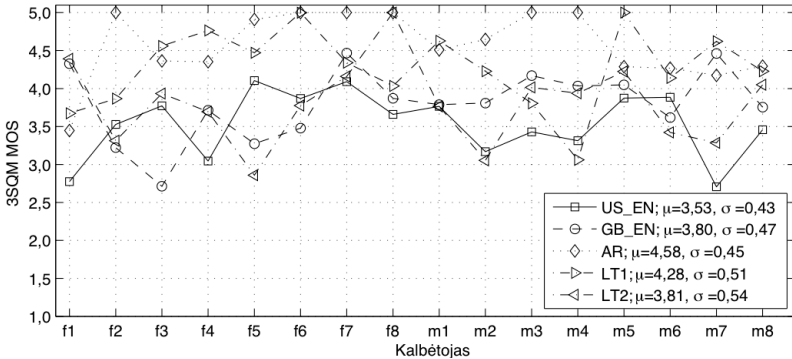
### 3. Lietuvių kalbos garsyno kūrimas

Darbe sudarytas lietuvių kalbos garsynas, kuris tenkina ITU-T P.50 rekomendacijos priedo reikalavimus (sakinių skaičius ir trukmė, diktorių skaičius ir pan.) ir atspindi bendrinės lietuvių kalbos ypatumus: dažniausiai vartojamas kalbos dalis, dažniausiai vartojamus žodžius ir pan.

Nustatyta, kad lietuvių kalboje dažniausiai vartojamos kalbos dalys yra daiktavardžiai (45 %), veiksmazodžiai (20 %) ir būdvardžiai arba dalyviai (8 %). Dažniausiai vartojami daiktavardžiai: *galva, laikas, darbas, metas, vanduo, gyvenimas, motina, miestas, vakaras, širdis, pasaulis, saulė, dangus, vėjas*; veiksmazodžiai: *būti, galėti, nebūti, turėti, žinoti, eiti, sakyti, norėti, reikėti, matyti, žiūrėti, pasakyti, negalėti, imti, gyventi, atrodyti, ateiti, suprasti, stovėti, sėdėti, kalbėti*; ir būdvardžiai: *aukštas, valstybinis, administracinis, pagrindinis, politinis, bendras, naujas, didelis, tarptautinis, konstitucinis, ekonominis, atskiras, socialinis, visuomeninis*.

Remiantis šiais tyrimais buvo sudaryti keturi sakiniai susidedantys iš 2–3 frazių:

1. *kalbos jos negalėjo suprasti.* (Lt\_f1.wav, Lt\_f2.wav, Lt\_m1.wav, Lt\_m1.wav).
2. *Taip galėjo sakyti tik žmogus, kurio širdis žinojo, kad šiame gyvenime galima pasitikėti tik savo tėvu.* (Lt\_f3.wav, Lt\_f4.wav, Lt\_m3.wav, Lt\_m4.wav).
3. *Vakare aplink jį susirinkę stovėjo vyrai, tarytum atėję iš tuščio pasaulio, kuriame nebuvo saulės ir dangaus.* (Lt\_f5.wav, Lt\_f6.wav, Lt\_m5.wav, Lt\_m6.wav).
4. *Po sunkaus darbo mieste motinai skaudėjo galvą, tačiau jos veidas visada buvo gražus, o balsas gyvas.* (Lt\_f7.wav, Lt\_f8.wav, Lt\_m7.wav, Lt\_m8.wav).



**3.1 pav.** 3SQM įverčiai pasiūlytų lietuvių kalbos (LT1 ir LT2) garsynų ir anglų kalbos amerikoms tarmės (US), anglų kalbos britų tarmės (GB), arabų (AR) garsynų iš ITU-T P.50 rekomendacijos priedo

Taikant sudarytus sakinius sudarytos ir įrašytos dvi lietuvių kalbos garsyno versijos: LT1 ir LT2. Įrašai atlikti aukštos kokybės įrašymo technika ir mikrofonu, tačiau skirtinguose aplinkos sąlygose. LT1 garsyno įrašai turi didesnę signalo ir triukšmo santykį, negu LT2 garsyno įrašai, ir yra kandidatai tapti lietuvių kalbos garsynu atitinkančiu ITU-T P.50 rekomendacijos priedą.

Siekiant objektyviai palyginti tarpusavyje abi garsyno versijas bei jų adekvatumą garsynams esantiems ITU-T P.50 rekomendacijos priede, atlikti kokybės vertinimo testai taikant 3SQM kalbos kokybės vertinimo algoritmą su LT1 ir LT2 lietuvių kalbos ir jau rekomendacijos priede esančiais anglų (amerikos ir britų tarmių) ir arabų kalbų garsynais.

Gauti rezultatai (3.1 pav.) rodo, kad tiek individualių įrašų, tiek vidurkių pagal kalbą kokybės įverčiai kinta gana plačiose ribose. Taigi, 3SQM algoritmas nėra tinkamas absoliučios kokybės įvertinimui, tačiau Lietuvių kalbos įrašų atveju galime teigti, kad LT1 garsynas yra tinkamesnis kandidatas negu LT2, nes vidutinė MOS įverčio vertė apie 0,4 balo aukštesnė ir mažesnė individualių MOS įverčių sklaida.

