VILNIAUS UNIVERSITETAS

Gražina Pyž

LIETUVIŠKŲ FONEMŲ DINAMINIŲ MODELIŲ ANALIZĖ IR SINTEZĖ

Daktaro disertacijos santrauka

Technologijos mokslai, informatikos inžinerija (07T)

Vilnius, 2013

Disertacija rengta 2009–2013 metais Vilniaus universiteto Matematikos ir informatikos institute.

Mokslinis vadovas
doc. dr. Vytautas Slivinskas (Lietuvos edukologijos universitetas, technologijos mokslai, informatikos inžinerija – 07 T).

Mokslinė konsultantė
doc. dr. Virginija Šimonytė (Lietuvos edukologijos universitetas, technologijos mokslai, informatikos inžinerija – 07 T).

Disertacija ginama Vilniaus universiteto Matematikos ir informatikos instituto Informatikos inžinerijos mokslo krypties taryboje:

Pirmininkas

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

Nariai:

prof. habil. dr. Juozas Augutis (Vytauto Didžiojo universitetas, fiziniai mokslai, informatika – 09 P),

prof. habil. dr. Antanas Čenys (Vilniaus Gedimino technikos universitetas, technologijos mokslai, informatikos inžinerija – 07 T),

prof. dr. Robertas Damaševičius (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija – 07 T),

prof. habil. dr. Kazys Kazlauskas (Vilniaus universitetas, fiziniai mokslai, informatika – 09 P).

Oponentai:

prof. dr. Eduardas Bareiša (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija – 07 T),

doc. dr. Olga Kurasova (Vilniaus universitetas, fiziniai mokslai, informatika – 09 P).

Disertacija bus ginama Vilniaus universiteto viešame Informatikos inžinerijos mokslo krypties tarybos posėdyje 2013 m. lapkričio mėn. 19 d. 13 val. Vilniaus universiteto Matematikos ir informatikos instituto 203 auditorijoje.

Adresas: Akademijos g. 4, LT-08663 Vilnius, Lietuva.

Disertacijos santrauka išsiuntinėta 2013 m. spalio mėn. 18 d.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje.

VILNIUS UNIVERSITY

Gražina Pyž

ANALYSIS AND SYNTHESIS OF LITHUANIAN PHONEME DYNAMIC SOUND MODELS

Summary of Doctoral Dissertation

Technological Sciences, Informatics Engineering (07T)

Vilnius, 2013

Doctoral dissertation was prepared at the Institute of Mathematics and Informatics of Vilnius University in 2009–2013.

Scientific Supervisor
Assoc. Prof. Dr. Vytautas Slivinskas (Lithuanian University of Educational Sciences, Technological Sciences, Informatics Engineering – 07 T).

Academic Consultant
Assoc. Prof. Dr. Virginija Šimonytė (Lithuanian University of Educational Sciences, Technological Sciences, Informatics Engineering – 07 T).

The dissertation will be defended at the Council of the Scientific Field of Informatics Engineering at the Institute of Mathematics and Informatics of Vilnius University:

Chairman

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

Members:

Prof. Dr. Habil. Juozas Augutis (Vytautas Magnus University, Physical Sciences, Informatics – 09 P),
Prof. Dr. Habil. Antanas Čenys (Vilnius Gediminas Technical University, Technological Sciences, Informatics Engineering – 07 T),
Prof. Dr. Robertas Damaševičius (Kaunas University of Technology, Technological Sciences, Informatics Engineering – 07 T),
Prof. Dr. Habil. Kazys Kazlauskas (Vilnius University, Physical Sciences, Informatics – 09 P).

Opponents:

Prof. Dr. Eduardas Bareiša (Kaunas University of Technology, Technological Sciences, Informatics Engineering – 07 T),
Assoc. Prof. Dr. Olga Kurasova (Vilnius University, Physical Sciences, Informatics – 09 P).

The dissertation will be defended at the public meeting of the Council of the Scientific Field of Informatics Engineering in the auditorium number 203 at the Institute of Mathematics and Informatics of Vilnius University, at 1 p. m. on 19th of November 2013.

Address: Akademijos st. 4, LT-08663 Vilnius, Lithuania.

The summary of the doctoral dissertation was distributed on 18th of October 2013.

A copy of the doctoral dissertation is available for review at the Library of Vilnius University.

1. Įvadas

Tyrimų sritis

„Tekstas į šneką“ (žymima TTS pagal anglišką šio termino atitikmenį „Text-to-speech“) sistema įvestų žodžių seką konvertuoja į šneką (SIL, 2004). TTS sistema gali būti naudojama daugeliu atvejų. Šios sistemos dėka, balsu gali būti perskaitytas bet koks tekstas iš tinklalapių, navigacijos, vertimo ir kitų programų. Sistema gali būti puiki priemonė mokantis taisyklingo žodžių tarimo, užsienio kalbų, lavina klausymo įgūdžius. TTS sistema leidžia atlikti kelias užduotis vienu metu, tuo tarpu dėmesys gali būti skiriamas skaitymo medžiagai. Tai yra puiki pagalbinė priemonė žmonėms turintiems skaitymo sunkumų, ar žmonėms su regėjimo sutrikimais.

Kalbos sintezatoriaus kūrimas yra be galo sudėtingas uždavinys. Įvairių šalių mokslininkai bando automatizuoti kalbos sintezę. Iki šiol nėra automatinės lietuviškos TTS sistemos atitinkančios žmogaus šneką. Komercinės TTS sistemos nepalaiko lietuvių kalbos. Lietuviško sintezatoriaus kūrimo poreikis išlieka. Lietuvišką kalbos sintezatorių yra sukūręs P. Kasparaitis (P. Kasparaitis, 2001). Jame yra realizuotas konkatenacinės sintezės metodas. Konkatenacinė sintezė remiasi į duomenų bazę įrašytais natūralios kalbos segmentais, kurie sintezės metu yra jungiami į žodžius. Vienas iš pagrindinių konkatenacinės sintezės trūkumų yra tas, kad duomenų bazė turi būti pakankamai didelė. Tuo tarpu tai reikalauja didelių kompiuterio resursų. Jei žodis nėra duomenų bazėje, tai jis negali būti susintezuotas. Sintezuoti garsai nepasiekia natūralios kalbos kokybės dėl trikdžių atsiradusių ant sujungimų ribų. Formantinėje sintezėje žmogaus šnekos įrašai nėra naudojami sintezavimo metu. Sintezuotos šnekos išėjimas yra sukuriamas naudojant adityvią sintezę ir akustinį modelį.

Formantiniai sintezatoriai turi privalumų prieš konkatenacinius. Sintezuota kalba yra pakankamai suprantama net sintezuojant dideliu greičiu. Galima kontroliuoti sintezuotos kalbos prozodijos aspektus: intonaciją, ritmą, kirtį. Pagrindinis formantinės sintezės trūkumas yra tas, kad garsai gauti sintezuojant šiuo metodu skamba nenatūraliai, panašiai kaip roboto šneka. Šiame darbe daroma prielaida, kad formantinio sintezatoriaus modeliai yra pernelyg paprasti. Siekiant sumažinti sintetinį skambėjimą, būtina kurti naujus kalbos garsų matematinius modelius. Balsių ir pusbalsių fonemų modeliai yra šios problemos dalis.

Problemos aktualumas

Garsai, sintezuoti formantinės sintezės metodu, skamba nenatūraliai (panašiai kaip roboto šneka). Siekiant sumažinti sintetinį skambėjimą, būtina kurti naujus kalbos garsų matematinius modelius, kurie gali būti naudojami kaip sintezatoriaus pagrindas.

Disertacijos tyrimo objektas

Disertacijos tyrimo objektas yra dinaminiai lietuviškos šnekos balsių ir pusbalsių fonemų modeliai.

Darbo tikslas ir uždaviniai

Darbo tikslas – sukurti lietuviškos šnekos balsių ir pusbalsių fonemų dinaminius modelius ir nustatyti perėjimus tarp fonemų tam, kad būtų galima sujungti šiuos modelius.

Siekiant iškelto tikslo buvo sprendžiami šie uždaviniai:

- susipažinti su kalbos gamybos aparatais, pagrindiniais kalbos sintezės metodais ir esamomis „tekstas į šneką“ sistemomis lietuvių ir kitoms kalboms;
- išanalizuoti lietuvių kalbos balsių ir pusbalsių garsus ir nustatyti jų pagrindines charakteristikas;
- nustatyti, kokie modeliai tinkamiausi lietuvių kalbos garsams aprašyti;
- sukurti lietuvių kalbos balsių ir pusbalsių fonemų matematinius modelius;
- sukurti perėjimus tarp fonemų, kad būtų galima sujungti balsių ir pusbalsių modelius;
- įvertinti siūlomų modelių tikslumą eksperimentiškai.

Tyrimo metodai

- Skaitmeninis signalų apdorojimas,
- sistemų teorija,
- optimizavimo metodai,
- matricų teorija,
- matematinė statistika,
- programavimas Matlab aplinkoje,
- programavimas C # kalba.

Darbo mokslinis naujumas

- Daugelio-įėjimų ir vieno-išėjimo (žymima MISO pagal anglišką šio termino atitikmenį „Multiple-Input and Single-Output”) sistema, kurios impulsinės charakteristikos aprašomos trečios eilės kvazipolinominiais modeliais, o amplitudės kinta laike, yra naudojama balsių ir pusbalsių fonemų modeliavimui.
- Sukurtas naujas parametrų įvertinimo iš sąsūkos duomenų algoritmas, pagrįstas Levenbergo ir Markvarto metodu.
- Pasiūlytas naujas pagrindinio tono patikslinimo algoritmas.
- Sukurtas naujas metodas, kuris leidžia automatiškai parinkti reprezentatyvų periodą.
- Sukurti perėjimai tarp balsių ir pusbalsių fonemų modelių.

