



VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

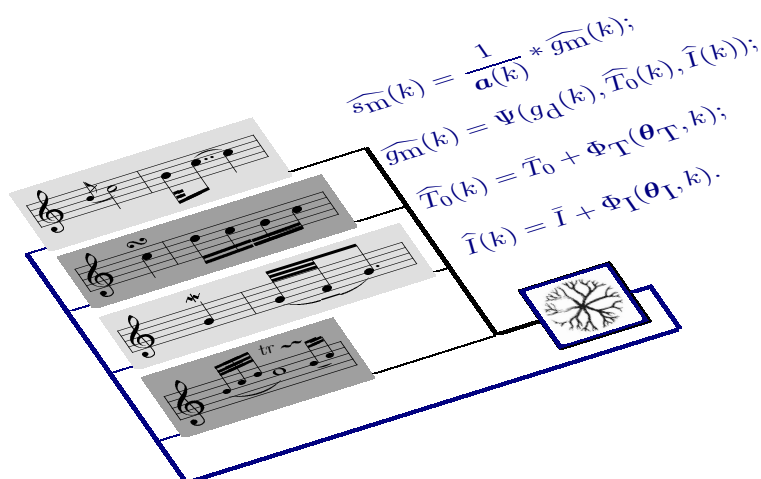


Romas LEONAVIČIUS

MELIZMŲ SINTEZĖ DIRBTINIŲ NEURONŲ TINKLAIS

DAKTARO DISERTACIJA

TECHNOLOGIJOS MOKSLAI, ELEKTROS IR ELEKTRONIKOS INŽINERIJA (01T)



VILNIUS, 2006

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Romas LEONAVIČIUS

**MELIZMŲ SINTEZĖ
DIRBTINIŲ NEURONŲ TINKLAIS**

Daktaro disertacija

Technologijos mokslai, elektros ir elektronikos inžinerija (01T)

Vilnius, 2006

Disertacija rengta 2002–2006 metais Vilniaus Gedimino technikos universitete.

Darbo vadovas

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

Reziumė

Dauguma šiuolaikinių kalbos sintezės metodų pagrįsti nekintančiu pagrindiniu tonu ir orientuoti į monotoniškos kalbos sintezę. Šie metodai netinka tokiems uždaviniams kaip dainos restauravimas, nes sintezuotoje kalboje trūksta natūralumo. Dirbant su viena sudėtingesnių dainos puošmeninių figūrų – melizma – pasižyminčia staigiai kintančiu pagrindiniu tonu ir intensyvumu, siekiama išnagrinėti kokybinius kalbos sintezės aspektus ir pasiūlyti metodus leisiančius pakelti kalbos sintezės kokybę.

Šio darbo tikslas – taikant dirbtinių neuronų tinklus sintezuoti lietuviškose liaudies dainose sutinkamas melizmas. Darbe siekta: apžvelgus melizmas ir signalų modeliavimo metodus parinkti esmines melizmų charakteristikas bei atrasti ir parodyti dirbtinių neuronų tinklų taikymo melizmomis sintezuoti galimybes; pasiūlyti melizmų sintezės metodiką ištariant tiesinės prognozės metodo taikymo galimybes; sukurti melizmų matematinius modelius tinkančius pagrindinių rūšių melizmų sintezei; sukurti melizmų įrašų rinkinį apimančių skirtingų rūšių melizmas bei įdainuotą keletos atlikėjų; atlikti sukurtų melizmų modelių eksperimentinę patikrą bei rasti tinkamus modelių dydžius ir jų parametrų reikšmes.

Disertaciją sudaro šeši skyriai iš kurių paskutinis – išvados. Pirmame įvado skyriuje suformuluojama problema, darbo tikslas ir uždaviniai, aptariami tyrimų metodai ir darbo mokslinis naujumas, pristatomas autoriaus indėlis bei disertacijos struktūra.

Antrame skyriuje nagrinėjama melizmos vieta netiesinių bei dinaminių sistemų modeliavime. Aprašoma melizma, jos tipai ir charakteristikos, parodoma kad melizma pasižymi netiesiškumu bei staigia dinamika. Apžvelgiamos dinaminės sistemos, jų struktūros. Aprašomi plačiausiai taikomi dinaminių sistemų identifikavimo ir patikros būdai. Paskutinėje antrojo skyriaus dalyje apžvelgiami dirbtinių neuronų tinklai ir parodomas jų ryšys su netiesinėmis ir dinaminėmis sistemomis – melizmų sintezės priemone.

Trečias skyrius skirtas apžvelgti melizmų sintezės metodiką. Nagrinėjamas tiesinės prognozės koeficientų metodas – pateikiami jo privalumai ir trūkumai. Atsižvelgiant į melizmos specifiką *siūlomas modifikuotas tiesinės prognozės metodas* ir jo pagrindu *sudaroma originali melizmos sintezės metodika*.

Ketvirtame skyriuje kuriami melizmų matematiniai modeliai: *pristatomas apibendrintas melizmos modelis*, parenkamas identifikavimo būdas ir *pristatomi originalūs pagrindinių melizmų rūšių – foršlago, grupeto, mordento ir trelės – modeliai*. Skyriaus pabaigoje pasiūlomi mažiausi atskirų rūšių melizmų modelių dydžiai.

Penktame skyriuje vykdoma sukurtų modelių taikymo melizmomis sintezuoti eksperimentinė patikra: aprašomas sudarytas melizmų rinkinys bei pateikiami atliktų vieno atlikėjo melizmų sintezės bei nuo atlikėjo nepriklausomų melizmų sintezės eksperimentų rezultatai. *Parodoma, jog sukurti melizmų modeliai yra tinkami vieno atlikėjo melizmų sintezei*.

Pagrindiniai disertacijos rezultatai paskelbti 6 mokslinėse konferencijose. Darbo tema paskelbti 5 moksliniai straipsniai: vienas straipsnis respublikiniame žurnale cituojamame Inspec duomenų bazėje, du straipsniai užsienio recenzuojamuose žurnaluose ir du straipsniai respublikinių konferencijų darbų medžiagoje.

Abstract

Modern methods of speech synthesis are oriented at slowly varying pitch signals. They are not suitable for restoration of song signals due to lack of vitality and intonation in the resulted sounds. Lithuanian ethnographic song records urgently require remastering. Melismas are inherited features of old songs and their preservation during remastering process needs special attention. When dealing with one of most complicated ornaments of songs – melismas – we seek here to elaborate qualitative speech synthesis aspects and to propose methods that will allow us to raise the quality of songs synthesis.

The aim of this work is to synthesize melismas met in Lithuanian folk songs, by applying Artificial Neural Networks. The tasks of work were as follows after reviewing melismas and signal modeling methods, to select the basic features of melismas and to establish the possibilities of applying Artificial Neural Networks to synthesize melismas; to propose the technique for synthesizing melismas after investigating the possibility to apply the Linear Prediction method; to construct mathematical models of melismas suitable for the synthesis of the main kinds of melismas; to compile a collection of melisma records ranging over different kinds of melismas and sung by several performers; to test the developed models of melismas experimentally and to find the proper size of models as well as the values of their parameters.

The thesis consists of 6 chapters where the last one is Conclusions. In the first chapter – Introduction – problem is stated, the aim and tasks are outlined, applied methods and scientific novelty are explained, the contribution of author and structure of dissertation are presented.

In the second chapter an analytical survey of rather a widespread literature is presented. First classification and comprehensive discussion of melismas are given, then brief but essential discussion of the theory of dynamic systems which will make the basis for studying melismas is presented and finally the relationship for modelling a melisma with nonlinear and dynamic systems is outlined.

The third chapter presents investigation of the most widely used Linear Prediction Coding (LPC) method and possibilities of its improvement. *The modification of original Linear Prediction method based on dynamic LPC frame positioning is proposed.* On its basis, *the new melisma synthesis technique is presented.*

In the fourth chapter *developed flexible generalized melisma model*, based on two Artificial Neural Networks – a Multilayer Perceptron and Adaline – as well as on two network training algorithms – Levenberg-Marquardt and the Least Squares error minimization – is presented. Moreover, *original mathematical models of Fortis, Gruppett, Mordent and Trill are created*, fit for synthesizing melismas, and their minimal sizes are proposed.

The fifth chapter concerns experimental investigation and corroborates application of the new mathematical models to melisma synthesis of one performer.

Major results of this work were presented at 6 conferences and five scientific papers have been published – among them one paper in the prestigious national journal quoted in the international database Inspec and two papers in periodical reviewed journals published abroad.

Žymėjimai

Symboliai

k	Laiko indeksas.
n	Analizės lango indeksas.
p	Pagrindinio tono periodo indeksas.
$a_1, \dots, a_{N_M}$	Tiesinės prognozės kodavimo koeficientai.
$A(q), \dots, D(q), F(q)$	Apibendrinto modelio daugianariai.
$g_d(k)$	Dainos gerklų signalas.
$\hat{g}_m(k)$	Melizmos gerklų signalas.
$f(\cdot)$	Neurono aktyvavimo funkcija.
F^{MEL}	MEL skalės dažnis.
F^{L}	Dažnis.
$\mathbf{H}$	Hesianas.
$I(k)$	Signalų intensyvumas.
$\mathbf{J}$	Jakobianas.
N_{L}	Signalų analizės lango trukmė atskaitomis.
$N_{\text{L}}^{(i)}$	Neuronų skaičius i -iame sluoksnyje.
$N^{\text{F,G,M,T}}$	Paslėptų neuronų skaičius foršlago, grupeto, mordento ir trelės modeliuose.
$\mathbf{p}_k$	Krypties vektorius k laiko momentu.
$\mathbf{q}_k$	Klaidos gradientas k laiko momentu.
$\mathbf{Q}_k$	Klaidos kryptis k laiko momentu.
R_{xy}	Kryžminė koreliacija.
$s(k)$	Originalus signalas.
$\hat{s}_A(k), \hat{s}_{\text{MLP}}(k)$	Sintezuotas signalas, taikant adalina ar taikant daugiasluoksnį perceptroną).
$s_m(k), s_b(k)$	Melizmos signalas, balsingos dainos dalies signalas.
$T_0(k)$	Signalų pagrindinis tonas.
$\bar{T}_0, \bar{I}$	Pastovaus pagrindinio tono ir intensyvumo reikšmės.
$\hat{T}_0, \hat{I}$	Modeliuotas pagrindinis tonas ir intensyvumas.
$\widehat{T}_0^{\text{F,G,M,T}}, \widehat{I}^{\text{F,G,M,T}}$	Foršlago, grupeto, mordento ir trelės modeliuoti pagrindinis tonas ir intensyvumas.
V_k	Dirbtinio gerklų signalo kilimo greitis.
V_n	Dirbtinio gerklų signalo nusileidimo greitis.
$\mathbf{w}^{(L)}$	L -tojo sluoksnio neuronų ryšių svoriai.
$W^{\leftarrow}(k), W^{\rightarrow}(k)$	Pasvėrimo (iš kairės ir iš dešinės) funkcija.
Δ_n	Melizmos analizės dinaminio lango poslinkis.
Ω_n	Melizmos analizės dinaminio lango trukmė.

$\theta_T, \theta_I,$	Melizmos pagrindinio tono ir intensyvumo modelių parametrai.
$\hat{\theta}_k$	Modelio parametrų įvertis k laiko momentu.
$\Psi_m(\cdot)$	Apibendrintas melizmos modelis.
$\Phi_T(k), \Phi_I(k),$	Melizmos pagrindinio tono ir intensyvumo modeliai.

Operatoriai ir funkcijos

$*$	Kompozicija.
$\cdot$	Sandauga.
$\hat{x}$	Žymi x įvertį.
$\mathbf{x}$	Žymi x vektorių.
$\mathbf{x}^T$	Žymi x transponavimą.
∂	Dalinė išvestinė.
q	Ankstinimo operatorius.
$ x $	Žymi x modulį.
$\lceil \cdot \rceil$	Apvalinamas iki didesnio sveiko skaičiaus.
$\lfloor \cdot \rfloor$	Apvalinamas iki mažesnio sveiko skaičiaus.

Santrumpos

ADCH	Amplitudės dažninė charakteristika.
AR	Autoregresinis (modelis).
ARARX	Autoregresinis-autoregresinis su išoriniu įėjimu (modelis).
ARMA	Autoregresinis slenkančio vidurkio (modelis).
ARMAX	Autoregresinis slenkančio vidurkio su išoriniu įėjimu (modelis).
ARX	Autoregresinis su išoriniu įėjimu (modelis).
BJ	Box-Jenkins (modelis).
DNT	Dirbtinių neuronų tinklas.
DP	Daugiasluoksnis perceptronas.
GFT	Greitoji Furjė transformacija.
MA	Slenkančio vidurkio (modelis).
NARX	Netiesinis autoregresinis su išoriniu įėjimu (modelis).
NDS	Netiesinė dinaminė sistema.
OE	Išėjimo klaidos (modelis).
TDS	Tiesinė dinaminė sistema.
VKK	Vidutinė kvadratinė klaida.
VNR	Vidutinės nuomonės rezultatas.

Turinys

Reziumė	iii
Abstract	iv
Žymėjimai	v
1. Įvadas	1
1.1. Darbo aktualumas	1
1.2. Problemos formulavimas	2
1.2.1. Tyrimų objektas	3
1.2.2. Darbo tikslas	4
1.2.3. Darbo uždaviniai	4
1.2.4. Tyrimų metodai	4
1.2.5. Mokslinis darbo naujumas	4
1.2.6. Autoriaus dalyvavimas mokslinėse programose	4
1.3. Autoriaus indėlis	5
1.3.1. Gynimui teikiama	5
1.3.2. Darbo aprobavimas	5
1.3.3. Rezultatų publikavimas	5
1.3.4. Publikacijų sąrašas	5
1.4. Disertacijos struktūra	6
2. Melizmos vieta netiesinių bei dinaminių sistemų modeliavime	7
2.1. Melizma ir jos charakteristikos	7
2.1.1. Melizma ir jos tipai	7

2.1.2. Laikinės charakteristikos	10
2.1.3. Dažninės charakteristikos	13
2.1.4. Laikinės-dažninės charakteristikos	15
2.2. Dinaminės sistemos, jų struktūros, identifikavimas ir patikra	16
2.2.1. Dinaminių sistemų struktūros	17
2.2.2. Dinaminių sistemų identifikavimas	23
2.2.3. Dinaminių sistemų patikra	25
2.3. Dirbtinių neuronų tinklų ryšys su netiesinėmis bei dinaminėmis sistemomis	26
2.3.1. Specialiosios netiesinės bei dinaminės struktūros	26
2.3.2. Specialiųjų netiesinių bei dinaminių struktūrų identifikavimas	29
2.4. Skyriaus apibendrinimas	30
3. Melizmos sintezės metodikos pagrindimas ir sudarymas	31
3.1. Tiesinės prognozės metodo tyrimas	31
3.1.1. Melizmos sintezė	32
3.1.2. Gerklų signalo modeliavimas	33
3.1.3. Gerklų signalo fragmentavimas	36
3.1.4. Pagrindinio tono modeliavimas	40
3.2. Modifikuoto tiesinės prognozės metodo tyrimas	43
3.2.1. Dinaminis tiesinės prognozės langų dėstymas	43
3.2.2. Dinaminio tiesinės prognozės langų dėstymo formalizavimas	45
3.2.3. Modifikuoto bei standartinio tiesinės prognozės metodų palyginimas	47
3.3. Melizmos sintezės metodikos sudarymas	48
3.3.1. Melizmos modeliavimo pagrindas	48
3.3.2. Melizmos modelio vieta	49
3.3.3. Melizmos gerklų signalo sintezės metodika	50
3.4. Skyriaus apibendrinimas	52
4. Melizmų modelių sudarymas	53
4.1. Apibendrinto melizmos modelio kūrimas	53
4.1.1. Gerklų signalo modeliavimas	54
4.1.2. Pagrindinio tono ir intensyvumo modeliavimas	56
4.1.3. Apibendrintas melizmos modelis	56
4.2. Melizmos modelio identifikavimo būdo parinkimas	57
4.2.1. Mažiausios vidutinės kvadratinės klaidos minimizavimo metodas	57
4.2.2. Levenberg-Marquardt minimizavimo metodas	58
4.3. Atskirų melizmos rūšių modelių sudarymas	63
4.3.1. Foršlago modelis	64
4.3.2. Grupeto modelis	65
4.3.3. Mordento modelis	66
4.3.4. Trelės modelis	67
4.4. Skyriaus apibendrinimas	68

5. Melizmos modelių taikymo sintezėje eksperimentinė patikra	69
5.1. Melizmų duomenų rinkinių sudarymas	69
5.1.1. Melizmų rūšys	69
5.1.2. Atlikėjų tipai	70
5.1.3. Išoriniai faktoriai. Triukšmų tipai	70
5.2. Vieno atlikėjo melizmų sintezė	72
5.2.1. Melizmų modelių identifikavimas taikant V1 ir M1 rinkinius	72
5.2.2. V1 ir M1 melizmų rinkinių sintezė	76
5.3. Nuo atlikėjo nepriklausomų melizmų sintezė	79
5.3.1. F06, G02 ir T02 melizmų modelių identifikavimas taikant B1 rinkinį	79
5.3.2. B1 rinkinio F06, G02 ir T02 melizmų sintezė	80
5.4. Skyriaus apibendrinimas	84
6. Galutinės išvados	85
Literatūra	87
Autoriaus publikacijos	93
Priedas	95
A. Melizmų modeliavimo įrankiai	95
A.1. Sukurtų Matlab funkcijų rinkinys	95
A.2. Surinktų melizmų rinkinys	96
B. Sąvokų žodynas	97
Dalykinė rodyklė	99

Įvadas

1.1. Darbo aktualumas

Skaitmeninis kalbos apdorojimas plačiai paplitęs ir neretai palengvinantis tiek žmonių bendravimą tarpusavyje, tiek ir su intelektualiosiomis elektroninėmis sistemomis. Šiuolaikiniame pasaulyje vis sunkiau surasti sričių kur nenaudojamos kalbos atpažinimo bei kalbos sintezės sistemos. Net tokia kasdiene tapusi paslauga kaip mobilusis ryšis taip pat naudoja kalbos sintezę, kuri leidžia sumažinti dažnių juostos plotį, o kartu ir teikiamų paslaugų kainą. Kita vertus, nors ir yra sukurta nemažai komercinė ir grynai mokslinę naudą nešančių produktų, visgi dauguma iš jų orientuoti į monotoniškos kalbos sintezę.

Dauguma šiuolaikinių kalbos apdorojimo priemonių skirta maksimaliai sumažinti kalbai perduoti reikalingos dažnių juostos plotį, todėl kalbai trūksta natūralumo bei intonacijos.

Visgi egzistuoja tokių sričių, kuriose reikalinga geresnė sintezės kokybė negu gali pasiūlyti šiuo metu paplitę sprendimai. Viena iš tokių yra dainų restauravimas. Šiuo metu yra nemažai muzikos archyvų, kurie vis dar įrašyti analoginėse laikmenose. Bandant įrašus pervesti į skaitmeninį formatą, visuomet iškyla tokių klausimų kaip signalo iškirtimo arba impulsinių triukšmų atstatymas. Šiuolaikinėmis kalbos sintezės priemonėmis to išspręsti nepavyksta, nes jos naudoja pastovų pagrindinį toną – parametą apibūdinantį kalbos aukštumą. Nors monotoniškoje dainoje jis išlieka pastovus, tačiau dainoje netgi vienos natos intervale kinta tiek pagrindinis tonas, tiek ir intensyvumas.

Šiame darbe pasirinkus dainos smulkia ritminės vertės melodinę puošmeninę figūrą – melizmą – siekiama ištyrinėti kokybinius kalbos sintezės aspektus ir pasiūlyti metodus, leisiančius pakelti sintezuojamos kalbos kokybę.

1.2. Problemos formulavimas

Šiuolaikinės kalbos sintezės priemonės paremtos pastoviu arba geriausiu atveju lėtai kintančiu pagrindiniu tonu. Tokio kitimo lokaliu apdorojimui naudojami tiesinio aproksimavimo metodai. Tuo tarpu dainoje pagrindinis tonas kinta labai greitai. Ribinis tokio kitimo pavyzdys būtų puošybinis dainos ornamentas – melizma. Pagrindinis tonas ir intensyvumas melizmoje kinta vienos natos ribose.

Dainos sintezė nėra plačiai ištirta. Pasaulio mastu atlikta labai nedaug tyrimų skirtų dainos sintezei. Šiuo metu vis dar nepasiūlytas universalus sprendimas, įgalinantis sintezuoti dainas. Lietuvoje, kaip ir kitose šalyse tėra keletas darbų skirtų melizmos sintezei. D. Navakauskas (VGTU) pirmas pasiūlė melizmų sintezei naudoti dirbtinių neuronų tinklus [26, 56]. Darbe buvo pasiūlytas originalus, specialios struktūros, sumažinto dydžio pynučių-kopetelių daugiasluoksnių perceptronų tinklas skirtas melizmos signalo restauravimui. Šiame darbe nebuvo nagrinėtas vokalinio trakto modeliavimas, nebuvo dirbama su kalbos generavimo mechanizmu – melizmos signalas buvo sintezuojamas tiesiogiai. Vėliau su kolega Š. Pauliku (VGTU) tas pats autorius nagrinėjo balsinių kalbos fragmentų atstatymą [61], kuomet žinoma tik kirčio vieta ir priegaidės tipas. Atstatymui buvo siūloma naudoti signalo restauravimą pasitelkus interpoliaciją ir filtravimą.

Melizmos lietuvių liaudies dainose taip pat nagrinėtos G. Raškinio (VDU). Viename pirmųjų darbų [69] tirta melizmų transkripcija naudojant signalo energiją ir pagrindinį toną. Autoriaus darbuose dirbtinių neuronų tinklai nebuvo naudojami.

Kita vertus, kalbos sintezės metodika pagrįsta tiesinės prognozės koeficientų skaičiavimu, jau seniai plėtojama [50] ir yra paplitusi pasauliniu mastu. Ji grindžiama prielaida, kad kalbos signalą galima išskaidyti į gerklų signalą bei vokalinio trakto modelį. Tokiu būdu kalbos sintezę siejant su kalbos generavimo mechanizmu.

Lietuvoje jau ilgą laiką dirbama ties kalbos atpažinimu [46]. V. Rudžionio (KTU) 1998 m. disertacijoje [75] nagrinėtas kalbos atpažinimas fonetinių vienetų pagrindu. Vėliau šio autoriaus dirbta ties lietuvių šnekamosios kalbos garsynu LTDIGITS [74] bei gilintasi į lietuvių kalbos sintezės ir atpažinimo taikymą automatikoje [73]. Didelį indėlį kalbos atpažinimo Lietuvoje tyrimuose yra įnešęs A. Lipeika (MII) ir jo kolegos. Viename ankstesniųjų darbų nagrinėtas tiesinės prognozės koeficientų pritaikomumas kalbos atpažinimui [43]. Vėliau nagrinėtas atskirų žodžių atpažinimas [44], paremtas dinaminio laiko mastelio keitimu, vektoriniu kvantavimu ir tiesinės prognozės koeficientų skaičiavimu. Taip pat nagrinėti formantiniai požymiai atpažįstant kalbą [45].

Viena iš perspektyvių kalbos atpažinimo kryptų yra paslėptųjų Markovo modelių panaudojimas. Vienas pirmųjų šią temą nagrinėjo G. Raškinis bei D. Raškinienė (VDU), pasiūlę metodiką atskirų žodžių atpažinimui, paremtą MEL dažnio kepsčio skaičiavimu [70]. Vėliau ši metodika patobulinta M. Filipovič (MII), apjungiant paslėptuosius Markovo modelius su daugiasluoksnių perceptronų tinklais [16]. Ši metodika leido pasiekti didesnę atskirų žodžių atpažinimo kokybę. D. Šilingas (VDU) nagrinėjo paslėptuosius Markovo modelius ištisinės kalbos atpažinime [28].

Su lietuvių kalbos sintezę dirbantis P. Kasparaitis (MII) dar 2000 m. nagrinėjo automatinį lietuvių kalbos kirčiavimą [33]. Darbe autorius parodė, kad kirčiavimui reikiamą informacijos kiekį įmanoma sumažinti padalinus lietuvių kalbos žodžius į tris grupes: daiktavardžius-būdvardžius, veiksmažodžius bei nelinksniuojamus žodžius. Kiekvieną iš

šių grupių autorius siūlė nagrinėti atskirai ir sintezės žodynuose saugoti tik reikiamą kirčiavimui informaciją. Vėliau autorius patobulino būdvardžių ir daiktavardžių kirčiavimą [34] bei 2005 m. pristatė vienai iš žinomiausių pasaulyje kalbos sintezavimo sistemų MBROLA formatui pritaikytas lietuvių kalbos sintezės bibliotekas. Darbe [35] nagrinėtos fonemų kombinacijos, būtinos lietuvių kalbos sintezei.

Kaip matome, kalbos sintezės, o ypač atpažinimo klausimai Lietuvoje yra nagrinėjami visos eilės specialistų, tačiau dainos sintezės klausimai nėra plačiai nagrinėti. Netgi pasauliniu mastu šia kryptimi nėra toli pažengta. Pagrindinė darbo kryptis vykdoma dainų klasifikavime, kuriame siūloma naudoti paslėptus Markovo procesus ir MEL skalės kepstro koeficientus [87]. Muzikos kūrinių klasifikavimui siūloma naudoti takto laikinę analizę [49], o klasifikatoriaus funkcijai pasitelkti atraminį vektorių mašinas bei Gauso mišinių tinklus. Kalbos ir dainos suvokimo palyginimai atliekami [17, 27] siekiant surasti esminius dainos ir kalbos sąlyčio taškus bei suprasti žmogaus dainos suvokimą. Kiti autoriai siekia automatinėmis priemonėmis išskirti kertinius dainos segmentus [88, 93, 94], pagal kuriuos galima atlikti dainos klasifikavimą pagal žanrą, taip pat siūloma semantinė dainos analizė [48], leisianti vykdyti paiešką dainų duomenų bazėse bei automatinę dainos klasifikavimą. Automatiniam dainos identifikavimui siūlomas maišymo motodas [52], leidžiantis identifikuoti dainą neturint jokios pirminės informacijos apie ją. Taip pat siūlomi būdai leidžiantys vykdyti muzikos pokyčių detektavimą [64] tam, kad atskirti dainas einančias viena po kitos. Pagrindinio tono išskyrimui dainoje tobulinti [91] siūloma naudoti juostinius filtrus su nuosekliu jungimu. Pateikti sprendimai leidžia išskirti dviejų ir netgi trijų tuo pat metu dainuojančių atlikėjų pagrindinius tonus. Bene vieninteliame užsienio šaltinyje [10] nagrinėjamas liaudies dainų atstatymas ir siūlomas ritminių šablonų klasifikavimas leidžiantis surasti duomenų bazėje trūkstamus dainos segmentus.

Dainos sintezė dar yra silpnai išnagrinėta sritis. Be to, dirbant su liaudies dainomis, būtina atsižvelgti į dainos savybes, kurios kiekvienoje kalboje yra specifinės. Todėl sprendimai, tinkantys vienos šalies dainoms, gali netikti kitos šalies dainoms.

Šiame darbe, taikant tiesinės prognozės metodą bei modeliuojant kalbos charakteristikas, mes galėtume sintezuoti melizmas. Melizmos charakteristikų, kurios yra netiesinės ir kinta laike, modeliavimui galėtume taikyti universalius aproksimatorius – dirbtinių neuronų tinklus. Tuomet, melizmos modelių radimas galėtų būti supaprastintas iki tinklo struktūros parinkimo ir apmokymo.

1.2.1. Tyrimų objektas

Darbo tyrimų objektas yra netiesinių bei dinaminių sistemų analizė ir taikymas dainos fragmentų – melizmų – sintezėje:

- tiesinės prognozės metodo taikymas melizmų sintezėje;
- melizmų charakteristikų aproksimavimas dirbtinių neuronų tinklais;
- foršlago, grupeto, mordento ir trelės matematinių modelių sandara ir jų eksperimentinė patikra;
- vieno atlikėjo bei nepriklausomų nuo atlikėjų melizmų sutinkamų lietuvių liaudies dainose sintezė.

1.2.2. Darbo tikslas

Darbo tikslas yra taikant dirbtinių neuronų tinklus sintezuoti lietuviškose liaudies dainose sutinkamas melizmas.

1.2.3. Darbo uždaviniai

- ♦ Apžvelgus melizmas ir signalų modeliavimo metodus, parinkti esmines melizmų charakteristikas bei atrasti (ir parodyti) dirbtinių neuronų tinklų taikymo melizmomms sintezuoti galimybes.
- ♦ Pasiūlyti melizmų sintezės metodiką, ištiriant tiesinės prognozės metodo taikymo galimybes.
- ♦ Sukurti melizmų matematinius modelius, tinkančius pagrindinių rūšių melizmų sintezei.
- ♦ Sukurti melizmų įrašų rinkinį, apimantį skirtingų rūšių melizmas bei įdainuotą keletos atlikėjų.
- ♦ Atlikti sukurtų melizmų modelių eksperimentinę patikrą bei rasti tinkamus modelių dydžius ir jų parametrų reikšmes.

1.2.4. Tyrimų metodai

Darbe taikomi kalbos signalų apdorojimo, netiesinių funkcijų aproksimavimo ir paieškinės optimizacijos metodai, dirbtinių neuronų tinklai, skaitinis modeliavimas.

1.2.5. Mokslinis darbo naujumas

- Sukurta nauja melizmų sintezės metodika, leidžianti taikyti apibendrintą melizmų modelį ir paremta dvikrypčiu apdorojimu, įtakojamu aproksimuotomis melizmų charakteristikomis.
- Pasiūlyto apibendrinto melizmų modelio pagrindu sukurti originalūs visų keturių melizmų rūšių matematiniai modeliai grįsti dirbtinių neuronų tinklais.
- Atlikus eksperimentus su virš 500 melizmomis, nustatytos originalių melizmų modelių parametrų reikšmės bei patvirtintas šių modelių taikymas vieno atlikėjo melizmų sintezei.

1.2.6. Autoriaus dalyvavimas mokslinėse programose

Autorius dalyvavo Švedijos Instituto remiamoje mokslinėje programoje pagal tarptautinį projektą „Nonlinear Dynamic Signal Processing“ (2002–2004 m.), taip pat Vilniaus Gedimino technikos universiteto kvalifikaciniuose darbuose: „Netiesinio skaitmeninių signalų apdorojimo technologijų tobulinimas“ (2005–2006 m.) ir „Vaizdo ir garso signalų skaitmeninio apdorojimo technologijų tobulinimas“ (2002–2004 m.).

1.3. Autoriaus indėlis

1.3.1. Gynimui teikiama

- ♦ *Melizmų sintezės metodika* paremta dvikrypčiu apdorojimu, taikant dainos aproksimuotas pagrindinio tono ir intensyvumo charakteristikas, originalų gerklų signalo fragmentą bei tiesinės prognozės koeficientus.
- ♦ *Apibendrintas melizmų modelis* pagrįstas dviem dirbtinių neuronų tinklais – daugiasluoksniu perceptronu ir adalina bei dviem tinklų mokymo algoritmais – Levenberg-Marquardt ir mažiausios vidutinės kvadratinės klaidos.
- ♦ *Originalūs foršlago, grupeto, mordento ir trelės matematiniai modeliai*, pasiūlyti jų minimalūs dydžiai bei eksperimentų metu rastos modelių parametrų reikšmės.
- ♦ Eksperimentų, naudojant virš 500 melizmų įrašus, vieno atlikėjo bei nuo atlikėjo nepriklausomų melizmų sintezės rezultatai.

1.3.2. Darbo aprobavimas

Pagrindiniai disertacijos rezultatai paskelbti šiose mokslinėse konferencijose:

- Tarptautinėje konferencijoje „Modern Information Technology“, 2004 ir 2005 m., Braslau, Baltarusija;
- Tarptautinėje konferencijoje „Elektronika“ 2004 ir 2005 m., Vilnius;
- Jaunųjų mokslininkų konferencijoje „Lietuva be mokslo – Lietuva be ateities“ 2001 ir 2003 m., Vilnius.

1.3.3. Rezultatų publikavimas

Pagrindiniai disertacijos rezultatai paskelbti 6 mokslinėse konferencijose. Darbo tema paskelbti 5 moksliniai straipsniai: vienas straipsnis respublikiniame žurnale cituojamame Inspec duomenų bazėje [A1], du straipsniai užsienio recenzuojamuose žurnaluose [A3, A2] ir du straipsniai respublikinių konferencijų darbų medžiagose [A5, A4]. Tyrimų rezultatai naudoti 3 kvalifikacinio mokslinio darbo ataskaitose.

1.3.4. Publikacijų sąrašas

- [A1] Leonavičius, R.; Navakauskas, D. Aspects of Melisma Synthesis. Elektronika ir Elektrotechnika, ISSN 1392-1215, 6(48), 2003, p. 18–21. Prieiga per internetą: <http://www.ktu.lt/lt/mokslas/zurnalai/elektr/z48>
- [A2] Leonavičius, R.; Navakauskas, D. Improvement of the Restoration of Melisma by a Signal Synthesis. Izvestija Beloruskoj Inzenernoj Akademii, 1(19), No. 1, 2005, p. 110–113.
- [A3] Leonavičius, R.; Navakauskas, D. Restoration of Melisma by Signal Synthesis. Izvestija Beloruskoj Inzenernoj Akademii, 1(17), No. 2, 2004, p. 64–67.

- [A4] Leonavičius, D. Netiesinių dinaminių sistemų identifikavimo metodų apžvalga. Iš *6-osios Lietuvos jaunųjų mokslininkų konferencijos „Lietuva be mokslo – Lietuva be ateities“ 2003 medžiaga*, 2003, p. 36-48. ISBN 9986-05-473-7.
- [A5] Leonavičius, R.; Navakas, D. Melizmų sintezės tyrimas. Iš *4-osios Lietuvos jaunųjų mokslininkų konferencijos „Lietuva be mokslo – Lietuva be ateities“ medžiaga*, 2001, p. 62-71. ISBN 9986-05-473-7.

1.4. Disertacijos struktūra

Disertaciją sudaro šeši skyriai, kurių pirmas yra įvadas, paskutinis – išvados, bei du priedai. Prieduose pateikiamas sukurtų Matlab funkcijų aprašas ir sąvokų žodynas. Papildomai disertacijoje yra pateikti naudotų žymėjimų ir santrumpų sąrašai bei dalykinė rodyklė.

Darbo apimtis yra 102 puslapiai, kuriuose pateikta: 90 formulių, 38 paveiksliukai, 6 lentelės, 2 algoritmai ir 1 pavyzdys. Disertacijoje remtasi 95 literatūros šaltiniais.

Melizmos vieta netiesinių bei dinaminių sistemų modeliavime

Melizma kaip reiškinys bei pats jį apibūdinantis terminas yra kilęs iš muzikologijos. Kitą vertus, ši disertacija yra skirta skaitinių metodų bei priemonių tinkamų melizmai sintezuoti kūrimui, todėl ji taip pat apima informatikos bei elektronikos mokslus. Taigi šio skyriaus tikslas yra pateikti analitinę, pakankamai plačios sklaidos, literatūros apžvalgą.

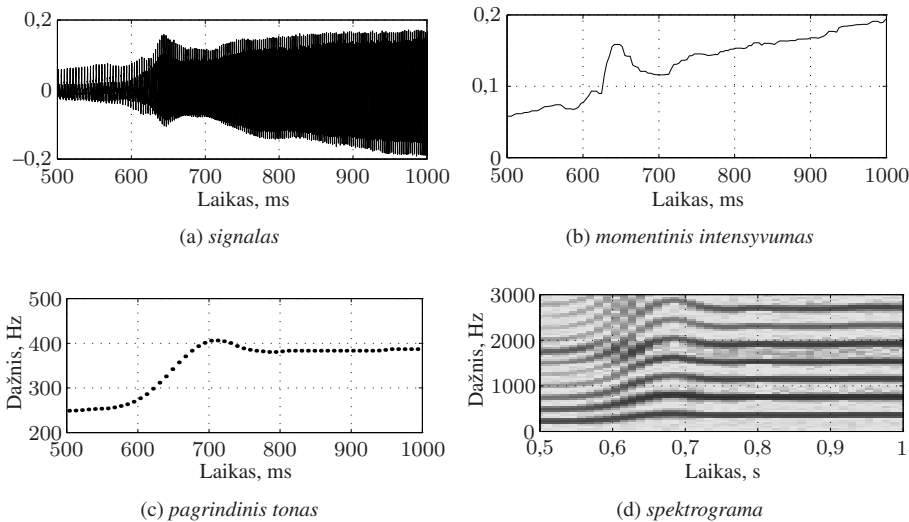
Siekiant pagrindinio tikslo, iš pradžių 2.1 skyrelyje suklasifikuosime bei visapusiškai aptarsime melizmas; vėliau 2.2 skyrelyje trumpai, tačiau iš esmės aptarsime dinaminių sistemų teoriją, kuri bus pagrindu melizmai studijuoti; pagaliau 2.3 skyrelyje pateiksime dirbtinių neuronų tinklų, kaip pasirinkto pagrindinio įrankio melizmai modeliuoti, sąryšį su netiesinėmis bei dinaminėmis sistemomis.

2.1. Melizma ir jos charakteristikos

Disertacijos tyrimų objektas yra melizma. Taigi tiek muzikologinis, tiek ir technologinis apibūdinimai ypač svarbūs tolimesnei studijai. Iš pradžių aptarsime melizmą ir jos rūšis, vėliau pereisime prie melizmą apibūdinančių charakteristikų, pastarąsias pakankamai apibendrintai suklasifikavę į laikines, dažnines ir laikines-dažnines. Šio skyrelio išdavoje bus sudarytas kiek įmanoma pilnesnis melizmos apibūdinimas, leisiantis toliau priimti esminius melizmos modeliavimo sprendimus.

2.1.1. Melizma ir jos tipai

Melizma – smulki ritminės vertės melodinė puošmeninė figūra [11, 57]. Signalų teorijos požiūriu melizma – nestacionarus signalas [59], t. y. pasižymintis staigiai kintančiais



2.1 pav. Foršlago pagrindinės charakteristikos

amplitude (intensyvumu) bei periodu (pagrindiniu tonu).