Atlikti kalbos kokybės tyrimai parodė, kad tiek individualūs, tiek vidutiniai kokybės įverčiai kinta plačiose ribose. Taigi, 3SQM algoritmas nėra tinkamas absoliučiam kalbos kokybės vertinimui.

## 4. Kodekų kokybės tyrimai koduojant įvairių kalbų garso signalus

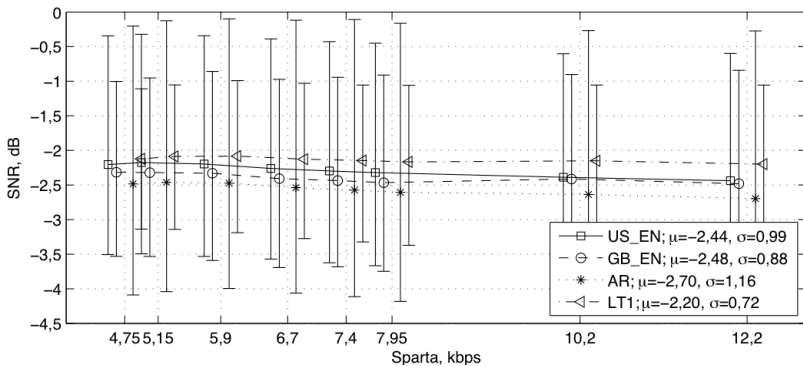
AMR ir Speex balso kodekų efektyvumas buvo vertintas pagal šiais kodekais transformuotos (koduotos ir dekodautos) kalbos kokybę. Pastarajai nustatyti pasirinkti du objektyvūs kokybės matai: spektriniai iškreipiai ir signalo ir triukšmo santykis. Tyrimuose naudoti anglų (amerikos ir britų tarmių) ir arabų kalbų garso įrašai iš ITU-T P.50 rekomendacijos priedo ir lietuvių kalbos garso įrašai iš LT1 garsyno. Garso įrašai buvo koduoti ir dekoduoti taikant AMR ir Speex balso kodekus prie skirtingų spartų. Taip kaip AMR kodeko spartos yra fiksuotos, buvo priderintos Speex kodeko spartos. Tokiu būdu galima tarpusavyje lyginti ne tik kalbos įtaką balso kodekams, bet ir pačius balso kodekus tarpusavyje.

Iš 4.1 ir 4.2 paveikslų matyti, kad SNR matas nėra tinkamas šiuo atveju įvertinti transformuoto kalbos signalo kokybę. Vidutinės SNR vertės praktiškai nekinta, kintat kodekų kodavimo spartai, tuo tarpu SNR rezultatų dispersija yra didelė. Tai gali būti paaiškinta tuo, kad net ir nedidelis dviejų signalų (originalo ir transformuoto) persislinkimas laike (dėl kodekose vykdomo signalo apdorojimo fiksuotos trukmės kadrais), iššaukia atsitiktinį signalų atskaitų nesutapimą (kartais ir labai didelį amplitūžių skirtumą) ir kartu didelę signalo ir triukšmo santykio (SNR) variaciją. Tuo tarpu žmogaus suvokimo požiūriu mažas kalbos signalo vėlinimas nesukelia jokio kalbos kokybės pablogėjimo.

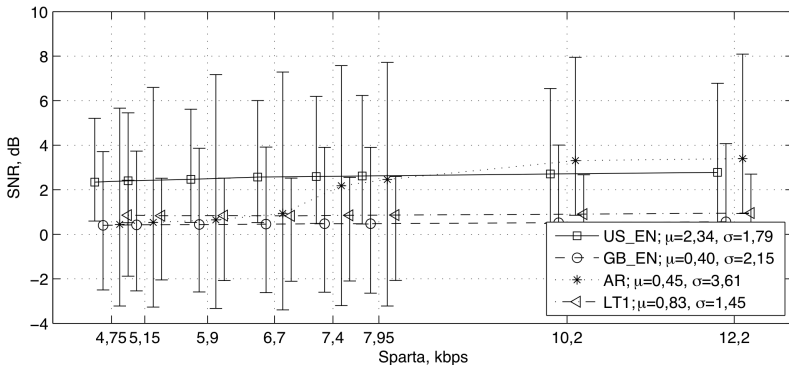
Spektriniai iškreipiai eliminuoja nesutapimo laike įtaką ir yra tinkamesni transformuoto kalbos signalo kokybei ir kartu kodeko efektyvumui įvertinti. Kaip matyti iš 4.3 ir 4.4 paveikslų, didėjant kodekų veikimo spartai, spektriniai iškreipiai mažėja. Taip pat pastebima, kad abiejų kodekų atveju ne anglų kalbos signalų spektriniai iškreipiai yra didesni ir/arba pasklidę platesniame intervale. Tai rodo, kad tiek AMR, tiek Speex kodekas stabiliau koduoja anglų kalbos garso signalus.