Sukurtos fonemų modeliavimo sistemos privalumas yra tas, kad ji yra automatinė ir nepriklauso nuo kalbėtojo ir fonemų.

Darbo rezultatų praktinė reikšmė

Pasiūlyti balsių ir pusbalsių fonemų dinaminiai modeliai gali būti panaudoti kuriant formantinį kalbos sintezatorių. Fonemų modeliai taip pat gali būti pritaikyti kitoms

problemoms spręsti, pavyzdžiui, gydant kalbos sutrikimus, mokantis užsienio kalbų ar taisyklingo žodžių tarimo.

Ginamieji teiginiai

1. Diskretaus laiko stacionari MISO sistema yra naudojama balsių ir pusbalsių fonemų modeliavimui. Kiekvieno kanalo impulsinė charakteristika yra aprašoma trečios eilės kvazipolinominiu modeliu.
2. Siekiant išgauti natūralesnį sintezuotų garsų skambesį, yra svarbu naudoti ne tik aukštos eilės modelius, bet ir sudėtingus įėjimų scenarijus.
3. Sintezuotų balsių ir pusbalsių fonemų kokybė yra pakankamai gera.
4. Žodis, sudarytas iš balsių ir pusbalsių, sintezuotas siūlomu metodu skamba gerai ir jį sunku atskirti nuo realaus. Sintezuoto žodžio kokybė žymiai pagerėjo dėl įvestų perėjimų tarp įėjimų.

Darbo rezultatų aprobavimas

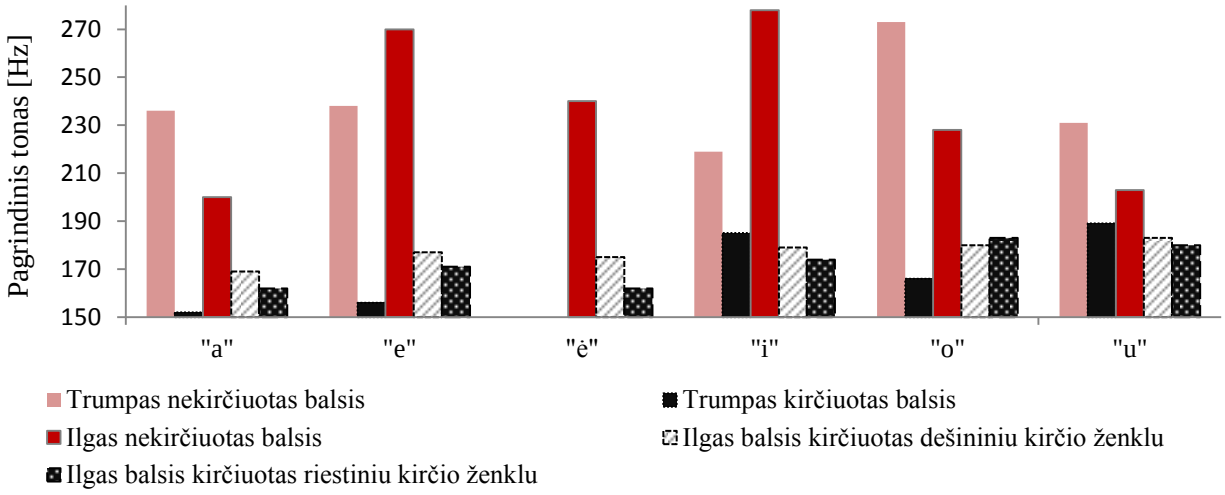
Tyrimų rezultatai publikuoti 6 periodiniuose recenzuojamuose mokslo žurnaluose. Tyrimų rezultatai pristatyti ir aptarti 21 nacionalinėje ir tarptautinėje konferencijoje Lietuvoje ir užsienyje.

Darbo apimtis

Disertaciją sudaro 5 skyriai, literatūros sąrašas ir priedai. Disertacijos skyriai: Įvadas, Kalbos sintezės pagrindai, Fonemų modeliavimo sistema, Eksperimentiniai tyrimai, Bendrosios išvados. Papildomai disertacijoje pateikti lentelių, paveikslų bei naudotų žymėjimų ir santrumpų sąrašai. Disertacijos apimtis 114 puslapių (be priedų), kuriuose pateikti 47 paveikslai, 78 formulės ir 19 lentelių. Disertacijoje remtasi 83 literatūros šaltiniais.

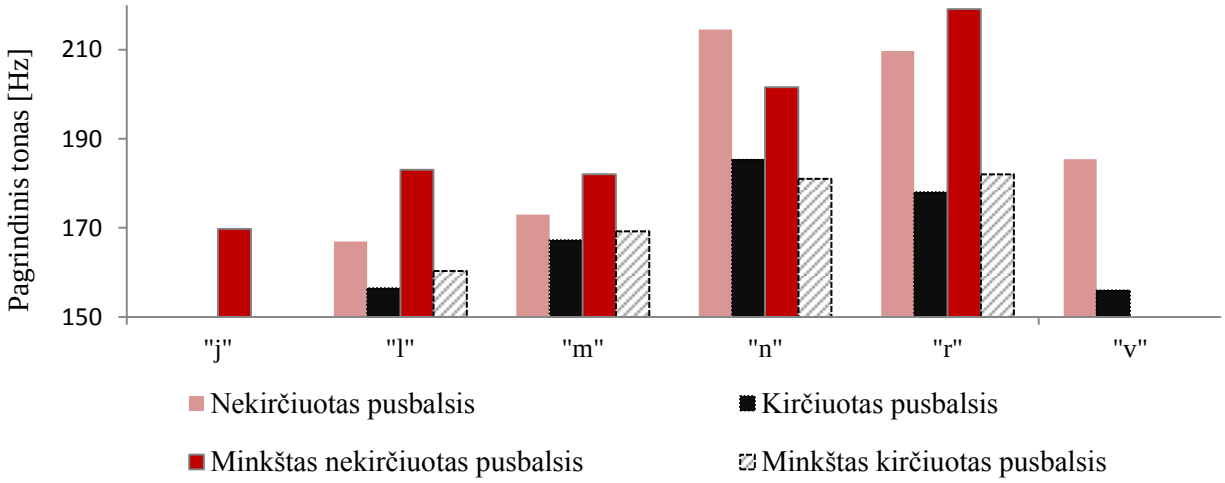
2. Kalbos sintezės pagrindai

Lietuvių kalboje išskiriami 92 kalbos vienetai, kurie vadinami fonemomis. Iš jų 28 yra balsių fonemos, 19 – pusbalsių fonemos. Visų balsių ir pusbalsių fonemų pagrindinio tono dydžiai buvo nustatyti panaudojus po 50 kiekvienos fonemos vyriškų ir po 50 moteriškų įrašų. Balsių fonemų pagrindinių tonų vidutinės reikšmės pavaizduotos 1 pav.



1 pav. Balsių fonemų pagrindinio tono kitimo tendencijos

Pusbalsių fonemų pagrindinių tonų vidutinės reikšmės pavaizduotos 2 pav.



2 pav. Pusbalsių fonemų pagrindinio tono kitimo tendencijos

Iš 1 pav. ir 2 pav. matome, kad pagrindiniai tonai nekirčiuotų balsių ir pusbalsių fonemų yra didesni už tų pačių kirčiuotų balsių ir pusbalsių fonemų pagrindinius tonus.

3. Fonemų modeliavimo sistema

Lietuviškų balsių ir pusbalsių fonemų modeliavimo sistema yra pateikta. Du sintezės metodai yra pasiūlyti: harmoninis ir formantinis.

Balsių ir pusbalsių fonemų signalo išskaidymas į harmonikas

Balsiai ir pusbalsiai yra periodiniai signalai, todėl juos galima išskaidyti į paprastesnės formos signalus. Tam, kad signalą išskaidytume į harmonikas, įvertinamas signalo pagrindinis tonas f_0. Įvertinus pagrindinį toną skaičiuojama signalo Furjė transformacija ir signalo amplitudinė charakteristika dalinama į dažnių juostas pavaizduotas lentelėje:

1 lentelė. Amplitudinės dažnuminės charakteristikos padalinimas į dažnių juostas

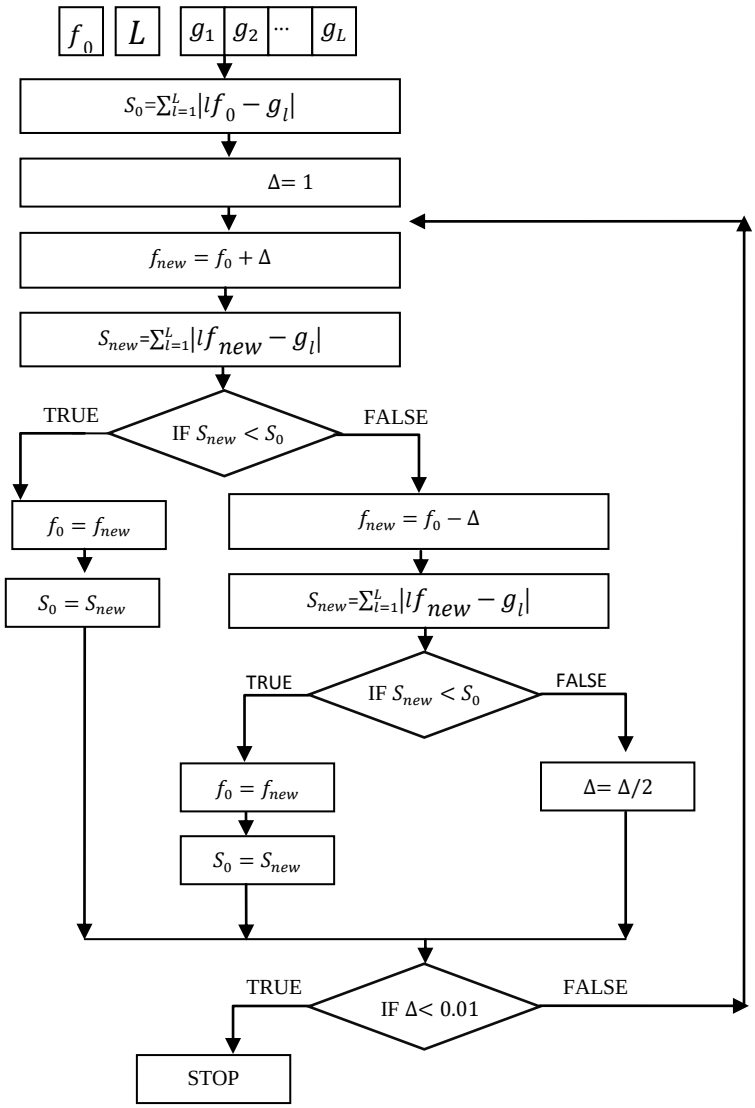
1-oji juosta	$[0.5f_0, 1.5f_0)$
2-oji juosta	$[1.5f_0, 2.5\,f_0)$
…	
L-oji juosta	$[\big((L-1)+0.5\big)f_0, (L+0.5)f_0)$

L yra didžiausias sveikas skaičius, kuriam galioja nelygybė $(L+0.5)f_0 \leq 6000$.

Atlikus padalinimą, jis yra patikslinamas. Nustatomi kiekvienos juostos maksimumai $a_1, a_2, \dots, a_L$ ir dažniai, atitinkantys šiuos maksimumus $g_1, g_2, \dots, g_L$. Nustatyti dažniai lyginami su pagrindinio tono kartotiniais dažniais $f_0, 2f_0, \dots, Lf_0$. Mūsų tikslas – rasti tokį pagrindinį toną f_0, kuris minimizuotų atstumą tarp dažnių:

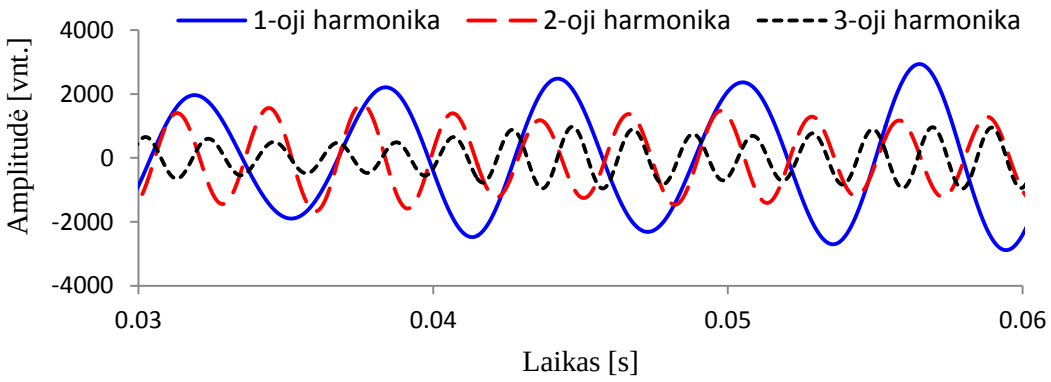
$$S_0 = \sum_{l=1}^{L} |lf_0 - g_k|. \tag{1}$$

Pagrindinio tono patikslinimo algoritmo blokinė diagrama pavaizduota 3 pav.



3 pav. Pagrindinio tono patikslinimo algoritmo blokinė diagrama

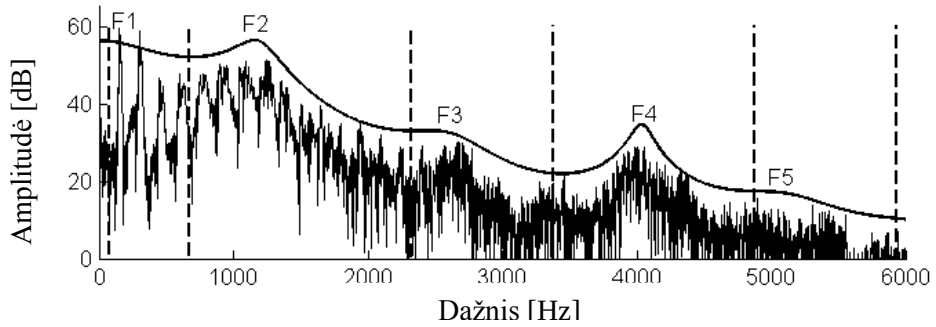
Gavus optimalią pagrindinio tono reikšmę $\widetilde{f_0}$, fonemos signalo amplitudinė dažnuminė charakteristika iš naujo padalinama į dažnių juostas. Skaičiuojama atvirkštinė Furjė transformacija. Pirmos trys fonemos /a:˜/ harmonikos pavaizduotos 4 pav.



4 pav. Pirmosios trys fonemos /a:˜/ harmonikos (kaip žodyje ãčiū)

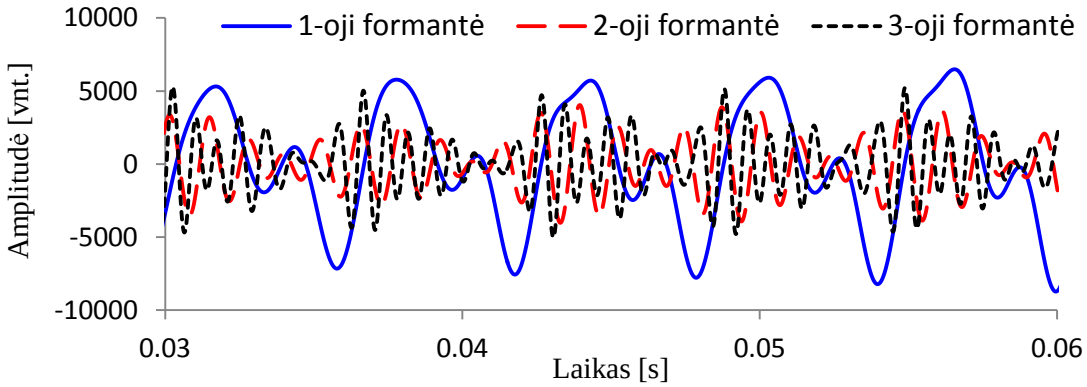
Balsių ir pusbalsių fonemų signalo išskaidymas į formantes

Formantė yra apibrėžiama kaip spektro gaubtinės maksimumas. Spektro gaubtinės maksimumas nustatyti naudojamas Tiesinės prognozės kodavimo (žymima LPC pagal angliską šio termino atitikmenį „Linear Predictive Coding “) metodas (Markel ir Gray, 1976). Dažnių reikšmės atitinkančios spektro gaubtinės minimumus yra laikomos padalinimo taškais. Labai svarbu, kad padalinimas į formantes sutaptų su padalinimu į harmonikas, t.y. viena harmonikos dalis negali priklausyti vienai formantei, o kita harmonikos dalis kitai formantei. Siūloma gretimas harmonikas apjungti į grupes. Padalinimas į formantes pavaizduotas 5 pav.



5 pav. Fonemos /a:˜/ spektras (kaip žodyje ãčiū)

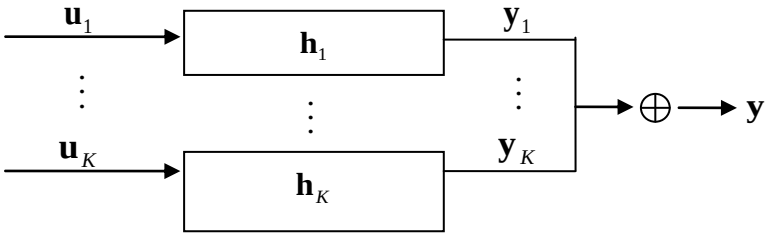
Pirmos trys fonemos /a:˜/ formantės pavaizduotos 6 pav.



6 pav. Pirmosios trys fonemos /a:˜/ formantes (kaip žodyje ãčiū)

Balsių ir pusbalsių fonemų modelis

Balsių ir pusbalsių fonemos modeliuojamos MISO sistema susidedančia iš lygiagrečiai sujungtų vieno-įėjimo ir vieno-įsėjimo (žymima SISO pagal anglišką šio termino atitikmenį „Single-Input and Single-Output") sistemų. SISO sistemų skaičius lygus komponenčių į kurias išskaidytas signalas skaičiui t. y lygus harmonikų skaičiui – harmoninės sintezės atveju, arba formančių skaičiui – formantinės sintezės atveju. MISO sistemos diagrama pavaizduota 7 pav.



7 pav. Fonemos signalo modelio diagrama

MISO sistemos išėjimas **y** yra lygus SISO sistemų išėjimų sumai:

$$\mathbf{y} = \mathbf{y}_1 + \ldots + \mathbf{y}_K , \tag{2}$$

čia

$$\mathbf{y}_k = \mathbf{u}_k * \mathbf{h}_k = \sum_{i=0}^{\infty} u_k(n-i)h_k(i) , \tag{3}$$

$\boldsymbol{h}_k(n)$ – k-ojo kanalo impulsinė charakteristika, $\boldsymbol{u}_k(n)$ – k-ojo kanalo įėjimų seka.