Priimta melizmas skirstyti į;

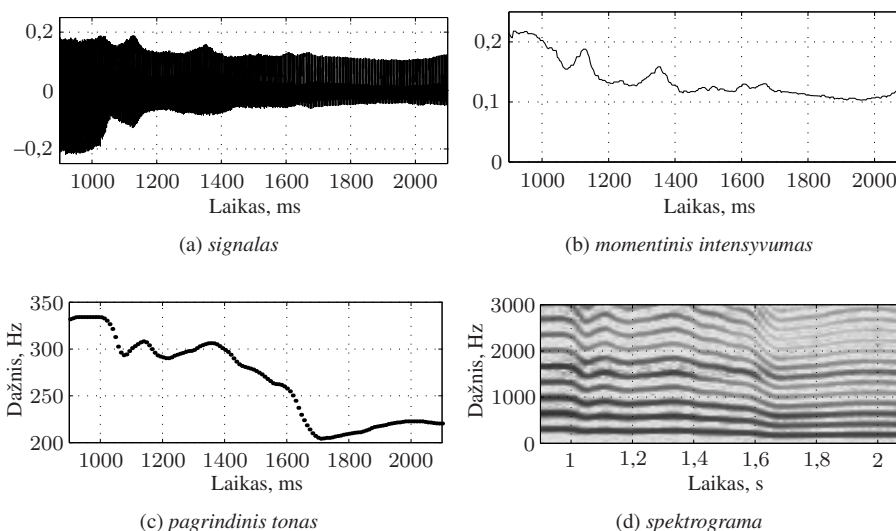
- foršlagus (2.1 pav.),
- grupetus (2.2 pav.),
- mordentus (2.3 pav.),
- treles (2.4 pav.).

2.1 apibrėžimas (Foršlagas). *Foršlagas apibūdinamas kaip sudėtingas, artikuliuotas ir aiškiai išreikštas balso stygų įtempimas.*

Foršlagai skirstomi į du tipus – į trumpą foršlagą ir ilgą foršlagą. *Trumpas foršlagas* yra sudarytas iš vieno ar kelių trumpų garsų, atliekamų ant jau esamo garso arba prailginimu pagražinant garso skambesį. Trumpas foršlagas išgaunamas dviem būdais: maža kryžmine nata arba grupe mažų šešioliktainių natų, viršijančių pagrindinę natą. *Ilgas foršlagas* susideda iš vieno garso ir yra atliekamas ankstesnio garso prailginimo sąskaita. Paprastai ilgojo foršlago trukmė yra dvigubai mažesnė už prieš tai sekusią natą. Tačiau, jeigu nata yra su tašku, tuomet ilgojo foršlago trukmė yra $2/3$ natos ilgio.

2.2 apibrėžimas (Grupetas). *Grupetas – tai grupė natų sudarančių melodinį ornamentą (reiškia supti pagal itališką žodį gruppo).*

Grupetas yra sukuriamas naudojant du pagalbinius garsus – vieną aukštesnį ir vieną žemesnį už pagrindinę natą. Yra keli grupeto tipai, kurie skiriasi tik pagal naudojamų garsų



2.2 pav. Grupeto pagrindinės charakteristikos

skaičių. Pirmas grupetų tipas yra sudaromas naudojant keturis vieną po kito einančius garsus: aukštesnį, pagrindinį, žemesnį ir vėl pagrindinį. Kitas grupetų tipas sudaromas iš penkių garsų: pagrindinio, aukštesnio, pagrindinio, žemesnio ir užbaigiama pagrindiniu.

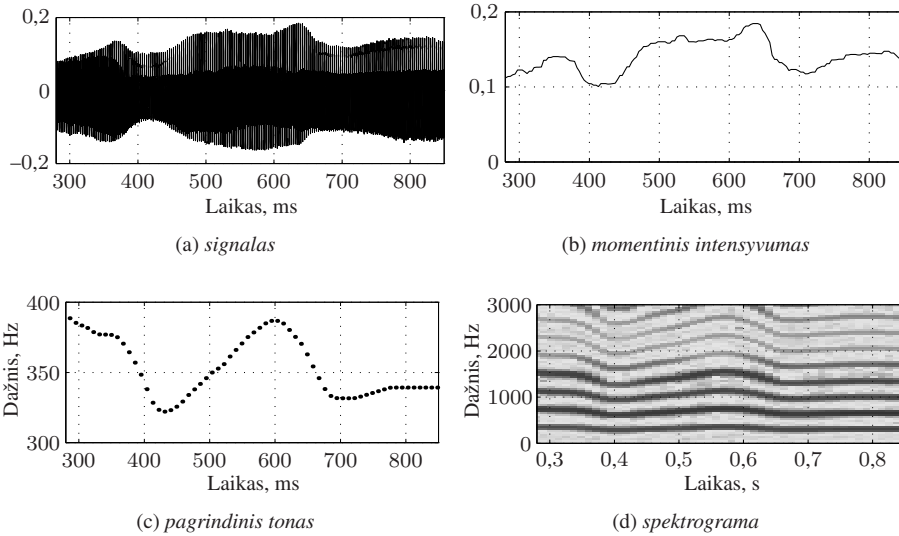
2.3 apibrėžimas (Mordentas). Mordentas – tai melodinis ornamentas sudarytas ant pagrindinės natos greitai keičiantis vienu intervalu aukštesnei arba žemesnei natai (nuo itališko mordente).

Mordentas yra sukuriamas naudojant vieną pagalbinį garsą, kuris nuo pagrindinio garso tonacijos skiriasi tonu arba pustoniu. Melodinė mordento figūra sudaroma iš trijų garsų: pagrindinio, pagalbino ir vėl pagrindinio. Jeigu pagalbino garso tonas aukštesnis už pagalbini, toks mordentas vadinamas *paprastu mordentu*. Jeigu pagalbini garsas yra žemesnis už pagrindinį – tuomet turime *kryžminį mordentą*. Mordentas atliekamas ilginant garso kuriame jis yra trukmę. Pagal trukmę mordentas gali būti įprastos arba *dvigubos* trukmės.

2.4 apibrėžimas (Trelė). Trelė yra melodinis ornamentas sudarytas iš greitai kintančių pagrindinės natos ir visu tonu arba pustoniu aukštesnės natos (nuo itališko trillo).

Trelės ilgis yra lygus garso ant kurio trelė atliekama ilgiui. Trelės klasifikuojamos į tris grupes, priklausomai nuo to iš kokio garso ji yra atliekama: iš žemesnio tono pagalbino, iš pagrindinio arba iš aukštesnio pagalbino.

Egzistuoja eilė charakteristikų kuriomis gali būti aprašytas kalbos signalas. Pereikime prie detalesnio techninio melizmos charakterizavimo laikinėmis, dažninėmis ir laikinėmis-dažninėmis charakteristikomis.



2.3 pav. Mordento pagrindinės charakteristikos

2.1.2. Laikinės charakteristikos

Laikinėmis vadinamos tokios charakteristikos, kurios priklauso tik nuo laiko. Tai vienos populiariausių ir dažniausiai atitikmenų muzikologijoje turinčios charakteristikos. Melizmai apibūdinti pagrindinėmis laikinėmis charakteristikomis priimta naudoti:

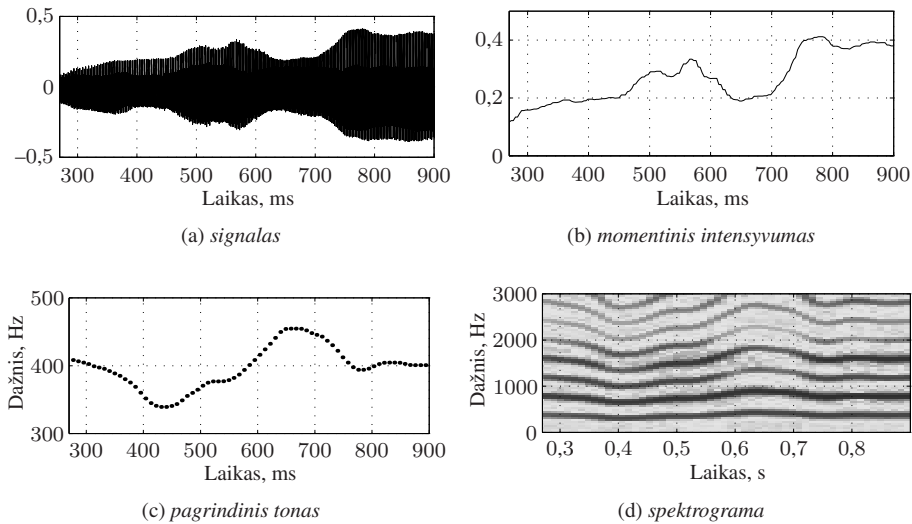
- intensyvumą (b dalis 2.1–2.2. paveiksluose) ir
- pagrindinį toną (c dalis 2.1–2.2. paveiksluose).

Akivaizdu, kad ir signalo momentinis intensyvumas (a dalis 2.1–2.2. paveiksluose) taip pat yra laikinė charakteristika.

Signalų intensyvumas gali būti skaičiuojamas visa eile metodų, t. y. momentinis intensyvumas $I(k)$ randamas paskaičiavus signalo $s(k)$ modulį. Tačiau momentinis intensyvumas nėra plačiai taikomas, nes kalbos signaluose gretimų amplitudžių atskaitos kinta pernelyg greitai. Vietoje jo naudojamas apvalintas momentinis intensyvumas. Tam, kad gauti apvalintą momentinį intensyvumą, galima naudoti slenkančio vidurkio (MA) filtrą:

$$I(k) = \sum_{j=0}^{N_L} s(k-j)b_j, \quad (2.1)$$

čia b_j yra MA filtro koeficientas. Visgi mūsų atveju, dirbant su kalbos signalu, tikslingiau būtų panaudoti integravimą, kuriam suskaičiuoti reikėtų pasirinkti integravimo lango plotį



2.4 pav. Trelės pagrindinės charakteristikos

ir slenkant per visą signalą sumuoti signalo intensyvumą kiekviename lange n :

$$I(n) = \sum_{j=0}^{N_L} s(Mn - j + 1). \quad (2.2)$$

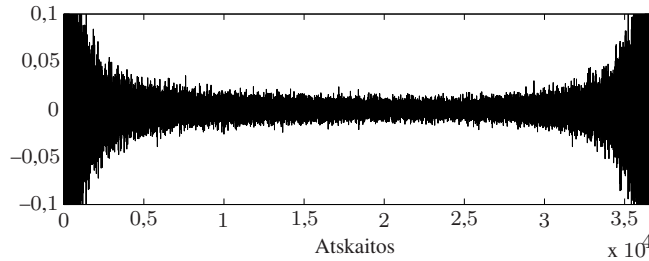
Pagrindinio tono skaičiavimas nėra naujas uždavinys – pagrindinio tono skaičiavimo niuansai detalai yra išnagrinėti [20, 38, 39] ir daugelyje kitų darbų. Šiuo metu sukurta visa eilė algoritmų pagrindinio tono skaičiavimui [4, 18, 21, 25]. Skaičiavimo metodai skiriasi naudojamų skaičiavimų sudėtingumu, kurie savo ruožtu sąlygoja pagrindinio tono radimo tikslumą. Tačiau net ir patys paprasčiausi duoda pakankamai tikslius rezultatus ($\pm 1\%$) [37, 54, 68]. Keli metodai buvo išbandyti [15, 53, 71], galiausiai parinktas anksčiau išbandytas metodas [15], kuris remiasi autokoreliaciniais skaičiavimais. Naudojami tokie parametrai: pirminio filtravimo dažnis 900 Hz, nukirtimo lygis 0,64, normuotos autokoreliacijos lygis 0,3.

Be minėtų charakteristikų kalbos signalus dar priimta apibūdinti šiomis išvestinėmis laikinėmis charakteristikomis:

- kepstru,
- tiesinės prognozės koeficientų kitimu,
- MEL skalės filtrų bankų išėjimo signalų kitimu.

A. Kepstras

Kepstras yra netiesinis signalų analizės metodas. Kalbos signalo kepstro koeficientai teikia vertingą informaciją apie pagrindinio tono arba balso stygų svyravimų dažnį. Šis



2.5 pav. Melizmos kompleksinis kepstras

periodiškumas matomas spektre ir atsiranda dėl periodinio balso stygų virpesio. Kompleksinis signalo kepstras yra apskaičiuojamas randant kompleksinę Furje transformacijos natūrinį logaritmą ir atlikus inversinę Furje transformaciją (2.5 pav.):

$$\hat{h}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[H(e^{j\omega})] e^{j\omega k} d\omega. \quad (2.3)$$

Realus kepstras skaičiuojamas logaritmuojant Furje transformacijos metu gautą ADCH, bei vėliau atliekant inversinę Furje transformaciją:

$$c_x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |H(e^{j\omega})| e^{j\omega k} d\omega. \quad (2.4)$$

Pirmieji kepstro koeficientai dažnai laikomi kalbos signalo požymiais, nes iš jų galima identifikuoti garsą, todėl kepstras sėkmingai naudojamas kalbos atpažinime [44, 45].

B. Tiesinės prognozės koeficientų kitimas

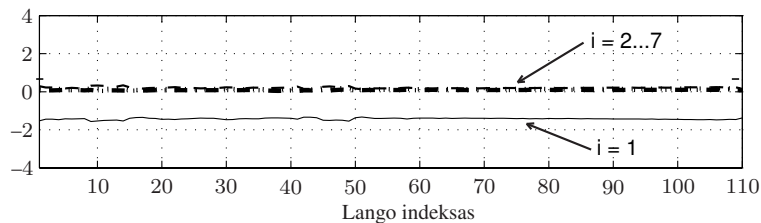
Tiesinės prognozės koeficientų metodas paremtas prielaida, kad kalbos signalas $s(k)$ momentu k gali būti aproksimuotas tiesine kombinacija p ankstesnių kalbos signalo reikšmių. Sistemos funkcija užrašoma kaip:

$$H(z) = \frac{1}{A(z)}. \quad (2.5)$$

Yra eilė metodų, naudojamų tiesinės prognozės koeficientų skaičiavimui. Labiausiai paplitę autokoreliacinis [50] ir kovariacinis, tačiau pastarasis, dėl sudėtingesnio skaičiavimo, naudojamas rečiau už autokoreliacinį. Naudojami N_m -tos eilės tiesinio autoregresinio proceso koeficientai ir stiprinimas pagal:

$$\hat{s}(k) = \sum_{i=1}^{N_m} -a_{i+1} \cdot \hat{s}(k-i). \quad (2.6)$$

Melizmoje (2.6 pav.) tiesinės prognozės koeficientai paprastai nekinta, arba kinta labai nežymiai. Taip gauname todėl, kad melizmos metu vokalinio trakto modelis neturėtų kisti, o turi keistis tiktai pagrindinis tonas. Tiesinės prognozės koeficientai plačiai naudojami



2.6 pav. TPK parametrų kitimas melizmoje

kalbos kodavimui [80] ir atpažinimui.

2.1.3. Dažninės charakteristikos

Galima būtų išskirti šias pagrindines dažnines charakteristikas:

- kalbos signalo amplitudės dažninė charakteristika;
- kalbos signalo fazės dažninė charakteristika;
- kalbos signalo tiesinės prognozės modelio amplitudės dažninė charakteristika;
- kalbos signalo tiesinės prognozės modelio fazės dažninė charakteristika.

Fazinės dažninės charakteristikos kalbos signalų teorijoje taikomos nedažnai. Esminis skirtumas tarp kalbos signalo amplitudės dažninių charakteristikų ir jo tiesinės prognozės modelio yra pateiktas 2.8 pav. Iš čia matyti, kad modelio taikymas gali padėti išskirti esminius kalbos signalo dažnius (rezonansus), tačiau tam turi didelę reikšmę tinkamas modelio eilės parinkimas. Parinkus per žemą eilę, kalbos signalo rezonansiniai dažniai išskirti nebus. Parinkus per aukštą eilę, modelio skaičiavimas bus lėtesnis taip pat modelis išskirs ne tik esminius rezonansinius dažnius, bet ir atskirus rezonanso pikus.

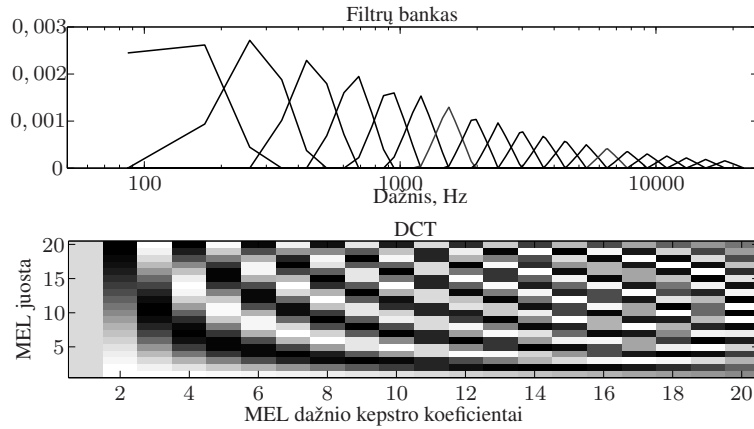
MEL skalės filtrų bankai

MEL skalės filtrų bankai leidžia analizuoti kalbos signalą filtrų režiais logaritminiame mastelyje. MEL skalės dažnis randamas pagal formulę:

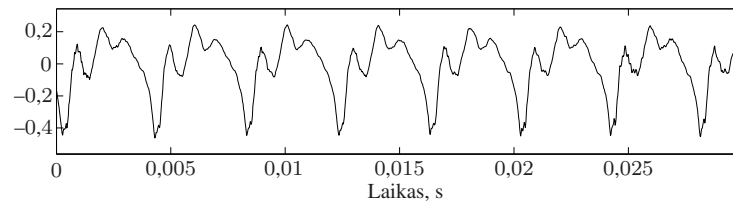
$$F^{\text{MEL}} = 2595 \log\left(1 + \frac{F^{\text{L}}}{700}\right), \quad (2.7)$$

čia F^{L} – linijinis dažnis. MEL skalės filtrai leidžia kalbos signalą analizuoti taip, kad kiekviena juosta turėtų vienodą įtaką kalbos supratimui, nes žmogaus klausa yra logaritmė. Ši skalė buvo sukurta dalinant dažnių skalę į akustiškai vienodai girdimas dalis. Pirminiai MEL filtrai nėra labai plačiai naudojami, tačiau jų pagrindu buvo sukurta eilė išvestinių parametrų, tokių kaip: MEL spektras ir MEL keptras. MEL dažnių idėja sėkmingai taikoma psichoakustiniais principais pagrįstuose kalbos suspaudimo įrenginiuose.

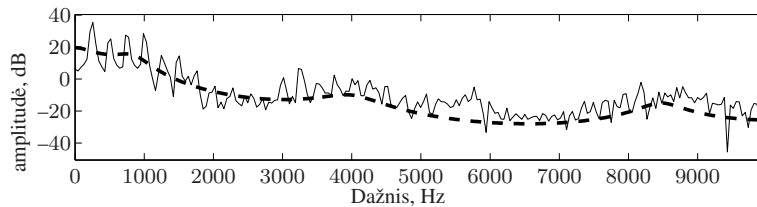
MEL skalės filtrai naudojami kalbos atpažinimo procese [16], kvantavimui išskiriant reikiamas savybes [81], bei daugelyje kitų kalbos apdorojimo sričių [51, 55].



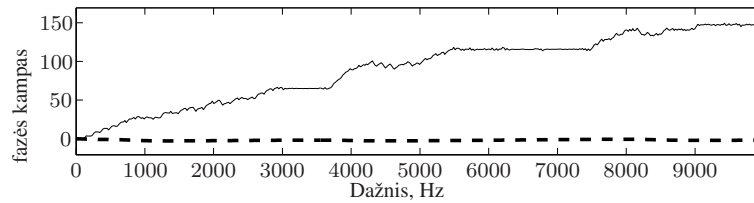
2.7 pav. Melizmos analizė MEL skalėje



(a) Melizmos signalas (trukmė = 30 ms)



(b) Signalo (ištininė linija) ir TPK modelio (punktyrinė linija) ADCH (lango plotis = 1024, TPK eilė 15)



(c) Signalo (ištininė linija) ir TPK modelio (punktyrinė linija) FDCH (lango plotis = 1024, TPK eilė 15)

2.8 pav. Melizmos signalo ir TPK modelio dažninių charakteristikų palyginimas

Kaip matome iš pateikto 2.8(b) pav., tiesinės prognozės modelio amplitudės dažninė charakteristika aproksimuoja kalbos signalo amplitudės dažninę charakteristiką (ADCH), išskirdama kalbos vokalinio trakto parametrus. Kalbos signalas gali būti laikomos kvazi-stacionariu tik trumpą laiko intervalą (apie 10–30 ms). Taigi tik šiame intervale dažninės charakteristikos gali būti skaičiuojamos, todėl žymiai didesnę praktinę vertę kalbos signalų analizei turi laikinės-dažninės charakteristikos.

2.1.4. Laikinės-dažninės charakteristikos

Laikinės-dažninės charakteristikos plačiai taikomos kalbos signalams apibūdinti. Jų privalumas glūdi universalume – vienoje charakteristikoje matyti tiek laikiniai, tiek ir dažniniai pokyčiai. Galima būtų išskirti šias pagrindines laikines-dažnines charakteristikas:

- kalbos signalo spektrogramą (dalis 2.1–2.2. paveiksluose),
- kalbos signalo tiesinės prognozės modelių spektrogramą,
- kalbos signalo tiesinės prognozės modelių šaknų kampų kitimą laike.

Spektrogramos vizualizacija priklauso nuo pasirinktų parametrų: analizės lango pločio, greitosios Furjė transformacijos (GFT) lango pločio, panaudotos lango funkcijos bei lango postūmio. Dažniausiai spektrograma pradedama skaičiuoti su standartiniais parametrais, bet vėliau priderinant parametrus, kad gauti geriausią vizualizaciją arba norint išskirti tam tikras signalo savybes. Melizmos spektrogramose svarbu išryškinti pagrindinio tono kitimą, todėl skaičiavimas atliktas su parametrais kurie geriausiai jį atskleidžia. Skaičiuojant buvo naudojamos melizmos su 44100 Hz diskretizavimo dažniu ir pasirinkti tokie parametrai:

- analizės lango plotis 1323 atskaitos (30 ms);
- GFT atskaitų skaičius 1024;
- signalas apdorojamas Haningo langu;
- analizės lango stumiamas per $1/3$ lango plotį – 441 atskaitą (10 ms).

Taip pat naudojami ir kitais principais paremti spektrogramų radimo būdai [79] naudojami specializuoto pobūdžio spektrogramoms rasti.

2.1 lentelėje nurodytos tirtų melizmų pagrindinės charakteristikos, bei pagrindinio tono ir intensyvumo kitimas. Pateikti skaičiai yra tik informaciniai, paskaičiuoti pagal vieną melizmos rūšies tipą. Charakteristikų kitimas priklauso nuo to kokį melizmos tipą nagrinėjame.

Remiantis pateiktomis charakteristikomis matyti, kad signalo pagrindinis tonas nėra pastovus. Daugumos šiuolaikinės sintezės sistemų veikla pagrįsta tuo, jog pagrindinį toną jos išlaiko pastovų. Pagrindinis tonas jose gali būti pakeistas, tačiau negali pasikeisti to paties garso ribose. Pagrindinio tono pagalba parenkamas sintezuojamos kalbos tonas: tai kaip kalba skambės – vyriškai, moteriškai ar vaikiškai. Nagrinėjamo grupeto pagrindinis tonas kinta nuo 200 Hz iki 330 Hz. Sudėtingesnių melizmų – trelių šis kitimas yra dar didesnis. Pagrindinio tono kitimas akivaizdžiai matomas spektrogramoje 2.4 pav. (d), kurios netolygumai sutampa su pagrindinio tono kitimo vietomis. Iš 2.4 pav. (c) matyti, kad

2.1 lent. Melizmų charakteristikų suvestinė

Melizma	Trukmė, ms	I	ΔI	T_0 , Hz	ΔT_0 , Hz
Foršlagas	1000 ± 100	$0,10 \pm 0,02$	$0,08 \pm 0,01$	340 ± 20	70 ± 10
Grupetas	2000 ± 200	$0,16 \pm 0,02$	$0,06 \pm 0,01$	270 ± 20	80 ± 10
Mordentas	900 ± 100	$0,14 \pm 0,02$	$0,05 \pm 0,01$	360 ± 20	40 ± 10
Trelė	900 ± 500	$0,22 \pm 0,02$	$0,07 \pm 0,01$	400 ± 20	60 ± 10

smarkiai kinta signalo intensyvumas. Sėkmingai melizmos sintezei reikia parinkti metodą, kuris leis sintezuoti greitus signalo pokyčius. Sintzei naudosime išvestinius signalus, charakteristikas ir jų aproksimacijas.

Baigiamosios pastabos

- ▼ Norint kuo plačiau aprėpti melizmas, tolesni tyrimai bus atliekami su visų keturių pagrindinių melizmų rūšių signalais.
- ▼ Esminės melizmų charakteristikos yra intensyvumas ir pagrindinis tonas, o vizualizuoti jas yra patogiausia skaičiuojant spektrogramą.
- ▼ Charakteristikos rodo melizmos greitą kitimą laike, kuri galima būtų modeliuoti dinaminėmis sistemomis.
- ▼ Charakteristikos rodo melizmos priklausomybę nuo pagrindinio tono, kurią galima būtų modeliuoti netiesine sistema.

Taigi pereikime prie sekančio skyrelio, kuriame apžvelgsime dinaminių sistemų teoriją, kuri bus pagrindas tolimesniam melizmos modeliavimui.

2.2. Dinaminės sistemos, jų struktūros, identifikavimas ir patikra

Netiesinių dinaminių sistemų (NDS) modeliavimas yra iššūkį keliantis uždavinys, nes proceso netiesiškumas iškelia ypatingus darbo reikalavimus. Pagrindinis NDS uždavinio tikslas yra universalumas – modelio pritaikomumas įvairioms struktūriškai skirtingoms sistemoms. NDS identifikavimas paprastai skaidomas į uždavio:

1. modelio įėjimų parinkimas,
2. žadinimo signalų parinkimas,
3. modelio architektūros pasirinkimas,
4. modelio dinamikos parinkimas,
5. modelio eilės parinkimas,
6. struktūros ir sudėtingumo parinkimas,

7. modelio parametrų skaičiavimas,
8. modelio patikra (validavimas).

Šiuo metu matematinis procesų analizavimo metodas vis labiau išstumia tikrąją eksperimentinę analizę. Tam yra begalė priežasčių, pagrindinės iš kurių – lankstumas, pigumas, universalumas. Taipogi nemažai procesų tiesiog neįmanoma sukurti laboratorijoje. Šiuolaikinės skaičiavimo technikos pajėgumai leidžia modeliuoti netgi branduolinį sprogimą. Nors bene didžiausi resursai skiriami orų prognozės tobulinimui. Kita sritis, kur taikomas matematinis sistemų modeliavimas – kontrolės ir valdymo įrenginiai. Iš tiesų tai net būtų sunku rasti tokias sritis, kur vienoje ar kitoje formoje nėra taikomas matematinis procesų modeliavimas.

Pereikime prie dinaminių sistemų apžvalgos. NAGRINĖSIME dinaminių sistemų struktūras, jų identifikavimo bei patikros klausimus, siekdami iš esmės apibūdinti tiesinių bei netiesinių dinaminių sistemų teoriją kaip būsimą melizmos modeliavimo terpę.

2.2.1. Dinaminių sistemų struktūros

Dinaminių sistemų struktūrų apžvalgą pradėkime nuo paprastesnio atvejo – tiesinės dinaminės sistemos (TDS).

A. Tiesinių dinaminių sistemų struktūros

Tiesinių dinaminių sistemų modelių aprašymui patogiu naudoti apibendrintą struktūrą, kuri aprėpia visas modelio variacijas pasak kurios, tam tikros tiesinės sistemos išėjimas $s(k)$ laiko momentu (k) gali būti paskaičiuotas filtruojant įėjimo signalą $g(k)$ tiesiniu filtru $G(q)$ (q žymi tiesioginio postūmio operatorių, t. y. $q^{-1}x(k) = x(k-1)$) [47]. Analogiškai filtruojant baltą triukšmą $v(k)$ tiesiniu filtru $H(q)$, galima modeliuoti bet kokią triukšmo charakteristiką. Apjungus įėjimo ir triukšmo perdavimo charakteristikas gauname apibendrintą bet kokios tiesinės sistemos modelio išraišką:

$$s(k) = G(q)g(k) + H(q)v(k) = \frac{B(q)}{F(q)A(q)}g(k) + \frac{C(q)}{D(q)A(q)}v(k), \quad (2.8)$$

arba pertvarkius:

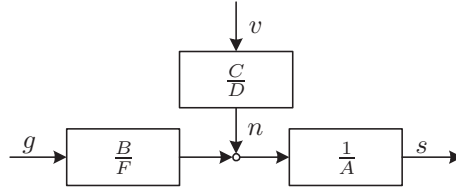
$$A(q)s(k) = \frac{B(q)}{F(q)}g(k) + \frac{C(q)}{D(q)}v(k). \quad (2.9)$$

Čia $s(k)$ yra tiesinės sistemos išėjimas laiko momentu k veikiant įėjimo signalui $g(k)$ ir triukšmui $v(k)$.

Apibendrintasis tiesinis dinaminis modelis (2.9 pav.) nėra naudojamas. Jis prastinamas ir gaunamos realios modelio struktūros. Žemiau bus pateikti dažniausiai naudojami modeliai. Taigi, pagal struktūrą modeliai skirstomi į [59]:

- laiko ašies modelius,
- modelius su grįžtamuoju ryšiu.

Laiko ašies modeliai signalą skleidžia laiko ašyje. Modeliavimo uždavinys šiuose modeliuose supaprastinamas iki modelio radimo naudojantis tiksliai prieš tai buvusiomis



2.9 pav. Apibendrintas tiesinės dinaminės sistemos modelis

laiko atskaitomis. Iš apibendrintos laiko ašies modelio struktūros:

$$s(k) = H(q)v(k) = \frac{C(q)}{D(q)} v(k), \quad (2.10)$$

gaunami trys modeliai:

- autoregresinis (AR),
- vidurkių (MA),
- autoregresinis vidurkių (ARMA).

Autoregresinio modelis gaunamas filtruojant baltąjį triukšmą $v(k)$:

$$s(k) = \frac{1}{D(q)} v(k). \quad (2.11)$$

Tokiu būdu modelio klaida aprašoma formule

$$e(k) = D(q)s(k). \quad (2.12)$$

Autoregresinių modelių radimui naudojamos įvairios metodikos, pvz., mažiausių kvadratų ar koreliacinis metodai. Plačiausiai AR modelių radimui naudojama Yule-Walker lygčių metodika.

Vidurkių modelis yra mažiau paplitęs už AR modelį. Jis užrašomas:

$$s(k) = C(q)v(k), \quad (2.13)$$

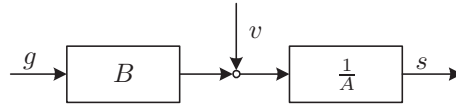
čia klaidos funkcija yra

$$e(k) = \frac{1}{C(q)} v(k). \quad (2.14)$$

Vidurkių modelis ypatingas tuo, kad savo parametrais jis yra netiesinis. Tai sąlygoja sudėtingesnę modelio skaičiavimą, nes MA modeliui negalima pritaikyti koreliacinio skaičiavimo metodikos, o tenka naudoti tiesioginį klaidos minimavimo metodą.

Trečiasis laiko ašies modelis – autoregresinis vidurkių modelis – išreiškiamas formule:

$$s(k) = \frac{C(q)}{D(q)} v(k). \quad (2.15)$$



2.10 pav. Autoregresinis su išoriniu įėjimu modelis

Tuomet prognozės klaida:

$$e(k) = \frac{D(q)}{C(q)} v(k). \quad (2.16)$$

ARMA modelių skaičiavimas susideda iš dviejų dalių: AR modelio radimo dažniausiai mažiausių kvadrato metodikos pagalba ir iteracinio skaičiavimo kuriuo aproksimuojamas klaidos signalas $v(k)$. Tokia metodika leidžia išvengti netiesinio optimizavimo. Tačiau geriausi optimizavimo rezultatai dirbant su AR, MA ir ARMA modeliais pasiekiami naudojant didžiausios tikimybės metodą kuris yra paremtas netiesinio optimizavimo algoritmais, tokiais kaip Levenberg-Marquardt metodika. Pradinius šios metodikos parametrus pakankamai sėkmingai duoda anksčiau aprašytas dvipakopis modelio radimas.

Antroji modelių grupė – modeliai su grįžtamu ryšiu. Šie modeliai, kaip ir laiko ašies modeliai, pagal struktūrą skirstomi į dalinių atvejų modelius:

- autoregresinis su išoriniu įėjimu (ARX),
- autoregresinis slenkančių vidurkių su išoriniu įėjimu (ARMAX),
- autoregresinis-autoregresinis su išoriniu įėjimu (ARARX),
- išėjimo klaidos (OE),
- Box-Jenkins (BJ),
- būsena erdvės.

Autoregresinis modelis su išoriniu įėjimu yra labiausiai paplitęs TDS modelis ir, kuriant modelį, dažniausiai būtent jis yra išbandomas pirmiausiai. Šio modelio privalumai – tikroviškas realių procesų modeliavimas ir nesudėtingas parametrų skaičiavimas, naudojant mažiausių kvadratų klaidos korekcijos metodiką. Taip pat šiam metodui pritaikomi rekursiniai skaičiavimo metodai. ARX modelis (2.10 pav.) aprašomas

$$A(q)y(k) = B(q)u(k) + v(k), \quad (2.17)$$

o jo klaidos funkcija yra

$$e(k) = A(q)y(k) + B(q)u(k). \quad (2.18)$$

Autoregresinis modelis ypač populiarus todėl, kad parametrų skaičiavimui galima pritaikyti mažiausių kvadratų klaidos korekcijos metodiką. Modelio trūkumas yra tas, kad parametruose dalyvauja nuolatinė dedamoji – jie yra nutolę per konstantą nuo savo opti-

malios vertės. Taip pat parametų skaičiavimas nėra pastovus – parametruose nuolatiniė dedamoji dalyvauja net tuomet, kai parametų skaičius artėja į begalybę.

Autoregresinis slenkančių vidurkių modelis su išoriniu įėjimu yra antras pagal populiarumą po autoregresinio modelio. Šis modelis išreiškiamas

$$A(q)s(k) = B(q)g(k) + C(q)v(k), \quad (2.19)$$

o jo klaidos funkcija

$$e(k) = \frac{A(q)}{C(q)}s(k) + \frac{B(q)}{C(q)}g(k). \quad (2.20)$$

Palyginus su ARX, matomas jo patobulinimas – triukšmo modelio naudojimas – kas padaro jį netiesinių parametų erdvėje. Tačiau prognozės klaidos skaičiavimą galima suvesti į pseudo-tiesinį pavidalą:

$$C(q)e(k) = A(q)y(k) - B(q)u(k), \quad (2.21)$$

perrašę gautume skirtuminę lygtį

$$\begin{aligned} e(k) = & a_1s(k-1) + \dots + a_ms(k-N_m) - \\ & - b_1g(k-1) - \dots - b_ms(k-N_m) - \\ & - c_1g(k-1) - \dots - c_ms(k-N_m), \end{aligned} \quad (2.22)$$

kuriai spręsti naudojama netiesinė parametų optimizacija arba rekursinis mažiausių kvadratų klaidos skaičiavimas.

Autoregresinis-autoregresinis su išoriniu įėjimu modelis pasižymi tuo, kad triukšmai filtruojami ne per slenkančio vidurkio filtrą $C(q)v(k)$ (kaip ARMAX), o per AR filtrą $1/D(q)v(k)$. Taigi ARARX modelis išreiškiamas:

$$A(q)s(k) = B(q)g(k) + \frac{1}{D(q)}v(k), \quad (2.23)$$

ir yra retai naudojamas, nes modelio sudėtingumas neatperka modelio naudingumo.

Išėjimo klaidos modelio ypatybė tame, kad triukšmas neveikia paties modelio, tačiau jis veikia tiesiogiai modelio išėjimą. Šie modeliai dažnai tikroviškiau atspindi realius modeliuojamus procesus. Tačiau išėjimo klaidos modeliai yra netiesiniai parametų erdvėje ir juos paskaičiuoti yra gerokai sunkiau negu kitus modelius. OE modelis užrašomas taip:

$$s(k) = \frac{B(q)}{F(q)}g(k) + v(k). \quad (2.24)$$

Box-Jenkins modelis taip pat priklauso prie išėjimo klaidos modelių klasės. Pagal BJ modelio išraišką

$$s(k) = \frac{B(q)}{F(q)}g(k) + \frac{C(q)}{D(q)}v(k), \quad (2.25)$$

matome kad šiame modelyje papildomai atsirado klaidos valdymo funkcija $C(q)/D(q)$.

Tarp visų tiesinių modelių Box-Jenkins modelis yra pats lanksčiausias, tačiau tuo pat metu norint jį paskaičiuoti, reikia skaičiuoti daug parametrų.

Būsenų erdvės modeliai yra gaunami jau aptartus įėjimo-išėjimo modelius perrašius kitokia forma. Išėjimo klaidos būsenų erdvės modelis tuomet būtų

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{b}g(k); \\ s(k) &= \mathbf{x}^T(k)\mathbf{x}(k) + v(k). \end{aligned} \quad (2.26)$$

Būsenų erdvės modelių pagrindiniai privalumai: pirminė informacija apie modelį gali būti įvertinta ir įtraukta į modelio pirminę struktūrą, taipogi regresoriai būsenų erdvės modeliuose paprastai yra mažesni negu įėjimo-išėjimo erdvės modeliuose. m -tos eilės būsenų erdvės modeliui reikės $m+1$ eilės regresoriaus, kai tuo tarpu įėjimo-išėjimo erdvės modeliui – $2m$ regresorių.

Aptarus pagrindines TDS struktūras bei jų savybes, pereikime prie netiesinio dinamių sistemų (NDS) struktūrų, kurios iš esmės remiasi TDS pagrindais, aptarimo.