Siekiant įvertinti kodekų efektyvumą atsižvelgiant į žmogaus suvokiamą kalbos kokybę pritaikytas PESQ kalbos kokybės vertinimo algoritmas. Šis algoritmas palygina originalų kalbos garso signalą su sulygiuotu laike transformuotu signalu taikydamas suvokimo modelį. PESQ rezultatas yra objektyvus MOS įvertis MOS skalėje tiesiogiai įvertinantis transformuotos kalbos kokybę. PESQ MOS atsižvelgiant į ITU-T P.862 rekomendaciją yra nuo 1,0 (blogiausia kokybė) iki 4,5 (geriausia kokybė).

Tiesa, ITU MOS skalė yra iki 5,0 balų, tačiau PESQ modeliuoja realų auditorinio klausimosi testą ir yra optimizuotas surasti vidutinį visų klausytojų suteiktą kokybės įvertį. Statistika rodo, kad geriausias vidutinis rezultatas, kurio galima tikėtis auditorinio klausimosi testo metu, yra ne 5,0 (puiku) bet 4,5.



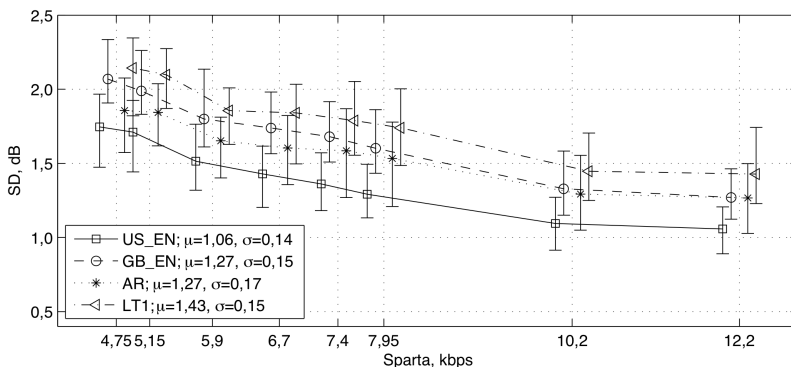
**4.1 pav.** SNR vidurkis ir maksimalus/minimalus nuokrypiai AMR kodekui



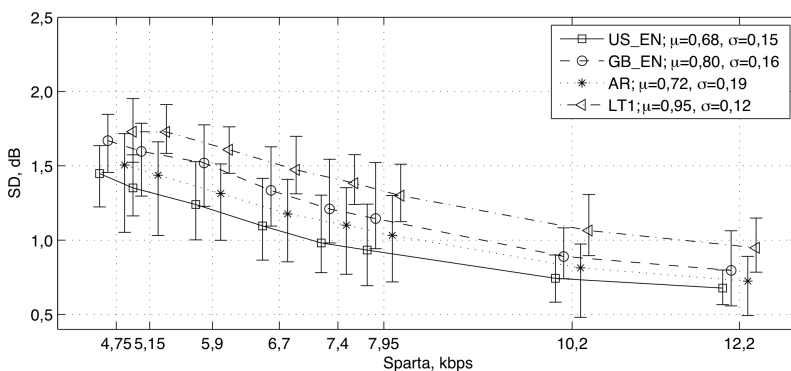
**4.2 pav.** SNR vidurkis ir maksimalus/minimalus nuokrypiai Speex kodekui

Šiame eksperimente taip pat panaudotas sukurtas lietuvių kalbos garso įrašų garsynas ir anglų (amerikos ir britų tarmių) ir arabų kalbų garso įrašai iš ITU-T P.50 rekomendacijos priedo. Dekoduoto kalbos signalo kokybę įvertinta jau minėtu PESQ algoritmu (4.5 ir 4.6 pav.).

Iš 4.5 ir 4.6 paveikslių matyti, kad prie aukštų kodekų kodavimo spartų MOS kokybės įvertiniai visoms kalboms yra artimi. Tačiau iš 4.6 paveikslo pastebime, kad Speex kodeko kokybė visumoje yra mažesnė koduojant ne anglų kalbos garso signalus negu AMR kodeko. Taip pat Speex kodekas ne toks geras prie žemų kodavimo spartų, MOS įvertis kinta net 0,5 balo ribose.