Impulsinė charakteristika aprašoma antro laipsnio kvazipolinominiu modeliu. Antro laipsnio kvazipolinomo matematinė išraiška pateikta žemiau:

$$q(t) = e^{\lambda t}(a_1 \sin(2\pi f t + \varphi_1) + a_2 t \sin(2\pi f t + \varphi_2) + a_3 t^2 \sin(2\pi f t + \varphi_3)) , \tag{4}$$

čia $t \in R^+ \cup \{0\}$, $\lambda < 0$ – gesimas, f – dažnis, a_k – amplitudė, $\varphi_k\ (-\pi \le \varphi_k < \pi)$ – fazė.

Tegul $\mathbf{y}^N = [y(0),\ y(1),\ y(2),\ \ldots,\ y(N-1)]^T$ fonemos signalo atskaitos. Apibrėžkime išraiška $\mathbf{\Psi}_k = \mathbf{\Psi}_k(\Lambda_k, \Omega_k)$ tokią $N \times 6$ dydžio matricą:

$$\mathbf{\Psi}_k = [\mathbf{\Psi}_{k1} \quad \mathbf{\Psi}_{k2} \quad \mathbf{\Psi}_{k3}], \tag{5}$$

čia

$$\mathbf{\Psi}_{ki} = \begin{bmatrix} \delta(i-1) & 0 \\ c_{ki} & s_{ki} \end{bmatrix} \qquad \left(\delta(i) = \begin{cases} 1, i = 0 \\ 0, i \neq 0 \end{cases} \right), \tag{6}$$

$$c_{ki} = \left[e^{\Lambda_k} \cos\Omega_k, \quad 2^{i-1} e^{2\Lambda_k} \cos 2\Omega_k, \quad \ldots, \quad (N-1)^{i-1} e^{(N-1)\Lambda_k} \cos(N-1)\Omega_k \right]^T, \tag{7}$$

$$s_{ki} = \left[e^{\Lambda_k} \sin\Omega_k, \quad 2^{i-1} e^{2\Lambda_k} \sin 2\Omega_k, \quad \ldots, \quad (N-1)^{i-1} e^{(N-1)\Lambda_k} \sin(N-1)\Omega_k \right]^T \tag{8}$$

ir išraiška $\boldsymbol{\alpha}_k = \boldsymbol{\alpha}(A_{k1}, A_{k2}, A_{k3}, \varphi_{k1}, \varphi_{k2}, \varphi_{k3})$ tokį 6×1 dydžio vektorių:

$$\boldsymbol{\alpha}_k = [A_{k1}\sin(\varphi_{k1}),\ A_{k1}\cos(\varphi_{k1}),\ A_{k2}\sin(\varphi_{k2}),\ A_{k2}\cos(\varphi_{k2}),\ A_{k3}\sin(\varphi_{k3}),\ A_{k3}\cos(\varphi_{k3})]^T. \tag{9}$$

Nesunkiai galima patikrinti, kad vektorius $\mathbf{y}^N$ gali būti užrašytas kaip bazinių signalų matricos $\mathbf{\Psi}$, priklausančios nuo dažnių ir gesimų, ir koeficientų vektoriaus $\boldsymbol{\alpha}$, priklausančio nuo amplitudžių ir fazių, sandauga:

$$\mathbf{y}^N = \mathbf{\Psi}_1\boldsymbol{\alpha}_1 + \ldots + \mathbf{\Psi}_K\boldsymbol{\alpha}_K = \mathbf{\Psi} \cdot \boldsymbol{\alpha}, \tag{10}$$

čia $\mathbf{\Psi} = [\mathbf{\Psi}_1 | \mathbf{\Psi}_2 | \ldots | \mathbf{\Psi}_K]$ ir $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T | \boldsymbol{\alpha}_2^T | \ldots | \boldsymbol{\alpha}_K^T]^T$.

Tegul $\mathbf{y}^M = [y(0),\ y(1),\ y(2),\ \ldots,\ y(M-1)]^T$ vieno fonemos signalo periodo atskaitos (čia M – periodo ilgis).

Disertacijoje daroma prielaida, kad impulsinė charakteristika užgęsta po trijų periodų. Įvedus $M \times 6K$ dydžio bazinių signalų sąsūkos matricą $\mathbf{\Phi} = \mathbf{\Phi}(\boldsymbol{\theta})$:

$$\mathbf{\Phi} = \mathbf{\Psi}(1\!:\!M, :) + \mathbf{\Psi}(M+1\!:\!2M, :) + \mathbf{\Psi}(2M+1\!:\!3M, :), \tag{11}$$

nesunkiai galima patikrinti, kad

$$\mathbf{y}^M = \mathbf{\Phi} \cdot \boldsymbol{\alpha}. \tag{12}$$

Kadangi paprastai išėjimas matuojamas su paklaida, prie išraiškos apibrėžtos formule (12) pridedamas baltas triukšmas:

$$\mathbf{y}^M = \mathbf{\Phi} \cdot \boldsymbol{\alpha} + \mathbf{e}. \tag{13}$$

Turime minimizuoti funkcionalą:

$$r(\mathbf{a}, \boldsymbol{\theta}) = \left\| \mathbf{y}^M - \boldsymbol{\Phi}(\boldsymbol{\theta})\boldsymbol{\alpha} \right\|^2. \quad (14)$$

Straipsnyje (Golub ir Pereyra, 1973) parodyta, kad funkcionalo $r(\mathbf{a}, \boldsymbol{\theta})$ minimizavimo uždavinys susiveda į tokio functionalo minimizavimą:

$$r_2(\boldsymbol{\theta}) = \left\| P^{\perp}_{\boldsymbol{\Phi}(\boldsymbol{\theta})} y^M \right\|^2, \quad (15)$$

čia

$$P^{\perp}_{\boldsymbol{\Phi}(\boldsymbol{\theta})} = \mathrm{I}_M - \boldsymbol{\Phi}\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}} = I_M - \boldsymbol{\Phi}\boldsymbol{\Phi}^{+}. \quad (16)$$

Functionalo apibibrėžto išraiška (15), minimizavimui yra naudojamas netiesinės optimizacijos Levenbergo ir Markvarto metodas. Kitame skyriuje pateikiamas naujas parametrų įvertinimo iš sąsūkos duomenų algoritmas. Parametrų įvertinimo iš nesąsūkos duomenų algoritmas buvo pateiktas darbe (Šimonytė ir Slivinskas, 2007).

Modelio parametrų įvertinimas

Levenbergo ir Markvarto metodo (Levenberg, 1944; Marquardt, 1963) iteracinė parametrų įvertinimo lygtis užrašoma tokia išraiška:

$$\boldsymbol{\theta}^{l+1} = \boldsymbol{\theta}^{l} - \left(\mathbf{V}^T\left(\boldsymbol{\theta}^l\right)\mathbf{V}\left(\boldsymbol{\theta}^l\right) + c_l \mathbf{I}_{2K}\right)^{-1}\mathbf{V}^T\left(\boldsymbol{\theta}^l\right)\mathbf{b}\left(\boldsymbol{\theta}^l\right), \quad l = 0, 1, \ldots, \quad (17)$$

čia

$$\mathbf{V}(\boldsymbol{\theta}) = \mathscr{D}\left(\mathbf{P}^{\perp}_{\boldsymbol{\Phi}(\boldsymbol{\theta})}\right)\mathbf{y}^M \quad (18)$$

simbolis $\mathscr{D}$ – diferencijavimo operacija, t. y. $\mathscr{D} = \dfrac{\partial}{\partial \boldsymbol{\theta}}$, $\boldsymbol{\theta}^l$ – parametro $\boldsymbol{\theta}$ reikšmė *l*-oje iteracijoje,

$$\mathbf{b}(\boldsymbol{\theta}) = \mathbf{P}^{\perp}_{\boldsymbol{\Phi}(\boldsymbol{\theta})}\mathbf{y}^M, \quad (19)$$

yra $M \times 1$ dydžio vektorius, $\mathbf{I}_{2K}$ – $2K \times 2K$ dydžio vienetinė matrica, c_l – Levenbergo ir Markvarto algoritmo konstanta *l*-oje iteracijoje.

Projektoriaus diferencialo skaičiavimas yra gana sudėtingas uždavinys. Mokslininkai G. Golub ir V. Pereyra savo straipsnyje (Golub ir Pereyra, 1973) parodė, kad:

$$\mathscr{D}\left(\mathbf{P}^{\perp}_{\boldsymbol{\Phi}(\boldsymbol{\theta})}\right) = -\mathbf{P}^{\perp}_{\boldsymbol{\Phi}(\boldsymbol{\theta})}\mathscr{D}(\boldsymbol{\Phi})\mathbf{B} - (\mathbf{P}^{\perp}_{\boldsymbol{\Phi}(\boldsymbol{\theta})}\mathscr{D}(\boldsymbol{\Phi})\mathbf{B})^T, \quad (57)$$

čia **B** yra matricos ***Φ*** apibendrinta atvirkštinė matrica.

Parametrų įvertinimo algoritmas

Pradiniai duomenys:

1. Indeksas *k* – komponenčių skaičius.

2. Charakteringojo periodo atskaitų seka **y**.
3. Pradinių parametrų vektorius $\boldsymbol{\theta}^0 = \left[\Omega_k^0 \quad \Lambda_k^0\right]^T$.
4. Pradinė Levenbergo ir Markvarto konstantos reikšmė $c_0 = 0.001$.
5. Pradinis skaičius $l = 0$.
6. Didžiausias leistinų iteracijų skaičius $l_{\max}$.
7. Parametrų įvertinimo santykinis tikslumas procentais $\varepsilon_{\min}$.
8. Pradinis parametrų įvertinimo santykinis tikslumas procentais ε_{-1}, pavyzdžiui $\varepsilon_{-1} = 100$.
9. Leistina gesimo koeficiento mažėjimo riba $\Lambda_{\lim}$, pavyzdžiui $\Lambda_{\lim} = -0.006$.
10. Didžiausia leistina Levenberg-Marquardt konstantos reikšmė $c_{\max}$, pavyzdžiui $c_{\max} = 10^{10}$.
11. Sustojimo kriterijus $\varepsilon_l < \varepsilon_{\min}$ arba $l \geq l_{\max}$ arba $c_l > c_{\max}$ arba $\Lambda_k^l > \Lambda_{\lim}$.