B. Netiesinių dinamių sistemų struktūros

Autoregresinis modelis su išoriniu įėjimu:

$$A(q)s(k) = B(q)g(k) + v(k). \quad (2.27)$$

Užrašius šią formulę N_m -tos eilės modeliui gautume

$$\hat{s} = b_1g(k-1) + \dots + b_{N_m}g(k-N_m) - a_1s(k-1) - \dots - a_{N_m}s(k-N_m). \quad (2.28)$$

Ši išraiška gali būti išplėsta į NARX (netiesinį ARX). Tuo tikslu tiesinė (2.28) priklausomybė keičiama į nežinomą netiesinę funkciją $F(\cdot)$ taip, kad:

$$\hat{s} = F(g(k-1), \dots, g(k-N_m), s(k-1), \dots, s(k-N_m)). \quad (2.29)$$

Tokiu būdu parametrų $\mathbf{b}$ ir $\mathbf{a}$ paieškos uždavinys išplečiamas iki funkcijos F aproksimavimo. Ši įėjimo-išėjimo modelių klasė vadinama modeliais su išorine dinamika. Nežiūrint į tai, kad išorinės dinamikos modeliai tenkina didelę dalį modelių, iškyla sunkumai norint aprašyti modeliuojamos sistemos būvius. Ypač tai pasireiškia kai didėja sistemos matmenų skaičius. Tokiu atveju neįmanoma tiksliai padalinti sistemą į atskiras dalis ir tenka procesų dinamiką įtraukti į pačią sistemą ir modeliuoti įvertinant vidinius būvius. Šie modeliai vadinami modeliais su vidine dinamika. Išorinės dinamikos modeliai yra plačiausiai naudojami. Pati sąvoka „išorinė dinamika“ reiškia, kad modelį galima padalinti į dvi dalis – netiesinį statinį aproksimatorių ir išorinių dinamių filtrų banką. Aproksimatoriumi $F(\cdot)$ gali būti pasirinkta bet kokia funkcija, tačiau aproksimatorius turi turėti galimybę apdoroti didelį matmenų skaičių. Todėl dažniausiai tam naudojami vėlinimo elementai. Vienas iš pagrindinių faktorių yra priklausomybė tarp įėjimo-išėjimo signalo ir aproksimatoriaus įėjimo erdvės. Reikia žinoti, kaip duotasis įėjimo signalas $g(k)$ ir išėjimo signalas $s(k)$ per laiką k atitinka duomenų pasiskirstymą aproksimatoriaus $F(\cdot)$ įėjime. Pirmos eilės sistemai šią įėjimų erdvę įmanoma pavaizduoti.

Kaip pavyzdį galima pateikti pirmos eilės laike vėluojančią sistemą:

$$s(k) = 0,1 \arctan(g(k-1) + 0,9s(k-1)). \quad (2.30)$$

Galime pastebėti, kad:

- aproksimatoriaus įėjimus negalime valdyti tiesiogiai – galime tik užduoti pradinį $g(k)$, o visi kiti užvėlinti aproksimatoriaus įėjimai ir išėjimai yra pradinio signalo pasekmė;
- kuo žemesnis dažnis, tuo duomenys bus arčiau statinio netiesiškumo;
- paprastai duomenys pasiskirsto tankiau prie statinio netiesiškumo, negu jo kraštuose;
- reikia labai dinamiško įėjimo žadinimo signalo norint padengti plačias įėjimo erdvės sritis duomenimis.

Remiantis šiomis sąlygomis, galime daryti išvadas, kad žadinimo signalo $g(k)$ pasirinkimui reikia skirti ypatingą dėmesį tam, kad iš sistemos išgauti kaip galima daugiau informacijos. Aukštesnės eilės sistemoms įėjimo erdvė tampa daugiamatė, tačiau duomenų pasiskirstymo charakteristikos iš principo lieka tos pačios. Kitas klausimas – kaip atrodo funkcija $F(\cdot)$. Pirmos eilės sistemoms ji gali būti užrašyta kaip $s(k) = F(g(k-1), s(k-1))$. Tačiau modeliuodami NDS su išorine dinamika galimi du modeliavimo atvejai – aproksimatoriaus $F(\cdot)$ bus prieš dinaminčius filtrus ar už jų [76]. Išskiriami du NDS modeliai – Hameršteino ir Vinerio.

Sukeitus Hameršteino sistemoje (2.30) statinio netiesiškumo ir dinaminčius blokus gausime:

$$s(k) = \arctan[0,1g(k-1) + 0,9 \tan(s(k-1))]. \quad (2.31)$$

Kaip matome – Hameršteino [2] sistemoje netiesiškumas yra tik prie įėjimo $g(k-1)$, o tuo tarpu Vinerio sistemoje netiesiškumas yra prie abiejų įėjimų. Todėl Vinerio sistemą daug sunkiau sintezuoti [66, 67]. Kita ypatybė – vieno žingsnio prognozavimo sistemose funkcijos $F(\cdot)$ netiesiškumas yra santykinai mažas, net jei įėjime yra didelis netiesiškumas. Tai leidžia šių sistemų sintezėje taikyti lokalų tiesinį modeliavimą.

Netiesiniai dinaminiai modeliai naudojami dviejose konfigūracijose – prognozavimui ir sintezei [7, 8]. Prognozavimas reiškia, kad pagal prieš tai buvusių įėjimus $g(k-i)$ ir proceso išėjimus $s(k-i)$ modelis spėja kelis žingsnius į priekį. Prognozavimui keliamas reikalavimas – darbo metu turi būti matuojami proceso išėjimai. Tuo tarpu sintezė remiasi tikrai praėjusiais proceso įėjimais $g(k-i)$. Taip pat sintezė nereikalauja matuoti proceso išėjimus darbo metu.

Modeliai mokomi minimizuojant klaidos funkciją, nuo kurios priklauso klaida $e(k)$. Prognozavimui ši klaida $e(k)$ vadinama lygties klaida, o lygiagrečiam sintezės modeliui matuojama išėjimo klaida. Antros eilės modeliui prognozavimas randamas pagal formulę:

$$\hat{s}(k) = F(g(k-1), g(k-2), s(k-1), s(k-2)), \quad (2.32)$$

o sintezė atliekama remiantis prieš tai buvusiais modelio išėjimais:

$$\hat{s}(k) = F(g(k-1), g(k-2), \hat{s}(k-1), \hat{s}(k-2)). \quad (2.33)$$

Vieno žingsnio prognozavimas atliekamas sudarius grįžtamąjį ryšį, kai tuo tarpu sintezė yra rekurentinė. Dviejų žingsnių prognozavimas atliekamas sujungus du vieno žingsnio prognozavimo įtaisus nuosekliai ir į antrojo įėjimą padavus prieš tai buvusio prognozaatoriaus išėjimą. Sintezavimo žingsnių didinimas atliekamas jungiant sintezatorius lygiagrečiai. Išorinės dinamikos modeliai ne visuomet sugeba perimti sistemos netiesines ir dinamines savybes. Tai ypač pasireiškia tuomet, kai didėja sistemos matmenų skaičius. Taip pat tuomet kai sistemos vidiniai ryšiai nėra aiškiai susiję su įėjimu. Tuomet tenka naudoti modelius su vidine dinamika. Tokie modeliai gali būti užrašyti tokioje formoje:

$$\begin{aligned} \hat{x}(k+1) &= h(\hat{x}(k), g(k)); \\ \hat{s}(k) &= t(\hat{x}(k)), \end{aligned} \quad (2.34)$$

čia $\hat{x}(k)$ aprašo vidinius modelio būvius, kurie neturi tiesioginio sąryšio su fizikiniais proceso būviais. Modelio būvių skaičius paprastai parenkamas didesnis negu dinaminė proceso eilė. Vidinės dinamikos modeliams, priešingai negu išorinės dinamikos, nėra būtini prieš tai buvę įėjimai ir išėjimai. Kadangi vidinės dinamikos modeliai neturi išorinių grįžtamųjų ryšių, tai šie modeliai tinka kurti lygiagrečias sistemas.

Tuo tarpu nuoseklios sistemos, tokios kaip vieno žingsnio prognozavimas, šiuose modeliuose yra netaikytinos. Vidinės dinamikos modelių sudaryme naudojamos rekurentinės struktūros tokios kaip atgalinio sklaidimo laike ir realaus laiko rekurentinis mokymas [32]. Be vidinės ir išorinės dinamikos, dar yra parametrų pakeitimų metodas kurio principas yra užfiksuoti netiesinio modelio parametrus viename taške tokiu būdu padarant sistemą tiesinę tame taške. Vėliau jau galima taikyti tiesinius identifikavimo metodus [19, 77].

Kaip matome, yra didelė įvairovė netiesinių dinaminų struktūrų, skirtingų savo sudėtingumo lygiu, klaidos paieškos metodais ir modeliavimo tikslumu. Svarbu sistemai parinkti tinkamą modelį, kuris sugebės modeliuoti būtinas charakteristikas. Tačiau parinktas modelis neturi būti per sudėtingas. Toliau trumpai aptarsime tiesinių ir netiesinių sistemų identifikavimą.

2.2.2. Dinaminių sistemų identifikavimas

Dinaminių sistemų teorijoje identifikavimu yra vadinamas pasirinktos struktūros sistemos parametrų radimas. Iš anksčiau pateikto dinaminų sistemų struktūrų aptarimo matyti, kad yra daug sudėtingų sistemų struktūrų, kurių parametrų skaičiavimas nėra trivialus uždavinys [22]. Kaip ir praeitame skyrelyje, pradėkime identifikavimo būdų analizę nuo paprastesnio – tiesinių dinaminų sistemų – atvejo.

A. Tiesinių dinaminų sistemų identifikavimas

Sistemos identifikavimas apibrėžiamas taip: tai matematinio modelio sukūrimas remiantis sistemos įėjimo-išėjimo duomenimis. Identifikavimas atliekamas keičiant modelio parametrus tol, kol modelio ir sistemos išėjimai bus pakankamai panašūs. Dažniausiai

naudojami panašumo kriterijai yra vidutinė kvadratinė klaida (VKK) arba klaidos modulis. Bet gali būti ir kitokie panašumo nustatymo metodai, pavyzdžiui, koreliacinė analizė. Ji paremta tuo, kad įėjimo signalas neturi būti koreliuotas su modelio klaidos signalu. Panašumo kriterijai gaunami paduodant į sistemą ir modelį tikrinimui parinktą duomenų seką. Geras modelis yra tas, kuris gerai dirbs ne tik su signalu, kuriuo buvo ši sistema identifikuojama, bet ir su kitais signalo fragmentais [59].

Kadangi tiesinė sistema yra netiesinės sistemos dalinis atvejis, pirmiausia aptarkime tiesinės sistemos modelį. Tiesinio modelio optimizavimui naudojamos šios metodikos: a) mažiausių kvadratų; b) rekursinė mažiausių kvadratų; c) tiesinė optimizacija su nelygomis ribinėmis sąlygomis.

Mažiausių kvadratų metodika yra dažniausiai taikomas tiesinių uždavinių sprendimo metodas. Imdami duomenų atskaitas $k = 1, \dots, N$ su duomenimis $\mathbf{g}(k), \mathbf{s}(k)$, taip pat įvardindami triukšmą v gauname uždavinį, kuriame reikia paskaičiuoti modelio išėjimą $\hat{s}$, kuris geriausiai – vidutinių kvadratų metodu – aproksimuoja sistemos tikrąjį išėjimą s . Esant v optimizuojamų parametru, turime $x_1, \dots, x_k$ regresorius ir $\theta_1, \dots, \theta_k$ optimizuojamus parametrus. Remiantis vidutinių kvadratų klaidos išraiška:

$$\mathbf{s} = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k, \quad (2.35)$$

gauname parabolinę mažiausių kvadratų funkciją:

$$I(\boldsymbol{\theta}) = \frac{1}{2} \boldsymbol{\theta}^T \mathbf{H} \boldsymbol{\theta} + \mathbf{h}^T \boldsymbol{\theta} + h_0, \quad (2.36)$$

čia $\boldsymbol{\theta}$ – optimizuojamas parametras, $\mathbf{H}$ – kvadratinė dedamoji – hesianas (klaidos funkcijos antros eilės išvestinė), $\mathbf{h}$ – tiesinė dedamoji ir h_0 – konstanta.

B. Netiesinių dinaminių sistemų identifikavimas

Identifikavimo tikslas – rasti mažiausią parametru $\boldsymbol{\theta}$ klaidos reikšmę. Netiesinis optimizavimas susietas su šiais sunkumais:

- egzistuoja daug lokalių minimumų;
- paviršius apie lokalų minimumą gali būti aproksimuojamas hiperbolinėmis funkcijomis;
- nėra analitinių sprendimų;
- tenka naudoti interaktyvų algoritmą;
- algoritmas sunkiai realizuojamas automatinėmis priemonėmis.

Tiesioginės paieškos algoritmai – paremti tiksliai klaidos funkcijos modifikavimu. Šie algoritmai nereikalauja kad egzistuotų klaidos funkcijos išvestinės. Tiesioginės paieškos algoritmai yra patys paprasčiausi ir skirstomi į: a) simpleksinę paiešką; b) Hooke-Jeeves metodą.

Simpleksinės paieškos tikslas – rasti N_m -tos eilės parametru vektorių kuris minimizuos klaidos funkciją. Ieškant N_m -tos eilės funkcijos minimumo, yra formuojamas $N_m + 1$ -os eilės daugiakampis. Pavyzdžiui, dviejų matmenų uždavinyje simpleksas yra trikampis. Trikampis uždedamas ant klaidos paviršiaus ir žiūrima kurioje iš viršūnių yra didžiausia

klaida. Tuomet šis didžiausios klaidos taškas yra trinamas, o priešingoje jo pusėje dedamas naujas taškas ir formuojamas naujas simpleksas. Ir vėl tikrinama klaida visose simplekso viršūnėse. Taip, žingsnis po žingsnio, simpleksas artėja prie minimumo. Tačiau turi būti numatytos tam tikros papildomos procedūros, kurios užtikrina proceso konvergavimą. Aptikus uždara ciklą, simplekso dydis yra mažinamas. Dažniausiai paieška pradedama su dideliu simpleksu ir kaskart, pasiekus uždara ciklą, jis mažinamas kol pasiekama minimali užduota simplekso reikšmė.

Antroji algoritmų grupė – gradientiniai algoritmai. Tai dažniausiai naudojamų algoritmų grupė kurios užduotis yra pakeisti parametraž θ_{k-1} , proporcingai žingsnio dydžiui η_{k-1} kryptimi p_{k-1} . Pati kryptis apibūdinama dviem parametrais: gradientu q_{k-1} pakreiptu ir padaugintu iš krypties matricos Q_{k-1} :

$$\theta_k = \theta_{k-1} - \eta_{k-1} p_{k-1}; \quad (2.37)$$

$$p_{k-1} = Q_{k-1} q_{k-1}. \quad (2.38)$$

Gradientiniai algoritmai būna: a) tiesinės paieškos (intervalų mažinimo ir intervalų paieškos metodai); b) riboto skirtumo; c) šlaitinio nusileidimo; d) Niutono; e) kvazi-Niutono; f) susietų gradientų.

2.2.3. Dinaminių sistemų patikra

Radus pasirinktai dinaminės sistemos struktūrai parametraž reikšmes, iškyla natūralus klausimas – ar gauta tinkama pasirinktam uždaviniui sistema, o konkrečiau – ar sistema veiks ne blogiau, jai pateikus nematytas to paties proceso signalus? Dinaminės sistemos patikra skirta rasti atsakymus į šiuos klausimus.

Atliekant modelio patikrą, susiduriama su dviejų tipų klaidomis: nuokrypiu (ang. Variance) ir sisteminė klaida (ang. Bias).

Sisteminė klaida gaunama kaip pasekmė to, kad modelis nepilnai atspindi sistemos sudėtingumą – tai yra modelis nepakankamai gerai aprašo sistemą. Ši klaida išlieka net tuomet, kai modelis turi optimalias parametraž reikšmes. Sisteminė klaida matuojama kaip neatitikimas tarp modelio ir sistemos atmetus triukšmų įtaką. Tokios klaidos atsiradimo pavyzdys būtų tuomet, kai bandytume penktos eilės daugianarį aprašyti trečios eilės daugianario modeliu. Taip pat verta atkreipti dėmesį į patį modeliuojamą procesą – modeliuojant aukščiau pateiktą pavyzdį penktos eilės modeliu, galime tikėtis nulinės paklaidos. Tačiau tik tada kai procesas bus tiesinis. Netiesinių procesų modeliavime, nepriklausomai nuo to kiek geras bus modelis, mes vienaip ar kitaip gausime paklaidą. Tiesinei sistemai galime gauti sisteminę klaidą lygią nuliui taikydami pakankamai aukštos eilės modelį. Netiesinės sistemos modeliavimui paprastai naudojami dirbtinių neuronų tinklai [1, 3, 23, 41, 95], neraiškios logikos sistemos, vilnelių modeliai, genetiniai algoritmai [2] ir t.t. leidžia tikėtis tik klaidos mažėjimo didinant modelio sudėtingumą – DNT atveju tai būtų didinant neuronų skaičių ar įvedant papildomus tarp neuronų ryšius (faktiškai keičiant neuronų tinklo struktūrą). Modelio sisteminė klaida tiesiogiai priklauso nuo modelio universalumo – kuo modelis universalnesnis tuo didesnė bus modelio klaida. Modeliavimo tikslas yra minimizuoti modelio sisteminę klaidą, tačiau taip pat reikia turėti omenyje ir

modelio sudėtingumą, nes dažnai yra geresnis paprastesnis modelis su didesne klaida negu sudėtingas, sunkiai realizuojamas modelis su maža klaida.

Antroji modelio klaidos dalis, su kuria tenka susidurti, tai nuokrypio klaida. Kadangi realioje sistemoje dirbama su užtriukšmintu signalu, tai parametrai paprastai nesutampa su optimalia reikšme. Nuokrypio klaida atsiranda dėl modelio parametrų paskaičiavimo netikrumo. Kuo modelis sudėtingesnis tuo nukrypimas gaunamas didesnis. Savo ruožtu mažinant signalo triukšmus ir didinant analizuojamo signalo atskaitų skaičių nukrypimo klaida mažėja. Jeigu fiksuotume duomenų skaičių ir signalo kokybę, tai nuokrypio klaida tiesiškai didėja didėjant modelio sudėtingumui.

Taip gauname dvi klaidos dedamąsias: modelio sisteminę klaidą ir nuokrypį. Sisteminė klaida didėjant modelio sudėtingumui, mažėja, o nuokrypis didėja. Paprastai optimaliu modelio sudėtingumu vadinamas toks kuriam esant šių dviejų klaidų suma yra minimali.

Baigiamosios pastabos

- ▼ Netiesinių bei dinaminių sistemų modeliavimas yra aktualus, bet praktiškai sudėtingas uždavinys.
- ▼ Konkrečiam uždaviniui spręsti reikia parinkti tinkamą NDS modelį, jį identifikuoti bei atlikti gauto modelio patikrą.
- ▼ Kiekvienas iš NDS modeliavimo etapų nėra vienareikšmis, todėl šių sistemų taikymas tampriai siejasi su netiesinės optimizacijos teorija.

Šiuolaikinis NDS įgyvendinimo būdas yra dirbtinių neuronų tinklais grįsti metodai, kuriuos ir aptarsime toliau.

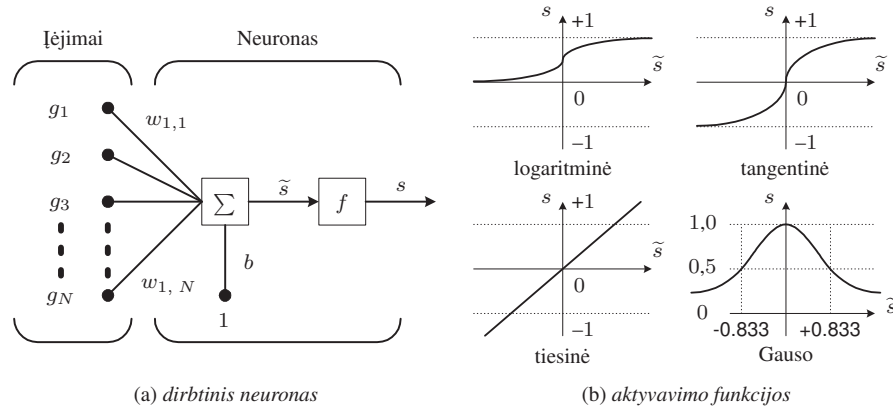
2.3. Dirbtinių neuronų tinklų ryšys su netiesinėmis bei dinaminėmis sistemomis

Netiesinių dinaminių sistemų identifikavimo uždavinys bus sprendžiamas dirbtinių neuronų tinklų (DNT) pagalba.

2.3.1. Specialiosios netiesinės bei dinaminės struktūros

Atliekant netiesinės sistemos modeliavimą, uždavinys dažnai skaidomas į tiesinę dinaminę sistemą ir netiesinę statinę. Dirbtinių neuronų tinklas gali būti taikomas kaip netiesinis elementas, arba galima neuronų tinklą panaudoti kaip visą netiesinę bei dinaminę sistemą.

Iš esmės, netiesinis elementas gali būti realizuojamas šiomis sistemomis: a) vilnelėmis, b) interpoliacinėmis funkcijomis, c) splainais, d) genetiniais algoritmais, e) skleidžiant Voltero eilutėmis, f) Markovo modeliais, g) dirbtinių neuronų tinklais. Neuronų tinklai dėl sugebėjimo modeliuoti netiesines sistemas, plačiai taikomi kalbos atpažinimo uždaviniuose [42, 65, 84]. Šiame skyriuje apsistosime ties modeliais, realizuotais naudojant dirbtinių neuronų tinklus.



2.11 pav. Dirbtinio neurono modelio iliustracija

A. Dirbtinis neuronas

Dirbtinio neurono modelis (2.11 pav.) susideda iš sumatoriaus, į kurį ateina signalai iš visų įėjimų, ir slenksčio funkcijos bei žadinimo funkcijos, kuri apsprendžia tai, kaip kis signalas neurono išėjime. Dažniausiai naudojamos tangentinės ir tiesinės žadinimo funkcijos. Tokiu būdu bet kokio neurono išėjime signalas užrašomas [26]:

$$s = f\left(\underbrace{\sum_{i=1}^N w_{1,i} \cdot g_i}_{\tilde{s}} + b\right) = f(\mathbf{W} \cdot \mathbf{g} + b), \quad (2.39)$$

čia $\mathbf{g}$ – įėjimo signalas, $\mathbf{W}$ – ryšių svoriai, b – slenksčio funkcija, f – aktyvavimo funkcija, N – įėjimų skaičius.

B. Daugiasluoksniai perceptronai

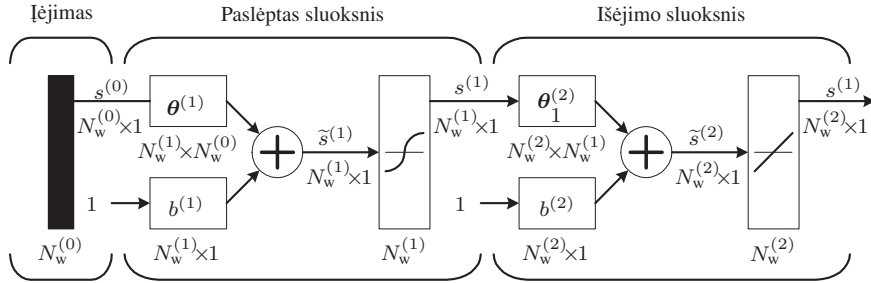
Neuronai su stačiakampe aktyvavimo funkcija vadinami perceptronais. Nuoseklių ryšių daugiasluoksniai dirbtinių neuronų tinklai taip pat vadinami perceptronų tinklais, tačiau šiuose tinkluose nebūtinai taikomos stačiakampės aktyvavimo funkcijos.

Daugiasluoksniai perceptronai (DP) vieni iš sėkmingiausiai naudojamų tinklų. Jie plačiai paplitę ir naudojami kalbos atpažinimo, aproksimavimo, identifikavimo uždaviniuose. Netiesinių aktyvavimo funkcijų naudojimas leidžia juos taikyti netiesiniuose uždaviniuose. Galimybė turėti neribotą paslėptų sluoksnių skaičių leidžia DP spręsti įvairaus sudėtingumo uždavinius (2.12 pav.).

Paprastai, norint parinkti paslėptų neuronų skaičių N_P , laikomasi taisyklės:

$$N_P = \log_2 N_D, \quad (2.40)$$

čia N_D – skirtingų pavyzdžių skaičius. Šia formule remiantis gauname, kad, jeigu turime



2.12 pav. Daugiasluoksnio perceptrono struktūra

8 skirtingus duomenų pavyzdžius, kuriuos tinklui reikia išmokyti, turėtume paimti tinklą su 3 paslėptais neuronais. Tačiau ši formulė tegali būti taikoma tik pradiniam pasirinkimui, nes tinklo sudėtingumas priklauso nuo uždavinio sudėtingumo ir duomenų pasiskirstymo.

C. Išvestiniai modeliai

Yra visa eilė netiesinių modelių, kurie kilę iš tiesinių pridedant netiesinę funkciją. Tai būtų netiesinis RIR (NRIR), netiesinis ARX (NARX), netiesinis OE (NOE), netiesinis ARMAX (NARMAX) ir netiesinis BJ (NBJ) modeliai [63]. Plačiai paplitęs yra NARX modelis (2.13 pav.).

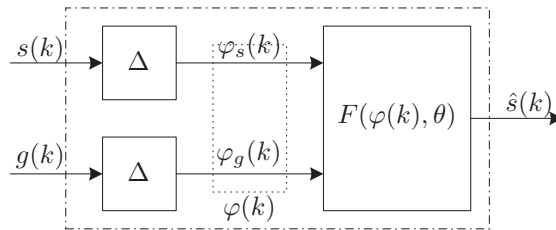
NARMA modelio realizavimui gali būti panaudoti rekurentiniai daugiasluoksniai perceptronų tinklai [32]. Tiesioginio sklidimo tinklai naudojami netiesinių sistemų identifikavime [6, 58, 63].

Tarkime, kad turime funkciją aprašytą

$$\varphi_k(t) = g_k(\varphi(t)) = \kappa(\varphi(t), \beta_k, \gamma_k), \quad (2.41)$$

ir sugrupuojame į vektorių

$$\varphi_k(t) = [\varphi_1(t), \dots, \varphi_n(t)]. \quad (2.42)$$



2.13 pav. NARX modelis (\$\Delta\$ – vėlinimo grandinės)

Tuomet vietoje to, kad tiesinę kombinaciją $\varphi_k(t)$ naudoti kaip modelio išėjimą, mes galime apibrėžti naują regresorių ir aprašyti jį kaip naują sluoksnį dirbtinių neuronų tinkle:

$$g(\varphi, \theta) = \sum \alpha_l \varphi_l \kappa(\varphi, \beta_l, \gamma_l). \quad (2.43)$$

Rekurentiniai tinklai yra paprastų neuronų tinklų patobulinimas įvedant į tinklą dinamiką:

$$\varphi_k(t) = g(\varphi(t - k), \theta). \quad (2.44)$$

2.3.2. Specialiųjų netiesinių bei dinaminų struktūrų identifikavimas

Netiesiniai bei dinaminiai modeliai naudojami dviejose konfigūracijose – prognozavimui ir sintezei [36]. Prognozavimas reiškia, kad pagal prieš tai buvusių įėjimų $g(k - i)$ ir proceso išėjimų $s(k - i)$ modelis spėja kelis žingsnius į priekį. Prognozavimui keliamas reikalavimas – darbo metu turi būti matuojami proceso išėjimai. Tuo tarpu sintezė remiasi tikrai praėjusiais proceso įėjimais $g(k - i)$. Taip pat sintezė nereikalauja matuoti proceso išėjimų darbo metu. Modeliai mokomi minimizuojant klaidos funkciją, nuo kurios priklauso klaida $e(k)$.

Netiesiniams optimizavimo uždaviniams spręsti vieni iš plačiausiai naudojamų yra gradientiniais skaičiavimais pagrįsti identifikavimo algoritmai [47]. Gradientą galima užrašyti per klaidos funkcijos $I(\hat{\theta})$ ir parametų vektoriaus $\hat{\theta}$ pokytį: $\hat{q} = \partial I(\hat{\theta}) / \partial \hat{\theta}$. Gradientiniais skaičiavimais paremto algoritmo kertinis tikslas yra keisti $\hat{\theta}_{k-1}$ reikšmę proporcingai žingsniui η_{k-1} judant kryptimi $\hat{p}_{k-1}$:

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \eta_{k-1} \hat{p}_{k-1}. \quad (2.45)$$

Skaičiavimo kryptis $\hat{p}_{k-1}$ yra užrašoma per gradientą $\hat{q}_{k-1}$, kuris yra normuojamas bei sukamas įvertinant krypties matricą $\hat{Q}_{k-1}$, t. y.:

$$\hat{p}_{k-1} = \hat{Q}_{k-1}^{-1} \hat{q}_{k-1}. \quad (2.46)$$

Optimizavimo uždavinio galutinis tikslas yra pasiekti, kad kiekvienas iteracinis žingsnis mažintų klaidą: $I(\hat{\theta}_k) < I(\hat{\theta}_{k-1})$. Norint, kad klaida mažėtų, reikia teisingai parinkti klaidos mažėjimo krypties matricą $\hat{Q}_{k-1}$. Yra eilė metodų leidžiančių ją apskaičiuoti, o plačiausiai taikomi tiesinio optimizavimo metodai yra šie: a) tiesinės paieškos (intervalų mažinimo ir intervalų paieškos); b) baigtinių skirtumų (ang. Finite Difference); c) greičiausio nusileidimo; d) susietų gradientų metodas; e) Niutono metodas.

Niutono metodas leidžia judėti tiesiai minimumo kryptimi. Krypties matrica $\hat{R}_{k-1}$ randama paskaičiavus klaidos funkcijos $\hat{\theta}_{k-1}$ hesiano $\hat{H}_{k-1}$ inversiją:

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \eta_{k-1} \hat{H}_{k-1}^{-1} \hat{q}_{k-1}. \quad (2.47)$$

Deja Niutono metodas yra kartu ir pats sudėtingiausias. Jis pasižymi eile ypatybių į kurias reikia atkreipti dėmesį, norint šį metodą naudoti: 1) reikalingos antros eilės išvestinės;

2) matricų inversija skaičiavimų sudėtingumą padidina kubiniu dėsniu; 3) atminties poreikis didėja kvadratinio dėsniu, nes reikia saugoti hesiano funkcijos reikšmes; 4) tiesinio uždavinio atveju viena vienintele iteracija pasiekiamas minimumas; 5) metodo neįtakoja tiesinės parametų transformacijos; 6) tinkamas mažiems uždaviniams spręsti (dešimties parametų eilės).

Niutono metodas realiuose optimizavimo uždaviniuose taikomas retai, nes jis yra per sudėtingas didelės apimties uždaviniams, tačiau šis optimizavimo metodas buvo panaudotas visai eilei išvestinių metodų sukurti, kurių daugumą taikoma netiesinių uždavinių sprendimui. Bendru atveju gradientiniai metodai neatsižvelgia į klaidos funkcijos ypatybes.

Baigiamosios pastabos

- ▼ Dirbtinių neuronų tinklų teorija sukūrė pamatus sudėtingų sistemų, tame tarpe bendros paskirties dinaminių bei netiesinių sistemų, pakankamai formalizuotam bei patogiam įgyvendinimui.
- ▼ Praktiškai bet koks sistemos netiesiškumas gali būti įgyvendintas DNT pagalba.
- ▼ Egzistuoja nemažai specializuotų DNT struktūrų, kurios tampriai siejasi su NDS teorija ir gali būti įgyvendintos taikant ne NDS identifikavimo, o DNT mokymo teorijos algoritmus.

Netiesinių bei dinaminių sistemų įgyvendinimui perspektyvūs yra dirbtinių neuronų tinklais grįsti metodai, todėl juos nagrinėsime toliau.

2.4. Skyriaus apibendrinimas

- ▲ Siekiant kuo plačiau aprėpti melizmas, tolimesni tyrimai bus atliekami su visomis keturiomis melizmų rūšimis – foršlagu, mordentu, trele bei grupetu.
- ▲ Pagrindinės charakteristikos, kuriomis apibūdinamos melizmos yra pagrindinis tonas bei intensyvumas – priklausomai nuo melizmos rūšies, šios charakteristikos turi savą kitimo „dėsningumą“.
- ▲ Melizmų charakteristikos pasižymi staigiais ne periodiniais bei kvazi-periodiniais kitimais, kurie galėtų būti modeliuojami netiesinėmis sistemomis.
- ▲ Melizmų esmė yra staigūs kitimai laike, todėl jas modeliuojanti sistema taip pat turėtų būti dinamini.
- ▲ Netiesinių dinaminių sistemų modeliavimas yra aktualus bei sudėtingas uždavinys, o dirbtinių neuronų tinklai šiuo metu yra perspektyvus bei naujas šių sistemų įgyvendinimo būdas.
- ▲ Specializuotos dirbtinių neuronų tinklų struktūros bus toliau tiriamos siekiant sukurti tinkamas melizmai modeliuoti, o šių struktūrų parametrus rasti bus ieškomas tinkamas DNT mokymo būdas.

Melizmos sintezės metodikos pagrindimas ir sudarymas

Charakterizavus melizmas bei apibrėžus melizmų modeliavimo pagrindų ratą pereikime prie konkretaus melizmos sintezės metodikos sudarymo.

Siekiant sukurti originalią metodiką šiame skyriuje tirsime kalbos signalų sintezėje plačiausiai taikomą tiesinės prognozės (kodavimo) metodą (TP(K)) ir jo tobulinimo galimybes. Sekančiame 3.1 poskyryje trumpai supažindinę su tiesine prognoze bei pateikę jos taikymo melizmai pavyzdį tirsime galimybę papildyti TP metodą gerklų signalų bei pagrindinio tono modeliavimo funkcijomis; 3.2 skyrelyje pasiūlysime originalų modifikuotą tiesinės prognozės metodą, kuris remiasi dinaminio TP langų dėstymu; pabaigoje, 3.3 skyrelyje pagrįsdami bei parodydami melizmos modelio vietą pateiksime naują melizmos sintezės metodiką.

3.1. Tiesinės prognozės metodo tyrimas

Kalbos sintezė turi labai plačią taikymo sritį. Dažniausiai ji naudojama telekomunikacijose siekiant sumažinti ryšio kanalo plotį, o taip pat padidinti perduodamos kalbos raiškumą. Tačiau atėjo tie laikai, kuomet kalbos sintezės sistemos turi ne tik padėti žmonių tarpusavio bendravime, tačiau jos pačios privalo perimti komunikavimo su žmonėmis funkcijas. Tokios sistemos atlieka kalbos sintezę. Bet joms iškeliami nauji reikalavimai – kalbos raiškumo ir supratingumo. To negalima pasiekti neperduodant intonaciją, akcentą. Iš šiuolaikinių kalbos sintetatorių pradedama reikalauti sintezuojamos kalbos gyvumo ir netgi emocingumo. Mūsų tyrimų rezultatų taikymo sritis yra senų audio įrašų atstatymas [56, 78].

Šiuolaikinė kalbos sintezė pagrįsta trimis pagrindiniais metodais:

- didelės apimties žodynais [85];
- fonemų sinteze [5, 35, 72, 83];
- mišriais sintezės būdais [13, 30, 82, 86, 92].

Šie metodai netinka sintezuoti dainas, nes kiekvienas žodis turi daug melodinių variacijų, metodai neperteikia atlikėjo asmeninių savybių, kurios būtinos kokybiškai dainos sintezei. Sintezuota kalba gaunama pakankamai monotoniška ir negyva.

Remiantis tradiciniu kalbos signalų modeliavimo metodu, melizmos signalą galima sintezuoti naudojant gerklų signalą bei burnos trakto modelį (filtrą). Tiesinės prognozės koeficientų metodas jau seniai yra taikomas kalbos atpažinime [43] kaip vienas iš būdų išskirti kalbėtojo parametrus. Eksperimentuose naudosisime tiesinės prognozės kodavimą, kurio eigoje bus randami N_m -tos eilės tiesinio autoregresinio proceso (2.2.1 skyrelis) koeficientai ir stiprinimas pagal:

$$\hat{s}(k) = -a_2\hat{s}(k-1) - a_3\hat{s}(k-2) - \dots - a_{k+1}\hat{s}(k-N_m). \quad (3.1)$$

Kadangi tiriamas signalo fragmentas turi būti stacionarus, tai melizma analizuojama 30 ms fragmentais perslenkant analizės langą kas 10 ms.

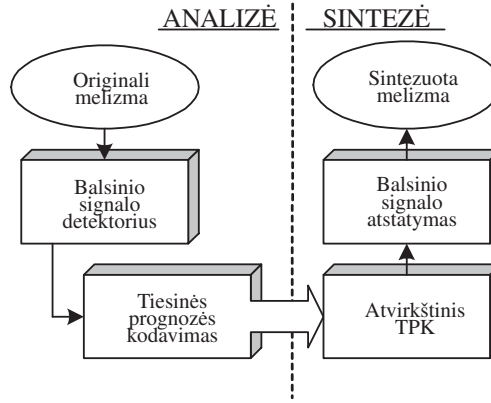
3.1.1. Melizmos sintezė

Melizmos sintezės schema pavaizduota 3.1 paveiksle. Pirmiausia melizmos signalas apdorojamas amplitudės detektoriumi, kuris pašalina pauzes prieš ir po melizmos. Toliau TPK pagalba randami koeficientai. Atvirkščio TPK proceso metu randamas gerklų signalas. Sintezės metu naudojant TPK koeficientus ir gerklų signalą atstatoma melizma. Visa analizė atliekama 30 ms fragmentais perstumiant langą kas 10 ms. Sintezuojant šie signalai sudedami taikant Hamingo langą. Skaičiavimų rezultate gauname visiškai identišką melizmą išskyrus jos kraštus, kuriuose signalas yra silpnas ir todėl neanalizuotas (3.2 pav.).

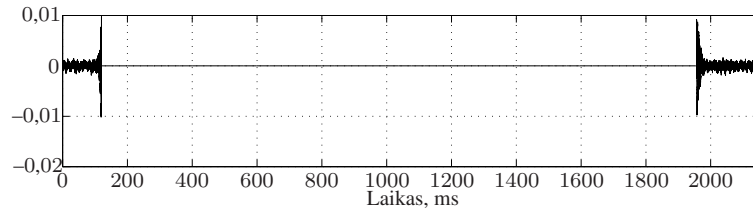
Eksperimentą atlikus su visomis melizmų rūšimis patvirtinta, kad TPK tinka kalbos analizei – melizmą galima išskaidyti į parametrus bei gerklų signalą ir vėliau iš jų atstatyti pradinį signalą. Taip pat spęstas vienas iš klausimų – kokios eilės TPK taikyti. Kuo TPK eilė yra didesnė, tuo geriau modelis perima melizmos ypatumus, tačiau iš kitos pusės – kuo didesnė modelio eilė, tuo labiau modelis perima kalbos signalo triukšmus. Idealus TPK modelis turėtų išskirti švarų gerklų signalą kuris neneštų savyje melizmos informacijos. Todėl gero parametrų išskyrimo matu paimta koreliacija – kuo ji mažesnė tuo geriau modelis aprašo sistemą.