**4.3 pav.** Spektrinių iškreipčių vidurkis ir maksimalus/minimalus nuokrypiai Speex kodekui

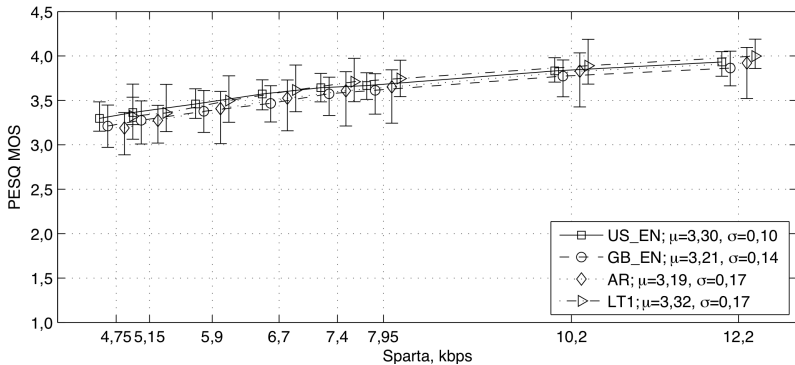


**4.4 pav.** Spektrinių iškreipčių vidurkis ir maksimalus/minimalus nuokrypiai Speex kodekui

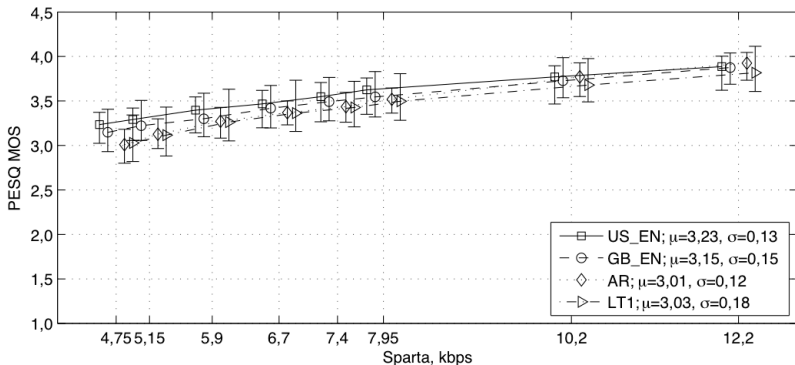
Taip pat, kaip jau buvo pastebėta, abu kodekai veikė stabiliau koduojant anglų kalbos garso signalus, o Speex kodeko atveju kokybė buvo apie 0,2 balo aukštesnė, negu koduojant lietuvių ir arabų kalbu garso signalus.

Galiausiai, siekiant įsitikinti, kad eksperimentų metu gauti rezultatai yra teisingi, buvo atliktas auditorinis klausimosi testas su lietuvių kalbos LT1 garsynu. Klausimosi teste dalyvavo šeši asmenys, kurių gimtoji kalba yra lietuvių.

Pradžioje klausytojams buvo pateikiama klausymuisi originalus kalbos garso įrašas, o paskui sugadintas, transformuotas AMR ir Speex kodekų.



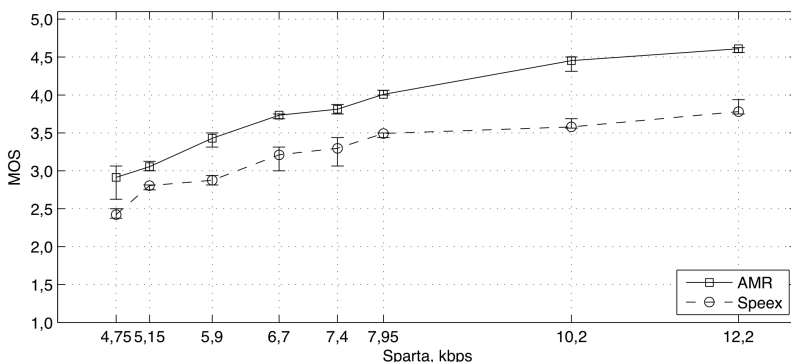
4.5 pav. PESQ MOS vidurkis ir maksimalus/minimalus nuokrypiai AMR kodekui



4.6 pav. PESQ MOS vidurkis ir maksimalus/minimalus nuokrypiai Speex kodekui

Sugadinto kalbos garso įrašo kokybė kiekvieno klausytojo buvo vertinama MOS skalėje. Suvidurkintas visų klausytojų MOS įvertis atskirai AMR ir Speex kodekams pateiktas 4.7 paveiksle.

Šio testo rezultatai taip pat patvirtina, kad AMR kodekas šiuo atveju yra geresnis negu Speex kodekas, nes juo dekodutos kalbos kokybė yra aukštesnė. Patys kokybės įverčiai ir jų kitimo kintant kodavimo spartai tendencija yra artimi MOS įverčiams gautiems taikant PESQ algoritmą.



4.7 pav. AMR ir Speex kodekų auditorinio klausymosi testo MOS įverčiai

## 5. Speex kodeko kodavimo kokybės gerinimas

Iš dviejų tirtų kodekų, tik Speex kodekas yra platinamas, kaip atviro kodo programinė įranga (C programavimo kalba). Tuo tarpu AMR kodekas yra prieinamas tik kaip uždara dvejetainio formato C programavimo kalbos biblioteka. Todėl darbe adaptavome tik Speex kodeko fiksuotas kodų lenteles, siekiant pagerinti lietuvių ir arabų kalbos kodavimo kokybę.