1 žingsnis. Apskaičiuojama bazinių vektorių matrica $\boldsymbol{\Psi}$ pagal formules (5)-(8).

2 žingsnis. Apskaičiuojama bazinių vektorių sąsūkos matrica $\boldsymbol{\Phi}$ pagal formulę (11)

3 žingsnis. Naudojant matricos $\boldsymbol{\Phi}$ QR dekompoziją, randama apibendrinta atvirkštinė matrica **B**.

4 žingsnis. Apskaičiuojamas projektorius į triukšmo erdvę $P_{\boldsymbol{\Phi}(\boldsymbol{\theta})}^{\perp}$ pagal formulę (16).

5 žingsnis. Randama duomenų vektoriaus **y** projekcija į triukšmo erdvę pagal formulę (19)

6 žingsnis. Nustatoma modelio paklaida $\varepsilon_l = r_2\left(\boldsymbol{\theta}^l\right)/\|y\|^2 \cdot 100\%$, čia $r_2\left(\boldsymbol{\theta}^l\right)$ apskaičiuojamas pagal formulę (15).

7 žingsnis. Jei ($\varepsilon_l < \varepsilon_{l-1}$) tada

$$c_l = c_{l-1}/10$$

priešingu atveju

$$c_l = c_{l-1} \cdot 10$$

$$\boldsymbol{\theta}^l = \boldsymbol{\theta}^{l-1}$$

pereinama prie *11 žingsnio*.

8 žingsnis. Apskaičiuojamos matricos $\boldsymbol{\Phi}$ dalinės išvestinės pagal Λ_k ir Ω_k.

9 žingsnis. Apskaičiuojama projektoriaus į triukšmo erdvę išvestinė $\mathscr{D}\left(\mathbf{P}_{\boldsymbol{\Phi}\left(\boldsymbol{\theta}^l\right)}^{\perp}\right)$.

10 žingsnis. Apskaičiuojama matrica $V(\boldsymbol{\theta}^l)$ pagal formulę (18).

11 žingsnis. Apskaičiuojamas parametrų vektorius $\boldsymbol{\theta}^{l+1}$ pagal formulę (17) ir grįžtama prie *1 Žingsnio*.

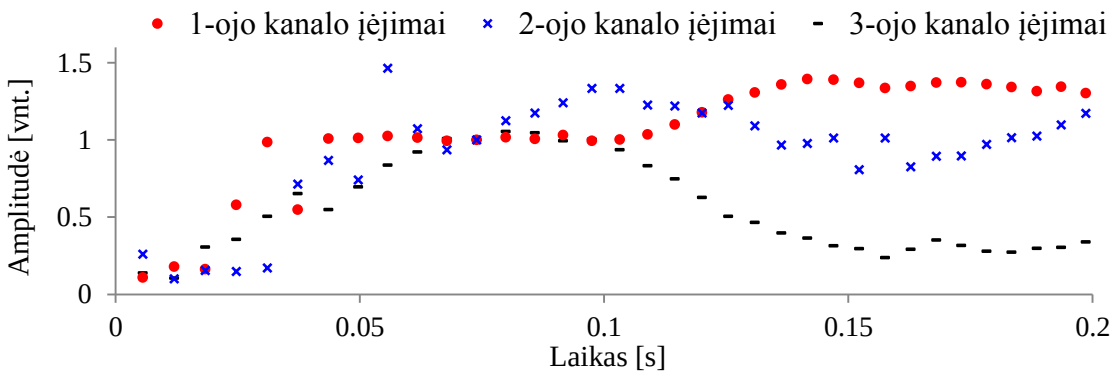
PABAIGA

Charakteringojo periodo išrinkimas

Periodas, kurio amplitudė didžiausia, laikomas charakteringuoju periodu. Charakteringojo periodo ieškoma iš 60 % viso signalo atskaitų (t. y. atmetama 20 % signalo atskaitų iš signalo pradžios ir 20 % - iš signalo pabaigos).

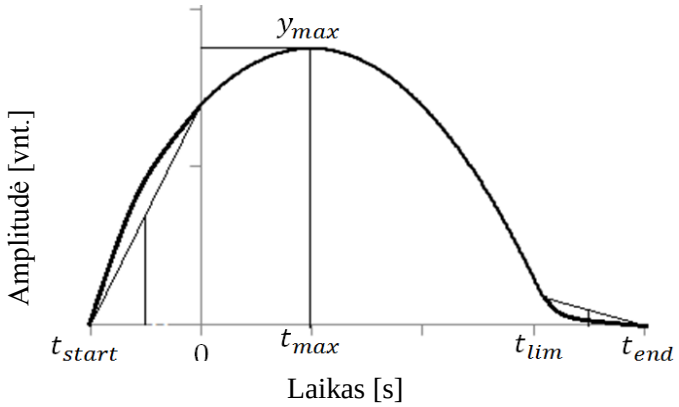
Sistemos įėjimų parinkimas

Jei į sistemos įėjimą paduosime vienetinius impulsus vienodais laiko tarpais, išėjime gausime signalą su identiškais periodais. Toks signalas skambės nenatūraliai. Tam, kad nustatytume sistemos įėjimus, fonemos signalas dalinamas į periodus ir užfiksuojami padalinimo taškai. Pagal užfiksuotus padalinimo taškus, visos komponentės dalinamos į tokius pat periodus. Perioduose surandami lokalūs maksimumo taškai. Rastos reikšmės surašomos į matricą. Matricos *k*-ojo stulpelio reikšmės yra *k*-os SISO sistemos įėjimai. Nustačius visus maksimumus, atliekamas matricos reikšmių normavimas. Tuo tikslu visos matricos stulpelių reikšmės dalinamos iš matricos eilutės, atitinkančios charakteringąjį periodą, reikšmių. Pirmų trijų fonemos /a:˜/ MISO sistemos kanalų įėjimai pavaizduoti 8 pav.



8 pav. Pirmų trijų fonemos /a:˜/ MISO sistemos kanalų įėjimai (kaip žodyje ãčiū)

Įėjimai aprašomi parabolėmis (9 pav.). Kiekviena parabolė charakterizuojama tokiais parametrais: maksimali įėjimo reikšmė, maksimalios reikšmės laiko momentas ir fonemos ilgis.



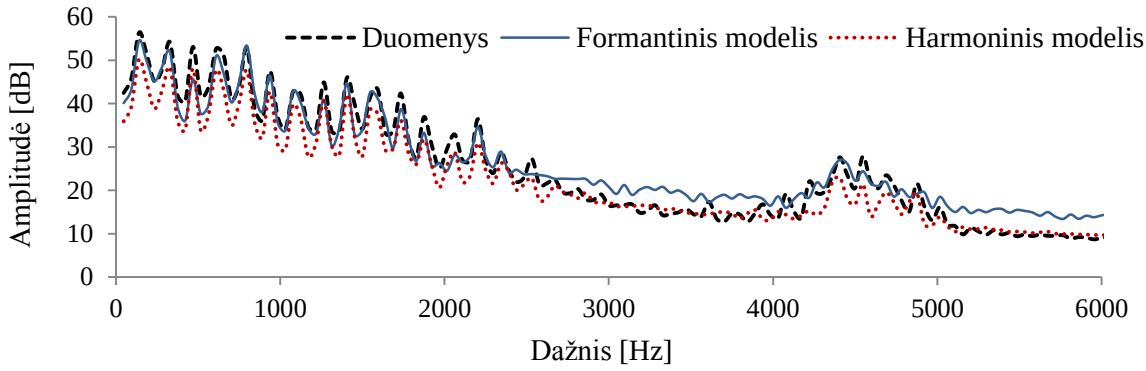
9 pav. Įėjimų kreivė

4. Eksperimentiniai tyrimai

Eksperimentuose naudojamos realių garsų atskaitos. Garsai buvo įrašyti į kompiuterį naudojant mikrofoną ir "Sound Recorder" programą. Garso formato parametrai: PCM 48 kHz, 16 bitų; stereo.