Atlikus eilę eksperimentų paaiškėjo, kad kuo melizma sudėtingesnė, tuo modelis sunkiau ją gali aprašyti, t. y. gerklų signalas tampa labiau koreliuotas su pačiu tiriamuoju melizmos signalu. Didinant modelio eilę koreliacija mažėja sudėtingų melizmų atkarpos (vietose kuriose pagrindinis tonas smarkiai kinta), tačiau paprastesnėms melizmoms (tai ypač būdinga foršlagui) koreliacija pradeda didėti. Eksperimentiškai buvo parinkta 7 modelio eilė. Ji ir naudojama visuose tolimesniuose eksperimentuose.

Toliau tirta, kaip melizmą įtakoja jos parametrų aproksimavimas ir keitimas kitais – dirbtinai sukurtais.



3.1 pav. Sintezė naudojant originalius melizmos parametrus



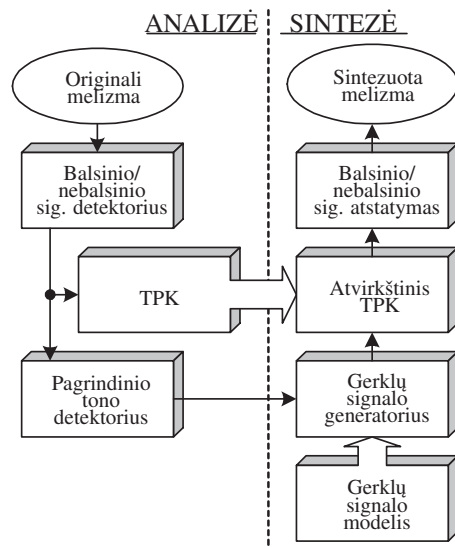
3.2 pav. Sintezės klaida naudojant originalius melizmos parametrus

3.1.2. Gerklų signalo modeliavimas

Gerklų signalas (3.4 pav. (a) dalis) pasižymi informacijos pertekliškumu. Šiuolaikinėse kalbos sintezės sistemose naudojamas dirbtinis gerklų signalas. Tai leidžia atlikti didelę duomenų spūdą ir leidžia ženkliai sumažinti informacinio kanalo plotį. Šis metodas taip pat naudojamas fonemomis grindžiamos sintezės metu. Paimtas dažniausiai taikomas [14] gerklų signalo modelis:

$$\hat{g}(k) = \begin{cases} \frac{1}{2} \left[1 - \cos \left(\frac{\pi k}{V_k} \right) \right], & 0 \leq k \leq V_k; \\ \cos \left(\frac{\pi(k - V_k)}{2(V_n - V_k)} \right), & V_k \geq k \geq V_n; \\ 0, & k > V_n, \end{cases} \quad (3.2)$$

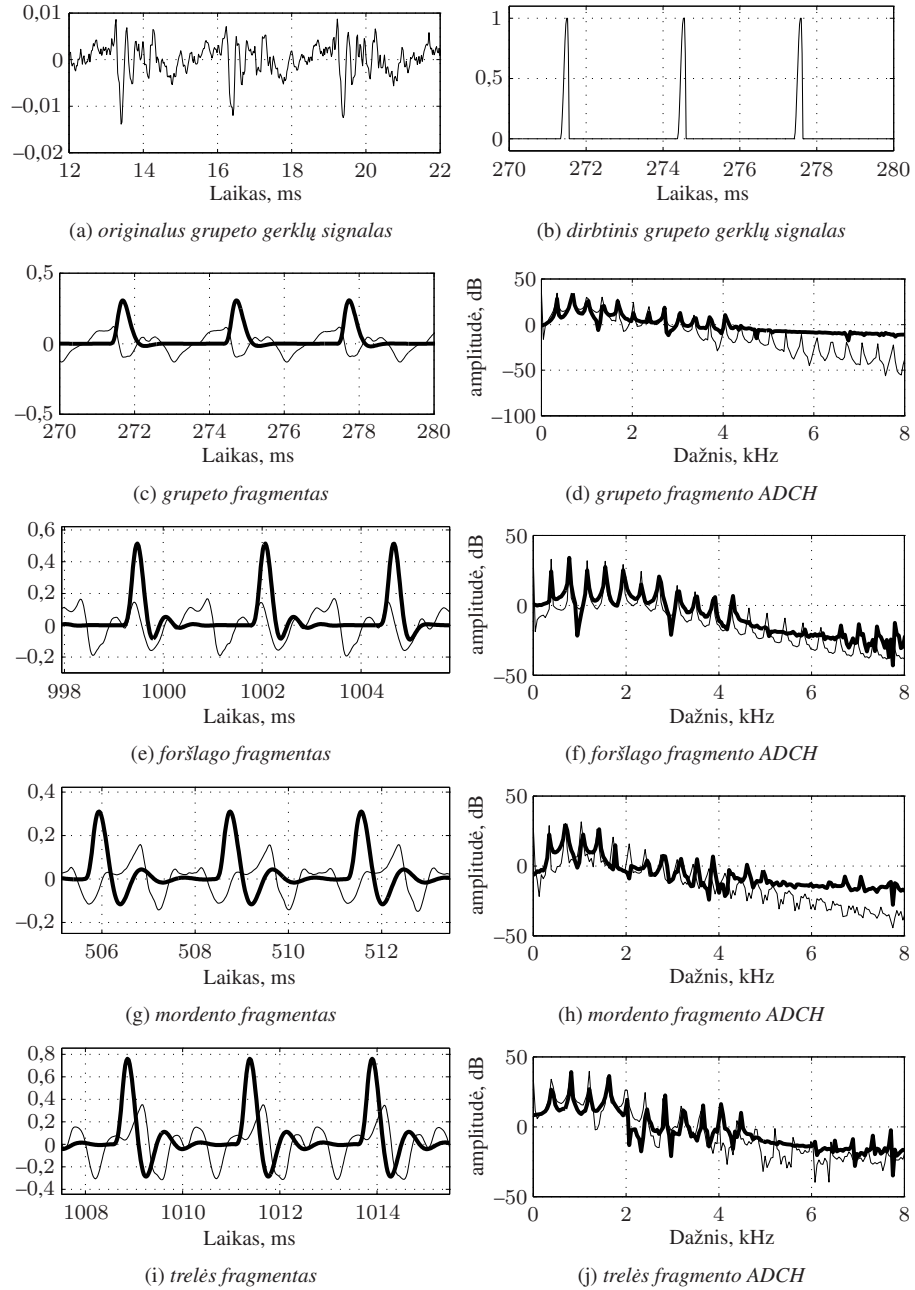
ir gautas dirbtinis gerklų signalas (3.4 pav. (b) dalis). Keičiant koeficientus V_k ir V_n galima keisti gerklų signalo formą. Koeficientas V_k aprašo gerklų signalo kilimo greitį, o V_n – nusileidimo.



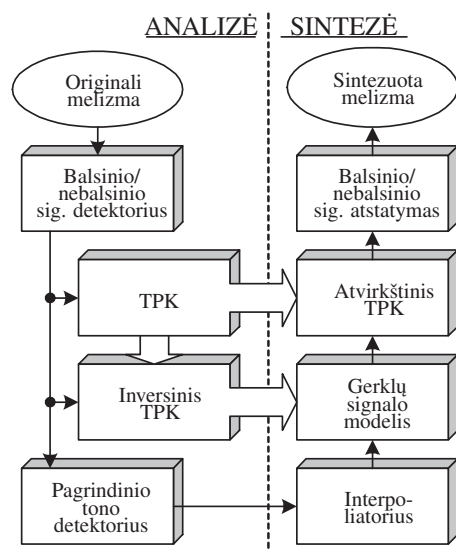
3.3 pav. Sintezė naudojant dirbtinį gerklų signalą

Melizmų sintezė atliekama pagal 3.3 pav. parodytą schemą. Kaip ir ankstesniame eksperimente, melizma per amplitudės detektorių paduodama į TPK bloką ir randami parametrai. Šiuo atveju be TPK parametrų naudojami stiprinimo koeficientai, kurie leidžia susieti atskirus 30 ms melizmos fragmentus. Norint sukurti dirbtinį gerklų signalą, iš melizmos išskiriamas pagrindinis tonas (3.4 pav. (a) dalis), po to jis apdorojamas interpoliatoriumi kuriame atliekama pagrindinio tono spūda (sutankinimas), nes pagrindinis tonas skaičiuojamas pastovios trukmės intervaluose. Pagrindinis tonas randamas filtruojant žadinimo signalą žemo dažnio filtru ir ieškant tarpusavyje koreliuotų fragmentų. Sutankinimas atliekamas aproksimuojant pagrindinio tono funkciją daugianariu. Tai leidžia naudojant anksčiau pateiktą (3.2) išraišką sukurti dirbtinį gerklų signalą susidedantį iš fragmentų pavaizduotų 3.4 pav. (b) dalyje.

Sintezės metu suformuotas dirbtinis gerklų signalas išdėstant (3.2) signalus pagal pagrindinio tono kitimo dėsnį. Gavus ištisinį gerklų signalą jis dalimis praleidžiamas per vokalinio trakto modelį ir įvertinus tikro signalo amplitudę formuojama sintezuota melizma. Tačiau gauta melizma (3.4 pav. (c) dalis) skiriasi ne tik vizualiai bet ir akustiškai – signalė trūksta aukšto dažnio dedamųjų. Tai matoma palyginus tikros ir sintezuotos melizmos amplitudės spektrus (3.4 pav. (d) dalis). Nors signalo amplitudė perteikta tačiau gautas garsas labai duslus ir mechaniškas. Taip yra todėl, kad gerklų signalas perteikia būtent tas kalbos savybes, kurios ją daro „žmogišką“. Keičiant (3.2) koeficientus V_k ir V_n galima pakeisti gerklų signalo formą taip, kad sintezuota melizma turėtų daugiau aukštų dažnių, tačiau tai teigiamo efekto neduoda, nes tada garsas tampa „metalinis“. Tokius pačius rezultatus (3.4 pav.) gavome sintezuojant kitas melizmas: forčiui – (e), (f); mordentui – (g), (h); trelei – (i), (j) dalys. Tikras ir dirbtinis gerklų signalas pernelyg skiriasi, tačiau gauta



3.4 pav. Melizmų sintezė su dirbtiniu gerklų signalu: (a-b) gerklų signalai; (c, e, g, i) melizmų signalai; (d, f, h, j) melizmų amplitudės spektrai. Plona linija atvaizduotas originalas, riebia – modelis



3.5 pav. Sintezė naudojant originalų gerklų signalą

kokybė dažniausiai tenkina standartinės kalbos sintezės reikalavimus. Palyginus sintezuotą melizmą ir plačiai naudojamų sintetorių garso kokybę, buvo pastebėta, kad ji yra labai panaši. Taip yra todėl, kad būtent šis metodas dažniausiai taikomas standartinėje kalbos sintezėje. Eksperimentų metu gautas signalas ypatingas tuo, kad jis turi kintantį pagrindinį toną – išskirtinį melizmos parametą. Apibendrinant galima teigti, kad melizmų sintezei taikytas dirbtinis gerklų signalas netinka.

3.1.3. Gerklų signalo fragmentavimas

Pateikti eksperimentų rezultatai rodo, kad gerklų signalas labai įtakoja atkuriamo garso kokybę. Iširsime galimybę naudoti tikro gerklų signalo fragmentą. Tam sintezę vykdysime pagal 3.5 pav. pateiktą schemą.

Skirtingai nuo 3.3 pav. iliustruoto metodo, šiuo atveju nenaudojama formulė (3.2), tačiau atliekamas inversinis TPK ir randamas gerklų signalo fragmentas. Ištinis gerklų signalas gaunamas gerklų signalo fragmentą komponuojant pagal pagrindinio tono dažnį. Kadangi fragmentas imamas vienas visai melizmai, tai reikia atkreipti dėmesį į tai iš kur jis bus paimtas. Negalima imti iš pačios signalo pradžios ir iš pabaigos, kadangi ten tikrai triukšmai. Be to, atlikus inversinį TPK gerklų signalas pradžioje yra nestacionarus. Šių eksperimentų metu dirbta su 30 ms trukmės gerklų signalo fragmentais. Kadangi melizmos signalas apdorojamas amplitudės detektoriumi, tai analizė vykdoma tikrai tose vietose, kur turime balsingą signalą, o ne pauzę. Visa kita yra identiška prieš tai buvusiam tyrimui. Eksperimento metu gauname sintezuotą melizmą (3.6(a) pav.). Sintezės metu

gauta melizma akustiškai beveik nesiskiria nuo tikrosios, tačiau trelėms pastebėtas skirtumas – melizmoje nebuvo garso tembro pokyčių. Tokioms melizmom kaip grupetas ar foršlagas tai beveik nepastebima, nes jose tembras išlieka beveik pastovus, kai tuo tarpu trelėje – jis kinta. Vadinasi gerklų signalą suvienodinti negalima – bendru atveju reikia imti kintantį gerklų signalą. Šiam teiginiui išbandyti sintezės metodas patobulintas.

Gerklų signalo fragmentas skaičiuojamas ne vieną kartą visai melizmai, bet kiekvienam 30 ms trukmės fragmentui. Norint gauti vidutinį gerklų signalą atspindintį visą 30 ms fragmentą, naudojamos koreliacinės funkcijos:

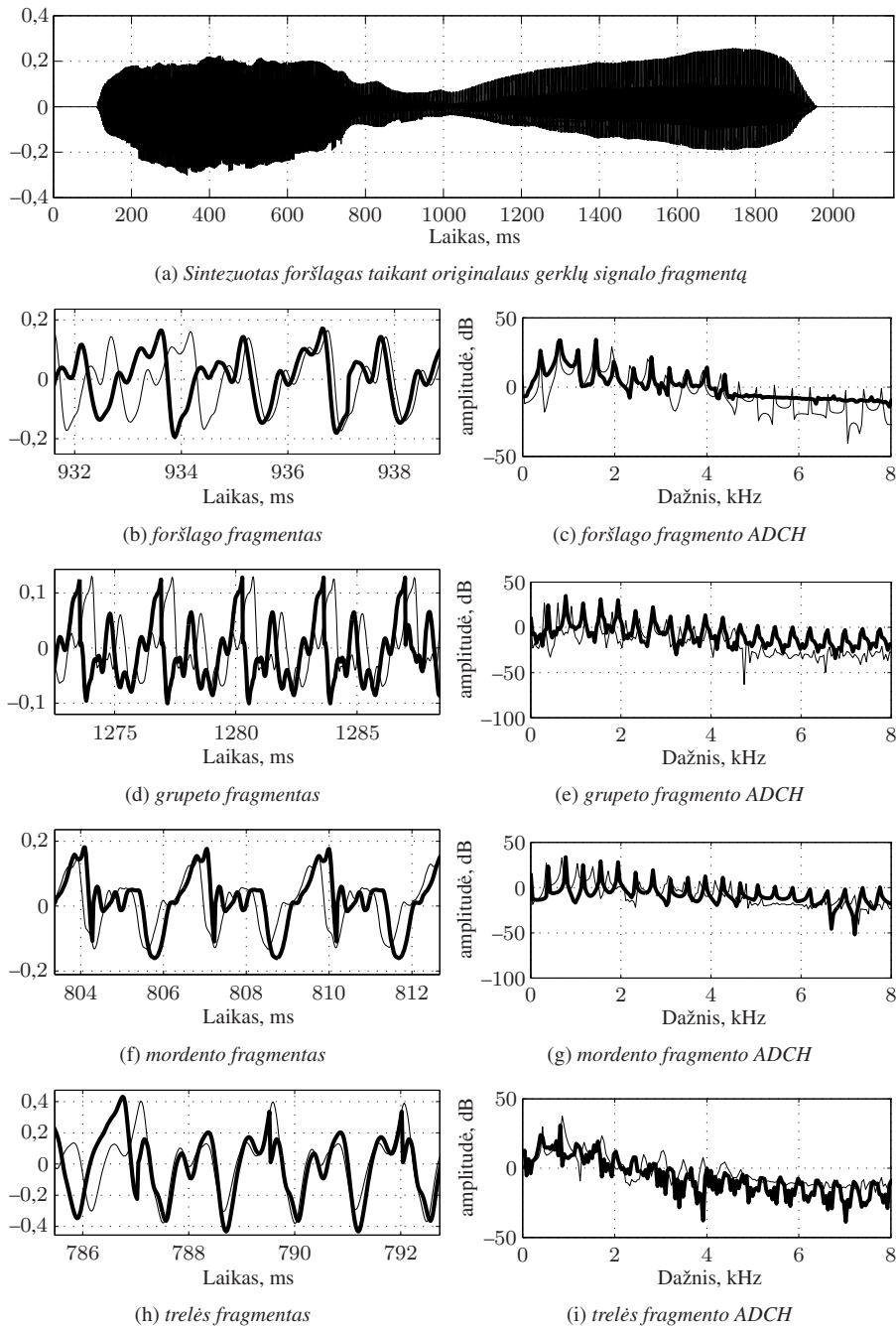
$$\hat{R}_{xy}^N(\tau) = \frac{1}{N} \sum y(k)x(k - \tau), \quad (3.3)$$

čia $x(k)$ ir $y(k)$ yra koreliuoti N trukmės signalai. Aprašytu būdu gautas gerklų signalo fragmentas, atspindintis suvidurkintą gerklų signalą. Jo gavimui iš gerklų signalo imama vidurinė dalis, kurioje signalas yra pastovus, ir jame ieškomas maksimumas. Šis signalas būdavo prieš tai lokaliai suvidurkinamas, siekiant per klaidą nepalaikyti maksimumu šuolių, kurie kaip taisyklė gaunami impulsinių trukdžių vietose. Atliekant TPK ir melizmą skaidant į gerklų signalą ir TPK koeficientus, impulsiniai trukdžiai per šiuos koeficientus neperduodami, nes jie nekoreliuoja su likusiu signalu. Todėl net ir nedideli, vos girdimi impulsiniai trukdžiai išryškėja gerklų signale – visas signalas sumažėja, o impulsiniai trukdžiai lieka beveik nepakitę. Šis metodas gali būti taikomas impulsinių trukdžių radimui ir filtravimui kalbos signale [89]. Suvidurkinus signalą šie šuoliai sumažėdavo ir geriau išryškėdavo gerklų signalo kitimo tendencija. Suradus gerklų signalo maksimumą, imamas fragmentas ir skaičiuojama fragmento koreliacija su likusia gerklų signalo dalimi. Pirmasis šuolis praleidžiamas (tai gerklų signalo fragmentas koreliuoja pats su savim), o kiti mažesni šuoliai yra tose vietose, kur aptikti kiti gerklų signalo periodai. Šiame eksperimente ieškomi penki gerklų signalo periodai ir suvidurkinus gaunamas signalas labiau atspindintis visą gerklų signalą, nei paimtas vienas fragmentas.

Sintezės metu formuojant gerklų signalą, parenkamas atitinkamas gerklų signalo fragmentas. Gerklų signalas formuojamas dviem būdais:

- kuriant vieną ištisinį gerklų signalą, o vėliau iš jo išrenkant 30 ms fragmentus;
- formuojant 30 ms fragmentus tiesiogiai.

Skirtumas tarp šių metodų – analizė ir sintezė vykdoma 2/3 persidengiančiais fragmentais. Jeigu turime 10 000 atskaitų sintezuojamą signalą, tai bendras gerklų signalo ilgis bus tris kartus didesnis. Pirmuoju atveju yra sukuriamas 10 000 atskaitų gerklų signalas ir vėliau iš jo išimami 30 ms fragmentai ir paduodami į sintezę. Šiuo atveju vienas fragmentas susidės iš trijų originalaus gerklų signalo fragmentų. Antruoju atveju ištisinio gerklų signalo nekuriame, o iškart sukuriame 30 ms fragmentai kurių ribose gerklų signalas vienodas. Abiem atvejais sintezuotose melizmosse tembras perteikiamas, tačiau girdima signalo vibracija. Ji atsirado dėl gerklų signalo maksimumų (fazių) gretimose 30 ms fragmentuose nesutapimų. To pasekoje, sumuojant gretimus fragmentus taikant Hamingo langą, galutinis signalas gaunamas iškraipytas. Nesutapimai atsirado dėl netobulo gerklų signalo fragmento radimo ir sutapdinimo. Tai galima iš dalies išspręsti skaičiuojant gerklų signalą tikrai tose vietose kur jis kinta, o pastoviose vietose palikti ankstesnį, bet tikrai prieš tai įvertinant to fragmento stiprinimą.



3.6 pav. Melizmų sintezė taikant originalaus gerklų signalo fragmentą: (a, d, f, h) melizmų signalai; (c, e, g, i) melizmų amplitudės spektrai. Plona linija atvaizduotas originalas, riebia – modelis

Gerklų signalo panašumo kriterijumi parinkta koreliacija. Skaičiuojama koreliacija tarp gretimų gerklų signalo fragmentų bei ieškomas santykis tarp koreliacijos maksimalios ir vidutinės vertės:

$$\frac{\max \hat{R}_{xy}^N(\tau)}{\text{mean } \hat{R}_{xy}^N(\tau)} > 10. \quad (3.4)$$

Ekspimentiškai gauta santykio vertė lygi 10. Jei viršijama ši riba, tai daroma išvada, kad signalai yra pakankamai panašūs ir naudojamas prieš tai buvęs gerklų signalo fragmentas įvertinus esamo fragmento stiprinimą. Sintezuotoje melizmoje nepavyko pilnai atsikratyti signalo virpėjimo. Jis girdimas ties gerklų signalo kitimo vietomis, tačiau jo įtaka žymiai sumažėjusi.

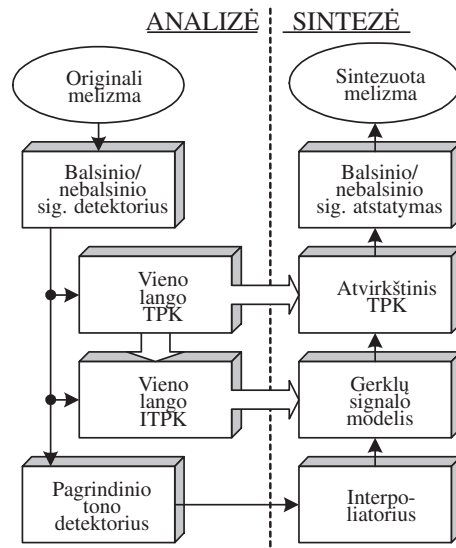
Taikant (3.4) nustatyta, kad melizmos vietose kuriose keičiasi pagrindinis tonas gerklų signalą reikia keisti dažniau – kas kelis 30 ms fragmentus, o vietose kur pagrindinis tonas pastovus – užtenka vieno gerklų signalo kelioms dešimtims fragmentų. Gerklų signalo parinkimui taip pat įtakos turi signalo amplitudė. Kuomet ji maža, gerklų signalas mažiau išsiskiria triukšmo fone ir tampa labiau atsitiktinis. To pasekoje gretimi gerklų signalai mažiau koreliuoti ir dėl to imami dažniau. Tai akivaizdžiai matoma 3.9 pav. pateikto signalo laiko tarpe nuo 0,7 s iki 1,1 s. Pastarąją prielaidą pavyko patvirtinti atlikus analogiškus eksperimentus su mordentu. Mordente pagrindinis tonas kinta pradžioje, todėl ir gerklų signalas dažniau buvo keičiamas melizmos pradžioje.

Tuomet, kai melizmos signalas tampa mažos amplitudės, gerklų signalas paskęsta triukšmuose, todėl nevisuomet teisingai pavyksta rasti gerklų signalo pradžią. Sumuojant atskirus fragmentus dažnai nesutampa signalų fazės, todėl sintezuojant mažos amplitudės signalus girdimas smarkus virpėjimas. Gerklų signalas turi periodinį kitimą, nors ir yra užgožtas triukšmo. Galima bandyti išfiltruoti gerklų signalą žemo dažniu filtru. Eksperimente naudotas Čebyševio I tipo skaitmeninis filtras:

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_1 + b_2 z^1 + \dots + b_{N_m+1} z^{-N_m}}{1 + a_2 z^1 + \dots + a_{N_m+1} z^{-N_m}}. \quad (3.5)$$

Atlikus eilę bandymų parinkta, kad tinkamiausias yra žemo dažnio filtras su 500 Hz nukirtimo dažniu. Gerklų signalą apdorojus šiuo filtru, išėjime gauname į harmoninį virpesį panašų signalą, kurio periodas atkartoja gerklų signalo periodą. Pamatavę šio signalo periodo trukmę gautume pagrindinio tono periodą. Panašiu principu veikia analizėje naudotas pagrindinio tono detektorius. Kadangi pašalinamos aukštadažnės dedamosios, tai signalas tampa žymiai silpnesnis, tačiau tai nemaišo, nes gerklų signalo fragmento pradžią imsime ten, kur filtruotas signalas bus minimalus. Taip galima vienareikšmiškai skirti 30 ms gerklų signaluose rasti reikiamus fragmentus, kurie tiksliau bus sutapdinti sintezės metu.

Kadangi viena svarbiausių melizmos charakteristikų yra pagrindinis tonas, tai jo perteikimui reikia skirti didelį dėmesį. Interpoliatorius, atliekantis gerklų signalo sutankinimą, veikia daugianarių pagrindu. Tai lemia paklaidas atsirandančias melizmos pradžioje ir pabaigoje. Jeigu skaičiuosime pagrindinį toną fragmentais, tuomet galime gauti iškraipytus fragmentų galuose ir todėl nesutaps vieno fragmento pabaiga su kito pradžia. Interpoliavimui galima naudoti 2.3 skyriuje pateiktą dirbtinių neuronų tinklą. Tinklas mokomas



3.7 pav. Sintezė naudojant originalaus gerklų signalo vieną periodą

su turimomis pagrindinio tono atskaitomis. Gauname apmokytą tinklą, į kurį padavę laiko atskaitą išėjime gausime naują pagrindinio tono reikšmę. Šis būdas patikrintas eksperimentiškai. Verta pastebėti, kad už melizmos ribų gavome DNT spėtą pagrindinio tono vertę. Toks rezultatas yra geresnis už daigianario interpoliavimo metu gautą pagrindinio tono reikšmę lygią nuliui. Interpoliavimo daigianariu atveju, esant pakankamai melizmos signalo amplitudei, tačiau per trumpam pagrindinio tono periodui, interpoliacija nepavyks.

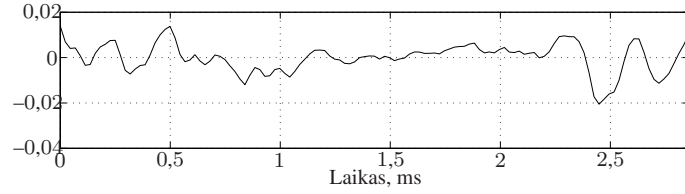
Atlikus sintezę gavome 3.6 pav. (b)–(i) dalyse parodytas melizmas. Panaudojus žemo dažnio filtrą ir DNT pagrindu padarytą interpoliatorių sintezės kokybė pagerėjo. Tai ypač jaučiama melizmos viduryje, kur yra maža amplitudė. Deja bendru atveju, nepavyko gauti universalaus sprendimo tinkančio visų melizmų rūšių sintezei.

Šiame skyrelyje aptarti ir ištirti melizmų sintezės metodai paremti gerklų signalo fragmento taikymu. Galime teigti, kad bendruoju atveju vieno gerklų signalo fragmento taikymas irgi nedavė norimų rezultatų.

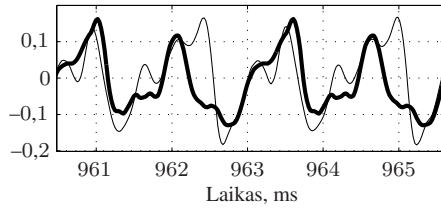
3.1.4. Pagrindinio tono modeliavimas

Ankstesnių skyrių rezultatai rodo, kad geros kokybės melizmos sintezei trukdo faziniai nesutapimai tarp gretimų fragmentų. Darbo eigoje pasiūlytas būdas išskirti iš gerklų signalo vieną periodą ir juo žadinti vokalinio trakto modelį (3.7 pav.).

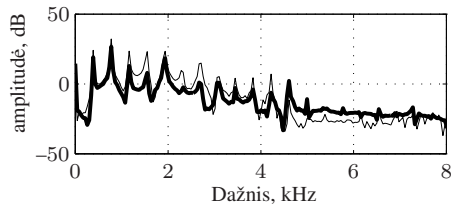
Esminis skirtumas nuo 3.5 pav. pateiktos schemos yra tame, kad 3.8(a) pav. pateiktoje schemoje naudojame viso labo vieną gerklų signalo periodą. Taip pat verta pažymėti, kad vokalinio trakto parametrai melizmos ribose kinta žymiai mažiau negu pagrindinis tonas.



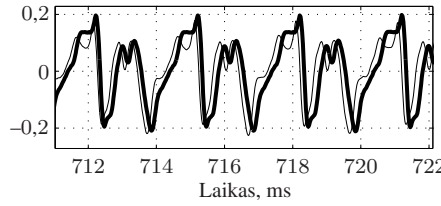
(a) sintezėje naudotas gerklų signalo periodas



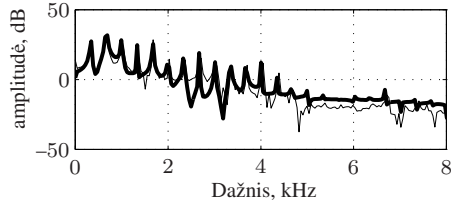
(b) originalios ir sintezuotos (riebi linija) melizmos fragmentas



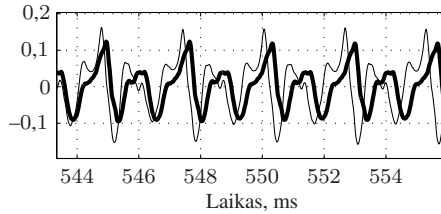
(c) originalios ir sintezuotos (riebi linija) melizmos fragmento GFT



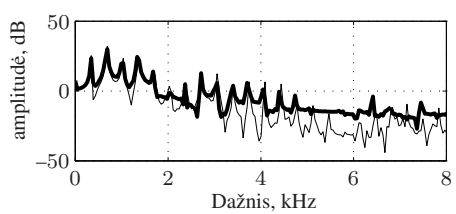
(d) originalios ir sintezuotos (riebi linija) melizmos fragmentas



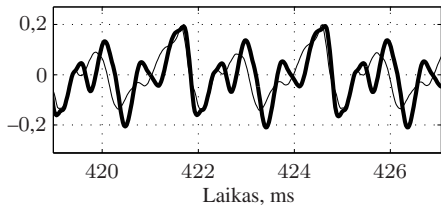
(e) originalios ir sintezuotos (riebi linija) melizmos fragmento GFT



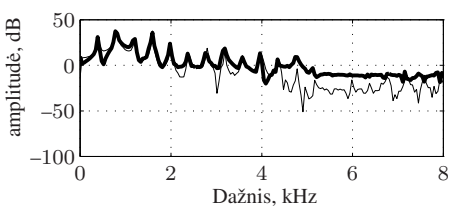
(f) originalios ir sintezuotos (riebi linija) melizmos fragmentas



(g) originalios ir sintezuotos (riebi linija) melizmos fragmento GFT

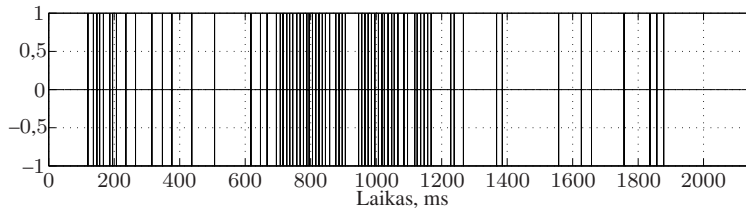


(h) originalios ir sintezuotos (riebi linija) melizmos fragmentas

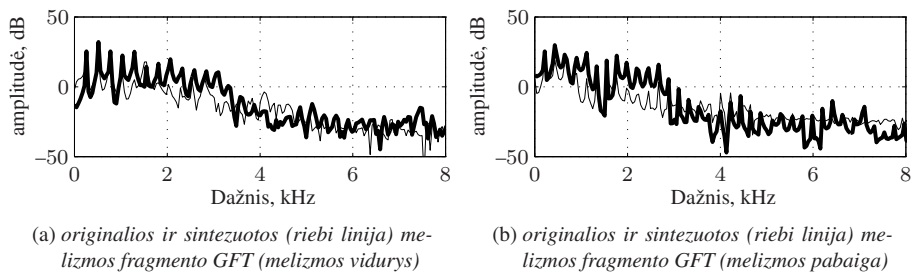


(i) originalios ir sintezuotos (riebi linija) melizmos fragmento GFT

3.8 pav. Melizmų sintezė taikant vieną gerklų signalo periodą: (a) gerklų signalas; (b, d, f, h) melizmų signalai; (c, e, g, i) melizmų amplitudės spektrai. Plona linija atvaizduotas originalas, riebia – modelis



3.9 pav. Gerklų signalo diskretizavimo dažnio keitimo vietos



3.10 pav. Grupeto GFT kitimas skirtingose melizmos atskaitose

Eksperimentiniai tyrimai taip pat patvirtino šią prielaidą. Todėl sintezėje naudoti ne visų 30 ms fragmentų vokalinio trakto parametrai (priešingai nei visuose ankstesniuose eksperimentuose), o tiksliai suvidurkintos vokalinio trakto modelio parametrų reikšmės. Kaip ir ankstesniame eksperimente – gerklų signalo diskretizavimo dažnis keičiamas atsižvelgiant į pagrindinio tono kitimo dėsnį (3.9 pav.).

Atlikus sintezę buvo gautos 3.8 pav. (b)–(i) dalyse pateiktos melizmos. Pastebėtas esminis šio metodo trūkumas – melizma vietomis skamba dirbtinai, nes skirtingose melizmos vietose gerklų signalo forma ženkliai skiriasi (3.10 pav.). Taip gauname todėl, kad gerklų signalo fragmentas buvo paimtas iš melizmos pirmojo trečdalo, tuo tarpu melizmos parametrai keičiasi laike. Klausantis šiuo metodu sintezuotų melizmų galime konstatuoti, kad melizmos pradžioje gauname labai kokybišką sintezę, tačiau vietose, kur pagrindinis tonas smarkiai pasikeičia – skambesys tampa „metaliniu“.

Sintezuota melizma neturi trukdžių dėl nesutampančių atskirų gerklų signalo fragmentų, kadangi sintezės metu sukurtas ištisinis gerklų signalas, o patys gerklų signalai nekeičiami. Keičiant gerklų signalą (3.9 pav.) susiduriama su vienu trūkumu – gerklų signalų radimui skirtinguose 30 ms languose labai didelę įtaką turi melizmos fragmento signalo pradinė fazė bei signalo energijos pasiskirstymas. Net esant gretimiesiems analizės langams galima gauti skirtingus TPK koeficientus, kadangi viename lange bus penki periodai o gretimame bus šeši signalo periodai (dėl pagrindinio tono kitimo melizmoje). Dirbant su standartiniu TPK radimo metodu to išvengti nepavyksta.

Verta pastebėti, kad dėl žmogaus klausos fiziologijos nedideli sintezuoto signalo amplitudės pokyčiai nepastebimi. Net ir smarkiai iškraipytas sintezuotas signalas yra supran-

tamas ir atpažįstamas, todėl gerai sintezei nebūtinai visiškas signalų sutapimas. Daug didesnę reikšmę turi sintezuotos melizmos dažninis vientisumas – net nedideli dažnio svyravimai, atsirandantys dėl gretimų signalo fragmentų persiklojimo, yra pastebimi.

Baigiamosios pastabos

- ▼ Melizmų analizei galima naudoti tiesinės prognozės kodavimo metodą.
- ▼ Tirtas dirbtinis gerklų signalas netinka melizmų sintezei, nes savyje neperneša informacijos apie atlikėjo ir dainos charakteristikas.
- ▼ Melizmos sintezė įmanoma taikant originalaus gerklų signalo fragmentą, tačiau šis būdas netinka visoms melizmų rūšims.
- ▼ Standartinis tiesinės prognozės koeficientų radimo metodas sintezuojant melizmas dažnai veikia blogai dėl smarkių signalo pokyčių.

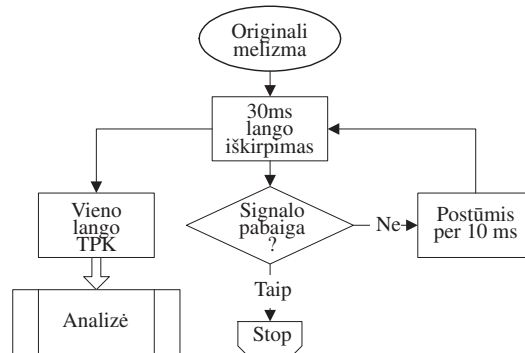
3.2. Modifikuoto tiesinės prognozės metodo tyrimas

Atliekant anksčiau aprašytus melizmos sintezės eksperimentus buvo pasiekti geri rezultatai melizmomis su nestipriai kintančiu pagrindiniu tonu. Tai paaiškinama tuo, kad esant nedideliems pagrindinio tono pokyčiams nebuvo smarkiai keičiamas gerklų signalo fragmento diskretizavimo dažnis. Tuo tarpu atliekant didesnius pagrindinio tono pokyčius gaunamas didelis sintezuotos kalbos signalo iškraipymas – ypač jaučiamas didinant gerklų signalo dažnį. Kita šio būdo silpnoji vieta – sintezė labai priklauso nuo gerklų signalo fragmento melizmos vietos. Remiantis gautais rezultatais turime pripažinti, kad šis būdas netinka visų melizmų tipų sintezei. Kita vertus toks būdas tinka sintezuoti lėtai kintančias melizmas ar atstatyti balsių signalus, nes tose vietose pagrindinis tonas nekinta arba kinta mažai.