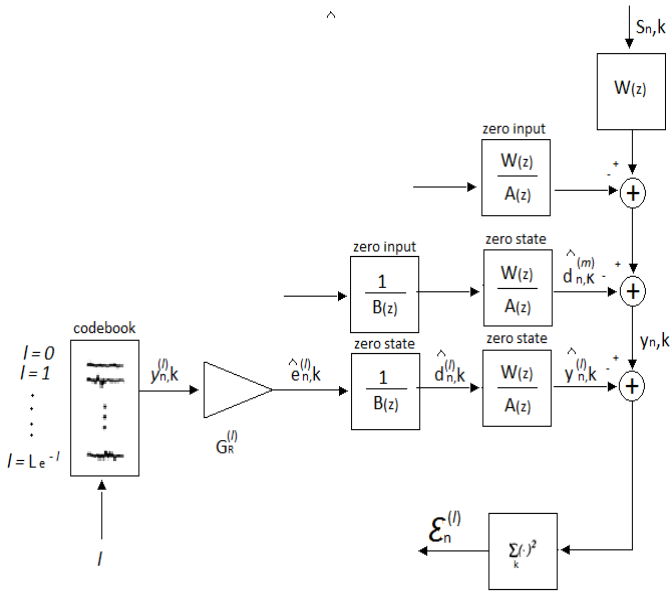
CELP kodavimo idėja pavaizduota 5.1 paveiksle. Čia ilgalaikės prognozės kvantuotas inversinis filtras (su  $2q + 1$  nuliui nelygiomis pakopomis)  $B(z)$ , atskaitų skaičiui  $n$  yra aprašomas taip:

$$B(z) = 1 - \sum_{k=-q}^q b_k z^{-(M+k)}. \quad (5.1)$$

Trumpalaikės kvantuotas inversinis filtras (kurio eilė yra  $p$ )  $A(z)$ , atskaitų skaičiui  $n$  yra parašomas kaip  $p$  eilės daugianaris:

$$A(z) = 1 - \sum_{k=1}^p a_k z^{-k}. \quad (5.2)$$

Tada svorio filtras pritaikytas žmogiškajam garso suvokimui ir padidinantis signalo ir triukšmo santykį tarp formančių yra išreiškiamas taip:



5.1 pav. CELP veikimo principas (LeBlanc *et. al.* 1990).

$$W(z) = \frac{A_p(z/\beta)}{A_p(z/\gamma)}. \quad (5.3)$$

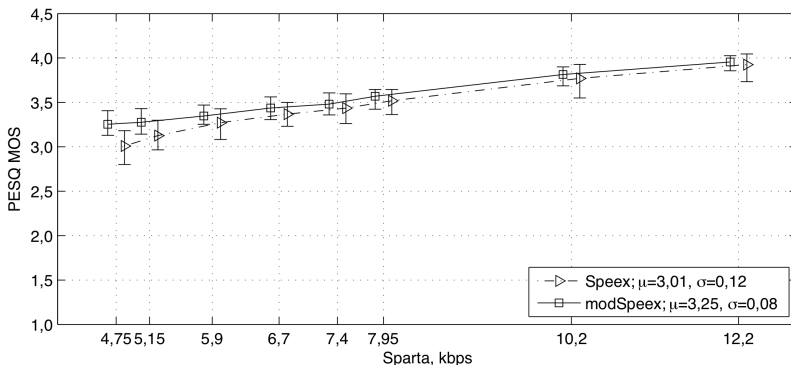
Čia  $\gamma$  ir  $\beta$  yra optimizuoti atsižvelgiant į subjektyvius kokybės įverčius,  $A_p(z)$  – optimalus nekvantuotas inversinis filtras atskaitų skaičiui  $n$  (langui).

Paprastai,  $A(z)$  yra toks filtras, kuris minimizuoja atviros kilpos skirtuminę energiją, o  $B(z)$  randamas uždaroje kilpoje minimizuojant pasvertą triukšmo paklaidą prieš nustatant fiksuotą kodų lentelę žadinimo signalą  $y_{n,k}$ .

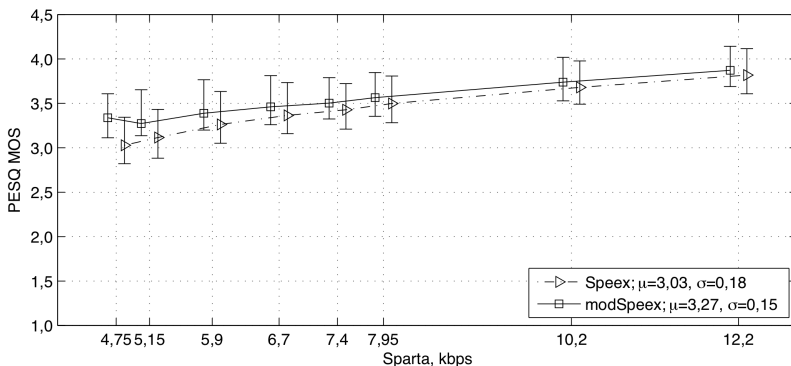
Pasverta vidutinė kvadratinė paklaida kiekvienam fiksuotos kodų lentelės įrašo indeksui  $l$  atskaitų skaičiui  $n$ , kurios ilgis  $K_C$  ir persidengimas  $K_O$  yra randama:

$$\begin{aligned} \epsilon_n^{(l)} = \left| \mathbf{y}_n - \hat{\mathbf{y}}_n^{(l)} \right|^2 = & \mathbf{y}_n^T \mathbf{y}_n - 2G_n^{(l)} \mathbf{y}_n^T \mathbf{H}_n \mathbf{v}^{(l)} \\ & + G_n^{(l)2} \mathbf{v}^{(l)T} \mathbf{H}_n^T \mathbf{H}_n \mathbf{v}^{(l)}. \end{aligned} \quad (5.4)$$





5.2 pav. PESQ MOS įverčių palyginimas tarp originalaus ir modifikuoto Speex kodekų koduojant arabiškos kalbos signalus



5.3 pav. PESQ MOS įverčių palyginimas tarp originalaus ir modifikuoto Speex kodekų koduojant lietuviškos kalbos signalus

Čia  $K_C + K_O$  ant  $K_C$  dimensijos apatinė trikampė Toeplitz matrica  $\mathbf{H}_n$ , žymi nulinę filtravimo būseną filtrui  $W(z)/A(z)$ .  $l$ -tasis žadinimo (stulpelis) vektorius  $\mathbf{v}^{(l)}$  yra  $K_C$  ilgio vektorius, o (stulpeliai) vektoriai  $\mathbf{y}_n$  ir  $\hat{\mathbf{y}}_n^{(l)}$  yra  $K = K_C + K_O$  ilgio.

Fiksuotos kodų lentelės radimo algoritmai remiasi apibendrintu Lloyd algoritmu (GLA), kuriam reikia gana didelės apmokymo sekos siekiant surasti patikimą kodų lentelę. Dėl  $1/B(z)$  ilgalaikės atminties šis algoritmas negarantuoja konvergavimo link minimumo. Todėl apmokymo vektorių  $\mathcal{T} = \{\mathbf{y}_n\}$  rinkiniai yra keičiami iteracijų būdu.

Optimali fiksuota kodų lentelė yra tokia, kuri minimizuoja (5.4) lygtį visiems apmokymo vektoriams:

$$\bar{\epsilon}_n^{(l)} = \sum_{n=0}^{N-1} \epsilon_n^{(l)} \quad (5.5)$$

Optimizuojant laikoma, kad apmokymo vektorius  $y_n$  nepriklauso nuo ankstesnių ir būsimų kodo lentelės įrašo vektorių (stulpelių).

Taikant turimus arabų kalbos garso įrašams iš ITU-T P.50 priedo ir lietuvių kalbos garso įrašams iš LT1 garsyno buvo adaptuota Speex kodeko kodų lentelė. Siekiant palyginti adaptacijos rezultatus, buvo atliktas modifikuoto Speex kodeko dekodutos kalbos kokybės vertinimas pagal spektrinius iškreipius ir MOS įvertį gautą PESQ algoritmu. Vertinimas buvo atliekamas taikant modifikuotą Speex kodeką koduojant ir dekoduojant arabų ir lietuvių garsyno kalbos balso įrašus.

5.2 ir 5.3 paveiksluose PESQ MOS įverčių palyginimas tarp originalaus ir modifikuoto Speex kodekų koduojant arabiškos ir lietuviškos kalbų signalus.

Rezultatai rodo, kad Speex kodeko fiksuotos kodų lentelės adaptavimas visumoje pagerina dekodutos kalbos kokybę, padidinant PESQ MOS įverčio vidurkį bei sumažinant skirtumą tarp maksimalios ir minimalios įverčio reikšmių (5.2 ir 5.3 paveiksluose pavaizduota rėžiais). MOS reikšmė vidutiniškai pagerėjo 0,2 balo, kartu pagerinant ir stabilumą.

Speex kodeku dekodutos kalbos kokybės pagerėjimas yra labiau pastebimas prie žemesnių kodeko kodavimo spartų.

## Bendrosios išvados

1. Atlikti CELP balso kodekų kokybės tyrimai rodo, kad jais dekodutos lietuviškos ir arabiškos kalbos kokybė priklauso nuo konkrečios panaudotos kalbos.
2. AMR ir Speex kodekai geriau koduoja angliškos kalbos signalus. Angliškai kalbai jų PESQ MOS kokybės įvertis yra apie 0,2 balo aukštesnis, o sukuriami spektriniai iškreipiai yra 0,05 balo žemesni.
3. AMR ir Speex kodekų subjektyvūs kokybės MOS įverčiai yra 0,22 balo žemesni lietuviškai ir 0,18 balo arabiškai kalboms lyginant su angliška kalba.

4. Fiksuotos Speex kodeko kodų lentelės adaptavimas lietuviškai arba arabiškai kalbai pagerina dekoduotos kalbos kokybę vidutiniškai 0,2 MOS įverčio ir 7–10 % sumažina kokybės sklaidą.