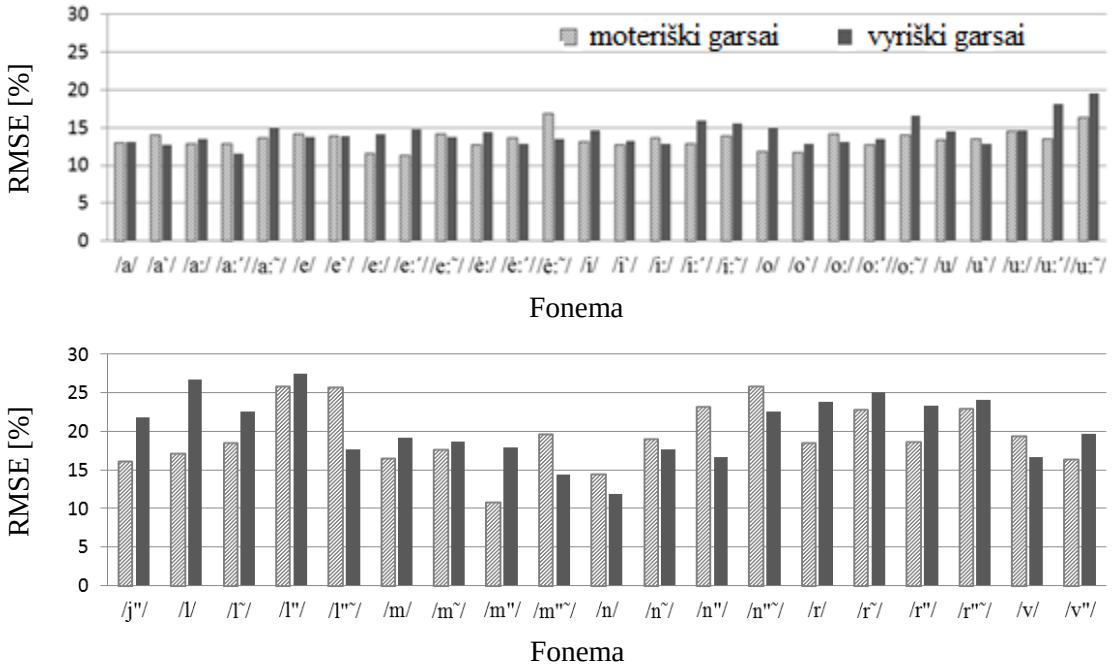
Balsių ir pusbalsių modeliavimas formantiniu ir harmoniniu metodais

Visų balsių ir pusbalsių fonemų modeliavimui buvo parinkta po 50 įrašytų moteriškų fonemų ir po 50 vyriškų fonemų. Garsai sintezuoti harmoniniu metodu buvo palyginti su garsais gautais sintezuojant formantiniu metodu. Tam, kad įvertinti sintezuotų garsų kokybę, skaičiuojamas vidutinis kvadratinis spektras. Tikro ir modelinių signalų spektrų palyginimas pavaizduotas 10 pav.

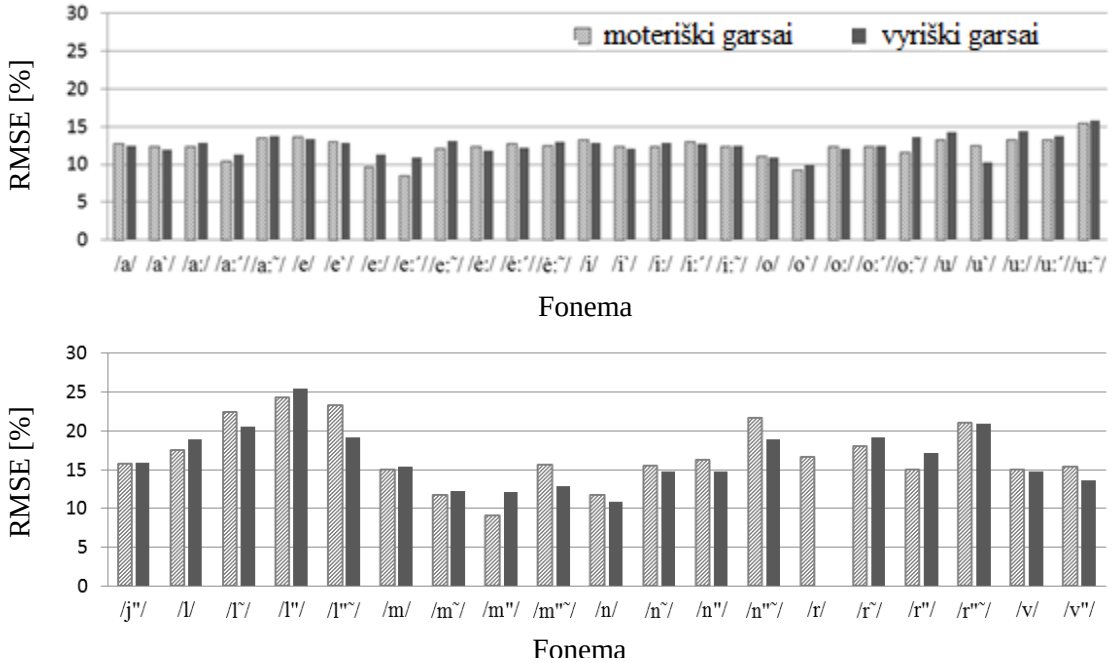


10 pav. Tikros fonemos /a/ ir jos modelių spektrai

Šaknų iš vidutinių kvadratinių paklaidų (žymima RMSE pagal angliską šio termino atitikmenį „Root Mean Square Error“) vidurkiai ir jų pasikliautinumo intervalai visoms balsėms ir pusbalsiams pavaizduoti 11 pav. ir 12 pav.



11 pav. Signalų spektrų RMSE vidurkiai (garsai sintezuoti formantiniu metodu): viršutinis grafikas – balsių fonemos, apatinis grafikas - pusbalsių fonemos



12 pav. Signalų spektrų RMSE vidurkiai (garsai sintezuoti harmoniniu metodu): viršutinis grafikas – balsių fonemos, apatinis grafikas - pusbalsių fonemos

Visų vyriškų ir moteriškų balsių signalų spektrų vidutinių kvadratinių paklaidų vidurkis yra lygus 13.9 % formantinio metodo atveju ir 12.4 % harmoninio metodo atveju. Visų vyriškų ir moteriškų pusbalsių signalų spektrų vidutinių kvadratinių paklaidų vidurkis yra lygus 19.9 % formantinio metodo atveju ir 16.7 % harmoninio metodo atveju.

Vidutinis visų vyriškų ir moteriškų balsių ir pusbalsių fonemų parametrų įvertinimo laikas lygus 16.1 s formantinio metodo atveju ir 37.2 s harmoninio metodo atveju. Vidutinis visų vyriškų ir moteriškų balsių ir pusbalsių fonemų sintezės laikas lygus 0.09 s sintezuojant formantiniu metodu, 0.44 s – sintezuojant harmoniniu metodu.

Balsių ir pusbalsių modelių sujungimas

Į mikrofoną, prijungtą prie kompiuterio, diktorė ištarė žodį „laimė”. Šio žodžio trukmė yra 1.32 s. Impulsinių charakteristikų parametrų įvertinimui kiekvienai fonemai parenkama po vieną periodą.

Įvertinus parametrus, skaičiuojamas parametrų įvertinimo tikslumas. Į sistemos įėjimus paduodami trys vienetiniai impulsai ir gautos paskutinio signalo periodo atskaitos lyginamos su tikro signalo periodo atskaitomis. Skaičiuojama vidutinė kvadratinė paklaida. Parametrų įvertinimo paklaidos pavaizduotos lentelėje žemiau.

2 lentelė. Parametrų įvertinimo paklaidos

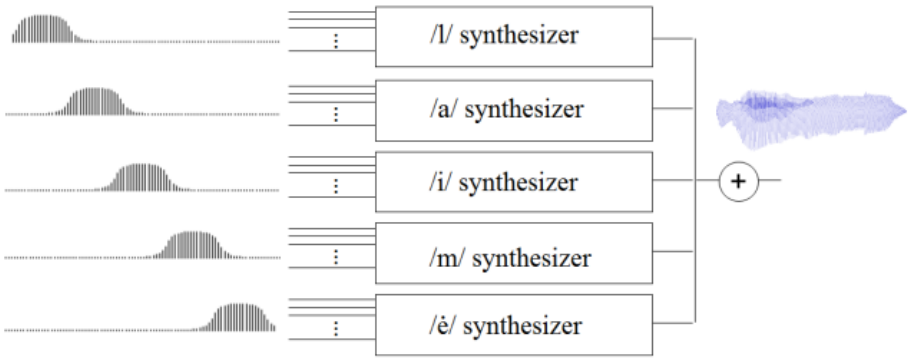
Fonema	RMSE
/l/	0.90 %
/a/	4.69 %
/i/	3.84 %
/m/	1.47 %
/ė/	0.66 %
/e/	2.00 %
/ę/	2.98 %
/j/	0.46 %

Išėjimo signalas $y(n)$ apskaičiuojamas pagal pateiktą sąsūkos formulę:

$$y(n) = \sum_{l=1}^{L} \sum_{k=1}^{K_l} \sum_{i=0}^{N} u_{kl}(n-i) h_{kl}(i), \tag{21}$$

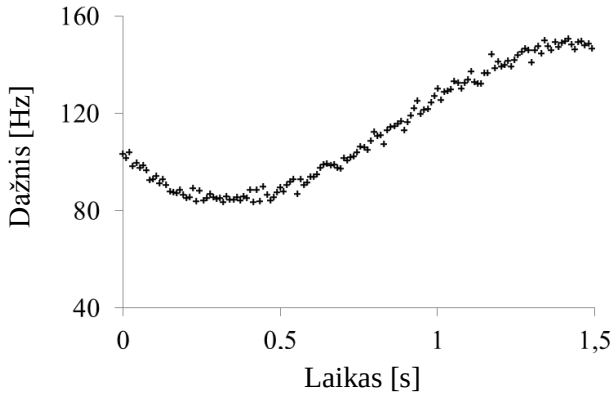
$n = 0, 1, \ldots$.

Žodžio „laimė“ MISO schema pavaizduota 13 pav.