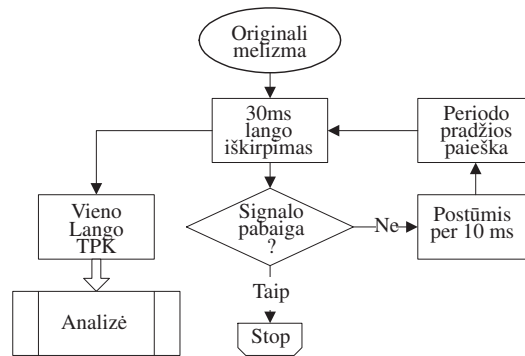
Galime daryti išvadą, kad melizmą skaidant TPK koeficientais, dalis informacijos lieka gerklų signale. Todėl, norint atlikti smarkiai kintančių melizmų sintezę, reikia naudoti nevieną gerklų signalo fragmentą ar jį keisti kas tam tikrą laikotarpį. Praktiniai eksperimentai parodė, kad paprasčiausias gerklų signalo fragmento keitimas kiekviename analizės lange arba kas tam tikrą laikotarpį netinkamas, nes gaunami dideli signalo iškraipymai.

3.2.1. Dinaminis tiesinės prognozės langų dėstymas

Eksperimentiniai tyrimai parodė tiesinės prognozės koeficientų radimo kokybės priklausomybę nuo analizės lango pradžios ir lango trukmės parinkimo. Egzistuoja eilė metodų kurie padeda sumažinti šią įtaką taikant papildomą filtravimą ar įvedant 20 dB triukšmą [24, 40, 60, 62, 72]. Deja, atlikus eksperimentinę patikrą įsitikinta, kad šie metodai neleidžia pasiekti norimų rezultatų. Įvedus triukšmą mes užtriukšminame patį signalą ir nors impulsinių triukšmų dydis sumažėja, tačiau bendras sintezuojamo signalo triukšmingumas dar labiau padidėja. Be to dirbant su melizmų sinteze stengiamasi pasiekti galimai geresnę kokybę, todėl nuspręsta, kad TPK radimo neapibrėžtumui sumažinti reikia kiekviename signalo fragmente TPK ieškoti su tomis pačiomis pradinėmis sąlygomis. Kadangi



3.11 pav. Standartinio TPK parametrų radimo schema

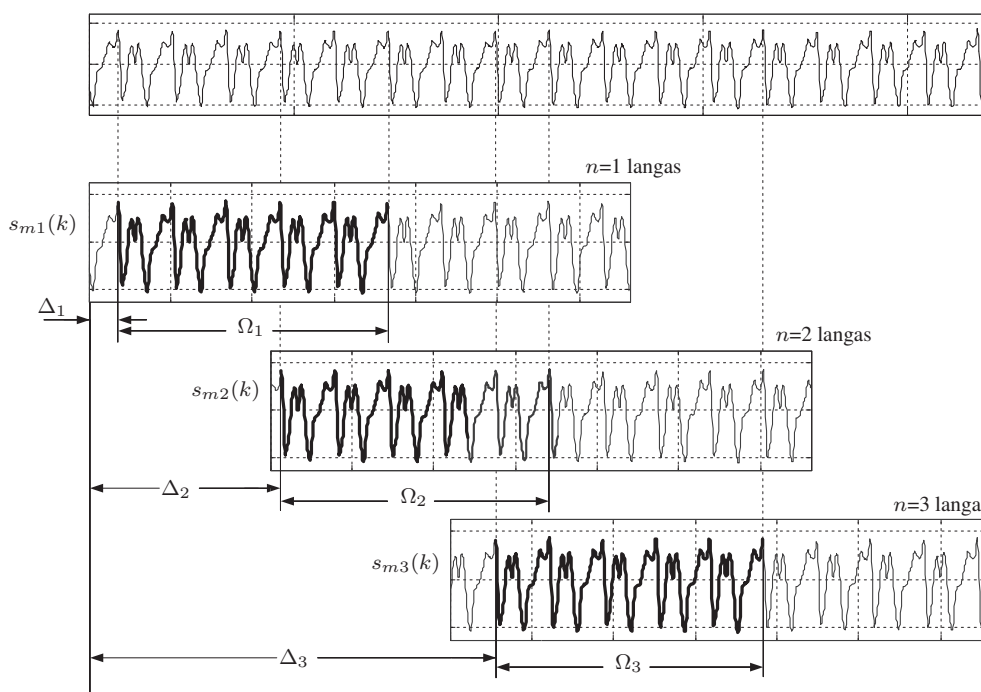


3.12 pav. Dinaminių langų TPK parametrų radimo schema

TPK radimas remiasi autoregresija, tai užfiksavus analizuojamų signalo periodų skaičių ir pradinę fazę, galima būtų tikėtis rezultatų pagerėjimo. Standartinio TPK radimo algoritme (3.11 pav.) taikytas fiksuotos 30 ms trukmės fragmentas.

Siekiant fiksuoti pradinę signalo fazę, buvo pasirinktas periodo pradžios taško radimas. Papildžius struktūrą gauname dinaminių langų TPK radimo algoritmą 3.12, o algoritmo veikimas vaizdžiai parodytas 3.13 paveiksle.

Melizmos signalas, kaip ir klasikiniame analizės variante, yra skaidomas į 30 ms fragmentus. Tačiau analizės metu naudojamas ne visas 30 ms fragmentas, o tik jo dalis. Patikimam TPK darbui užtenka turėti 3–5 signalo periodus. Dar viena priežastis apriboti periodų skaičių yra ta, kad esant fiksuotam 30 ms fragmentavimui bei žemo dažnio pagrindiniam tonui, lieka mažiau nei 6 gerklų signalo periodai. Siekiant užtikrinti vienodas kiekvieno analizuojamo melizmos fragmento sąlygas, fragmento pradžią lyginsime su periodo maksimumu. Maksimumas ieškomas pirmajame melizmos fragmento trečdalyje. Suradus signalo pradžią autokoreliaciniu metodu skaičiuojama pagrindinio tono reikšmė. Taip randamas lokalus pagrindinis tonas, pagal kurį galima rasti analizės lango pabaigą. Tačiau dėl melizmos pagrindinio tono kitimo surasta pabaiga dažniausiai nesutampa su signalo mak-

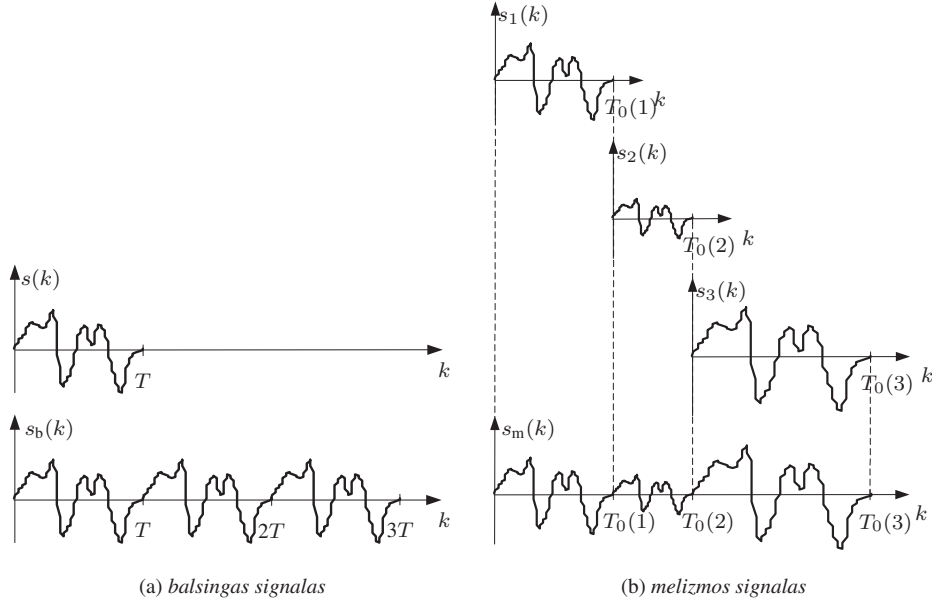


3.13 pav. Tiesinės prognozės metodo modifikavimo iliustracija

simumu. Šio taško aplinkoje vykdoma maksimumo paieška. Surasta maksimali atskaita tampa paskutine analizės lange. Toliau veiksmai kartojami ir surandamas sekantis analizės langas. Melizmos analizės metu gauti dinaminiai analizės lango pradžios ir pabaigos indeksai vėliau naudojami skaičiuojant TPK. Dinaminių langų būdas eliminuoja pagrindinio tono pokyčius, nes nepriklausomai nuo jo kitimo visuomet analizuojamas fiksuotas periodų skaičius. TPK taikomi melizmos fragmentai yra dinamiškai kintantys – lango trukmė yra skirtingas kiekvienam melizmos fragmentui, tačiau periodų skaičius ir jo pradine fazė visuomet išlieka ta pati.

3.2.2. Dinaminio tiesinės prognozės langų dėstymo formalizavimas

Dinaminių langų algoritmą užrašysime analitine išraiška. Tarkime, kad turime T trukmės vieno periodo $s(k)$ signalą (3.14 pav. (a) dalis), kai $s(k) \neq 0 \forall k =]0, T]$, o T yra pastovus periodas. Balsingas $s_b(k)$ signalas tada bus išreikštas kaip $s(k)$ signalo M periodų suma ir $s_b(k) \neq 0 \forall k =]0, M \cdot T]$ (3.14 pav. (a) dalies apačia).



3.14 pav. Signalų sandaros iliustracija

Tuomet balsingą signalą galime užrašyti šia išraiška:

$$s_b(k) = \sum_{m=1}^M s(k - mT). \quad (3.6)$$

3.1 pavyzdys (Balsingo signalo užrašymas)

Tarkime, kad balsingas signalas trunka tris periodus ($M = 3$), tuomet šis signalas išreiškiamas taip:

$$s_b(k) = \sum_{m=1}^3 s(k - mT) = s(k - T) + s(k - 2T) + s(k - 3T). \quad (3.7)$$

Jeigu įvesime kintamą pagrindinio tono periodą – $T_0(m)$ bei kintamą signalo amplitudę – $s_m(k)$, tai melizmos signalą galėsime sukurti iš tokių M pavienių signalų (3.14 pav. (b) dalis). Tada melizmos signalas $s_m(k) \neq 0 \forall k =]0, \sum_{m=1}^M T_0(m)]$ ir užrašomas išraiška:

$$s_m(k) = \sum_{m=1}^M s_m \left(k - \sum_{p=0}^{m-1} T_0(p) \right), \quad (3.8)$$

čia p periodo indeksas analizuojamame melizmos signale.

3.2 pavyzdys (Melizmos signalo užrašymas)

Trijų kvazi-periodų ($M = 3$) melizmos signalas užrašomas:

$$s_m(k) = s_1(k) + s_2(k - T_0(1)) + s_3(k - T_0(1) - T_0(2)). \quad (3.9)$$

Siekiant $s_m(k)$ melizmos signalą analizuoti langais (3.13 pav.), užrašysime melizmos signalą n -tame lange kaip:

$$s_{mn}(k) = s_m(k - \Delta_n), \quad \text{kai } k =]1, \Omega_n]. \quad (3.10)$$

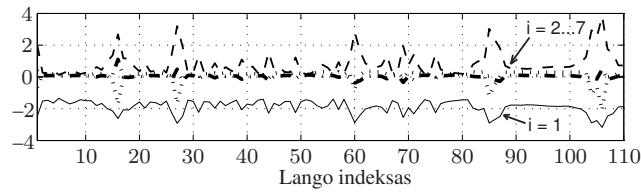
Čia Δ_n yra n -tojo lango poslinkis; o Ω_n yra n -tojo lango trukmė atskaitomis. Šie kintamieji randami taikant šias išraiškas:

$$\Delta_n = \sum_{p=0}^{(n-1)S} T_0(p); \quad \Omega_n = \sum_{p=(n-1) \cdot P}^{nP-1} T_0(p), \quad (3.11)$$

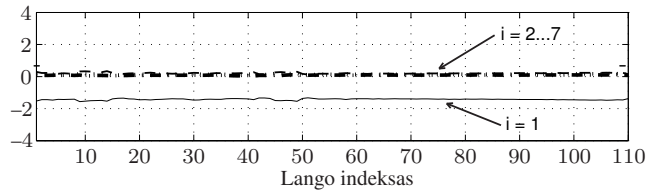
čia S – analizės lango poslinkis išreikštas pagrindinio tono periodų skaičiumi; P – analizės lango plotis irgi išreikštas pagrindinio tono periodų skaičiumi. Atkreipkime dėmesį į tai, kad $\Delta_1 = 0$ bei $T_0(0) = 0$.

3.2.3. Modifikuoto bei standartinio tiesinės prognozės metodų palyginimas

TPK skaičiavimo rezultatai, taikant klasikinių ir modifikuotą metodus, pavaizduoti 3.15 pav. (a) ir (b) dalyse atitinkamai. Matyti, kad melizmos analizei taikant klasikinių



(a) TPK



(b) modifikuotas TPK

3.15 pav. TPK parametrų priklausomybė nuo laiko

TPK, gaunami dideli TPK koeficientų svyravimai. Modifikuoto TPK koeficientų svyravimai žymiai mažesni ir sutampa su pagrindinio tono pokyčiais.

Baigiamosios pastabos

- ▼ Melizmos signalo sintezė galima taikant originalaus gerklų signalo vieną periodą.
- ▼ Vieno originalaus gerklų signalo fragmento kokybiškai melizmai sintezuoti neužtenka – gerklų signalo fragmentas turi būti atnaujinamas.
- ▼ Klasikinis TPK metodas, dėl melizmoje staigiai kintančio pagrindinio tono, lemia TPK koeficientų reikšmių žymius svyravimus.
- ▼ Pasiūlytas standartinės TPK radimo metodikos patobulinimas remiasi dinamiškai keičiama analizės lango trukme ir vieta bei leidžia tiksliau melizmą išskaidyti į gerklų signalą ir vokalo trakto modelį.

3.3. Melizmos sintezės metodikos sudarymas

Matyti, kad melizmos analizėje ypač svarbūs periodiškumo ir intensyvumo valdymas. Šiame skyrelyje pateiksime melizmos metodiką kuri remsis apibendrinto melizmos modelio naudojimu. Tai atliksime iš pradžių pagrindę modeliavimą bei išanalizavę kaip ir kur modelis galėtų būti taikomas.

3.3.1. Melizmos modeliavimo pagrindas

Įsitikinta TPK metodikos tinkamumu analizuoti melizmas. Patikrinta melizmos sintezė naudojant gerklų signalo fragmentą išskirtą iš melizmos pradžios ir keičiant diskretizavimo dažnį pagal pagrindinio tono kitimo dėsnį. Tačiau vieno gerklų signalo periodo visos melizmos sintezei neužtenka, nes gerklų signalas kinta keičiantis pagrindinio tono dažniui. Siekiant tiksliau sintezuoti melizmą, reikia koreguoti gerklų signalą.

Pagrindinis trukdis tokiai melizmos sintezei yra nevisuomet korektiškai veikiantis standartinis TPK metodas, nes jis neįvertina staigių melizmos parametrų kitimo. Gaunami faziniai sintezuotos melizmos fragmentų nesutapimai tose vietose, kur keičiasi pagrindinis tonas. Pasaulyje paplitę sprendimai netinkami, nes įveda triukšmą į gerklų signalą, o kartu ir į sintezuotą melizmą.

Pasiūlytas modifikuotas TPK paremtas dinamiškai kintančiais analizės lango trukme ir vieta, leidžia eliminuoti melizmos pagrindinio tono įtaką atliekant melizmos skaidymą į gerklų signalą ir vokalinio trakto modelį. Metodo modifikacija leidžia gauti pastovius vokalinio trakto parametrus visoje melizmoje. Taigi, melizmos modeliavimas galėtų remtis gerklų signalo sintezavimu ir melizmos parametrų: pagrindinio tono ir intensyvumo, modeliavimu. Gerklų signalo sintezei tinka vienas originalus gerklų signalo periodas kompojuojamas pagal pagrindinio tono kitimą bei stiprinamas pagal intensyvumo kitimo modelį. Pradinės sąlygos melizmos sintezei – iš anksto žinomas melizmos tipas bei signalo prieš melizmą ir po jos reikšmės.

3.3.2. Melizmos modelio vieta

A. Signalo modelis

Netiesiniai adaptyvūs modeliai plačiai paplitę teorinėje signalų analizėje [12]. Dirbant su netiesinėmis sistemomis modeliavimas dažnai skaidomas į tiesinę dinaminę sistemą ir netiesinę statinę. Netiesinis elementas gali būti realizuojamas vilnelėmis, interpoliacinėmis funkcijomis, splainais, genetiniais algoritmais, Markovo modeliais, DNT ir t.t. Taip pat galimas modeliavimas „juodosios dėžės“ principu. Netiesiniai adaptyvūs modeliai sunkiai realizuojami ir reikalauja daug skaičiavimo technikos resursų.

Taikyti netiesinius melizmų sintezei tiesiogiai netikslinga dėl šių priežasčių:

- melizmos signalas stipriai kinta, todėl gauti melizmos signalo modeliai bus ypač dideli ir nepraktiški;
- didėjant modelio apimčiai tampa sudėtinga užtikrinti šių modelių parametrų radijo tikslumą.

B. Tiesinės prognozės modelis

Atlikus eilę eksperimentų, buvo pasiekti geri melizmos modeliavimo rezultatai balsingiems signalams ir melizmom su nestipriai kintančiu pagrindiniu tonu. Klasikiniai metodai leidžia gerai sintezuoti balsinius garsus su nekintančiu pagrindiniu tonu. Melizmom su smarkiai išreikštu pagrindinio tono kitimu sintezės kokybės ženkliai prastesnės klasikinį TPK įtakoja pagrindinio tono pokyčiai. Pasiūlyta eilė metodų kurie leidžia sumažinti iškraipymus [9, 31, 40, 60, 90], tačiau šie patobulinimai taip pat įtakoja sintezės kokybę – nors pagrindinio tono pokytis į melizmos TPK skaičiavimą sumažėja, tačiau įvedami triukšmai į gerklų signalą.

Klasikinių metodų tenka atsisakyti, nes tikslas yra pasiekti kaip galima geresnę sintezuotos melizmos kokybę. Triukšmo lygio padidėjimas gerklų signale, taip pat sukelia signalo kokybinių įverčių neapibrėžtumą, kadangi sunku lyginti originalios melizmos signalą su triukšmais paveikta sintezuota melizma. Klasikiniai metodai leidžia paprastai pasiekti norimų rezultatų dirbant su nekintančio pagrindinio tono signalu [62] arba tuomet kai kokybė nėra lemiamas faktorius.

Modifikuoto TPK metodo taikymas naudojant kintamos vietos ir trukmės signalo langą leido gerokai sumažinti triukšmų įtaką ir pasiekti labai gerus rezultatus dirbant su maža pagrindinio tono variaciją turinčiomis melizmomis, tačiau esant didesniems pagrindinio tono pokyčiams gavome signalo iškraipymus. Iškraipymų atsiradimo priežastis – TPK metodas neperkelia visos informacijos į parametrus. Dalis informacijos lieka gerklų signale. To nepavyksta išvengti naudojant aukštesnės eilės TPK modelius – didinant eilės aukštesnės eilės parametrai neįtakoja signalo kokybės. Klasikiniuose sintezės metoduose [62, 86] naudojamas sintetinis gerklų signalas arba taikomi gerklų signalo bankai. Sintezės modeliuose skirtuose perduoti kalbančiojo intonaciją ir balso specifiką naudojami gerklų signalų bankai atrenkant tinkamą pagal kalbančiojo gerklų signalo ypatybes. Tokie metodai naudojami GSM ryšių sistemose ir leidžia pasiekti pokalbiui priimtina santykį tarp perduodamos informacijos kiekio ir kokybės.

Remiantis atliktais eksperimentais galime teigti, kad TPK koeficientų modeliavimo nepakanka siekiant pilnai apibūdinti melizmą. Melizmos savybės pasilieka gerklų signalą.

TPK išskiria kalbos modelio parametrus, tačiau melizmos metu vokalinio trakto modelis kinta labai nežymiai. Tuo tarpu pagrindinis kitimas vyksta pagrindiniame tone. Pagrindinis tonas ne tik keičia dažnį, bet ir gerklų signalo formą. Modifikuotas TPK metodas parodė, kad nevisuomet užtenka pakeisti pagrindinio tono dažnį – smarkiai keičiant gerklų signalo fragmento pagrindinį toną, gerklų signalas iškraipomas.

C. Signalų parametrų modelis

Iš ankstesnių skyrelių matome, kad melizmoje kinta kalbos signalų parametrai – pagrindinis tonas ir intensyvumas. Vokalinio trakto modelio koeficientai keičiasi nežymiai. Melizmos atveju ypatingai svarbus yra gerklų signalo kitimas, kuris keičiasi priklausomai nuo melizmos pagrindinio tono į kurį reikia atsižvelgti. Gerklų signalo modeliavimas leistų pasiekti žymiai kokybiškesnės melizmos sintezės.

Melizmos sintezėje naudojame kalbos signalo parametrų modeliavimą. Pritaikę modifikuotą TPK metodą gauname gerklų signalą ir vokalinio trakto parametrus. Tuomet, kai analizuojami langai yra arti vienas kito, gerklų signalai tiek savo forma tiek ir amplitudė yra artimi. Tuo tarpu toliant analizuojamoms atskaitoms – gerklų signalo forma ženkliai keičiasi.

Keičiantis pagrindiniam tonui keičiasi gerklų signalo lango trukmė ir vieta, nes pagal pagrindinį toną modifikuotas TPK metodas dinamiškai parenka lango ilgį kad kompensuoti pagrindinio signalo pokytį. Tačiau vien pakeisti signalo diskretizavimo dažnio neužtenka, nes tokiu atveju prarandame signalo formą. Mes žinome gerklų signalo ribines reikšmes, tačiau neturime jo tarpe tarp jų. Jeigu atstumas tarp analizuojamų atskaitų yra mažas – tiek forma tiek ir amplitudė nesikeičia. Tačiau toliant signalo atskaitoms, gretimų atskaitų gerklų signalas pradeda ženkliai skirtis. Todėl tarpuose tarp atskaitų mes turėtume sukurti tam tikrą išvestinį gerklų signalą kuris bus aproksimacija tarp gretimų gerklų signalo atskaitų. Ši aproksimacija turėtų atsižvelgti į tai arčiau kokios atskaitos yra gerklų signalas tam įvedant svorio koeficientus.

3.3.3. Melizmos gerklų signalo sintezės metodika

Melizmos sintezės metodika pateikta 3.1 algoritme. Detaliai aptarkime kiekvieną šios metodikos žingsnį.

3.1 algoritmas (Melizmos sintezės metodika)

A. Pradiniai duomenys

- dainos signalas, $s_d(k)$;
- melizmos sintezės ribos: pradžia – k_p , galas – k_g ;
- melizmos tipas ir jos modelis, $\Psi_m(\Theta)$.

B. Inicializavimas

- radimas dainos signalo pagrindinio tono periodo ir intensyvumo: $\bar{T}_0(k)$, $\bar{I}(k)$;
- radimas dainos signalo tiesinės prognozės koeficientų, $\mathbf{a}(k)$.

C. Gerklų signalo sintezė taikant tiesioginį ir atgalinį signalo apdorojimą

- gerklų signalo skaičiavimas dainos fragmentuose, $g_d^{\leftarrow}(k)$ ir $g_d^{\rightarrow}(k)$
- melizmos modelio taikymas atstatant melizmos gerklų signalą tiesiogine kryptimi:

$$\widehat{g_m^{\leftarrow}}(k) = \Psi_m(g_d^{\leftarrow}(k), \widehat{T_0}(k), \widehat{I}(k));$$
- melizmos modelio taikymas atstatant melizmos gerklų signalą atgaline kryptimi:

$$\widehat{g_m^{\rightarrow}}(k) = \Psi_m(g_d^{\rightarrow}(k), \widehat{T_0}(k), \widehat{I}(k));$$
- melizmos gerklų signalo skaičiavimas pagal tiesioginį ir atgalinį rezultatus, $\widehat{g_m}(k)$.

$$\widehat{g_m}(k) = \widehat{g_m^{\leftarrow}}(k)W^{\leftarrow}(k) + \widehat{g_m^{\rightarrow}}(k)W^{\rightarrow}(k),$$

čia

$$W^{\leftarrow}(k) = \frac{1}{2} \left[1 \pm \cos \left(\frac{\pi k - \pi k_p}{k_g - k_p} \right) \right].$$

D. Melizmos signalo sintezė

$$\widehat{s_m}(k) = \frac{1}{a(k)} * \widehat{g_m}(k).$$

Pirmame žingsnyje apsibrėžkime pradinis duomenis. Tarkime, kad turime dainos signalą $s_b(k)$ kuriame išvis nėra melizmos arba ji yra tačiau sugadinta ir dėl to ją reikia atstatyti. Žinome signalo ribas ties kuriomis melizma turi būti: melizmos pradžią k_p ir pabaigą k_g . Taip pat turime informaciją apie melizmos, kurią reikia sintezuoti, modelį $\Psi_m(\Theta)$.

Sekančiame žingsnyje turime išsiskaičiuoti melizmos sintezei reikalingą informaciją. Tam tikslui rasime pagrindinį toną $T_0(k)$ ir intensyvumą $I(k)$. Pagrindinis tonas signale be melizmos turi būti konstanta arba kisti labai nežymiai. Tačiau tuomet kai mes atstatinėsim melizmą o ne sintezuosime galime tikėtis skirtingos pagrindinio tono reikšmės sintezės pradžios ir galo taškuose. Tai susiję su tuo, kad melizmos gali prasidėti ir baigtis ties skirtingomis natomis. Melizmos intensyvumas $I(k)$ leis sutaptinti sintezuotos melizmos amplitudę su signalo amplitude. Taip pat melizmos sintezei reikia surasti tiesinės prognozės koeficientus. Pirmiausia signalas skaidomas į 30 ms trukmės fragmentus. Pritaikius jau anksčiau aprašytą modifikuotą TPK metodą randame analizės pradžios tašką ir analizuojamo lango ilgį. Skaičiuojame šiame lange tiesinės prognozės koeficientus. Tai leidžia tiksliau nustatyti TPK koeficientus ir eliminuoti signalo pradžios ir krašto efekto įtaką jo radimui. TPK radimo tikslas – signale su nekintančiu pagrindiniu tonu surasti vidutinę TPK reikšmę, nes kaip jau buvo išsiaiškinta – taikant modifikuotą TPK metodą galima išvengti pagrindinio tono įtakos ir visa informacija apie signalą lieką pagrindiniame tone ir gerklų signale. Jeigu signale su kuriuo reikia sintezuoti melizmą pagrindinis pradžios ir pabaigos taškuose skiriasi, tuomet mes turėsime skirtingą pagrindinio tono reikšmę kurios vidurkinti nereikia, nes šia informacija galėsime vėliau pasinaudoti sintezuojant gerklų signalą pradžios ir galo taškuose.

Atlikę inicializavimo darbus pereiname prie pagrindinio tono skaičiavimo. Kaip jau buvo minėta – melizmoje TPK kinta mažai arba išvis nekinta. Kaip tuo tarpu kinta pagrindinis tonas ir gerklų signalas. Kokybiškai melizmos sintezei neužtenka paimti tiksliai vieną gerklų signalo fragmentą ir keičiant diskretizavimo dažnį pagal pagrindinio tono ki-

timą. Tuomet mes prarandame dainos skambėjimo natūralumą. Signale mes paskaičiuojame pagrindinį toną, o iš jo pritaikius inversinį TPK filtravimą galime rasti gerklų signalą $g(k)$. Tačiau tai bus ne melizmos pagrindinis tonas. Melizmos modelis $\Psi_m(\Theta)$ mums duoda du dalykus: pagrindinio tono ir gerklų signalo kitimą melizmos metu. Pagrindinio tono kitimas nusakomas tuo per kiek oktavų ir kada pasikeitė pagrindinio tono dažnis, o melizmos pagrindinio tono kitimas leis prognozuoti gerklų signalą melizmos vietoje. Kadangi turime gerklų signalo kitimo tendenciją, tai kuo toliau nuo signalo sintezuojame, tuo didesnę paklaidą gauname. Paklaidos sumažinimui sintezę atliekame dviem etapais tiesioginiu būdu ir atbuliniu būdu. Tiesioginės sintezės metu atskaitos tašku laikome signalo ant kurio atliekame sintezę originalų gerklų signalą melizmos pradžios taške k_p . Sintezės metu gauname melizmos modelio gerklų signalą $\widehat{g}_m^{\leftarrow}(k)$. Atgalinio sklidimo sintezės metu naudojame originalų gerklų signalo fragmentą melizmos galo taške k_g ir sintezę atliekame judėdami link pradžios taško k_p . Taip mes gauname melizmos gerklų signalo modelį $\widehat{g}_m^{\leftarrow}(k)$ atgaline kryptimi. Tiesiogine ir atgaline kryptimis gauti melizmos gerklų signalo modeliai sujungiami į vieną suminį gerklų signalo modelį $\widehat{g}_m(k)$. Tokiu būdu sumažinę sintezės paklaidą galime tikėtis geresnių sintezės rezultatų.

Baigiamosios pastabos

- ▼ Melizmos modeliavimą galime atlikti naudojant sintezuotą gerklų signalą bei vokalinio trakto modelį.
- ▼ Vokalinio trakto modeliai skaičiuojami pagal dainos fragmentą esantį prieš melizmą.
- ▼ Norint pasiekti geresnę gerklų signalo sintezės kokybę, siūloma naudoti dvikryptį modeliavimą.

3.4. Skyriaus apibendrinimas

- ▲ Melizmos signalo tiesioginis modeliavimas nėra perspektyvus dėl savo sudėtingumo bei dėl to kylančio riboto taikomumo.
- ▲ Didelis dėmesys skirtas tiesinės prognozės metodo taikymo melizmai modeliuoti tyrimams parodė, klasikinio tiesinės prognozės metodo taikymo ribas bei jo neefektyvumą melizmos atveju.
- ▲ Sukurta nauja tiesinės prognozės metodo modifikacija paremta dinaminio tiesinės prognozės langų dėstymu yra tinkama melizmų analizei.
- ▲ Pasiūlyta originali melizmos sintezės metodika pagrįsta tiesinės prognozės kodavimu bei gerklų signalo modeliavimu tiesiogine ir atgaline kryptimis.

Melizmų modelių sudarymas

Sukurta ir praeitame skyriuje pateikta melizmos sintezės metodika leidžia apibendrintai taikyti bet koki melizmos modelį. Melizmos sintezės metodikos tinkamumą galima įvertinti tik konkretaus modelio taikymo atveju.

Šio skyriaus tikslas yra pristatyti originalius melizmos modelius. Iš pradžių, 4.1 skyrelyje pasiūlysiame apibendrintą melizmos modelį, kuriuo remsimės kurdami individualius konkrečių melizmos rūšių modelius. Siekiant sukurti išbaigtą melizmos modeliavimo aplinką, 4.2 skyrelyje pasiūlysiame tinkamus melizmų modelių identifikavimo, grindžiamo dirbtinių neuronų tinklų teorija, būdus. Pagaliau, 4.3 skyrelyje sudarysiame visų keturių melizmos rūšių – foršlago, grupeto, mordento ir trelės – matematinius modelius bei apytikriai nustatysime šių modelių dydį lemiančių parametrų kitimo ribas. Sekančiame skyriuje, modeliuodami realias melizmas, rasime modelių parametrų tikrąsias reikšmes.

4.1. Apibendrinto melizmos modelio kūrimas

Ankščiau nutarėme melizmą sintezuoti taikant gerklų signalo modeliavimą. Taigi, dar kartą, remiantis 3.1 algoritmu, apibendrintai (nekreipiant dėmesio į apdorojimo kryptį) užrašykime melizmos gerklų signalo išraišką:

$$\widehat{g}_m(k) = \Psi_m(g_d(k), \widehat{T}_0(k), \widehat{I}(k)). \quad (4.1)$$

Matome, kad melizmos gerklų signalas $\widehat{g}_m(k)$ gali būti randamas taikant funkciją Ψ_m , kuri priklauso nuo:

- $\widehat{g}_d(k)$ – be melizmos (toliau jį vadinsime dainos signalu) gerklų signalo,

- $\widehat{I}(k)$ – melizmos signalo intensyvumo,
- $\widehat{T}_0(k)$ – melizmos signalo pagrindinio tono.

4.1.1. Gerklų signalo modeliavimas

Dainos gerklų signalo radimui taikysime klasikinį inversinį tiesinės prognozės kodavimo modelį. Pirmuoju žingsniu, dainos signalui pritaikę tiesinės prognozės kodavimo modelį, rasime šio modelio parametrų reikšmes. Antruoju žingsniu tą patį dainos signalą apdorosime gauto TPK inversiniu modeliu. Šio apdorojimo rezultatas bus nagrinėjamos dainos gerklų signalas.

(4.1) pateiktą funkciją Ψ_m galima išreikšti įvairiai, tačiau jos esmė glūdės rasto gerklų signalo apdorojimo (priklausomo nuo melizmos charakteristikų – pagrindinio tono ir intensyvumo) būdo aprašyme.

Siūloma Ψ_m išreikšti remiantis 3.2.2 skyrelyje aprašyto dinaminių langų dėstymo principais. Melizmos sintezės atveju, sintezės lango postūmis bus lygus lango pločiui ($S \equiv P$), todėl S kintamasis nebus nagrinėjamas. Taip pat įvertinsime tai, kad šiuo atveju apdorosime ne kalbos signalą, o gerklų signalą. Taigi, remiantis (3.10) ir (3.11) funkcija Ψ_m išreiškiama taip:

$$\widehat{g}_m(k) = \sum_{n=1}^N \sum_{p=\kappa_n}^{n \cdot P} \frac{\widehat{I}(k)}{\widehat{I}_d(k)} g_{d\kappa_n} \left(\left\lfloor \frac{\widehat{T}_0(p)}{\widehat{T}_{0d}(\kappa_n)} \cdot k \right\rfloor - \sum_{h=\kappa_n}^{n \cdot P-1} \widehat{T}_0(h) - \Delta_n \right), \quad (4.2)$$

čia N – sintezės langų skaičius; P – pagrindinio tono periodų sintezės lange skaičius; $g_{d\kappa_n}(k)$ – κ_n -tasis dainos gerklų signalo pagrindinio tono periodas; $\widehat{T}_{0d}(\kappa_n)$ – κ_n -tojo dainos pagrindinio tono periodo trukmė atskaitomis; $\widehat{T}_0(p)$ – melizmos pagrindinio tono periodo trukmė atskaitomis; κ_n – n -tojo sintezės lango indeksas pagrindinio tono periodais. Išraiškoje (4.2) vykdomus skaičiavimus iliustruoja 4.1 pav. Iš šio paveikslėlio matyti, kad $\widehat{g}_m(k)$ sintezuojamas stumiant sintezės langą dydžiu:

$$\Delta_n = \sum_{p=0}^{(n-1)P} \widehat{T}_0(p), \quad (4.3)$$

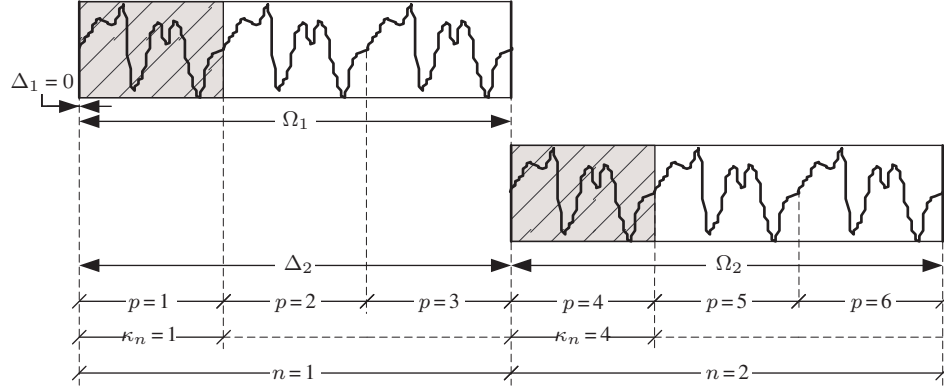
tuo pačiu išlaikant pastovų sintezės lange esančių pagrindinio tono periodų skaičių P . Kita vertus, sintezės lango plotis atskaitomis tampa kintamu:

$$\Omega_n = \sum_{p=(n-1)P}^{nP-1} \widehat{T}_0(p). \quad (4.4)$$

Kiekvienas sintezės lango postūmis vykdomas remiantis κ_n išreiškiamu pagrindinio tono indeksu, t.y.:

$$\kappa_n = (n-1)P + 1. \quad (4.5)$$

Sintezuojamos melizmos gerklų signalas konkrečiame sintezės lange remiasi tik vienu dai-



4.1 pav. Apibendrinto melizmos modelio veikimo iliustracija

nos gerklų signalo pagrindinio tono periodu. Visi P sintezuojami pagrindinio tono periodais gaunami vykdant parinkto gerklų signalo apdorojimą, „valdomą“ remiantis melizmos pagrindinio tono ir intensyvumo įverčiais.