## **Autoriaus publikacijų disertacijos tema sąrašas**

### **Recenzuojamuose mokslo žurnaluose**

Paulikas, Š.; Itani, M. 2010. Lithuanian Speech Records Database for Voice Codecs Quality Assessment. *Information Technology and Control* 39(1): 38–42. ISSN 1392–124X. (Thomson ISI Web of Science).

Itani, M, Paulikas, Š. 2008. Influence of languages on CELP codecs performance. *Information Technology and Control* 37(2):141–144. ISSN 1392–124X. (Thomson ISI Web of Science). (online).

### **Kituose recenzuojamuose leidiniuose**

Paulikas, Š.; Itani, M. 2009. Assessment of CELP Codecs Quality in Multi-Lingual Environment. *The 9th IEEE International Symposium on Signal Processing and Information Technology*. December 14–17, 2009 – Ajman, UAE. 55–60. ISBN 978-1-4244-5950-6. (IEEEExplore, Inspec).

### **Trumpos žinios apie autorių**

Mohamad Itani gimė 1982 metais gegužės 25 dieną Libano sostinėje Beirute. 2000 metais įstojo AUL Libano universitetą. 2003 metais ten įgijo bakalauro, o 2005 metais ir magistro laipsnį. 2006 metais atvyko studijuoti į Lietuvą ir įstojo į Vilniaus Gedimino technikos universiteto doktorantūrą telekomunikacijų inžinerijos katedroje. Šiuo metu Mohamad Itani taip pat dirba AUL Libano universitete instruktoriumi.

## **ANALYSIS AND IMPROVEMENT OF MULTILINGUAL SPEECH CODING**

**Topicality of the problem.** Technologies of telecommunication networks in these days migrate to multidimensional heterogeneity as they are widely used for multimedia applications. Until now communication by speech remains the most widely used for exchange of information among humans. The QoS of speech communication strongly depends on in electronic telecommunication system employed voice codec and its parameters.

However, human speeches (sounds) differ from one language to another. The English language has had a dominating influence in the advance of telecommunications. With many of the major developments coming from primarily English speaking areas there is the risk that these advances may not be linguistically robust.

In modern electronic communication systems in the design of low-bit-rate speech coding algorithms, language variability is often considered to be of secondary importance. It is surprising to find that no comprehensive studies have ever been carried out on the role of language in speech coding. Codec operation principles are based on human voice tract and voice producing mechanism. There is already noticed that voice codecs perform differently for different speeches. For example, codecs quality estimates for Japanese language is around 0.5 MOS score lower than for English language. However, more detailed investigations on language influence on voice codecs used in electronics communication systems quality is done only for few languages such as: English, German, French, Finish and Swedish. While Arabic and other languages of smaller nationalities, such as Lithuanian, are left aside of investigations.

Today in electronic telecommunications sector a lot of companies that design voice communication products and regulatory institutions admit importance of language on performance of voice communication systems. For example, over 20 languages (mostly from European countries) that are listed in ITU-T Rec. P.50 Appendix I were extensively used in tests of AMR codec that were chosen by 3GPP as default voice codec in 3G mobile communications.

Unfortunately, Lithuanian languages, as other languages of minor nationalities, has left aside of this investigation. One of the reasons for not including these languages in codecs performance and other tests is lack of appropriate speech records databases. It should be noticed that in Lithuania there is a few Lithuanian speech records databases created by Institute of Mathematics and Informatics, Lithuanian Radio News (LRN0 and LRN0.1), Vytautas Magnus University (VDU-RTG). Also, Kaunas University of Technology together with University of Vilnius created LITGIS database (records of Lithuanian digits sequences). However, these speech records databases are intended to use for Lithuanian speech recognition but not for testing of voice communication systems.

So, there arises the need for creating a Lithuanian speech record database suitable for analysis of the performance of communication systems. Also it is important to investigate the voice codecs performance over different non-English languages in order to find out the influence of specific language on quality of coded speech and there is the problem to develop of methods that

provide ways for improving the voice codecs performance for specific language. It is essential to modern electronic voice communication systems that will give possibilities for users to obtain better QoS of voice communication services.

***The object of research.*** The research object is analysis of CELP codecs and speech quality produced by them.

***Aim and tasks of the work.*** The aim of the work is to investigate and propose ways for improvement of the performance of CELP voice codecs for coding multilingual speech.

In order to achieve the goal, the following problems had to be solved:

1. Perform analytical review on aspects of multilingual CELP voice coding.
2. Create the speech records of Lithuanian language database according to ITU-T Rec. P.50 App. 1. and suitable for tests of performance voice codecs.
3. Investigate the performance AMR and Speex codecs for coding Arabic and Lithuanian speech.
4. Propose and prove techniques for improvement codecs' quality of in multilingual speech coding.

***Methodology of research.*** To investigate the object, following methodologies are applied: statistical analysis and analytical approach; computer modeling and simulation Matlab software and C++ programming language; active experimental tests.