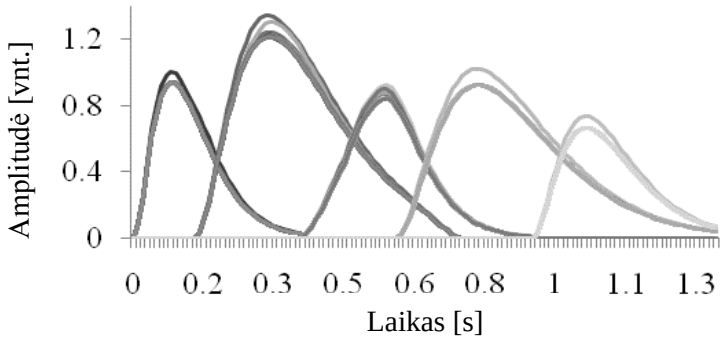
13 pav. MISO modelis žodžio “laimė” modeliavimui

Nustatomas žodžio „laimė“ pagrindinio tono kitimas laike. Pagrindinio tono kitimo trajektorija pavaizduota 14 pav.



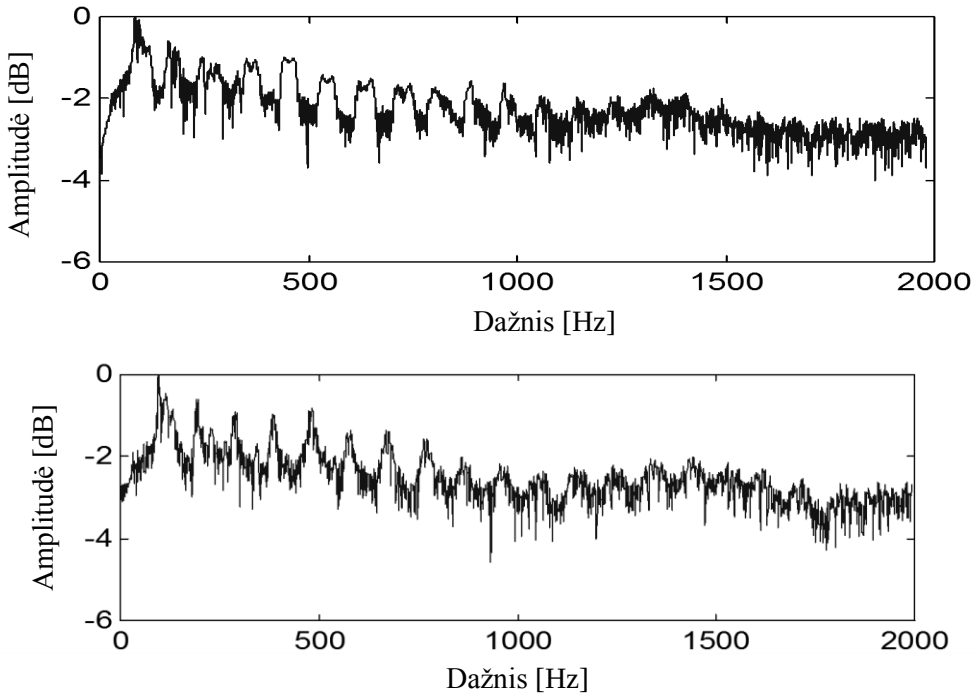
14 pav. Žodžio „laimė“ pagrindinio tono kitimo trajektorija

Pagrindinis tonas nurodo atstumus tarp įėjimų. Taip pat yra būtina žinoti kokius įėjimus parinkti kiekvienos fonemos komponentei. Norint išgauti sistemos įėjimus, signalas praleidžiame pro stačiakampius filtrus, komponentėms nustatytomis dažnių juostomis. Kiekviename fonemos signalą atitinkančiame periode ieškoma lokalių maksimumų taškų. Šio reikšmės yra normuojamos ir pateikiamos kaip sistemos įėjimai. Perėjimai tarp fonemų aprašomi parabolėmis pavaizduotomis 9 pav. 15 pav. parodyta, kaip kiekviena fonema yra žadinama.



15 pav. Žodžio „laimė“ MISO sistemos įėjimai

Tikro ir modelinio žodžio laimė amplitudinės dažnuminės charakteristikos pavaizduotos 16 pav.



16 pav. Tikro (viršutinis brėžinys) ir modelinio (apatinis brėžinys) žodžio „laimė“ amplitudinės dažnuminės charakteristikos

5. Bendrosios išvados

Disertacijos tyrimo objektas yra dinaminiai lietuviškos šnekos balsių ir pusbalsių fonemų modeliai. Tam, kad sukurti balsių ir pusbalsių fonemų modelius, buvo nustatytos pagrindinės šių garsų charakteristikos. Garsams aprašyti buvo pasiūlyta modeliavimo sistema pagrįsta balsių ir pusbalsių fonemų matematiniu modeliu bei pagrindinio tono ir įėjimų nustatymo automatine procedūra. Disertacijoje du sintezės metodai sukurti: harmoninis ir formantinis.

Šioje disertacijoje atlikti tyrimai leido padaryti tokias išvadas:

1. Lietuvių kalboje yra devyniasdešimt dvi fonemos. Dvidešimt aštuonios iš jų yra balsių fonemos, devyniolika – pusbalsių fonemos. Balsiai ir pusbalsiai yra periodiniai signalai.
2. Nekirčiuotų balsių ir pusbalsių fonemų pagrindiniai tonai yra didesni už tų pačių kirčiuotų balsių ir pusbalsių fonemų pagrindinius tonus.
3. Pagrindinio tono įverčiai, gauti MUSIC metodu, yra mažiau išsibarstę aplink savo vidurkį, palyginus su pagrindinio tono įverčiais, gautais DFT metodu. Tyrimas buvo atliktas su vyriškomis balsių fonemomis /a/, /i/, /o/, /u /. Mažiausias standartinis nuokrypis 2.36 Hz buvo gautas MUSIC metodu balsiui /i/, didžiausias – 5.59 Hz DFT metodu balsiui /a/.
4. Harmoninis metodas naudoja aukštesnės eilės modelius su didesniu parametrų skaičiumi palyginus su formantiniu metodu, tačiau garsai sintezuoti harmoniniu metodu skamba natūraliau. Visų vyriškų ir moteriškų balsių signalų spektrų vidutinių kvadratinių paklaidų vidurkis yra lygus 13.9 % formantinio metodo atveju ir 12.4 % harmoninio metodo atveju. Visų vyriškų ir moteriškų pusbalsių signalų spektrų vidutinių kvadratinių paklaidų vidurkis yra lygus 19.9 % formantinio metodo atveju ir 16.7 % harmoninio metodo atveju.
5. Vidutinis visų vyriškų ir moteriškų balsių ir pusbalsių fonemų parametrų įvertinimo laikas lygus 16.1 s formantinio metodo atveju ir 37.2 s harmoninio metodo atveju. Vidutinis visų vyriškų ir moteriškų balsių ir pusbalsių fonemų sintezės laikas lygus 0.09 s sintezuojant formantiniu metodu, 0.44 s – sintezuojant harmoniniu metodu.
6. Dvibalsiai modeliuojami gana tiksliai. Klausantis garso įrašų, beveik neįmanoma atskirti tikrus ir modeliuotus dvigarsius įvairiuose lietuviškuose žodžiuose. Nedidelius skirtumus tarp tikrų ir modeliuotų dvibalsių galima pamatyti palyginus šių įrašų impulsines charakteristikas.
7. Eksperimentiniai rezultatai parodė, kad žodžio, susidedančio iš balsių ir pusbalsių, sintezuoto siūlomu metodu kokybė yra pakankamai gera ir jį sunku atskirti nuo tikro žodžio. Sintezuoto garso kokybė žymiai pagerėjo dėl įvestų perėjimų tarp įėjimų.
8. Sukurta automatinė kalbos garsų modeliavimo sistema nepriklauso nuo kalbėtojo ir fonemų.

Autoriaus mokslinių publikacijų disertacijos tema sąrašas

Straipsniai recenzuojamuose periodiniuose mokslo žurnaluose

1. V. Šimonytė, G. Pyž, V. Slivinskas (2009). Application of the MUSIC method for estimation of the signal fundamental frequency. *Lietuvos matematikos rinkinys. LMD darbai*, T. 50, p. 391-396, ISSN 0132-2818.
2. G. Pyž, V. Šimonytė, V. Slivinskas (2011). Modelling of Lithuanian Speech Diphthongs. *Informatica*, Vol. 22 (3), p. 411-434. ISSN 0868-4952 [ISI Web of Science].
3. G. Pyž, V. Šimonytė, V. Slivinskas (2011). Joining of Vowel and Semivowel Models in Lithuanian Speech Formant-based Synthesizer. *Proc. of the 6th International Conference on ECT-2011*, p. 114-119, ISSN 1822-5934 [ISI Proceedings].
4. G. Pyž, V. Šimonytė, V. Slivinskas (2012). Lithuanian Speech Synthesizing by Computer Using Additive Synthesis. *Elektronika ir elektrotechnika*, Vol. 18 (8), p. 77-80. ISSN 1392-1215 [ISI Web of Science].
5. G. Pyž, V. Šimonytė, V. Slivinskas (2012). An automatic system of Lithuanian speech formant synthesizer parameter estimation. *Proc. of the 7th International Conference ECT-2012*, p. 36-39, ISSN 1822-5934.
6. V. Slivinskas, V. Šimonytė, G. Pyž (2013). Control of Computer Programs by Voice Commands. *Proc. of the 8th International Conference ECT-2013*, p. 37-40, ISSN 1822-5934.

Metodinė priemonė

V. Šimonytė, G. Pyž, V. Slivinskas (2010). Signals and their parameter estimation, Vilnius: BMK, ISBN 978-9955-88-44-4 [in Lithuanian].

Santraukoje cituota literatūra

Golub, G., Pereyra, V. (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. SIAM Journal or Numerical Analysis, 2(10), 413 – 432.