4.3 pavyzdys (Melizmos sintezės užrašymas)

Tarkime, kad sintezuosime melizmą dvejais langais ($N = 2$), kiekviename lange naudosime tris pagrindinio tono periodus ($P = 3$). Tada šis sintezuojamas gerklų signalas užrašomas:

$$\widehat{g}_m(k) = \sum_{n=1}^2 \sum_{p=\kappa_n}^{3n} \frac{\widehat{I}(k)}{\widehat{I}_d(k)} g_{d\kappa_n} \left(\left\lfloor \frac{\widehat{T}_0(p)}{\widehat{T}_{0d}(\kappa_n)} \cdot k \right\rfloor - \sum_{h=\kappa_n}^{3n-1} \widehat{T}_0(h) - \Delta_n \right) = \quad (4.6)$$

$$= \sum_{p=\kappa_1}^3 \frac{\widehat{I}(k)}{\widehat{I}_d(k)} g_{d\kappa_1} \left(\left\lfloor \frac{\widehat{T}_0(p)}{\widehat{T}_{0d}(\kappa_1)} \cdot k \right\rfloor - \sum_{h=\kappa_1}^2 \widehat{T}_0(h) - \Delta_1 \right) + \quad (4.7)$$

$$+ \sum_{p=\kappa_2}^6 \frac{\widehat{I}(k)}{\widehat{I}_d(k)} g_{d\kappa_2} \left(\left\lfloor \frac{\widehat{T}_0(p)}{\widehat{T}_{0d}(\kappa_2)} \cdot k \right\rfloor - \sum_{h=\kappa_2}^5 \widehat{T}_0(h) - \Delta_2 \right).$$

Įvertinus (4.5) ir (4.3) išraiškas, gauname:

$$\kappa_1 = 1; \quad \kappa_2 = 4; \quad \Delta_1 = 0; \quad \Delta_2 = \sum_{h=1}^3 \widehat{T}_0(h). \quad (4.8)$$

Įstatę šias reikšmes į (4.6) gauname šią galutinę išraišką:

$$\widehat{g}_m(k) = \sum_{p=1}^3 \frac{\widehat{I}(k)}{\widehat{I}_d(k)} g_{d1} \left(\left\lfloor \frac{\widehat{T}_0(p)}{\widehat{T}_{0d}(1)} \cdot k \right\rfloor - \sum_{h=1}^2 \widehat{T}_0(h) \right) + \quad (4.9)$$

$$\begin{aligned}
& + \sum_{p=4}^6 \frac{\widehat{I}(k)}{\widehat{I}_d(k)} g_{d4} \left(\left\lfloor \frac{\widehat{T}_0(p)}{\widehat{T}_{0d}(4)} \cdot k \right\rfloor - \sum_{h=4}^5 \widehat{T}_0(h) - \sum_{h=1}^3 \widehat{T}_0(h) \right) = \\
& = \sum_{p=1}^3 \frac{\widehat{I}(k)}{\widehat{I}_d(k)} g_{d1} \left(\left\lfloor \frac{\widehat{T}_0(p)}{\widehat{T}_{0d}(1)} \cdot k \right\rfloor - \sum_{h=1}^2 \widehat{T}_0(h) \right) + \\
& + \sum_{p=4}^6 \frac{\widehat{I}(k)}{\widehat{I}_d(k)} g_{d4} \left(\left\lfloor \frac{\widehat{T}_0(p)}{\widehat{T}_{0d}(4)} \cdot k \right\rfloor - \sum_{h=1}^5 \widehat{T}_0(h) \right).
\end{aligned} \tag{4.10}$$

4.1.2. Pagrindinio tono ir intensyvumo modeliavimas

Melizmos pagrindinį toną modeliuosime taip:

$$\widehat{T}_0(k) = \bar{T}_0 + \Phi_T(\theta_T, k), \tag{4.11}$$

čia $\bar{T}_0$ – dainos pagrindinio tono įvertis, θ_T – pagrindinio tono modelio parametrai, k – laikas.

Panašiai modeliuosime melizmos intensyvumą:

$$\widehat{I}(k) = \bar{I} + \Phi_I(\theta_I, k), \tag{4.12}$$

čia $\bar{I}$ – dainos intensyvumo įvertis, θ_I – intensyvumo modelio parametrai, k – laikas.

Nežinomas, Φ_T pagrindinio tono ir Φ_I intensyvumo funkcijas modeliuosime parinktais dirbtinių neuronų tinklais. Įvertinant minėtų melizmos charakteristikų sudėtingą pobūdį taikysime dviejų sluoksnių perceptroną [26], išreiškiamą:

$$\Phi(\theta, k) = \sum_{i=1}^{N_w^{(1)}} \theta_i^{(2)} f(\theta_i^{(1)} k), \tag{4.13}$$

čia $\theta = [\theta^{(1)}, \theta^{(2)}]$. Žinant, kad kai kurių melizmos rūšių charakteristikos nestipriai kinta, taikysime ir paprastesnį adalinos [26] modelį užrašomą:

$$\Phi(\theta, k) = \theta^{(2)} f(\theta^{(1)} k). \tag{4.14}$$

4.1.3. Apibendrintas melizmos modelis

4.1 apibrėžimas (Apibendrintas melizmos modelis). Sintezuotos melizmos gerklų signalas, taikant N sintezės langus, P pagrindinio tono periodus sintezės lange, žinant $g_{d\kappa_n}(k)$ dainos gerklų signalo pagrindinio tono periodą, $\widehat{T}_{0d}(\kappa_n)$ dainos ir $\widehat{T}_0(p)$ melizmos pagrindinio tono periodo trukmes bei modeliuojant $\widehat{T}_0(k)$ pagrindinį toną ir $\widehat{I}(k)$ intensyvumą,

išreiškiamas:

$$\widehat{g}_m(k) = \sum_{n=1}^N \sum_{p=\kappa_n}^{n \cdot P} \frac{\widehat{I}(k)}{\widehat{I}_d(k)} g_{d\kappa_n} \left(\left\lfloor \frac{\widehat{T}_0(p)}{\widehat{T}_{0d}(\kappa_n)} \cdot k \right\rfloor - \sum_{h=0}^{n \cdot P - 1} \widehat{T}_0(h) \right); \quad (4.15a)$$

$$\kappa_n = (n - 1)P + 1; \quad (4.15b)$$

$$\widehat{T}_0(k) = \Phi_T(\bar{T}_0, \theta_T, k); \quad (4.15c)$$

$$\widehat{I}(k) = \Phi_I(\bar{I}, \theta_I, k). \quad (4.15d)$$

Baigiamosios pastabos

- ▼ Melizmos modelio pagrindu parinktas gerklų signalas, o pastarojo aprašymui siūloma taikyti pagrindinio tono ir intensyvumo melizmos charakteristikų įvėčius bei dainos gerklų signalo fragmentą.
- ▼ Sukurtas apibendrintas melizmos modelis, kurio įgyvendinimui pasiūlyti ir aprašyti adalinos bei daugiasluoksnio perceptrono dirbtinių neuronų tinklai.
- ▼ Apibendrintame melizmos modelyje sudaryta galimybė parinkti vieną iš dviejų dirbtinių neuronų tinklų melizmai modeliuoti ateityje leis sukurti lanksčius bei mažo dydžio modelius konkrečiai melizmos rūšiai.

4.2. Melizmos modelio identifikavimo būdo parinkimas

Melizmos modeliui identifikuoti ar parinktiems dirbtinių neuronų tinklams mokytį gali būti taikomi įvairūs metodai, plačiau aptarti 2.3.2 skyrelyje. Iš dalies galimų mokymo metodų ratą susiaurina pasirinktos DNT struktūros. Kita vertus išlieka natūralus noras parinkti našų bei greitą mokymo metodą. Taigi toliau detalai aptarsime mūsų požiūriu du perspektyviausius kuriamų melizmos modelių identifikavimo metodus.

4.2.1. Mažiausios vidutinės kvadratinės klaidos minimizavimo metodas

Mažiausių kvadratų klaidos minimizavimas yra dažniausiai taikomas tiesinių uždavinių sprendimo metodas. Imdami $s(k)$ duomenų atskaitas laiko momentu $k = 1, \dots, N_D$ su duomenimis taip pat įvardindami triukšmą v gauname uždavinį, kuriame reikia paskaičiuoti modelio išėjimą $\hat{s}(k)$, kuris geriausiai – vidutinių kvadratų prasme – aproksimuoja sistemos tikrąjį išėjimą $s(k)$. Turime $x_1, \dots, x_{N_m}$ regresorius ir $\theta_1, \dots, \theta_{N_m}$ optimizuojamus parametrus. Modelio išėjimas kinta tiesiniu dėsnio ir yra priklausomas nuo N_D parametrų θ :

$$\hat{s} = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_{N_m} x_{N_m} = \sum_{i=1}^{N_m} \theta_i x_i, \quad (4.16)$$

Užrašius regresorių $x_1, \dots, x_{N_m}$ vektoriniame pavidale bei apjungus į bendrą regresijos matricą $\mathbf{X}$ gauname:

$$\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_m}. \quad (4.17)$$

Tuomet modelio išėjimas vektoriniame pavidale užrašomas

$$\mathbf{s}(\hat{k}) = \boldsymbol{\theta} \mathbf{X}. \quad (4.18)$$

Mažiausios vidutinės kvadratinės klaidos atveju klaidos funkcija užrašoma

$$I(\boldsymbol{\theta}) = \frac{1}{2} \mathbf{e}^T \mathbf{e}, \quad (4.19)$$

čia modeliavimo klaida $\mathbf{e}$ randama lyginant tikrąjį išėjimą $\mathbf{s}$ su modeliavimo išėjimu $\hat{\mathbf{s}}$:

$$\mathbf{e} = \mathbf{s} - \hat{\mathbf{s}} = \mathbf{s} - \boldsymbol{\theta} \mathbf{X} \quad (4.20)$$

Ištačius (4.20) į (4.19), gauname apibendrintą klaidos funkciją:

$$I(\boldsymbol{\theta}) = \frac{1}{2} \boldsymbol{\theta}^T \mathbf{H} \boldsymbol{\theta} + \mathbf{h}^T \boldsymbol{\theta} + h_0, \quad (4.21)$$

čia $\mathbf{H}$ – hesiano funkcija – antros eilės klaidos funkcijos išvestinė:

$$\mathbf{H} = \mathbf{X}^T \mathbf{X}; \quad (4.22)$$

$\mathbf{h}$ – tiesinė klaidos funkcijos dedamoji:

$$\mathbf{h} = -\mathbf{X}^T \mathbf{s}; \quad (4.23)$$

h_0 – konstanta

$$h_0 = \mathbf{s}^T \mathbf{s}. \quad (4.24)$$

Pagal (4.20) ir (4.19) klaidos gradientas lygus nuliui. Mažiausios vidutinės kvadratinės klaidos minimizavime modelio parametrai randami pagal:

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{s}. \quad (4.25)$$

4.2.2. Levenberg-Marquardt minimizavimo metodas

Dirbtinių neuronų tinklams mokyti vieni iš plačiausiai naudojamų yra gradientų skaičiavimais pagrįsti mokymo algoritmai. Gradientą galima užrašyti per klaidos funkcijos $I(\hat{\boldsymbol{\theta}})$ ir parametrų vektorius $\hat{\boldsymbol{\theta}}$ pokytį: $\hat{\mathbf{r}} = \partial I(\hat{\boldsymbol{\theta}}) / \partial \hat{\boldsymbol{\theta}}$. Gradientų skaičiavimais paremto algoritmo kartinis tikslas yra keisti $\hat{\boldsymbol{\theta}}_{k-1}$ reikšmę proporcingai žingsniui η_{k-1} judant kryptimi $\hat{\mathbf{p}}_{k-1}$:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} - \eta_{k-1} \hat{\mathbf{p}}_{k-1}. \quad (4.26)$$

Skaičiavimo kryptis $\hat{\mathbf{p}}_{k-1}$ yra užrašoma naudojant gradientą $\hat{\mathbf{r}}_{k-1}$, kuris yra normuojamas bei sukamas įvertinant krypties matricą $\hat{\mathbf{R}}_{k-1}$, t. y.:

$$\hat{\mathbf{p}}_{k-1} = \hat{\mathbf{R}}_{k-1} \hat{\mathbf{r}}_{k-1}. \quad (4.27)$$

Optimizavimo uždavinio tikslas yra siekti, kad kiekviename iteraciniame žingsnyje klaida mažėtų: $I(\hat{\boldsymbol{\theta}}_k) < I(\hat{\boldsymbol{\theta}}_{k-1})$. Siekiant šio tikslo reikia teisingai parinkti klaidos mažėjimo krypties matricą $\hat{\mathbf{R}}_{k-1}$. Žinoma visa eilė metodų leidžiančių ją apskaičiuoti, tačiau plačiausiai taikomi šie optimizavimo metodai:

- tiesinės paieškos (intervalų mažinimo ir intervalų paieškos),
- baigtinių skirtumų (ang. Finite Difference),
- greičiausio nusileidimo,
- susietų gradientų,
- Niutono.

Geriausi skaičiavimo rezultatai pasiekiami Niutono metodu – jis leidžia judėti tiesiai minimumo kryptimi. Krypties matrica $\hat{\mathbf{R}}_{k-1}$ randama paskaičiavus klaidos funkcijos $\hat{\boldsymbol{\theta}}_{k-1}$ hesiano $\hat{\mathbf{H}}_{k-1}$ inversiją:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} - \eta_{k-1} \hat{\mathbf{H}}_{k-1}^{-1} \hat{\mathbf{r}}_{k-1}. \quad (4.28)$$

Deja Niutono metodas yra kartu ir pats sudėtingiausias. Jis pasižymi eile ypatybių į kurias reikia atkreipti dėmesį norint šį metodą naudoti:

- reikalingos antros eilės išvestinės,
- matricų inversija skaičiavimų sudėtingumą padidina kubiniu dėsniu,
- atminties poreikis didėja kvadratinio dėsniu, nes reikia saugoti hesiano funkcijos reikšmes,
- tiesinio uždavinio atveju vienintele iteracija pasiekiamas minimumas,
- metodo neįtakoja tiesinės parametrų transformacijos,
- tinkamas mažiems uždaviniams spręsti (dešimties parametrų eilės).

Niutono metodas realiuose optimizavimo uždaviniuose taikomas retai, nes jis yra per sudėtingas didelės apimties uždaviniams. Tačiau šis optimizavimo metodas buvo panaudotas visai eilei išvestinių metodų sukurti. Dauguma taikoma netiesinių uždavinių sprendimui. Bendru atveju gradientiniai metodai neatsižvelgia į klaidos funkcijos ypatybes.

Tiesiniam optimizavimo uždaviniui turime kvadratinę klaidos funkciją kuri skaičiuojama pakeliant kvadratu klaidos reikšmę e ir pritaikant svorių koeficientą q , t. y.:

$$I(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^N q_i e^2(i, \hat{\boldsymbol{\theta}}). \quad (4.29)$$

Netiesinio optimizavimo uždaviniui šią išraišką galime perrašyti:

$$I(\hat{\theta}) = \sum_{i=1}^N \xi^2(i, \hat{\theta}). \quad (4.30)$$

Šią išraišką perrašę vektoriniame pavidale, turėsime:

$$I(\hat{\theta}) = \hat{\xi}^T \hat{\xi}. \quad (4.31)$$

Gradientas j -tajai atskaitai bus:

$$r_j = 2 \frac{\partial I(\hat{\theta})}{\partial \theta_j} = 2 \sum_{i=1}^N \xi_i \frac{\partial \xi_i}{\partial \theta_j}. \quad (4.32)$$

Užrašę jakobianą matriciniame pavidale

$$\hat{J} = \begin{bmatrix} \partial \xi_1 / \partial \theta_1 & \cdots & \partial \xi_1 / \partial \theta_n \\ \vdots & & \vdots \\ \partial \xi_N / \partial \theta_1 & \cdots & \partial \xi_N / \partial \theta_n \end{bmatrix}, \quad (4.33)$$

gauname gradientą, išreikštą per jakobiano matricią:

$$\hat{r} = 2 \hat{J}^T \hat{\xi}. \quad (4.34)$$

Netiesiniuose optimizavimo uždaviniuose Niutono metodas irgi gali būti taikomas, tačiau, sudėtingėjant skaičiuojamam uždaviniui, hesiano skaičiavimas tampa pernelyg sudėtingas. Kadangi gradientą galime išreikšti per (4.34) jakobianą, tuomet hesiano funkcija gali būti aproksimuojama – $\hat{H} \approx \hat{J}^T \hat{J}$. Netiesinio optimizavimo metodas su aproksimuota hesiano funkcija vadinamas Gauso-Niutono metodu. Tuomet (4.28) išraiška tampa:

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \eta_{k-1} \left(\hat{J}_{k-1}^T \hat{J}_{k-1} \right)^{-1} \hat{J}_{k-1}^T \hat{\xi}_{k-1}. \quad (4.35)$$

Šio metodo pranašumas yra tame, kad skaičiuojant jakobianą nereikia antros eilės išvestinės, kuri dažnai nėra žinoma, o sudėtingesnei klaidos funkcijai ją paskaičiuoti yra sunku. Matricos inversija $(\hat{J}_{k-1}^T \hat{J}_{k-1})^{-1}$ dažniausiai neskaiciuojama tiesiogiai, bet yra sudaroma n -tos eilės tiesinė lygčių sistema krypties vektoriui $\hat{p}_{k-1}$ surasti:

$$(\hat{J}_{k-1}^T \hat{J}_{k-1}) \hat{p}_{k-1} = \hat{J}_{k-1}^T \hat{\xi}_{k-1}. \quad (4.36)$$

Yra ir kitų metodų šiai lygčių sistemai išspręsti, tačiau Gauso-Niutono metodas tampa nepatikimas tuomet, kai matrica $\hat{J}_{k-1}^T \hat{J}_{k-1}$ yra silpnai išreikšta arba netgi artėja prie vietinės matricos. Neigiamus padarinius galima sumažinti taikant krašto regresijos (ang. Ridge Regression) teoriją. Išraišką (4.35) papildžius reguliarizacijos dėmenimis, gauname

Levenberg-Marquardt netiesinio optimizavimo metodą:

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \eta_{k-1} \left(\hat{\mathbf{J}}_{k-1}^T \hat{\mathbf{J}}_{k-1} + \alpha_{k-1} \hat{\mathbf{I}} \right)^{-1} \hat{\mathbf{J}}_{k-1}^T \hat{\xi}_{k-1}, \quad (4.37)$$

kuriame, kaip Gauso-Niutono metode (4.35), pritaikome netiesioginį inversijos skaičiavimą krypties vektoriui rasti:

$$\left(\hat{\mathbf{J}}_{k-1}^T \hat{\mathbf{J}}_{k-1} + \alpha_{k-1} \hat{\mathbf{I}} \right) \hat{\mathbf{p}}_{k-1} = \hat{\mathbf{J}}_{k-1}^T \hat{\xi}_{k-1}. \quad (4.38)$$

Hesiano aproksimacijos $\hat{\mathbf{J}}_{k-1}^T \hat{\mathbf{J}}_{k-1}$ papildymas $\alpha_{k-1} \hat{\mathbf{I}}$ nariu yra ne kas kita kaip reguliarizacija, kuri yra taikoma tiesiniams optimizavimo uždaviniais spręsti mažiausių kvadratų metode.

Tuomet, kai žingsnis α_{k-1} yra mažas, tai Levenberg-Marquardt algoritmas artėja prie *Gauso-Niutono* metodo. Esant dideliui žingsniui – artėjama prie *greičiausio nusileidimo* metodo, kuris užtikrina judėjimą link minimumo, tačiau neužtikrina tikslaus minimumo radimo, kuris yra būdingas Gauso-Niutono metodui. Paprastai naudojant Levenberg-Marquardt optimizavimo metodą iš pradžių yra nustatoma pradinė α_{k-1} žingsnio teigiamą reikšmė. Toliau kiekvienos iteracijos metu žingsnis yra mažinamas tam tikru užsibrėžtu žingsniu. Taip artėjama prie optimalios reikšmės, kuriai esant pradeda veikti Gauso-Niutono metodas. Tačiau jeigu α_{k-1} mažinimas iššaukia klaidos padidėjimą (optimizavimas pajudėjo neteisinga linkme), tuomet žingsnis vėl yra didinamas, kol pasiekiamas klaidos mažėjimas.

Praktiniai aspektai

Levenberg-Marquardt metodas yra plačiai paplitęs DNT mokyme, nes garantuoja klaidos mažėjimą su kiekviena iteracija. Žemiau 4.4 pavyzdyje pateikiamas Matlab kodo fragmentas iliustruojantis Levenberg-Marquardt optimizavimo metodo panaudojimą.

4.4 pavyzdys (Levenberg-Marquardt metodo taikymas)

```

1 >% Duomenys
2 >p = [-1 -1 2 2;...
3     0 5 0 5];
4 >t = [-1 -1 1 1];
5
6 >% Tinklas
7 >net = newff(minmax(p), [3, 1], {'tansig', 'purelin'}, 'trainlm');
8 >net.trainParam.show = 5;
9 >net.trainParam.epochs = 300;
10 >net.trainParam.goal = 1e-5;
11
12 >% Mokymas
13 >[net, tr] = train(net, p, t);
14 TRAINLM, Epoch 0/300, MSE 2.7808/1e-05, Grad. 7.77931/1e-10
15 TRAINLM, Epoch 4/300, MSE 3.67935e-08/1e-05, Grad. 0.808272/1e-13
16 TRAINLM, Performance goal met.
17

```

```

18 >% Atsakas
19 >a = sim(net, p)
20 ans
21      -1.0000 -1.0000 1.0000 0.9996

```

Pateiktame pavyzdyje naudojamas Levenberg-Marquardt mokymas su standartinėmis parametų reikšmėmis. 4.1 algoritme pateiktas detalus Matlab įgyvendinimo išaiškinimas.

4.1 algoritmas (Levenberg-Marquardt metodo Matlab įgyvendinimas)

Sintaksė

```
[net,TR] = trainlm(net, Pd, Tl, Ai, Q, TS, VV, TV)
```

Įėjimo parametrai

- net – Neuronų tinklas
- Pd – Įėjimų vektorius
- Tl – Tikslų sluoksnio vektorius
- Ai – Įėjimo vėlinimo sąlygos
- Q – Duomenų masyvo dydis
- TS – Laiko žingsniai
- VV – Tuščia matrica arba validavimo masyvas
- TV – Tuščia matrica arba validavimo masyvas

Išėjimo parametrai

- net – Apmokytas neuronų tinklas
- TR – Parametų vektorius su masyvu parametų kiekvienai mokymosi epochai:
 - TR.epoch – Epochos numeris
 - TR.perf – Mokymosi klaidos vertė
 - TR.vperf – Validavimo klaidos vertė
 - TR.tperf – Testavimo klaidos vertė
 - TR.mu – Adaptyvaus parametro α_{k-1} reikšmė

Baigiamosios pastabos

- ▼ Turint konkrečias DNT struktūras bei siekiant našumo ir greitumo, melizmos modeliui identifikuoti, parinkti:
 - ▼ mažiausios vidutinės kvadratinės klaidos minimizavimo metodas – skirtas madalinos tinklui mokyti,
 - ▼ Levenberg-Marquardt metodas – skirtas daugiasluoksniui perceptronui mokyti.
- ▼ mažiausios vidutinės kvadratinės klaidos minimizavimo metodas leis rasti geriau-

šias madalinos tinklo reikšmes, tačiau jų radimui pagreitinti reikės parinkti tinkamus mokymo parametrus,

- ▼ Levenberg-Marquardt metodas negarantuoja geriausio mokymo rezultato radimo, todėl siekiant rasti geresnį sprendinį, tą patį tinklą teks mokyti keletą kartų (skirtingai inicializuojant), ir remiantis gautais rezultatais bus išrinktas „geriausiai“ išmokytas tinklas.

4.3. Atskirų melizmos rūšių modelių sudarymas

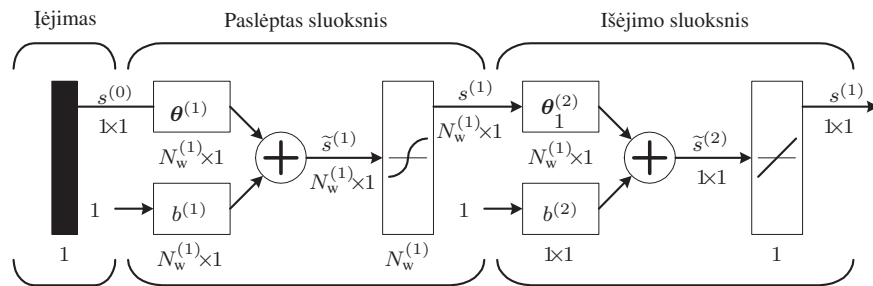
Kaip matome iš 3 skyriaus, pagrindinis tonas smarkiai kinta, ir tai yra esminis skirtumas tarp melizmos ir paprastos kalbos, nes kalboje pagrindinis tonas išlieka pastovus. Nuo pagrindinio tono dažnio priklauso kalbos skambesys – moterų ir vaikų jis yra aukštesnis, o vyrų – žemesnis. Paprastai jis varijuoja nuo 100 iki 500 Hz. Kitas melizmų parametras – intensyvumas – taip pat dažniausiai smarkiai kinta. Kadangi šios abi charakteristikos yra esminės, tai pabandydysime jas sintezuoti.

Tam tikslui naudosime dviejų sluoksnių daugiasluksnį perceptroną (4.2 pav.). DP svorių koeficientų matricos dydis yra $N_I \cdot N_H$, čia N_I – įėjimų skaičius, o N_H – neuronų skaičius DNT sluoksnyje. Kadangi dar taikomas slenksčio koeficientas b , tai bendras koeficientų skaičius bus:

$$N = N_H(N_I + 1). \quad (4.39)$$

Atskiru paslėptuoju neuronu galima aproksimuoti vieną kreivės krypties pasikeitimą. Tuomet, subraižę linijas šalia mums reikiamų aproksimuoti kreivių, gautume apytikrų neuronų paslėptame sluoksnyje skaičių. Melizmų atveju pažymėtina, kad dažninis nukrypimas žymiai labiau pastebimas už amplitudinį. Vadinasi melizmos parametrų sintezėje pirmumas turi būti teikiamas pagrindinio tono modeliavime.

Remiantis muzikologiniais melizmų apibrėžimais, pateiktais 2.1 skyrelyje, melizmas pagal sudėtingumą (pradedant nuo paprasčiausios) galime išdėstyti tokia tvarka: 1) foršlagas; 2) mordentas; 3) grupetas; 4) trelė. Mūsų turimuose testiniuose duomenyse buvo sudėtingas grupetas, tačiau pasitaikė paprasta trelė, todėl žemiau pateiktuose teoriniuose melizmų modeliuose grupetas yra sudėtingesnis už trelę.



4.2 pav. Parinkto daugiasluksnio perceptrono struktūra

4.3.1. Foršlago modelis

4.2 apibrėžimas (Foršlago modelis). Adalina modeliuojamas foršlago pagrindinis tonas:

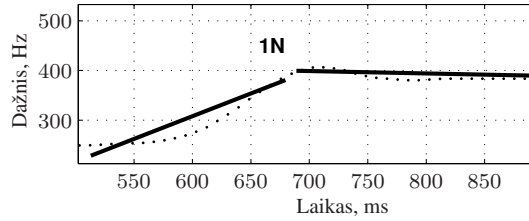
$$\widehat{T}_0^F(k) = \bar{T}_0^F + \theta_T^{(2)} f\left(\theta_T^{(1)} k\right), \quad (4.40a)$$

daugiasluoksniu perceptronu modeliuojamas intensyvumas:

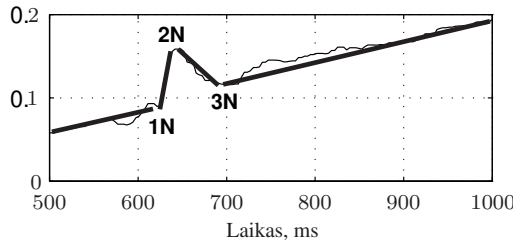
$$\widehat{I}^F(k) = \bar{I}^F + \sum_{j=1}^{N_I^F} \theta_{Ij}^{(2)} f\left(\theta_{Ij}^{(1)} k\right), \quad (4.40b)$$

čia $\bar{T}_0$ ir $\bar{I}$ – dainos pagrindinio tono ir intensyvumo įverčiai; $[\theta_1; \theta_2] = \theta_T$ – adalinos ryšių svoriai; $[\theta_{I1}^{(1)}, \dots, \theta_{IN}^{(1)}; \theta_{I1}^{(2)}, \dots, \theta_{IN}^{(2)}] = \theta_I$ – daugiasluoksniu perceptrono ryšių svoriai; $f(\cdot)$ – tangentinė neurono aktyvavimo funkcija; k – laiko indeksas intervale $[k_1, k_2]$; N_I^F – neuronų skaičius modelyje.

Foršlago charakteristikų modeliavimo iliustracijos pateiktos 4.3 paveiksle. Iš čia matyti, jog intensyvumo charakteristikos modeliavimui minimalus neuronų daugiasluoksniame perceptrone skaičius bus trys, o pagrindinio tono modeliavimui tinka adalina tinklas.



(a) pagrindinis tonas



(b) intensyvumas

4.3 pav. Foršlago charakteristikos ir jų modeliavimo minimalaus dydžio dirbtinių neuronų tinklais iliustracija (riebi linija – modelis)

4.3.2. Grupeto modelis

4.3 apibrėžimas (Grupeto modelis). Daugiasluoksniu perceptronu modeliuojamas grupeto pagrindinis tonas:

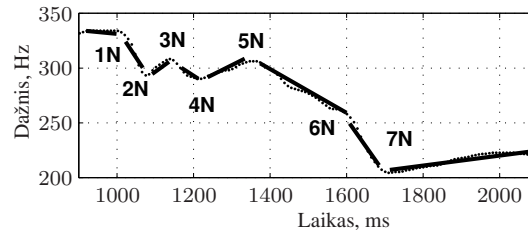
$$\widehat{T}_0^G(k) = \bar{T}_0^G + \sum_{j=1}^{N_T^G} \theta_{Tj}^{(2)} f\left(\theta_{Tj}^{(1)} k\right), \quad (4.41a)$$

ir intensyvumas:

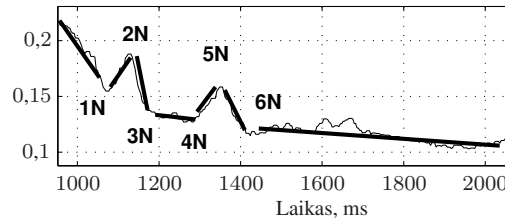
$$\widehat{I}^G(k) = \bar{I}^G + \sum_{j=1}^{N_I^G} \theta_{Ij}^{(2)} f\left(\theta_{Ij}^{(1)} k\right), \quad (4.41b)$$

čia $\bar{T}_0$ ir $\bar{I}$ – dainos pagrindinio tono ir intensyvumo įvertis; $[\theta_{T1}^{(1)}, \dots, \theta_{TN}^{(1)}; \theta_{T1}^{(2)}, \dots, \theta_{TN}^{(2)}] = \theta_T$ ir $[\theta_{I1}^{(1)}, \dots, \theta_{IN}^{(1)}; \theta_{I1}^{(2)}, \dots, \theta_{IN}^{(2)}] = \theta_I$ – daugiasluoksnių perceptronų ryšių svoriai; $f(\cdot)$ – tangentinė neurono aktyvavimo funkcija; n – laiko indeksas intervale $[n_1, n_2]$; N_T^G ir N_I^G – neuronų skaičius atitinkamose modelio dalyse.

Grupeto charakteristikų modeliavimo iliustracijoje pateiktose 4.4 paveiksle matome, kad charakteristikų kitimas yra pats sudėtingiausias iš visų keturių melizmų rūšių. Pagrindinio tono modeliavimui minimalus neuronų skaičius bus septyni, o intensyvumo modeliavimui minimalus neuronų skaičius bus šeši.



(a) pagrindinis tonas



(b) intensyvumas

4.4 pav. Grupeto charakteristikos ir jų modeliavimo minimalaus dydžio dirbtinių neuronų tinklais iliustracija (riebi linija – modelis)

4.3.3. Mordento modelis

4.4 apibrėžimas (Mordento modelis). Daugiasluoksniais perceptronais modeliuojamas mordento pagrindinis tonas:

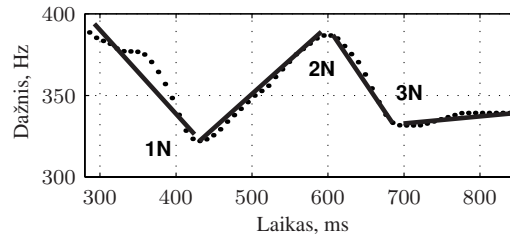
$$\widehat{T}_0^M(k) = \bar{T}_0^M + \sum_{j=1}^{N_T^M} \theta_{Tj}^{(2)} f\left(\theta_{Tj}^{(1)} k\right), \quad (4.42a)$$

ir intensyvumas:

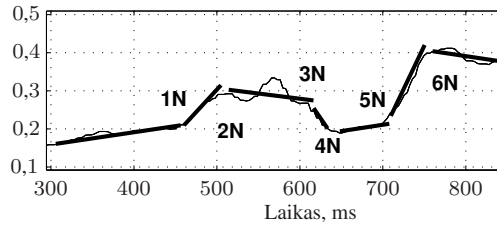
$$\widehat{I}^M(k) = \bar{I}^M + \sum_{j=1}^{N_I^M} \theta_{Ij}^{(2)} f\left(\theta_{Ij}^{(1)} k\right), \quad (4.42b)$$

čia $\bar{T}_0$ ir $\bar{I}$ – dainos pagrindinio tono ir intensyvumo įvertis; $[\theta_{T1}^{(1)}, \dots, \theta_{TN}^{(1)}; \theta_{T1}^{(2)}, \dots, \theta_{TN}^{(2)}] = \theta_T$ ir $[\theta_{I1}^{(1)}, \dots, \theta_{IN}^{(1)}; \theta_{I1}^{(2)}, \dots, \theta_{IN}^{(2)}] = \theta_I$ – daugiasluoksnių perceptronų ryšių svoriai; $f(\cdot)$ – tangentinė neurono aktyvavimo funkcija; n – laiko indeksas intervale $[n_1, n_2]$; N_T^M ir N_I^M – neuronų skaičius atitinkamose modelio dalyse.

Iš 4.5 pav. matome, kad mordentas yra žymiai sudėtingesnis už trelę. Pagal DNT teoriją pagrindinio tono modeliavimui reikėtų trijų neuronų, o intensyvumo modeliui – šešių. Verta pastebėti, kad intensyvumo kreivė gali smarkiai keistis ne tik skirtingų atlikėjų melizmomis, bet ir to paties atlikėjo melizmų signaluose.



(a) pagrindinis tonas



(b) intensyvumas

4.5 pav. Mordento charakteristikos ir jų modeliavimo minimalaus dydžio dirbtinių neuronų tinklais iliustracija (riebi linija – modelis)

4.3.4. Trelės modelis

4.5 apibrėžimas (Trelės modelis). Daugiasluoksniais perceptronais modeliuojamas trelės pagrindinis tonas:

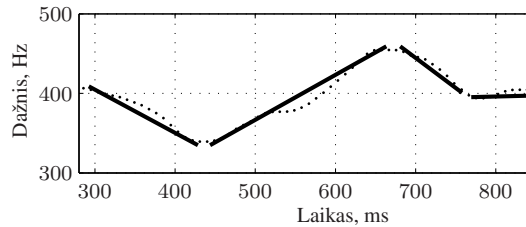
$$\widehat{T}_0^T(k) = \bar{T}_0^T + \sum_{j=1}^{N_1^T} \theta_{Tj}^{(2)} f\left(\theta_{Tj}^{(1)} k\right), \quad (4.43a)$$

ir intensyvumas:

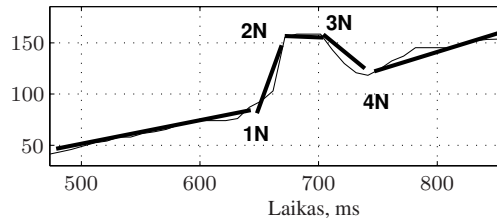
$$\widehat{I}^T(k) = \bar{I}^T + \sum_{j=1}^{N_I^T} \theta_{Ij}^{(2)} f\left(\theta_{Ij}^{(1)} k\right), \quad (4.43b)$$

čia $\bar{T}_0$ ir $\bar{I}$ – dainos pagrindinio tono ir intensyvumo įvertis; $[\theta_{T1}^{(1)}, \dots, \theta_{TN}^{(1)}; \theta_{T1}^{(2)}, \dots, \theta_{TN}^{(2)}] = \theta_T$ ir $[\theta_{I1}^{(1)}, \dots, \theta_{IN}^{(1)}; \theta_{I1}^{(2)}, \dots, \theta_{IN}^{(2)}] = \theta_I$ – daugiasluoksnių perceptronų ryšių svoriai; $f(\cdot)$ – tangentinė neurono aktyvavimo funkcija; n – laiko indeksas intervale $[n_1, n_2]$; N_T^T ir N_I^T – neuronų skaičius atitinkamose modelio dalyse.

Pagal 4.6 pav. matome, kad trelės pagrindinio tono modelis turėtų būti paprastas. Kita vertus teorinis trelės modelis gautas naudojant pačią paprasčiausią trelės variaciją. Dainose sutinkamos trelės gali būti žymiai ilgesnės, nes trelė neturi apibrėžto pagrindinio tono bangavimų skaičiaus – tuo ji ir skiriasi nuo mordento, kuriame tėra tik dvi natos.



(a) pagrindinis tonas



(b) intensyvumas

4.6 pav. Trelės charakteristikos ir jų modeliavimo minimalaus dydžio dirbtinių neuronų tinklais iliustracija (riebi linija – modelis)

4.1 lent. Melizmų modelių pirminis apibūdinimas

Melizma	Pagrindinio tono modelis			Intensyvumo modelis		
	Tinklas	$\min N$	$\min \Theta$	Tinklas	$\min N$	$\min \Theta$
Foršlagas	A	1	2	DP	4	11
Grupetas	DP	7	17	DP	5	13
Mordentas	DP	3	9	DP	6	15
Trelė	DP	3	9	DP	6	16

Baigiamosios pastabos

- ▼ Vienintelio foršlago modelis (dėl melizmos paprastumo) remiasi adalinos bei daugiasluoksnio perceptrono tinklais, tuo tarpu kitų melizmų – grupeto, mordento bei trelės – modeliai remiasi daugiasluoksniais perceptronais.
- ▼ Preliminariai nustatytas visų sukurtų melizmų modelių minimalus neuronų ir parametrų skaičius (4.1 lentelė), kuris rodo, jog sudėtingiausia melizmų rūšis yra grupetas.

4.4. Skyriaus apibendrinimas

- ▲ Užrašytas apibendrintas melizmos modelis paremtas dinaminio langų dėstymo principu.
- ▲ Melizmų sintezei bus naudojamas pagrindinio tono ir intensyvumo modeliavimas bei gerklų signalo sintezė.
- ▲ Atlikus melizmų preliminarą analizę sukūrėme visų melizmų rūšių pagrindinių charakteristikų – pagrindinio tono ir intensyvumo – matematinius modelius.
- ▲ Foršlagas turi paprasčiausią pagrindinio tono kitimą, todėl jam modeliuoti naudosi-me adalina tinklus.
- ▲ Foršlago intensyvumas, skirtingai nuo pagrindinio tono, yra sudėtingesnis, todėl jo modeliavimui bus naudojamas daugiasluoksnis perceptronas.
- ▲ Kitų melizmos rūšių: grupeto, mordento ir trelės charakteristikų modeliavimui bus naudojamas daugiasluoksnis perceptronas.
- ▲ Melizmų modeliavime bus naudojamas daugiasluoksnis perceptronas su vienu įėjimu ir vienu išėjimu, tinklas turės vieną paslėptų neuronų sluoksnį su tangentine aktyvavimo funkcija, o išėjimo neuronas turės tiesinę aktyvavimo funkciją.
- ▲ Adalina tinklui apmokyti bus naudojamas mažiausios vidutinės kvadratinės klaidos, o DP – Levenberg-Marquardt minimizavimo metodas.