***Scientific novelty.*** The aspects of scientific novelty on theoretical and experimental investigation of the performance of CELP voice codecs over different languages are as follows:

1. Assessment of quality in sense of decoded speech obtained during experimental tests of AMR and Speex codecs' for decoding speech of English, Lithuanian and Arabic languages.
2. PESQ and 3SQM speech quality evaluation algorithms assessment obtained during experimental tests of their suitability to evaluate a quality of speech.
3. The Speex codec codebook adaptation technique that improves its performance in sense of speech quality while coding Lithuanian and Arabic speech.

**Practical value.** Created speech records of Lithuanian language database according to ITU-T Rec. P.50 App. 1. is applicable for tests of performance voice communication systems.

The experimental test results of performance of CELP codecs may be used for specific voice codec performance evaluation and adaptation.

The adapted Speex codec's codebook could be employed for use in VoIP communication software.

The results of performed research were used in one scientific project supported by Lithuanian State Science and Study Fund, Investigation of Lithuanian Language Influence on Coding of Speech signals, Scientific project Nr. T-89/09, (Reg. Nr. T-09022).

### ***Defended statements***

1. The quality of Arabic and Lithuanian speech coding using CELP based voice codecs depends on specific language.
2. AMR and Speex codecs quality evaluated using objective and subjective measures show bias towards English language.
3. Fixed codebook adaptation for Lithuanian and Arabic language improves quality of speech coded by Speex codec and Speex codec's performance stability.

### ***Approval of the results***

3 scientific articles have been published in reviewed scientific publications by the author within scope of research:

- 2 in reviewed ISI Web of Science journals;
- 1 in other reviewed scientific editions.

The author has made 4 presentations at scientific conferences:

- The 9th IEEE International Symposium on Signal Processing and Information Technology. December 14–17, 2009 – Ajman, UAE.
- International conference “ELECTRONICS” in 2007 and 2008.
- Conference “Science – Future of Lithuania” 2007.

***The scope of the scientific work.*** The dissertation consists of introduction, five chapters and conclusions. Dissertation includes 102 pages, 20 figures, 9 tables and one annex.

Chapter 1 reviews the work done in the field of influence of language on coding, by introducing CELP based codes, and previous researches about the topic in hand. As well, it concludes in formulating of main objective and tasks of present investigation. Chapter 2 introduces the quality measuring techniques

categorized by subjective and objective. Moreover, it contains our primary experiment on the codecs and quality assessment techniques. Chapter 3 presents the construction of the Lithuanian speech database. After concluding the results in Chapter 2, we found ourselves in need of a well-formed speech database for Lithuanian language to carry out our experiments on. Chapter 4 presents the experimental study on the performance of CELP based codes under different languages. Chapter 5 suggests a solution the codecs to perform better under different languages. General conclusions as well as recommendations for further research summarize the present study. It is followed by an extensive list of references and a list of 3 publications by the author on the topic of the dissertation.

### ***General conclusions***

1. Performed investigations on performance of voice codecs show that the quality of Lithuanian and Arabic speech coding using CELP based voice codecs depends on specific language.
2. AMR and Speex codecs quality show bias towards English language, by exhibiting about 0.2 points higher average PESQ MOS values and about 0.05 lower spectral distortions.
3. AMR and Speex codecs subjective quality MOS is on average 0.22 lower for Lithuanian and 0.18 for Arabic languages in comparison with English language.
4. Fixed (innovation) codebook adaptation for Lithuanian and Arabic languages improves Speex codecs quality while coding Lithuanian and Arabic speech on average about 0.2 points together reducing its deviation about 0.1 points or 7–10 %.

### **About the author**

Mohamad Itani was born on the 25th of May 1982 in Beirut the capital of Lebanon. In 2000 he entered the AUL university in Lebanon and in 2003 obtained Bachelor degree and in 2005 Master degree. In 2006 Mohamad Itani came to Lithuania and entered VTGU to continue doctoral studies at Department of Telecommunication Engineering. Also, currently he is working in the AUL University of Lebanon as instructor.

Mohamad ITANI

ANALYSIS AND IMPROVEMENT  
OF MULTILINGUAL SPEECH CODING

Summary of Doctoral Dissertation  
Technological Sciences, Electrical and Electronic Engineering (01T)

Mohamad ITANI

ĮVAIRIŲ KALBŲ KALBOS SIGNALŲ  
KODAVIMO ANALIZĖ IR GERINIMAS

Daktaro disertacijos santrauka  
Technologijos mokslai, elektros ir elektronikos inžinerija (01T)

2011 02 07. 1,5 sp. l. Tiražas 70 egz.  
Vilniaus Gedimino technikos universiteto  
leidykla „Technika“,  
Saulėtekio al. 11, 10223 Vilnius,  
<http://leidykla.vgtu.lt>  
Spausdino UAB „Ciklonas“,  
J. Jasinskio g. 15, 01111 Vilnius