Kasparaitis, P. (2001). Text-to-Speech Synthesis of Lithuanian Language. Doctoral dissertation, Vilnius University, Vilnius [in Lithuanian].

Levenberg, K. (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares. The Quarterly of Applied Mathematics 2, pp. 164–168.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. SIAM Journal on Applied Mathematics 11, 431–441.

Markel, J. D., Gray, A. H. (1976). Linear Prediction of Speech. Springer Verlag, Berlin.

SIL International, 2004. Accessed at: http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/WhatIsAPhone.htm

Slivinskas, V., Šimonytė, V. (1990). Minimal realization and formant analysis of dynamic systems and signals. Mokslas. Vilnius 168 p. [in Russian] (republished by Booksurge, USA, 2007).

Trumpos žinios apie autorių

Gražina Pyž 2007 metais Vilniaus pedagoginio universiteto Matematikos ir informatikos fakultete įgijo matematikos bakalauro laipsnį. 2009 metais tame pačiame fakultete baigė informatikos magistratūros studijų programą ir įgijo magistro kvalifikacinį laipsnį. 2009-2013 metais studijavo informatikos inžinerijos krypties doktorantūroje Vilniaus universiteto Matematikos ir informatikos institute. Gražina Pyž yra Lietuvos kompiuterininkų sąjungos ir Lietuvos jaunųjų mokslininkų sąjungos narė.

ANALYSIS AND SYNTHESIS OF LITHUANIAN PHONEME DYNAMIC SOUND MODELS

Research Context and Challenges

A Text-to-speech (TTS) system is defined as a system that takes a sequence of words as input and converts it into speech (SIL, 2004). The speech synthesizer can be useful in many cases. TSS system can read aloud any texts from web pages, navigation, translation and other applications. It helps with pronunciation and learning foreign languages, promoting listening skills. The speech synthesizer allows multi-tasking so that attention can be given to reading materials when time would otherwise not permit. It helps people with reading challenges, or visual impairment.

Construction of speech synthesizer is a very complex task. Researchers are trying to automate speech synthesis. Yet there is no automatic Lithuanian TTS system equivalent to human speech. The commercial TTS systems have not yet supported Lithuanian language. The problem of developing Lithuanian synthesizer arises. There exists a Lithuanian synthesizer developed by P. Kasparaitis (P. Kasparaitis, 2001). It is based on concatenation speech synthesis type. Concatenation synthesis relies on speech sounds recorded in advanced database. One of the main drawbacks of concatenation synthesis is that the database has to be sufficiently large. That, however, requires extensive computer resources. If a word is not in the database, then it could not be synthesized. The synthesized speech quality does not achieve the natural speech quality since glitches occur on the concatenation boundaries. Formant synthesis does not require a sound database. Formant synthesizers have advantages against the concatenative ones. The speech produced by them can be sufficiently intelligible even at high speed. They can control prosody aspects of the synthesized speech (intonation, rhythm, stress). The main drawback of formant synthesis is that the sounds obtained by this synthesis type sound unnaturally, robot-like. In this work an assumption is made that the models of formant synthesizer are too simple. In order to reduce synthetic sounding, it is necessary to develop new mathematical models for speech sounds. The vowel and semivowel phoneme models are a part of this problem.

Problem Statement

The sounds obtained by formant synthesis type sound unnaturally, robot-like. In order to reduce synthetic sounding, it is necessary to develop new mathematical models for speech sounds, which could be used as a base of speech synthesizer.

Object of Research

The research object of the dissertation is Lithuanian vowel and semivowel phoneme models.

The Objective and Tasks of the Research

The objective of the thesis is to develop Lithuanian speech vowel and semivowel phoneme dynamic models, and create transition between phonemes in order to join these models.

In order to achieve the objective, the following tasks are stated:

- to acquaint with speech production apparatus, main speech synthesis methods and the existing text-to-speech systems of Lithuanian and other languages.
- to analyse main characteristics of Lithuanian speech vowel and semivowel sounds.
- to ascertain what models are suited best for Lithuanian speech sound description.
- to develop mathematical models of Lithuanian speech vowel and semivowel phonemes.
- to create transitions between phonemes in order to join vowel and semivowel models.
- to evaluate the proposed models accuracy experimentally.

Methodology of Research

- Digital signal processing,
- System theory,
- Optimization methods,
- Matrix algebra,
- Mathematical statistics,
- Programming in Matlab environment,
- Programming in C# language.

Scientific Novelty

- For vowel and semivowel phonemes modelling MISO system whose impulse response of each channel is described as a third order quasipolynomial and input amplitude impulse vary in time is proposed.
- A new parameter estimation algorithm for convoluted data, based on Levenberg-Marquardt approach, has been derived.
- A new fundamental frequency refining algorithm is proposed.

- A new method that allows one to select the representative period automatically is given.
- The transitions between vowel and semivowel phoneme models have been derived.

The advantage of my developed phoneme modelling framework is that anyone can use it and it can synthesize any phoneme of vowel and semivowel for any speaker.

Practical Significance of the Results

The proposed vowel and semivowel phoneme models can be used for developing a TTS formant synthesizer. The phoneme models can also be adapted to other similar problems, for example, treating language disorders, helping with pronunciation and learning of foreign languages.

Defended Propositions

1) For vowel and semivowel phoneme modelling a discrete time linear stationary system with multiple-input and single-output (MISO) is used. Impulse response of each MISO system channel is described as a third order quasipolynomial.
2) In order to obtain more natural sounding of the synthesized speech, it is important to use not only high-order models, but complex input sequence scenarios as well.
3) The vowel and semivowel phonemes synthesis quality is sufficiently good.
4) The word consisting of vowels and semivowels obtained with the proposed synthesis methods is enough and it is difficult to distinguish it from the real one. The quality of the synthesized sound was significantly improved due to input transitions.

Approbation and Publications of the Research

The main results of the dissertation were published in 6 articles in the periodical scientific publications. The main results of the work have been presented and discussed at 21 national and international conferences.

Outline of the Dissertation

The dissertation consists of 5 chapters, references and appendices. The total scope of the dissertation without appendices – 114 pages containing 78 formulas, 47 pictures and 19 tables.

The Introduction (Chapter 1) reveals research context and challenges, describes the problem statement, the object of research, the tasks and objective of the dissertation, methodology of research, presents scientific novelty, practical significance of results, defends propositions and approbation of obtained results.

In Chapter 2 an overview of Lithuanian speech engineering is given. Detailed information about Lithuanian speech phonemes and diphthongs is presented.

In Chapter 3 Lithuanian vowel and semivowel phoneme modelling framework is submitted. Within this framework two synthesis methods are proposed: harmonic and formant.

Chapter 4 provides the results of experimental researches.

Conclusions (Chapter 5) present the main conclusions of the dissertation.

Appendices present a list of Lithuanian phonemes with the examples of their usage, the plots of the vowel and semivowel phoneme signals.

Conclusions

The research object of the dissertation is Lithuanian vowel and semivowel phoneme models. In order to develop models for vowels and semivowels, the main characteristics of these sounds have been identified. A phoneme synthesis framework that is based on a vowel and semivowel phoneme mathematical model and an automatic procedure of estimation of the phoneme fundamental frequency and input determining has been proposed. Within this framework two synthesis methods have been given: the harmonic method and formant method.

The research completed in this thesis has led to the following conclusions:

1. Lithuanian language has ninety two phonemes. Twenty eight of them are pure vowel phonemes, nineteen – semivowel phonemes. In general case, the character of vowel and semivowel signals is periodic.
2. The fundamental frequencies of the stressed vowels and semivowels are lower than those of the unstressed ones.
3. The estimates of the fundamental frequency obtained by the MUSIC method are less scattered around their average if compared with the ones obtained by the DFT method. The methods for Lithuanian male vowels /a/, /i/, /o/, /u / were applied. The smallest standard deviation 2.36 Hz was obtained by the MUSIC method for the vowel /i/, and the largest – 5.59 Hz – by the DFT method for the vowel /a/.
4. The harmonic method uses a higher-order model with a larger number of parameters in comparison with the formant method but the sounds synthesized by the harmonic method sound more naturally. The average RMSE for the estimated signal spectrum for all the male and female vowels is equal to 13.9 % in the formant method case and 12.4 % in the harmonic method case. The average RMSE for the estimated signal spectrum for all the male and female semivowels is equal to 19.9 % in the formant method case and 16.7 % in the harmonic method case.
5. The average time of the phoneme parameter estimation for all the male and female vowels and semivowels is equal to 16.1 s in the formant method case and 37.2 s in the harmonic method case. The average time of the phoneme synthesis

for all the male and female vowels and semivowels is equal to 0.09 s in the formant method case and 0.44 s in the harmonic method case.

6. The accuracy of diphthong modelling is high. It is almost impossible to distinguish between the real and simulated diphthongs in various Lithuanian words with a help of audiotesting. Only the magnitude response of the whole signal of the simulated diphthong differs a little from the magnitude response of the recorded data in some frequency bands.
7. Simulation has revealed that the word consisting of vowels and semivowels obtained with the proposed synthesis method is good enough and it is difficult to distinguish it from the real one. The quality of the synthesized sound was significantly improved due to input transitions.
8. The created automatic system will suit any speaker, any vowel and semivowel phoneme.

Gražina Pyž

ANALYSIS AND SYNTHESIS OF LITHUANIAN PHONEME DYNAMIC SOUND MODELS

Doctoral Dissertation

Technological Sciences,
Informatics Engineering (07T)

Gražina Pyž

LIETUVIŠKŲ FONEMŲ DINAMINIŲ MODELIŲ ANALIZĖ IR SINTEZĖ

Daktaro disertacija

Technologijos mokslai,
Informatikos inžinerija (07T)