Melizmos modelių taikymo sintezėje eksperimentinė patikra

Šiame skyriuje aptarsime atliktos eksperimentinės patikros rezultatus, siekiant ištirti sudarytą melizmos sintezės metodiką ir sukurtus atskirų melizmų modelius. Todėl sekančiame 5.1 skyrelyje pateiksime melizmų duomenų rinkinių sudarymo eigą. Vėliau, 5.2 skyrelyje aptarsime vieno atlikėjo melizmų sintezės eksperimentų eigą bei rezultatus. Pabaigoje, 5.3 skyrelyje – aptarsime nuo atlikėjo nepriklausomų melizmų sintezės eksperimentų eigą bei rezultatus.

5.1. Melizmų duomenų rinkinių sudarymas

Melizmų duomenų rinkinių sudarymo eigą pateiksime etapais: aptarsime melizmų įrašymui parinktą techninę įrangą; apibūdinsime melizmas įdainavusius atlikėjus, išorinius įrašymo faktorius ir triukšmus; pateiksime paruoštus eksperimentams melizmų rinkinius, kurie apima visas melizmų rūšis ir dažniausiai pasitaikančius melizmų tipus.

5.1.1. Melizmų rūšys

Melizmos yra skirtomos į treles, grupetus, foršlagus ir mordentus. Tačiau kiekviena iš šių rūšių galima skaidyti į tipus aprašančius papildomas melizmų muzikines ypatybes. Pavyzdžiui, foršlagas gali prasidėti žema nata, o baigtis aukšta, arba prasidėjęs nuo aukštos, baigtis žema nata. Foršlagas taip pat skirstomas pagal tai per kiek oktavų peršoka pagrindinio tono garsas. Kitoms melizmų rūšims taip pat galioja panašūs skirstymai. Pavyzdžiui, grupetas sudarytas iš grupės viena po kitos einančių natų, skirstomas į daug tipų, nes yra daug perėjimų tarp natų būdų. Sudarant melizmų duomenų rinkinius stengtasi surinkti

5.1 lent. Melizmų įrašymui naudoto kompiuterio charakteristikos

Komponento pavadinimas	Charakteristikos
Kompiuteris	Standartinis, PC
Procesorius	AMD athlon XP2500
Operatyvioji atmintis, MB	512
Operacinė sistema	Windows 2000
Garso plokštė	Creative Audigy
Garso plokštės signalo-triukšmo santykis, dB	-95,6
Garso plokštės ADCH netolygumai (40 Hz ... 15 kHz), dB	+0,01 ... -0,05
Garso plokštės dinaminis diapazonas, dB	94,4
Garso plokštės netiesiniai iškraipymai, %	0,0077
Garso plokštės kanalų atskyrimas, dB	-92,8
Įrašams naudotas mikrofonas	PHILIPS SBCMC8650
Mikrofono varža, Ω	600
Mikrofono įrašomų dažnių juosta, Hz, Ω	100 ... 10 000
Mikrofono jautrumas, dB	-55
Mikrofono netiesiniai iškraipymai, %	< 0,5

ne tik visų keturių melizmų rūšių signalus, bet ir surinkti skirtingų melizmų tipų signalus. Įrašytas melizmų grupes pavadinome šiais vardais: F01, F02, F06, G01, G02, M01, M02, T02, T04. Čia raidė žymi melizmos rūšį (F – foršlagas, G – grupetas, M – morden-tas, T – trelė), o skaičius rodo melizmos tipą. Melizmų signalų įrašymas buvo atliktas su standartinė kompiuterine įranga, kurios charakteristikos pateiktos 5.1 pav. Naudota kokybiška garso įvesties-išvesties plokštė pasižyminti nedideliais vidiniais triukšmais ir pusiau profesionalus mikrofonas.

5.1.2. Atlikėjų tipai

Melizmų įdainuoti pakviesti 4 atlikėjai. Du vyrai (žymimi V raide), ir dvi moterys (žymimos M raide). Visi melizmų atlikėjai turi muzikinį išsilavinimą, o du iš jų (vienas vyras ir viena moteris) yra vokalistai. Toks atlikėjų pasirinkimas leido aprėpti platesnį melodinį spektrą. Melizmos įdainuotos vokalistų yra geresnės muzikinės kokybės, nes nevokalistui įdainuoti sudėtingesnes melizmas (pavyzdžiui trelę), išlaikant reikiamą pagrindinio tono kitimą, yra labai sunku. Taigi rinkinyje įdainuota trelė tik profesionalių vokalistų.

Ne visos melizmos naudotos modeliavimo eksperimentuose – dalis jų atmestos. Teko atmesti nemažai nevokalistų melizmų, nes jų pagrindinio tono kitimas neatitiko meliz-moms reikiamo kitimo dėsnio arba kalbos signale buvo trūkių. Šis papildomas melizmos rinkinių „apdorojimas“ leido padidinti DNT mokymosi kokybę bei greitį.

5.1.3. Išoriniai faktoriai. Triukšmų tipai

Melizmų rinkinys surinktas įrašant dainas tylioje aplinkoje. Vokalistų melizmos įdai-nuotos studijoje, nevokalistų – namų sąlygomis. Melizmų rinkinio sudarymui nebuvo keliami ypatingi signalo „švarumo“ reikalavimai, nes visiškas triukšmo nebuvimas nėra

5.2 lent. Sukurto melizmų rinkinio apibūdinimas

Melizmų rūšis	Melizmų tipas	Melizmų sk. rinkinyje				Viso
		V1	V2	M1	M2	
Foršlagas	F01	35	7	20	8	186
	F02	40	-	-	6	
	F06	35	-	35	-	
Grupetas	G01	35	8	40	10	139
	G02	17	9	20	-	
Mordentas	M01	29	8	45	10	147
	M02	21	-	25	9	
Trelė	T02	25	-	25	-	90
	T04	40	-	-	-	
Viso		277	32	210	43	562

būtinai melizmos parametrų (pagrindinio tono, intensyvumo ir gerklų signalo) išskyrimui ir pačios melizmos sintezei. Atliekant eilę pirmųjų melizmos sintezės eksperimentų (3.1.1 skyrius) parodyta, kad be triukšmo fono sintezuota melizma dažnai skamba „mechanškai“ ir nenatūraliai.

Melizmų įrašymas atliekamas atskirais fragmentais – kiekvieno atlikėjo ir kiekvienos melizmos tipui buvo kuriamas viena arba kelios bylos. Tokiu būdu vienoje byloje saugoma keletas ar keliasdešimt to paties tipo melizmų, kurios vėliau buvo fragmentuojamos (iškerpamos ir saugomos atskirose bylose). Fragmentavimo metu atskiros melizmos prasidėdavo panašiu, bet ne identišku metu – melizmos gautos su faziniu poslinkiu. Kalbos signalui yra labai sunku priderinti atskirų kalbos fragmentų fazes, todėl toks to paties atlikėjo melizmų išsibastymas buvo ignoruojamas. Paderinant kiekvienos melizmos pradžios laiką, melizmų fazinį neatitikimą galima minimizuoti, tačiau tai nedaryta, kadangi naudojantis dirbtinių neuronų tinklais bus siekiama gauti modelį, kuris turėtų minimalią klaidą apdorojant laike išsibarsčiusias melizmas. Melizmos modelis turi perimti bendrą charakteristikų kitimą. Melizmos buvo autoriaus fragmentuojamos ne automatinėmis priemonėmis, todėl tos pačios rūšies ir tipo melizmos gavosi skirtingos trukmės. Siekiant supaprastinti melizmų apdorojimą, visų melizmų signalų trukmė buvo suvienodinta remiantis pačia trumpiausia konkreto atlikėjo melizma.

5.2 lentelėje pateikti apibendrinti duomenys apie surinktų melizmų kiekius. Matyti, kad viso surinktos 562 melizmos, iš jų daugiausia foršlagų, maždaug po lygiai grupetų ir mordentų, o trelių surinkta mažiausia. Didžiausi yra V1 ir M1 (profesionalių vokalistų) rinkiniai. Trelės muzikine prasme yra sudėtingiausios iš melizmų, todėl jas teisingai įdainuoti sugebėjo tik vokalistai.

Baigiamosios pastabos

- ▼ Melizmų rinkinių sudarymui pakviesti 4 atlikėjai (po lygiai moterų ir vyrų), iš kurių 2 profesionalūs atlikėjai, bei parengta melizmų įrašymui reikiama techninė įranga.
- ▼ Sukurta 562 melizmų signalų duomenų bazė apimanti visas 4 melizmų rūšis ir daž-

niausiai pasitaikančius 9 melizmų tipus.

- ▼ Atmestos netinkamos eksperimentams melizmos bei, suvienodinant tos pačios rūšies ir tipo melizmų trukmes, duomenų bazė parengta eksperimentams.

5.2. Vieno atlikėjo melizmų sintezė

Vieno atlikėjo melizmų sintezės eksperimente naudojamos to paties atlikėjo melizmų charakteristikos (pagrindinis tonas bei intensyvumas) ir atlikėjo be melizmos gauto gerklų signalo fragmentas. Melizmų sintezė atlikta dviem žingsniais: modeliuojant melizmos charakteristikas ir jas taikant gerklų signalo sintezei.

Vieno atlikėjo melizmų sintezės eksperimentai atlikti su visų melizmų rinkiniu. Eksperimentai atlikti su kiekviena iš V1, V2, M1, M2 duomenų imčių atskirai, kadangi šio tipo eksperimentuose būtina melizmos gerklų signalo išskyrimui, mokymui, mokymo patikrai, sintezei ir sintezės rezultatų patikrai naudoti tik to paties atlikėjo melizmas.

Nors eksperimentai atlikti su visų melizmų signalų rinkiniais (5.2 lent.), tačiau glausumo dėlei toliau rezultatai pateikiami tik vokalistų rinkiniams (V1 ir M1). Vokalistų rinkiniai pasirinkti todėl, kad tos pačios rūšies ir tipo melizmų yra daugiau. Didesnis melizmų skaičius leis tiksliau identifikuoti melizmas ir atskleisti jų ypatybes melizmų sintezėje.

5.2.1. Melizmų modelių identifikavimas taikant V1 ir M1 rinkinius

A. Eksperimento sąlygos ir eiga

Melizmų modeliavimo tikslas yra gauti charakteristikų – pagrindinio tono ir intensyvumo modelius naudojant dirbtinių neuronų tinklus. Eksperimento eigoje siekiama surasti patį mažiausią melizmos sintezei tinkamą DNT modelį.

Modeliavimui naudoti V1 ir M1 duomenų rinkiniai dalinami į duomenų imtis: 50 % rinkinio naudojama dirbtinių neuronų tinklui mokyti, 30 % rinkinio skirta DNT mokymo patikrai (validavimui) ir likę 20 % rinkinio skirta gauto melizmos modelio testavimui. Duomenys sunormuoti režiuose nuo –1 iki 1. Normavimas šiuose režiuose leidžia geriausiai išnaudoti DNT naudotos tangentinės aktyvavimo funkcijos diapazoną.

Melizmos modeliavimo pradinėms sąlygoms nustatyti naudoti 4.3 skyriuje sudaryti atskirų melizmų rūšių modeliai. Atsižvelgus į 4.1 lentelėje pateiktus teorinius melizmų charakteristikų – pagrindinio tono ir intensyvumo – modelius, foršlago pagrindinio turėtų būti modeliuojamas adalina tinklais, o intensyvumas modeliuojamas daugiasluoksniu perceptronu. Kitų melizmų rūšių – grupeto, mordento ir trelės – charakteristikų modeliavimui turės būti naudojamas daugiasluoksnis perceptronas.

Svarbu atkreipti dėmesį į tai, kad 4.1 lentelėje pateiktos charakteristikos yra teorinės, nes melizmų pagrindinio tono, o ypač intensyvumo charakteristikos gali smarkiai kisti. Modeliavimas atliekamas pradedant nuo teorinio modelio, tačiau, gavus prastą modelį, eksperimente buvo numatyta galimybė eksperimentą kartoti kaskart didinant modelio sudėtingumą, įvedant papildomus neuronus daugiasluoksniame perceptrono modelyje. Eksperimento metu buvo įvertinta galimybė, kad teorinis modelis galėjo būti gautas sudė-

tingesnis negu reikia melizmomis modeliuoti. Šiam tikslui buvo atliktas modeliavimas su paprastesniu modeliu negu teorinis.

Įvertinant skirtingą atskirų melizmų rūšių charakteristikų kitimą, modeliavimo eksperimentai buvo atliekami pradedant nuo paprasčiausio tinklo – adalina, vėliau taikomas daugiasluoksnių perceptronų modelis, pradedant nuo vieno neurono paslėptajame sluoksnyje bei didinant neuronų paslėptajame sluoksnyje skaičių iki 12. Su kiekvienu iš dirbtinių neuronų tinklu buvo atliekama po 10 eksperimentų. Tokiu būdu kiekvienas melizmos rūšies tipas modeliuotas 130 eksperimentų.

Eksperimentai atlikti naudojant Matlab paketo dirbtinių neuronų modulį. Mokymas vykdomas nuosekliai paduodant į dirbtinių neuronų tinklą mokymo imties melizmas. Adalina tinklo mokymui naudotas mažiausios vidutinės kvadratinės klaidos minimizavimo algoritmas. Daugiasluoksnių perceptronų tinklui mokyti taikytas Levenberg-Marquardt minimizavimo metodas. Darbo metu buvo sukurtos Matlab funkcijos (A priedas) leidusios eksperimentus atlikti automatiniai režime – nurodžius duomenų rinkinį bei charakteristiką (pagrindinį toną ar intensyvumą) atliekama 10 modeliavimo eksperimentų su adalina tinklu, 10 eksperimentų su DP turinčiu vieną paslėptą neuroną ir t.t. kartojant tol, kol DP turės 12 neuronų paslėptajame sluoksnyje. Vėliau buvo imama sekanti melizma ir eksperimentai kartojami. Modeliavimo kriterijumi parinkta pasvertoji vidutinė kvadratinė klaidą leidžia palyginti skirtingų melizmų modelių kokybę. Kiekvieno modeliavimo rezultatai lyginami su testavimo duomenų imtimi. Rezultatai vidurkinami ir išsaugomi kompiuterio rinkmenoje. Laikoma, kad modeliavimo klaida neturi viršyti 0,1. Tai sudaro 5 % nuo melizmos normavimo rėžių. Atrenkami tik patys geriausi iš 10 atliktų su tos pačios rūšies melizmos tipu eksperimentų rezultatai.

B. V1 rinkinio taikymo rezultatai

Atliekant modeliavimo eksperimentus su V1 rinkiniu buvo atmestas modelis realizuotas adalina – nei viename iš eksperimentų šis tinklas nesugebėjo reikiamose klaidos rėžiuose išmokyti melizmos charakteristikų. Tai sąlygojo kelios priežastys. Visų pirma, paprasčiausios melizmos – foršlago – pagrindinio tono kitimas V1 rinkinyje buvo sudėtingesnis už teorinį. Antra priežastis – melizmos modeliai sukurti atsižvelgiant į vienos melizmos charakteristikos kitimą, tačiau eksperimentuose naudojama visa melizmų imtis. Atskiros melizmos turi parametrų išsibirstymą tiek atskaitų ašyje, tiek ir dažnių (pagrindiniam tonui) bei amplitudžių (intensyvumui). Siekiant sukurti modelį pagal išsibarsčiusias charakteristikas reikia naudoti sudėtingesnius tinklus – daugiasluoksnius perceptronus. Analogiški rezultatai gauti visų duomenų rinkinių modeliavimų metu, todėl rezultatų lentelėse adalina tinklo rezultatai nerodomi.

V1 rinkinio modeliavimo daugiasluoksniu perceptronu rezultatai pateikti 5.3 lentelėje. Čia N žymi neuronų skaičių paslėptajame DP sluoksnyje. Lentelėje parodytas pats geriausias rezultatas iš 10 eksperimentų. Modeliavimo rezultatai priklauso nuo melizmos tipo. Pavyzdžiui, foršlago F02 pagrindinio tono modeliui užtenka vieno paslėptojo sluoksniu neurono, tačiau F06 reikia net 5 neuronų. Taip pat verta atkreipti dėmesį į didelį modelių skirtumą tarp trelių T02 ir T04 – trelės T04 pagrindinis tonas smarkiau kinta, be to pastarosios trelės didesnis pagrindinio tono charakteristikų išsibirstymas. Grupeto pagrindinio tono modeliavimui reikia modelio su dviem paslėptais neuronais, o mordento M01 – 3, M02 – 4 neuronų.

5.3 lent. VI atlikėjo pagrindinio tono modeliavimas

Melizma	VKK pagrindinio tono modeliui $\times 10^{-3}$											
	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$	$N = 10$	$N = 11$	$N = 12$
F01	64,7	76,3	37,2	44,0	42,4	55,4	52,1	66,3	69,1	69,9	69,8	67,8
F02	38,6	38,4	36,3	35,9	33,3	38,3	40,0	38,1	37,9	38,1	36,5	36,7
F06	89,2	47,7	49,8	62,1	35,6	38,6	46,0	71,6	134,4	104,1	62,9	64,0
G01	86,7	47,0	44,0	41,6	43,3	36,5	28,4	38,1	36,5	32,4	33,5	33,1
G02	73,7	26,9	24,9	25,1	24,7	25,0	25,0	25,0	24,9	24,8	24,9	25,0
M01	151,9	125,3	69,9	78,8	76,2	69,8	58,1	47,6	68,5	64,4	66,6	65,0
M02	195,9	172,0	78,2	65,9	71,8	66,6	63,3	62,9	58,6	55,2	61,0	56,2
T02	135,9	144,1	69,2	106,2	92,3	106,7	94,5	106,2	109,3	92,9	89,7	101,9
T04	132,7	132,1	119,1	115,0	115,6	124,6	128,9	125,7	148,8	138,4	134,7	146,2
Melizma	VKK intensyvumo tono modeliui $\times 10^{-3}$											
	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$	$N = 10$	$N = 11$	$N = 12$
F01	71,9	222,2	240,9	161,9	195,7	176,1	247,9	203,7	234,5	300,4	267,4	258,5
F02	178,7	148,5	147,1	127,4	126,0	126,9	133,1	131,5	128,0	131,2	132,4	130,1
F06	147,5	135,1	128,4	132,4	82,0	79,9	144,7	160,0	127,3	155,8	97,4	149,6
G01	155,6	131,5	133,4	132,4	133,9	127,6	134,4	124,5	133,6	133,6	134,3	132,1
G02	32,5	28,8	29,2	26,2	26,0	25,5	25,4	25,7	27,5	24,9	26,9	27,1
M01	246,5	180,3	212,0	212,7	115,8	193,0	203,0	149,1	175,7	222,2	219,8	224,9
M02	57,2	61,7	45,2	49,5	49,8	45,6	46,4	42,0	42,1	41,6	41,8	48,3
T02	133,2	113,3	96,1	94,2	103,6	76,8	86,0	90,0	112,3	101,1	107,5	101,9
T04	37,2	31,6	28,1	28,0	27,7	27,2	27,9	26,9	28,1	28,2	26,9	27,9

5.4 lent. M1 atlikėjo pagrindinio tono modeliavimas

Melizma	VKK pagrindinio tono modeliui $\times 10^{-3}$											
	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$	$N = 10$	$N = 11$	$N = 12$
F01	166,3	158,1	138,9	148,7	151,4	134,6	137,2	147,5	155,8	165,0	173,7	171,3
F06	195,4	34,6	195,5	45,8	33,4	38,8	20,7	33,1	28,9	30,1	18,0	25,6
G01	52,1	50,5	28,4	25,7	25,9	26,2	25,8	25,7	25,9	26,0	25,9	26,0
G02	146,9	131,4	70,4	51,9	50,7	55,7	50,8	50,0	50,8	50,1	51,4	54,1
M01	114,8	108,3	59,4	40,3	54,1	54,3	64,9	51,4	51,1	64,5	68,1	67,2
M02	184,0	163,1	38,5	38,3	36,8	37,4	37,3	35,6	36,8	36,9	36,2	37,1
T02	16,8	14,1	9,1	8,8	8,9	8,7	9,0	9,2	9,3	9,1	9,0	9,0
Melizma	VKK intensyvumo tono modeliui $\times 10^{-3}$											
	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$	$N = 10$	$N = 11$	$N = 12$
F01	120,8	88,7	104,8	105,7	103,6	104,6	98,0	106,3	106,4	107,3	107,3	106,5
F06	207,0	184,3	175,8	176,1	173,4	169,1	172,5	168,3	170,4	168,9	171,5	163,0
G01	46,1	43,3	39,6	34,0	33,4	32,2	32,1	32,0	32,3	32,2	33,4	32,0
G02	36,5	28,6	26,8	24,6	25,4	24,5	25,3	25,4	26,2	25,8	26,3	26,3
M01	66,9	54,2	65,4	63,6	64,6	64,6	62,5	62,0	61,0	58,3	63,1	61,4
M02	85,7	58,2	60,4	221,5	191,3	195,2	240,9	184,5	221,3	186,9	202,8	239,9
T02	89,5	61,5	38,7	39,7	66,0	74,6	69,9	71,9	50,4	72,2	60,3	78,9

Intensyvumo modeliavimo rezultatai ženkliai prastesni. Vienintelio foršlago F01 paprastas intensyvumo modelis – vienas paslėptasis neuronas. F02, G01 ir M01 melizmų intensyvumo modeliai prasti – jų vidutinės kvadratinė klaida viršija užduotą ribą 0, 1. Tokių rezultatų pagrindinė priežastis – vokalistui žymiai lengviau išlaikyti reikiamą melizmos pagrindinio tono kitimą skirtingose melizmos, negu intensyvumo kitimą.

Lentelėje parodyti geriausi modeliavimo rezultatai išskirti pariebintu šriftu. Svarbu atkreipti dėmesį į tai, kad ne visuomet buvo parenkamas rezultatas su mažiausia VKK reikšme – esant nežymiam modelių VKK reikšmių skirtumui parenkamas paprastesnis modelis.

C. M1 rinkinio taikymo rezultatai

M1 rinkinio (vokalistės moters) melizmos modeliavimo rezultatai pateikti 5.4 lentelėje. Pagrindinio tono modeliavimui dažniausiai tinka trys paslėptus neuronus turintys DP. Šio atlikėjo foršlago F06 pagrindinio tono modeliavimas pavyko žymiai geriau – mažesnio melizmos signalo išsibarstymas leido pasiekti reikiamas modeliavimo tikslumą su dviem paslėptais neuronais. Tačiau foršlago F01 pagrindinio tono modeliavime nepavyko pasiekti reikiamos VKK reikšmės. Kitų pagrindinio tono charakteristikų modeliavimas rezultatai mažai skiriasi nuo V1 atlikėjo.

Intensyvumo modeliavimas pavyko sėkmingai visoms melizmomis išskyrus foršlagą F06. Verta išskirti grupeto G01 intensyvumo modelį – modeliavimo rezultatų patikra atskleidė, kad sintezei reikia parinkti šešių neuronų modelį, nes mažesnių neuronų skaičių turintys modeliai prastai aproksimuoja intensyvumo kitimą.

D. Apibendrinimas

Atlikus sukurtų teorinių melizmų modelių eksperimentinę patikrą su V1 ir M1 rinkiniais matyti, kad:

1. adalina netinka melizmų modeliavimui, kadangi teoriniai modeliai nevertino fakto, kad melizmos modeliai bus kuriami pagal charakteristikų grupę;
2. daugiasluoksnis perceptronas tinka melizmų modeliavimui, tačiau esant dideliams parametru išsibarstymui modelio kokybė prastėja;
3. bendru atveju intensyvumo modelis prastesnis už pagrindinio tono, kadangi vokalistui žymiai lengviau išlaikyti reikiamą melizmos pagrindinio tono kitimą skirtingose melizmos, negu intensyvumo kitimą.

5.2.2. V1 ir M1 melizmų rinkinių sintezė

A. Eksperimento sąlygos ir eiga

Eksperimento tikslas yra sintezuoti melizmas. Sintezė atlikta pagal 3.3.3 skyrelyje sudarytą metodiką. Melizma sintezuojama naudojant pagrindinio tono ir intensyvumo modelius bei gerklų signalo vieną fragmentą. Melizmos sintezė atliekama su melizmų rinkinio testavimo imtimi, kuri sudaro 20 % viso rinkinio dydžio.

Gerklų signalo periodai išskiriami iš melizmos pradžios ir pabaigos. Laikoma, kad šiose vietose melizmos nėra. Gerklų signalo sintezė taikant atstatymą tiesiogine ir atbuline kryptimi leidžia pasiekti geresnį gerklų signalo sintezavimą.

5.5 lent. Parinkta melizmų VNR skalė

Balas	Aprašymas
5	Sintezuota melizma akustiškai nesiskiria nuo originalios
4	Vietomis silpnai girdimas skirtumas tarp origino ir sintezės
3	Vietomis aiškiai girdimas skirtumas tarp origino ir sintezės
2	Didesnėje dalyje aiškiai girdimas skirtumas tarp origino ir sintezės
1	Sintezuota ir originali melizma aiškiai skiriasi

Sintezė atlikta naudojant Matlab programą. Melizmos sintezuotos vieną kartą. Melizmos sintezės kokybiniam įvertinimui taikyta vidutinės nuomonės rezultato (VNR) metodika (angl. Mean Opinion Score). Ji naudojama IP telefonijoje bei teksto vertimo į kalbą kokybei įvertinti [29].

Kiekvienu konkrečiu atveju yra parenkami specifiniai kriterijai, pagal kuriuos vertinama kalbos kokybė. Paprastai naudojama 5 balų kokybės skalė, kur mažiausias balas reiškia blogą kokybę, o aukščiausias – puikią. Melizmų sintezės atveju reikėjo įsivesti kokybės kriterijus, kuriais galima būtų vertinti melizmų kokybę. Anksčiau minėtose sistemose paprastai 5 balai skiriami signalams, kuriuos galima neįtempiant klausos patogiai klausytis, o 1 balas skiriamas visiškai nesuprantamiems sintezės rezultatams įvertinti. Šiuo atveju, dirbama su geros kokybės sinteze, iš kurios tikimasi minimalaus sintezuotos ir originalios melizmos skirtumo. Todėl vidutinės nuomonės rezultatams rasti buvo pasirinkti žymiai aukštesni kriterijai (5.5 lent.).

Sintezės metu skaičiuotas signalo fragmento prieš melizmą vokalinio trakto modelis. Tam tikslui buvo naudotas tiesinės prognozės koeficientų skaičiavimo metodikos patobulinimas – dinaminis analės langų dėstymas. Tai leidžia paskaičiuoti ne tik vokalinio trakto modelį, bet taip pat rasti gerklų signalo periodus prieš melizmą ir po jos su vienodomis sąlygomis.

Gerklų signalo sintezė atliekama dėstant pagrindinio tono gerklų signalo periodus. Gerklų signalo perioduose keičiamas diskretizavimo dažnis atsižvelgiant į melizmos pagrindinio tono modelį. Sekančiame žingsnyje pritaikius pasvėrimo funkciją gaunamas jungtinis gerklų signalo periodas artimesnis gerklų signalui prieš melizmą (tuo atveju kai sintezuojama arčiau melizmos pradžios) arba po melizmos (kuomet sintezuojama dalis yra arčiau melizmos pabaigos). Įvertinus intensyvumo modelį gaunamas sintezuotas melizmos gerklų signalo vienas periodas. Veiksmai kartojami iki pasiekiamas pabaigos laiko indeksas. Tokiu būdu gautas sintezuotas melizmos gerklų signalas apdorojamas vokalinio trakto modeliu ir gaunama sintezuota melizma.

B. V1 melizmų rinkinio sintezės rezultatai

Sintezės rezultatus pagal VNR kriterijus vertino 10 klausytojų. V1 atlikėjo sintezuotos 9 tipų melizmos. Sintezuotos tikrai tos melizmos, kurios naudojamos testinėje duomenų sekoje. Kadangi testinei sekai buvo paimta po 20 procentų duomenų iš kiekvieno melizmų tipų, gavome skirtingą sintezuotų melizmų skaičių.

Kiekvienas klausytojas vertino melizmos kokybę tokiu būdu: perklausė originalią melizmą iš testinės sekos ir po vienos sekundės pauzės einančią sintezuotą melizmą. Kiekviena melizma buvo vertinama atskirai. To paties vartotojo tos pačios melizmos tipo VNR

5.6 lent. V1 ir M1 melizmų rinkinių sintezės rezultatai

Melizma	VNR įvertis atlikėjo V1 melizmų rinkinių sintezei												Rinkiniui 4,1±0,5
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	Tipui	Rušiai	
F01	4,1	4,3	4,1	4,4	4,1	4,0	4,2	4,5	4,2	4,3	4,2±0,3	4,1±0,4	
F02	4,3	4,5	4,1	4,4	4,5	4,3	4,4	4,2	4,1	4,5	4,3±0,2		
F06	3,7	4,0	3,8	3,7	3,9	3,8	3,7	3,9	4,1	4,0	3,9±0,2		
G01	4,1	4,4	4,3	4,6	4,3	4,5	4,4	4,5	4,6	4,2	4,4±0,3	4,3±0,3	
G02	4,1	4,4	4,3	4,0	4,3	4,3	4,1	4,2	4,4	4,1	4,2±0,2		
M01	4,0	4,1	3,9	3,9	4,2	4,0	4,0	4,1	4,2	4,1	4,1±0,2	4,1±0,3	
M02	3,9	4,2	4,3	3,9	4,3	4,4	4,3	4,2	4,4	4,0	4,2±0,3		3,9±0,4
T02	4,3	3,9	3,8	4,3	4,2	4,1	4,1	3,8	4,3	4,1	4,1±0,3		
T04	3,8	3,9	3,6	3,7	3,7	3,8	3,6	3,8	3,7	3,9	3,8±0,2		
Melizma	VNR įvertis atlikėjo V1 melizmų rinkinių sintezei												Rinkiniui 3,9±0,4
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	Tipui	Rušiai	
F01	4,1	4,0	3,8	3,9	3,8	4,0	3,7	4,0	3,8	3,7	3,9±0,2	3,9±0,2	
F06	3,9	4,1	4,0	3,7	3,9	4,0	3,9	3,8	3,7	3,9	3,9±0,2		
G01	4,1	3,9	4,3	4,2	4,1	4,0	4,2	4,2	4,1	4,2	4,1±0,2	4,1±0,2	
G02	4,0	4,1	3,9	4,1	4,3	4,2	4,0	4,2	4,1	3,9	4,1±0,2		3,8±0,4
M01	3,9	4,0	3,6	4,1	3,9	4,2	3,9	3,9	3,8	3,7	3,9±0,3		
M02	3,9	3,7	3,6	3,7	3,8	3,9	3,6	3,8	3,7	3,5	3,7±0,2		
T02	3,9	4,0	3,8	4,0	3,8	4,1	3,9	4,1	3,7	4,0	3,9±0,2	3,9±0,2	

įverčio rezultatai buvo vidurkinami. Sintezės rezultatai tiesiogiai priklausė nuo pasirinkto gerklų signalo fragmento ir nuo pagrindinio tono kitimo. Foršlago F01 ir F02 sintezės rezultatų vidurkis yra virš 4 balų. Tai paaiškinama tuo, kad ši melizma yra paprasčiausia ir jos modeliavimas taip pat yra paprasčiausiai įgyvendinamas. Tuo tarpu F06 foršlagui gavome VNR kriterijų tik 3 balus. Tai įvyko todėl, kad ši melizma yra sudėtinga ir netipinė foršlagui. Dėl savo sudėtingumo melizmos duomenų banke buvo gautas didelis signalo pagrindinio tono ir amplitudės išsibarstymas.

C. M1 melizmų rinkinio sintezės rezultatai

M1 melizmų rinkinio sintezę pagal VNR kriterijus vertino taip pat 10 klausytojų. M1 atlikėjui buvo sintezuojamos 7 tipų melizmos. Moters melizmos sintezė yra prastesnės kokybės, nes moteriško gerklų signalo kitimas yra žymiai didesnis. Nors santykinai pagrindinis tonas kinta tiek pat, tačiau absoliučia reikšme kitimas yra 2–2,5 karto didesnis. Todėl tos melizmos, kurių sintezė atlikėjui V1 buvo vertinama 4 balais, M1 atlikėjui buvo 3,7 balo. Kaip ir vyro melizmų sintezės atveju, buvo gauti dideli melizmos F06 išsibarstymai, kurie sąlygojo prastesnę melizmos sintezę.

Baigiamosios pastabos

- ▼ Melizmų išsibarstymai neigiamai veikia melizmos modelio radimo kokybę.
- ▼ Melizmų sintezės VNR įverčiui labai didelę įtaką turi gerai parinktas gerklų signalas.
- ▼ Amplitudės iškraipymai labai nežymiai įtakoja melizmos kokybę, lemiamą įtaką turi dažnio neatitikimai.

5.3. Nuo atlikėjo nepriklausomų melizmų sintezė

Nuo atlikėjo nepriklausomų melizmų sintezėje buvo naudojami visų atlikėjų bendri melizmų parametrai (pagrindinis tonas bei intensyvumas) ir konkrečios sintezuojamos melizmos gerklų signalo periodas prieš melizmą ir po jos. Melizmų sintezė atlikta dviem etapais: melizmos parametrai – pagrindinio tono ir intensyvumo – modeliavimu ir melizmos sinteze, naudojant melizmos parametrus ir gerklų signalo fragmentą.

Nuo atlikėjo nepriklausomų melizmų sintezė vykdoma tik su vokalistų melizmomis – V1 ir M1 rinkiniais. Šie du rinkiniai apjungti į vieną bendrą B1 rinkinį.

Ekspperimentai atlikti su trimis melizmų tipais: foršlagu F06, grupetu G02 ir trele T02.

5.3.1. F06, G02 ir T02 melizmų modelių identifikavimas taikant B1 rinkinį

A. Eksperimento sąlygos ir eiga

Eksperimento sąlygos ir eiga yra analogiški vieno atlikėjo melizmų identifikavimo eksperimentams. Skiriasi tik duomenų rinkinys – dirbtinių neuronų tinklai mokomi bendru

rinkinį kurį sudaro V1 ir M1 rinkiniai.

Vyro V1 ir moters M1 rinkiniuose pagrindinis tonas skiriasi, tačiau parametrų modeliavimą pavyko įgyvendinti, nes modelis mokomas ne absoliučiomis pagrindinio tono reikšmėmis.

B. F06 melizmų modelių identifikavimo rezultatai

Eksperimento metu sukurti du dirbtinių neuronų tinklai: vienas skirtas pagrindinio tono modeliavimui, o kitas intensyvumui.

5.7 paveiksle pateikti modeliavimo rezultatai yra dviprasmiški: pagrindinio tono modeliavimas pavyko sėkmingai – gautas DP modelis su 4 neuronais paslėptajame sluoksnyje, tačiau intensyvumo modeliavimas nepavyko. Gauta didelė VKK klaida, kuri nemažėja sudėtinginant modelį. Detalesnė analizė rodo, kad intensyvumo kitimas V1 rinkinyje yra priešingas M1 rinkinio intensyvumo kitimui. Modelis nesugeba apibendrinti ženkliai besiskiriančius duomenis.

C. G02 melizmos modelio identifikavimo rezultatai

Gautas geras pagrindinio tono ir intensyvumo modelis. Abu atlikėjai – tiek V1, tiek ir M1, šią melizmą įdainavo labai panašiai. Modeliai sudėtingesni už pavienių rinkinių V1 ir M1 modelius, nes gautas bendrojo rinkinio B1 parametrų išsibarstymas yra didesnis.

D. T02 melizmos modelio identifikavimo rezultatai

Trelei buvo gautas patenkinamas intensyvumo modelis, tačiau prastas pagrindinio tono modelis. Prastas pagrindinio tono modelis gautas todėl, kad melizma skirtingai įdainuota – V1 rinkyje melizma yra signalo pradžioje, o M1 rinkinyje – signalo pabaigoje.

5.3.2. B1 rinkinio F06, G02 ir T02 melizmų sintezė

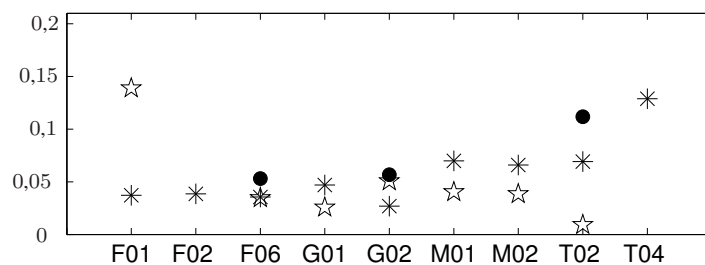
Sintezėje naudoti anksčiau sukurti melizmų parametrų modeliai ir sintezuojamos melizmos originalaus gerklių signalo periodai prieš ir po melizmos. Gerklių signalo periodas gautas naudojant modifikuotą tiesinę prognozės koeficientų metodą paremtą dinaminio analizės langų dėstymu. Melizmos sintezės kokybiniam įvertinimui taikyta vidutinės nuomonės rezultatų (VNR) metodika.

A. Eksperimento eiga ir rezultatai

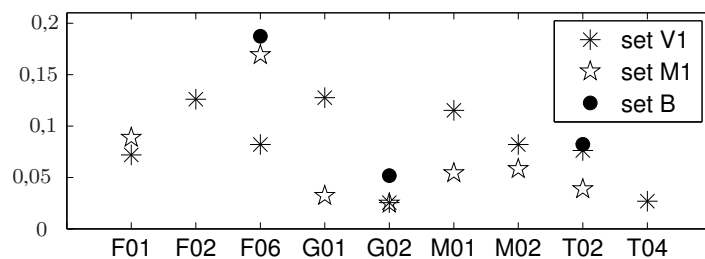
Rinkinio B1 sintezės rezultatai priklauso nuo melizmos rūšies. Geriausiai sintezuotas grupetas – VNR vidurkis 3, 95 balo. Šiuos rezultatus lemia geras pagrindinio tono modelis. Prastas intensyvumo modelis turi žymiai mažesnę įtaką sintezės kokybei.

Blogiausiai sintezuota trelė T04. Šiuos rezultatus lėmė rinkiniuose V1 ir M1 skirtingai įdainuotos melizmos. Norint kokybiškai atlikti nuo atlikėjo nepriklausomų melizmų sintezę, melizmos sudarančios bendrą rinkinį neturi ženkliai skirtis.

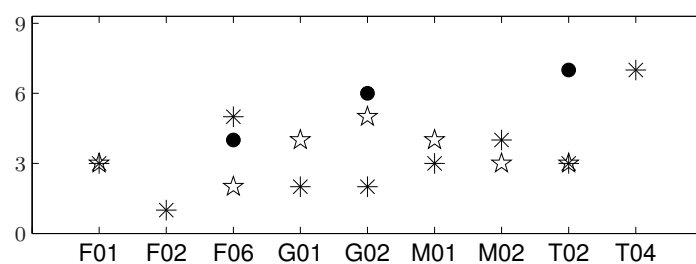
Gauti prasti rezultatai, tačiau kita vertus – nuo atlikėjo nepriklausomos sintezės VNR yra 3, 6 balo. Tai nėra priimtinas rezultatas melizmų restauravimui, nes aiškiai girdimas skirtumas tarp sintezuotos ir originalios melizmos, tačiau šią sintezės metodiką dar galima gerinti.



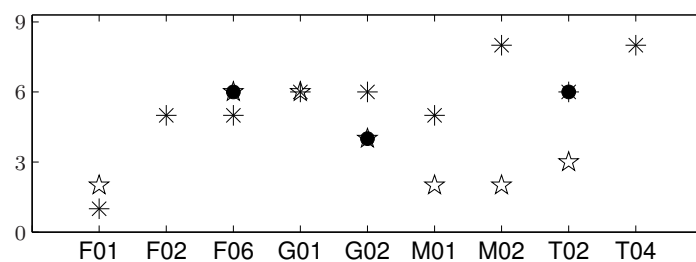
(a) pagrindinio tono VKK



(b) intensyvumo VVK

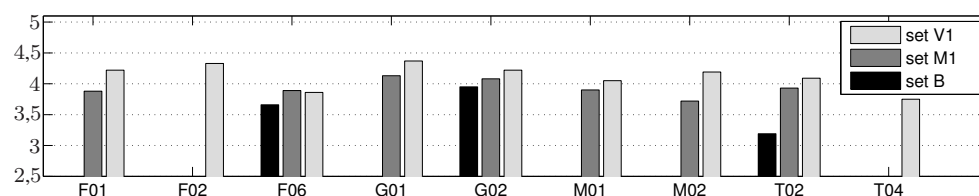


(c) pagrindinio tono DP paslėptųjų neuronų sk.



(d) intensyvumo DP paslėptųjų neuronų sk.

5.1 pav. Pagrindinio tono ir intensyvumo modeliavimo rezultatai



5.2 pav. Melizmos sintezės vidutiniai VNR įverčiai

Baigiamosios pastabos

- ▼ Pagrindinio tono modelis (5.1(a) pav.) geresnis už intensyvumo modelį (5.1(b) pav.) – tai sąlygoja faktas, kad vokalistui žymiai lengviau išlaikyti reikiamą melizmos pagrindinio tono kitimą skirtingose melizmos negu intensyvumo kitimą.
- ▼ Melizmų modeliavimui tinka daugiasluoksnių perceptronas, turintis nuo 1 iki 8 neuronų paslėptajame sluoksnyje.
- ▼ Modelio sudėtingumą lemia sintezuojamos melizmos tipas ir parametrų išsibarstymas – pagrindinio tono modeliavimui reikia paprastesnio DP modelio (5.1(c) pav.), o intensyvumui reikia sudėtingesnio modelio (5.1(d) pav.).
- ▼ Teoriniai adalina modeliai netinka melizmų modeliavimui, kadangi jie nevertino fakto, kad melizmos modeliai bus kuriami pagal charakteristikų grupę.
- ▼ Pagrindinio tono modelis turi didesnę įtaką melizmų sintezės VNR įverčiui negu intensyvumo modelis – amplitudės iškraipymai labai nežymiai įtakoja melizmos kokybę, lemiamą įtaką turi dažnio neatitikimai.
- ▼ Vyro V1 rinkinio vidutinis VNR yra net 4, 1 balo, moters M1 rinkiniui VNR mažesnis – 3, 9 balo, o bendro B1 rinkinio vidutinis VNR yra 3, 6 balo.
- ▼ V1 rinkinio dauguma atvejų VNR yra didesnis už M1 rinkinio rezultatus (5.2 pav.) – geresnius vyro sintezės rezultatus lėmė mažesnis pagrindinio tono kitimas. Santykinai pagrindinis tonas vyro ir moters melizmosse kinta tiek pat, tačiau absoliutiniu dydžiu moters pagrindinis tonas keičiasi apie 2 kartus daugiau negu vyro, o tai sąlygoja mažesnius sintezuojamo gerklų signalo iškraipymus.
- ▼ B1 rinkinio VNR yra mažesnis už V1 ir M1 rinkinių rezultatus – lemiamą įtaką rezultatų mažėjimui turi melizmų išsibarstymai.
- ▼ Siekiant kokybiškai sintezuoti nuo atlikėjo nepriklausomas melizmas, reikalingas duomenų rinkinys su mažu parametrų išsibarstymu.
- ▼ Melizmų išsibarstymai mažiau įtakoja vieno atlikėjo melizmų sintezės kokybę negu nuo atlikėjo nepriklausomą sintezę.
- ▼ Melizmų sintezės VNR įverčiui labai didelę įtaką turi gerai parinktas gerklų signalas – dvikryptis apdorojimas leidžia pasiekti kokybiškesnę melizmos sintezę.

5.4. Skyriaus apibendrinimas

- ▲ Eksperimentiškai patvirtintas sukurto melizmos sintezės metodikos, taikant naujus modelius, tinkamumas sintezuojant vieno atlikėjo melizmas.
- ▲ Parametrų modeliavimo kokybei lemiamą įtaką turi parametrų išsibarstymas – siekiant geresnio modeliavimo parametrų išsibarstymas turi būti kaip galima mažesnis.
- ▲ Melizmos sintezės kokybei didesnę įtaką turi pagrindinio tono neatitikimas negu intensyvumo.
- ▲ Eksperimentai parodė, kad nuo atlikėjo nepriklausomų melizmos modeliavimo rezultatai priklauso nuo rinkinio kokybės – geras rezultatas gaunamas tik tuomet, kai skirtingų atlikėjų melizmų išsibarstymas rinkinyje yra kaip galima mažesnis.

Galutinės išvados

Atlikus melizmų sintezės dirbtinių neuronų tinklais tyrimus ir išsprendus melizmų sintezės problemą gauti šie rezultatai:

1. Pasiūlyta tiesinės prognozės metodo modifikacija įvertinanti greitą pagrindinio tono kitimą ir tinkanti melizmų analizei.
2. Sukurta originali melizmų sintezės metodika leidžianti taikyti apibendrintą melizmos modelį bei paremta dvikrypčiu apdorojimu taikant dainos aproksimuotas pagrindinio tono ir intensyvumo charakteristikas, originalų gerklų signalo fragmentą ir tiesinės prognozės koeficientus.
3. Sukurtas lankstus apibendrintas melizmų modelis pagrįstas dviem dirbtinių neuronų tinklais – daugiasluoksniu perceptronu ir adalina bei dviem tinklų mokymo algoritmais – Levenberg-Marquardt ir mažiausios vidutinės kvadratinės klaidos minimizavimu.
4. Sukurti originalūs foršlago, grupeto, mordento ir trelės matematiniai modeliai tinkami melizmomis sintezuoti, pasiūlyti jų minimalūs dydžiai bei eksperimentų metu rastos modelių parametrų reikšmės: foršlago pagrindinio tono modeliavimui reikia adalina tinklo, o intensyvumo, bei visų kitų melizmų tiek intensyvumui tiek ir pagrindiniam tomui modeliuoti reikia daugiasluoksniu perceptrono.
5. Eksperimentiškai patvirtintas sukurto melizmų sintezės metodikos taikant naujus matematinius modelius tinkamumas vieno atlikėjo melizmų sintezei. Vertinant pagal vidutinės nuomonės rezultatų kriterijų vyro melizmų sintezės vidurkis – 4,1 balo, o moters – 3,9 balo, t.y. tik vietomis girdimas nesutapimas su originaliomis melizmomis

Tolimesni tyrimo darbai sietini su pasiūlytos melizmų metodikos tobulinimu siekiant pagerinti nepriklausomų nuo atlikėjų melizmų sintezės rezultatus.

Literatūra

- [1] Acernese, F.; Barone, F.; Eleuteri, A.; Garufi, F.; Milano, L.; Tagliaferri, R. A hierarchical Neural Network-based approach to VIRGO noise identification. // *International Joint Conference on Neural Networks*, Vol. 6, 2000, p. 53–62.
- [2] Akramizadeh, A.; Farjami, A. A.; Khaloozadeh, H. Nonlinear Hammerstein Model Identification Using Genetic Algorithm. // *Artificial Intelligence Systems, IEEE International Conference*, 2002, p. 67–73.
- [3] Anjou, A.; Torrealdea, F.; Sarasola, C. Control of Chaotic Behaviour with ANN. // *International Joint Conference on Neural Networks*, Vol. 6, 2000, p. 127–130.
- [4] Barnard, E.; Cole, R. A.; Wea, M. P.; Alleva, F. A. Pitch Detection with a Neural-Net Classifier. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 39, No. 2, 1991, p. 298–308.
- [5] Black, A. W.; Taylor, P. A. CHATR: A generic speech synthesis system. Technical report, COLING'94, Kyoto, Japan, 1994.
- [6] Botto, M. A.; da Costa, J. S. A comparison of nonlinear predictive control techniques using neural network models. // *Journal of Systems Architecture*, 1998, p. 597 – 616.
- [7] Bukovsky, I.; Redlapalli, S.; Gupta, M. Quadratic and Cubic Neural Units for Identification and Fast State Feedback Control of Unknown Non-Linear Dynamic Systems. // *International Symposium on Uncertainty Modeling and Analysis*, 2003.
- [8] Carotenuto, R.; Franchina, L.; Coli, M. A Nonlinear System Predictor from Experimental Data using Neural Networks. // *Southeastern Symposium on System Theory*, 1996.

- [9] Chanand, C.-F.; Law, K.-W. Thinned lattice filter for LPC analysis. // *Acoustics, Speech, and Signal Processing*, Vol. 1, 1992, p. 117 – 120.
- [10] Chul, Y. Y.; Jong, T. S.; Jin, W. K.; Hang, J. K. Korean folk song retrieval using rhythm pattern classification. // *Signal Processing and Its Applications*, Vol. 1, 1999, p. 123–126.
- [11] Collins. English Dictionary, 7 edition. Harpercollins Trade Sales Dept, 2005.
- [12] Constantin, N.; Dumitrache, I. Advanced Adaptive Control for Complex Nonlinear Processes. // *Mediterranean Conference on Control and Automation*, 1999, p. 399 – 407.
- [13] de Lamareand, R.; Alcaim, A. Effects of adaptive postfilters on the LSF quantisation for low bit rate speech coders in tandem connections. // *Acoustics, Speech, and Signal Processing, 2003*, Vol. 1, 2003, p. 393 – 396.
- [14] Deller, J. R.; Proakis, J. G.; Housen, J. H. L. Discrete-Time Processing of Speech Signals. New York: Macmillan, 1993.
- [15] Dubnowski, J. J.; Schafer, R. W.; Rabiner, L. R. Real-Time Digital Hardware Pitch Detector. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24, 1976, p. 2–8.
- [16] Filipovic, M.; Lipeika, A. Development of HMM/Neural Network-Based Medium-Vocabulary Isolated-Word Lithuanian Speech Recognition System. // *Informatica*, Vol. 15(4), 2004, p. 465–474.
- [17] Gerhard, D. Perceptual features for a fuzzy speech-song classification. // *Acoustics, Speech, and Signal Processing*, Vol. 4, 2002, p. 56–60.
- [18] Gillman, R. A. A Fast Frequency-Domain Pitch Algorithm. *Journal of Acoustics Society of America*, 58, 1975 S62(A), p.
- [19] Glad, T. Extensions of the RGA concept to nonlinear systems. Technical report, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Mar. 2000.
- [20] Gold, B. Computer Program for Pitch Extraction. *Journal of Acoustics Society of America*, 34, 1962, p. 916–921.
- [21] Gold, B.; Rabiner, L. R. Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain. *Journal of Acoustics Society of America*, 46, 1969, p. 442–448.
- [22] Goodman, S.; Liao, Y. Application of Neural Networks to Nonlinear Dynamical Systems. // *Southeastern Symposium on System Theory*, 1995.
- [23] Guo, S.; Shieh, L.; Lin, C.; Chandra, J. Adaptive Control for Nonlinear Stochastic Hybrid Systems with Input Saturation. // *Hawaii International Conference on System Sciences*, Vol. 2, 2001, p. 203–212.

- [24] Hai, J.; Joo, E. M. Improved Linear Predictive Coding Method for Speech Recognition. // *International Conference on Information, Communications and Signal Processing*, 2003.
- [25] Harris, C. M.; Weiss, M. R. Pitch Extraction by Computer Processing of High Resolution Fourier-analysis. *Journal of Acoustics Society of America*, 35, 1963, p. 339–343.
- [26] Haykin, S. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, Inc., Upper Saddle River, N.J., 2nd edition, 1999.
- [27] Herre, J.; Allamanche, E.; Ertel, C. How similar do songs sound? Towards modeling human perception of musical similarity. // *Applications of Signal Processing to Audio and Acoustics*, 2003, p. 83–86.
- [28] Šilingas, D.; Telksnys, L. Specifics of Hidden Markov Model Modifications for Large Vocabulary Continuous Speech Recognition. // *Informatika*, Vol. 15(1), 2004, p. 93–110.
- [29] Itoh, Y.; Tajima, K.; Kuwabara, N. Measurement of subjective communication quality for optical mobile communication systems by using mean opinion score. // *Personal, Indoor and Mobile Radio Communications*, Vol. 2, 2000, p. 1330–1334.
- [30] Johnson, W.; Narayan, S.; Whitney, R.; Dasand, R.; Bulut, M.; LaBore, C. Limited domain synthesis of expressive military speech for animated characters. // *Speech Synthesis*, 2002, p. 163 – 166.
- [31] Jungand, C.-J.; Hamand, M.-K.; Bae, M.-J. On a pitch alteration technique of speech using the asymmetry weighted window. // *Military Communications Conference Proceedings*, Vol. 2, 1999, p. 1439 – 1443.
- [32] Kambhampati, C.; Garces, F.; Warwick, K. Approximation of non-autonomous dynamic systems by continuous time recurrent neural networks. // *International Joint Conference on Neural Networks*, Vol. 1, 2000, p. 64 – 69.
- [33] Kasparaitis, P. Automatic stressing of the lithuanian text on the basis of a dictionary. // *Informatika*, Vol. 11(1), 2000, p. 19–40.
- [34] Kasparaitis, P. Automatic Stressing of the Lithuanian Nouns and Adjectives on the Basis of Rules. // *Informatika*, Vol. 12(2), 2001, p. 315–336.
- [35] Kasparaitis, P. Diphone Databases for Lithuanian Text-to-Speech Synthesis. // *Informatika*, Vol. 16(2), 2005, p. 193–202.
- [36] Kim, T.; Adali, T. Fully Complex Multi-Layer Perceptron Network for Nonlinear Signal Processing. *Journal on VLSI Signal Processing Systems*, 32, No. 1-2, 2002, p. 29–43. ISSN 0922-5773.
- [37] Krubsack, D. A.; Niederjohn, R. J. Comparison of Pitch Tracking Methods for Speech in Additive White Gaussian Noise. // *Proceedings of Midwest Symposium on Circuits Systems*, 1987, p. 1262–1265.

- [38] Krubsack, D. A.; Niederjohn, R. J. An Autocorrelation Pitch Detector and Voicing Decision with Confidence Measures Developed for Noise-Corrupted Speech. *IEEE Transactions on Speech Processing*, 39, No. 2, 1991, p. 319–330.
- [39] Lahat, M.; Niederjohn, R. J.; Krubsack, D. A. A Spectral Autocorrelation Method for Measurement of the Fundamental Frequency of Noise-Corrupted Speech. *IEEE Transaction on Acoustics, Speech, and Signal Processing*, 35, No. 6, 1987.
- [40] Lahouti, F.; Fazel, A. R.; Khandani, A. K. Single and Double Frame Coding of Speech LPC Parameters Using a Lattice-based Quantization Scheme. Technical report, *IEEE Transaction on Speech and Audio Processing*, 2004.
- [41] Lima, C.; Coelho, A.; Zuben, F. Mixture of Experts Applied to Nonlinear Dynamic Systems Identification: A Comparative Study. // *VII Brazilian Symposium on Neural Networks*, 2002 162, p.
- [42] Limand, C. P.; Wooand, S. C.; Lohand, A. S.; Osman, R. Speech recognition using artificial neural networks. // *Web Information Systems Engineering, 2000*, Vol. 1, 2000, p. 419 – 423.
- [43] Lipeika, A.; Lipeikienė, J. Speaker Recognition Based on the Use of Vocal Tract and Residue Signal LPC Parameters. // *Informatica*, Vol. 10(4), 1999, p. 377–388.
- [44] Lipeika, A.; Lipeikienė, J.; Telksnys, L. Development of Isolated Word Speech Recognition System. // *Informatica*, Vol. 13(1), 2002, p. 37–46.
- [45] Lipeika, A. L. Formantiniai požymiai atpažįstant kalbą. // *Information sciences*, Vol. 34, 2005, p. 215–219.
- [46] Lipeikienė, J.; Lipeika, A. Language EGINEERING in Lithuania. // *Informatica*, Vol. 9(4), 1998, p. 449–456.
- [47] Ljung, L. System Identification - Theory for the User. Prentice-Hall, Upper Saddle River, N.J., 2 edition, 1999.
- [48] Logan, B.; Kositsky, A.; Moreno, P. Semantic analysis of song lyrics. // *Multimedia and Expo*, Vol. 2, 2004, p. 827–830.
- [49] Maddage, N. C.; Changsheng, X.; Ye, W. Singer identification based on vocal and instrumental models. // *Multimedia and Expo*, Vol. 2, 2004, p. 375–378.
- [50] Makhoul, J. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63, No. 4, 1975, p. 561–580.
- [51] Manxia, T.; Dusheng, W.; Changxin, F. A novel variable-rate MELP speech coder. // *Signal Processing Proceedings, 2000*, Vol. 2, 2000, p. 693 – 696.
- [52] Mapelli, F.; Lancini, R. Audio hashing technique for automatic song identification. // *Information Technology: Research and Education*, 2003, p. 84–88.
- [53] Markel, J. D. The SIFT Algorithm for Fundamental Frequency Estimation. *IEEE Transactions on Audio Electroacoustics*, 20, 1972, p. 367–377.

- [54] Miller, R. L. Performance Characteristics of an Experimental Harmonic Identification Pitch Extraction (HIPEX) System. *Journal of Acoustics Society of America*, 47, 1970, p. 1593–1601.
- [55] Motlicek, P. Application of Mel-scale Filter bank for Noise Estimation in Speech Processing. // *12th International Czech-Slovak Scientific conference Radioelektronika'2002*, 2002 4, p.
- [56] Navakas, D. *Artificial Neural Networks for the Restoration of Noise Distorted Songs Audio Records* Doctoral dissertation, Vilnius Gediminas Technical University, September 1999 No. 434.
- [57] Navickaite-Martinonienė, E. *Elementarioji muzikos teorija*. Vilnius: Vaga, 1979.
- [58] Neji, Z.; Beji, F.-M. Neural Network and Time Series Identification and Prediction. // *International Joint Conference on Neural Networks*, Vol. 4, 2000, p. 44–61.
- [59] Nelles, O. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Springer-Verlag, 2001.
- [60] Paliwal, K. K.; Kleijn, W. B. Quantization of LPC parameters in Speech Coding and Synthesis. Eds. New York: Elsevier Science, 1995.
- [61] Paulikas, Š.; Navakas, D. New Implementation Scheme for the Restoration of Voiced Speech Signals. // *Informatica*, Vol. 14(3), 2003, p. 349–356.
- [62] Pereira, W. Modifying LPC Parameter Dynamics to Improve Speech Coder Efficiency. McGill University, Canada, 2001.
- [63] Petrovic, I.; Baotic, M.; Peric, N. Model Structure Selection for Nonlinear System Identification Using Feedforward Neural Networks. // *International Joint Conference on Neural Networks*, Vol. 1, 2000, p. 53–57.
- [64] Plotz, T.; Fink, G. A.; Husemann, P. Automatic Detection of Song Changes in Music Mixes Using Stochastic Models. // *Pattern Recognition*, Vol. 3, 2006, p. 665–668.
- [65] Polur, P. D.; Zhou, R.; Yang, J.; Adnani, F.; Hobson, R. S. Isolated speech recognition using artificial neural networks. // *Engineering in Medicine and Biology Society, 2001*, Vol. 2, 2001, p. 1731 – 1734.
- [66] Pupekis, R. Identification of Piecewise Affine Wiener Systems Using Data Partition. Technical report, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, May 2003.
- [67] Pupekis, R.; Navakas, D.; Ljung, L. Identification of Wiener systems with hard and discontinuous nonlinearities. Technical report, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Mar. 2003.
- [68] Rabiner, L. R.; Cheng, M. J.; Aaron E. Rosenberg; McGonegal, C. A. A Comparative Performance Study of Several Pitch Detection Algorithms. *IEEE Transaction on Acoustics, Speech, and Signal Processing*, 24, No. 5, 1976, p. 399–418.

- [69] Raškinis, G. Preprocessing for Folk Song Acoustic Records for Transcription into Music Scores. *Informatica*, 9, No. 3, 1998, p. 1–9. ISSN 0868-4952.
- [70] Raškinis, G.; Raškinienė, D. Building Medium-Vocabulary Isolated-Word Lithuanian HMM Speech Recognition System. // *Informatica*, Vol. 14(1), 2003, p. 75–84.
- [71] Ross, M. J. et al. Average Magnitude Difference Function Pitch Extractor. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 22, 1974, p. 353–362.
- [72] Rossi, P.; Palmieri, F.; Cutugno, F. Inversion of F_0 model for natural-sounding speech synthesis. // *Acoustics, Speech, and Signal Processing, 2003*, Vol. 1, 2003, p. 520 – 523.
- [73] Rudžionis, A.; Rudžionis, V.; Ratkevičius, K.; Kasparaitis, P. Lietuvių kalbos sintezės ir atpažinimo taikymai automatikoje. // *Automatika ir valdymo technologijos*, 2001, p. 13–18.
- [74] Rudžionis, A.; Rudžionis, V.; Žvinys, P. Lietuvių šnekamosios kalbos garsynas LTDIGITS: rezultatai ir problemos. // *Informacinės technologijos*, 2000, p. 162–166.
- [75] Rudžionis, V. E. *Kalbos atpažinimas fonetinių vienetų pagrindu* Daktaro disertacija, Kauno Technologijos universitetas, Spalis 1998.
- [76] Schön, T. On Computational Methods for Nonlinear Estimation. Technical Report Licentiate Thesis no. 1047, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Oct. 2003.
- [77] Schön, T.; Gustafsson, F. Particle filters for system identification of state-space models linear in either parameters or states. Technical report, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Apr. 2003.
- [78] Schuller, D. Data density versus data security: formats suitable for archival purposes., Vol. 85-97 43 FIAF Congress: Archiving the Audio-Visual Heritage, 1987.
- [79] Shiand, Y.; Chang, E. Spectrogram-based formant tracking via particle filters. // *Acoustics, Speech, and Signal Processing, 2003*, Vol. 1, 2003, p. 168 – 171.
- [80] So, S.; Paliwal, K. A comparative study of LPC parameter representations and quantisation schemes for wideband speech coding. *Digital Signal Processing*, , 2006.
- [81] So, S.; Paliwal, K. K. Improved noise-robustness in distributed speech recognition via perceptually-weighted vector quantisation of filterbank energies. // *INTERSPEECH-2005*, 2005, p. 941–944.
- [82] Speech, E. Elan Text to Speech© and ProVerbe Speech Unit©. Technical report, 1996. Prieiga per internetą <<http://www.elanTTS.com>>.
- [83] Sproat, R. Multilingual Text-to-Speech Synthesis: The Bell Labs Approach. Technical report, Kluwer Academic Publishers, Boston., 1997.
- [84] Stegmayer, G.; Chiotti, O. Identification of Frequency-Domain Volterra Model Using

- Neural Networks. // *International Conference on Artificial Neural Networks*, Vol. 2, 2005, p. 465–471.
- [85] Sutton, S.; Novick, D.; Cole, R.; Fanty, M. Building 10.000 spoken-dialogue systems. Technical report, Proceedings of the International Conference on Spoken Language Processing, Philadelphia, PA, 1996.
- [86] Thibault, F.; Depalle, P. Adaptive processing of singing voice timbre. // *Electrical and Computer Engineering*, 2004, Vol. 2, 2004, p. 871 – 874.
- [87] Trawicki, M. B.; Johnson, M. T.; Osiejuk, T. S. Automatic Song-Type Classification and Speaker Identification of Norwegian Ortolan Bunting (*Emberiza Hortulana*) Vocalizations. // *Machine Learning for Signal Processing*, 2005, p. 277–282.
- [88] Tzanetakis, G. Song-specific bootstrapping of singing voice structure. // *Multimedia and Expo*, Vol. 3, 2004, p. 2027–2030.
- [89] Vaseghi, S. V. *Advanced Signal Processing and Digital Noise Reduction*. John Wiley & Sons Ltd, 1996.
- [90] Woutersand, J.; Macon, M. Spectral modification for concatenative speech synthesis. // *Acoustics, Speech, and Signal Processing*, 2000, Vol. 2, 2000, p. 941 – 944.
- [91] Yamaguchi, M.; Tadokoro, Y. Pitch estimation of polyphonic songs using comb filters and singular value decomposition. // *Conference on Convergent Technologies for Asia-Pacific Region*, Vol. 1, 2003, p. 375–379.
- [92] Yangguo, T. J. K. Multi-Source Based Acoustic Model for Speech Synthesis. // *International Conference on Signal Processing*, 2004.
- [93] Yongwei, Z.; Kankanhalli, M. Key-based melody segmentation for popular songs. // *Pattern Recognition*, Vol. 3, 2004, p. 862–865.
- [94] Zhang, Y.-B.; Zhou, J.; Bian, Z.-Q. Samples-based automatic key segment extraction for popular songs. // *Machine Learning and Cybernetics*, Vol. 8, 2005, p. 4891–4897.
- [95] Zhu, Y.; Comaniciu, D.; Ramesh, V.; Schwartz, S. Parametric Representations for Nonlinear Modeling of Visual Data. // *IEEE International Conference on Computer Vision and Pattern Recognition*, 2001, p. 553–560.

AUTORIAUS PUBLIKACIJOS

- [A1] Leonavičius, R.; Navakauskas, D. Aspects of Melisma Synthesis. *Elektronika ir Elektrotechnika*, ISSN 1392-1215, 6(48), 2003, p. 18–21. Prieiga per internetą: <http://www.ktu.lt/lt/mokslas/zurnalai/elektr/z48>
- [A2] Leonavičius, R.; Navakauskas, D. Improvement of the Restoration of Melisma by a Signal Synthesis. *Izvestija Beloruskoj Inzenernoj Akademii*, 1(19), No. 1, 2005, p. 110–113.

- [A3] Leonavičius, R.; Navakauskas, D. Restoration of Melisma by Signal Synthesis. *Izvestija Beloruskaj Inženernoj Akademii*, 1(17), No. 2, 2004, p. 64–67.
- [A4] Leonavičius, D. Netiesinių dinaminių sistemų identifikavimo metodų apžvalga. Iš *6-osios Lietuvos jaunųjų mokslininkų konferencijos „Lietuva be mokslo – Lietuva be ateities“ 2003 medžiaga*, 2003, p. 36-48. ISBN 9986-05-473-7.
- [A5] Leonavičius, R.; Navakauskas, D. Melizmų sintezės tyrimas. Iš *4-osios Lietuvos jaunųjų mokslininkų konferencijos „Lietuva be mokslo – Lietuva be ateities“ medžiaga*, 2001, p. 62-71. ISBN 9986-05-473-7.



Melizmų modeliavimo įrankiai

A.1. Sukurtų Matlab funkcijų rinkinys

DinamLpc0x.m	TPK analizės dinaminio langų funkcija
dlpcm03.m	Dinaminė sintezė naudojant originalius TPK parametrus ir originalų gerklų signalą
dlpcm06.m	Sintezė naudojant suvidurkintą TPK ir gerklų signalo fragmentą
dlpcm0x.m	Sintezė su dirbtiniu gerklų signalu
ExtrMelPar.m	Melizmos parametrų DNT mokymui išskyrimas
GirbtGerklSig.m	Dirbtinio gerklų signalo sukūrimas
LoadParamsF.m	Sintezės parametrų įkėlimas
MelAvgTrain.m	DNT mokymo funkcija
MosPlot.m	Sintezės braižymo funkcija
NormalizeMelisma.m	Signalų paruošimo prieš perduodant į DNT funkcija
PitchPlot.m	Modeliavimo braižymo funkcija
ReadAllWav.m	Melizmų signalų įrašymas į masyvą
TrimMelisma.m	Melizmos signalo pirminis apdorojimas
WorkAroundIntensity.m	Intensyvumo DNT apmokymo funkcija
WorkAroundPitch.m	Pagrindinio tono DNT apmokymo funkcija

A.2. Surinktų melizmų rinkinys

M1

/Skaidyto/KRI/F01	Atlikėjo M1 foršlago F01 melizmos
/Skaidyto/KRI/F06	Atlikėjo M1 foršlago F06 melizmos
/Skaidyto/KRI/G01	Atlikėjo M1 grupeto G01 melizmos
/Skaidyto/KRI/G02	Atlikėjo M1 grupeto G02 melizmos
/Skaidyto/KRI/M01	Atlikėjo M1 mordento M01 melizmos
/Skaidyto/KRI/M02	Atlikėjo M1 mordento M02 melizmos
/Skaidyto/KRI/T02	Atlikėjo M1 trelės T02 melizmos

M2

/Skaidyto/AUSRA/F1	Atlikėjo M2 foršlago F01 melizmos
/Skaidyto/AUSRA/F2	Atlikėjo M2 foršlago F02 melizmos
/Skaidyto/AUSRA/G1	Atlikėjo M2 grupeto G01 melizmos
/Skaidyto/AUSRA/M1	Atlikėjo M2 mordento M01 melizmos
/Skaidyto/AUSRA/M02	Atlikėjo M2 mordento M02 melizmos

V1

/Skaidyto/ADM/F01	Atlikėjo V1 foršlago F01 melizmos
/Skaidyto/ADM/F02	Atlikėjo V1 foršlago F02 melizmos
/Skaidyto/ADM/F06	Atlikėjo V1 foršlago F06 melizmos
/Skaidyto/ADM/G01	Atlikėjo V1 grupeto G01 melizmos
/Skaidyto/ADM/G02	Atlikėjo V1 grupeto G02 melizmos
/Skaidyto/ADM/M01	Atlikėjo V1 mordento M01 melizmos
/Skaidyto/ADM/T02	Atlikėjo V1 trelės T02 melizmos
/Skaidyto/ADM/M02	Atlikėjo V1 mordento M02 melizmos
/Skaidyto/ADM/T04	Atlikėjo V1 trelės T04 melizmos

V2

/Skaidyto/IRMNT/F1	Atlikėjo V2 foršlago F01 melizmos
/Skaidyto/IRMNT/G1	Atlikėjo V2 grupeto G01 melizmos
/Skaidyto/IRMNT/M1	Atlikėjo V2 grupeto G02 melizmos
/Skaidyto/IRMNT/M02	Atlikėjo V2 mordento M01 melizmos

B

Sąvokų žodynas

- A** *amplitudė* – magnitude
 analitinės – analytical
 analizės langas – frame
 atnaujinti – restore
 atskaita – sample
 atstatyti – reconstruct
 autokoreliacija – autocorrelation
- B** *balsinis* – voiced
 balsis – vowel
 begalinė impulsinė reakcija – infinite impulse response
- D** *dažninė charakteristika* – frequency response
 daugianaris – polynomial
- F** *filtras* – filter
 fonema – phoneme
 foršlagas – fortis
- G** *gerklos* – glottis
 greitoji Furjė transformacija – fast Fourier transformations
 grupetas – gruppett

- I** *impulsinė reakcija* – impulse response
- K** *kalbos atpažinimas* – speech recognition
 kalbos kodavimas – speech coding
 kalbos segmentacija – speech segmentation
 kalbos sintezė – speech synthesis
 koeficientas – coefficient
 koreliacija – correlation
- M** *mažiausiųjų kvadratų* – least-squares
 melisma – melisma
 mordentas – mordent
- N** *nuo laiko nepriklausomas* – time-invariant
 nuokrypis – variance
- P** *periodinis* – periodical
 prastinimas – simplification
 priebalsis – consonant
 prognozatorius – predictor
- R** *realioji dalis* – real part
 ribota impulsinė reakcija – finite impulse response
- S** *sintezatorius* – synthesiser
 sisteminė klaida – bias
 spektras – spectrum
 sąsūka – convolution
 suminė kvadratinė klaida – sum-square-error
- Š** *šaknis* – root
- T** *tiesinė laikinė sistema* – linear time-varying system
 tiesinė prognozė – linear prediction
 trilė – trill
- V** *vidutinė kvadratinė klaida* – mean sum-square-error
 vidutinės nuomonės rezultatas – mean opinion score
 vokalinis traktas – vocal tract
- Ž** *žadinimo signalas* – excitation signal

Dalykinė rodyklė

A

ADCH	13
apibrėžimas	
apibendrintas melizmos modelis	56
foršlagas	8
foršlago modelis	64
grupetas	8
grupetas modelis	65
melizma	7
mordentas	9
mordento modelis	66
trelė	9
trelės modelis	67
AR	18
ARARX	19, 20
ARMA	18
ARMAX	19, 20
ARX	19
atgalinio sklidimo tinklas	27
atlikėjų tipai	70

B

BJ	19, 20
----	--------

C

charakteristika	
ADCH	13
dažninė	13

FDCH	13
intensyvumas	10
kepstras	11
laikinė	10
laikinė-dažninė	15
MEL skalės spektras	13
pagrindinis tonas	11
spektrograma	15
TPK kitimas	12
TPK modelio ADCH	13
TPK modelio FDCH	13
TPK spektrograma	15
TPK šaknų kampai	15

D

daugiasluoksnis perceptronas	27, 28, 56
dinaminis TPK	
algoritmas	43
formalizavimas	45
dirbtinis gerklų signalas	33
dirbtinis neuronas	27

F

FDCH	13
foršlagas	8
ilgas	8
modelis	64
rinkinys	69

trumpas	8	melizmų tipai	69
G		metodika	
gerklų signalas		išaiškinimas	51
modeliavimas	54	minimizavimas	
gerklų signalo fragmentavimas	36	Levenberg-Marquardt	58
gerklų signalo modeliavimas	33	mažiausia vidutinė kvadratinė klaida	57
grupetas	8	praktiniai aspektai	61
keturių garsų	9	modeliavimas	
modelis	65	gerklų signalas	54
penkių garsų	9	lango postūmis	54
rinkinys	69	modelis	
I		apibendrintas	53
identifikavimas	16, 23	AR	18
intensyvumas	10	ARARX	19, 20
modeliavimas	56	ARMA	18
išoriniai faktoriai	70	ARMAX	19, 20
K		ARX	19
Kepstras	11	būsenų	19, 21
L		BJ	19, 20
Levenberg-Marquardt		foršlagas	64
minimizavimas	58	grupetas	65
pavyzdys	61	Hameršteino	22
M		išvestinis	28
MA	18	MA	18
Matlab		mordentas	66
algoritmo įgyvendinimas	62	NARMAX	28
pavyzdys	61	NARX	28
mažiausia vidutinė kvadratinė klaida		NBJ	28
minimizavimas	57	neuronas	27
MEL skalės spektras	13	NOE	28
melizma		NRIR	28
metodika	31	OE	19, 20
modelio identifikavimas	72	pagrindas	48
sintezė	76	paslėptųjų neuronų skaičius	63
sintezės metodika	52	signalas	49
melizmos sintezė		signalų parametrai	50
metodika	48	struktūra	21
melizmos sintezės		TPK	49
modeliavimo pagrindas	48	trelė	67
melizmos sintezės metodika	50	Vinerio	22
melizmų		modifikuotas TPK	43
modeliavimas	7–30	mordentas	9
melizmų rūšys	69	dvigubos trukmės	9
melizmų rinkiniai	69	kryžminis	9
melizmų rinkinys	70	modelis	66
		paprastas	9
		rinkinys	69

N

NDS

identifikavimas	29
patikra	25
struktūros	26
neuronas	27
neurono aktyvavimo funkcija	27
nuokrypis	25

O

OE	19, 20
----------	--------

P

pagrindinio tono modeliavimas	40
pagrindinis tonas	11
modeliavimas	56
patikra	16

R

rezultatai

sintezė B1	79, 80
sintezė F06	79
sintezė G02	79
sintezė M1	79
sintezė T02	79
sintezė V1	77
rinkinių sudarymas	69

S

sintezė

adalina	73
atbulinis apdorojimas	50
daugiasluoksnis perceptronas	73
dirbtinis gerklų signalas	33
metodika	50
MOS	76
tiesioginis apdorojimas	50
vidutinės nuomonės rezultatas	76
vieno atlikėjo melizmų	72

sistema

dinaminė	16, 17, 21, 23–26
identifikavimas	23, 24
netiesinė	16, 21, 24, 26
patikra	25
struktūra	17
tiesinė	17, 23
sisteminė klaida	25
spektrograma	15
struktūros	26

T

tiesinė prognozė	31
TPK kitimas	12
TPK spektrograma	15
TPK šaknų kampų kitimas	15
trelė	9
žemesnio tono	9
aukštesnio tono	9
modelis	67
pagrindinio tono	9
rinkinys	69
triukšmas	70

V

vidutinės nuomonės rezultatas	76
-------------------------------------	----

MELIZMŲ SINTEZĖ DIRBTINIŲ NEURONŲ TINKLAIS

Daktaro disertacija

Technologijos mokslai, elektros ir elektronikos inžinerija (01T)

© Romas Leonavičius, 2006

romas.leonavicius@el.vtu.lt



ELEKTRONINIŲ SISTEMŲ KATEDRA

Elektronikos fakultetas

Vilniaus Gedimino technikos universitetas

Naugarduko 41, LT-03227 Vilnius-6

Parengta L^AT_EX2_ε sistema



VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
ELEKTRONIKOS FAKULTETAS