VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
KOMPIUTERIJOS KATEDRA

Baigiamasis magistro darbas

Balso tono valdymas lietuvių kalbos sintezėje

Atliko: 2 magistratūros kurso studentas
Vitalijus Agejevas (parašas)

Darbo vadovas:
assoc. prof. Agirdas Bastys (parašas)

Vilnius
2006

Turinys

Įvadas

Šiuo metu viena labiausiai nagrinėjamų problemų kalbos sintezės sistemose yra balso tono valdymas. Kalbos sintezė jau pakankamai pažengusi, kad klausytojas galėtų nesunkiai suprasti kas yra sakoma, tačiau neįmanoma nepastebėti, kad sintezuotos kalbos intonacija kol kas neprilygsta natūraliai. Aišku viena: tam, kad dirbtinė kalba skambėtų natūraliai, jai reikia suteikti natūralią intonaciją.

Magistrinio darbo tyrimo objektas – sintetinio balso tono valdymas dvigarsiais pagrįstoje lietuvių kalbos sintezėje. Kitais žodžiais tariant bus tiriama kaip sintetiniam balsui suteikti natūralumo, nes, kaip yra žinoma, dvigarsiais pagrįstas balsas sukuriamas įrašant neutraliai žmogaus tariamus garsus. Iš tokių garsų sukuriama dvigarsių duomenų bazė, kuri naudojama kalbos sintezei. Tokiu būdu sintetinė kalba gaunasi monotoniška ir netikroviška. Natūralumo kalbai galima suteikti varijuojant garsų trukmę bei toną.

Dvigarsis (angl.: Diphone) – tai kalbos vienetas, kuris prasideda vieno garso pastovios srities viduryje ir baigiasi kito garso pastovios srities viduryje.

Darbo tikslas – pamėginti sumodeliuoti lietuvių kalbos balso toną. Tam, kad pasiekti užsibrėžtą tikslą, racionaliausia skaidyti tiriamą objektą į smulkesnes autonomines dalis ir jas nagrinėti. Kalba gali būti skaidoma į sakinius, sakiniai - į frazes, frazės - į garsų (fonemų) sekas. Frazės skirstomos į pagrindinius tris tipus: konstatuojamąjį, klausiamąjį ir šaukiamąjį. Skirtingo tipo frazės skiriasi ne tik jų paskirtimi (klausiamoji frazė skirta klausti ir t.t.) bet ir jų tariamu tonu, kuris leidžia spręsti kokio tipo tai frazė. Pavyzdžiui konstatuojamosios frazės tonas frazės gale nusileidžia, o klausiamosios atvirkščiai – pakyla.

Darbo eigoje siekiama išanalizuoti balso tono kitimą natūralios lietuvių kalbos balso įrašuose bei pagal tyrimo rezultatus mėginama sumodeliuoti konstatuojamojo frazės toną bei kirčių toną.

1. Kalbos sintezė

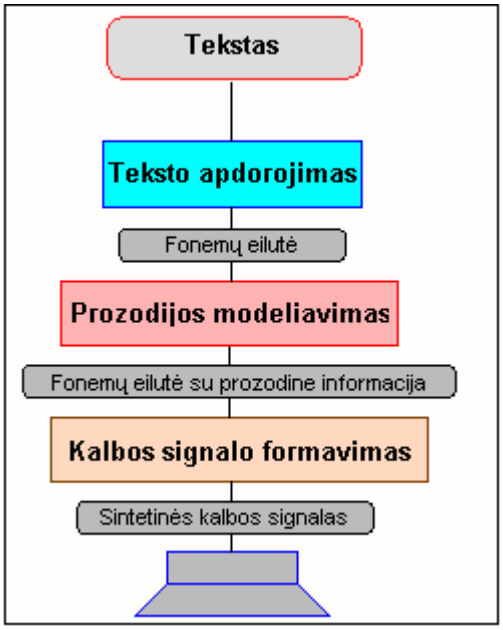
Kalbos sintezė – tai dirbtinis žmogaus kalbos generavimas [Wik06a]. Tai atliekanti sistema vadinama kalbos sintezatoriumi, kuris gali būti realizuotas tiek programiškai, tiek aparatūriškai. Kalbos sintezatoriai dažnai vadinami *text-to-speech* (TTS) sistemomis, taip nurodant jų gebėjimą konvertuoti tekstą į kalbą. Vis dėl to, dažnai kalbos sintezatorių įvesties duomenys yra ne paprastas kalbos tekstas, o kalbos fonetinė transkripcija su prozodine informacija.

Aprašant sintezuojamos kalbos kokybę naudojamos dvi charakteristikos: kalbos natūralumas bei suprantamumas. Natūralumas nusako kiek kalbos skambesys panašus į tikrą žmogaus kalbą. Suprantamumo kriterijus nusako kaip lengvai kalba gali būti suprantama. Idealaus kalbos

sintezatoriaus produktas – suprantama ir natūrali kalba. Visos kalbos sintezavimo technologijos siekia to pačio tikslo – maksimizuoti šiuos du parametrus.

1.1. Kalbos sintezė iš teksto

1 paveikslėlyje pavaizduota kalbos sintezės pagal tekstą funkcinė diagrama. Kalbos sintezės pagal tekstą procesas susideda iš trijų pagrindinių etapų: teksto apdorojimo, prozodijos modeliavimo bei kalbos signalo formavimo [Dre06].



1 pav. kalbos sintezės pagal tekstą funkcinė diagrama

1 etapas: Teksto apdorojimas

Kalbos teksto apdorojimo etape tekste esantys skaičiai, sutrumpinimai, datos ir kiti specialūs simboliai konvertuojami į juos atitinkančius ištisus žodžius (pvz.: santrumpa „prof.“ keičiama į žodį „profesorius“).

Atliekama sintaksinė analizė, kurios metu surandama teksto sintaksinė struktūra, kuri paprastai naudojama tekstą skaidant į sakinius, o šiuos į frazes.

Morfologinės bei kontekstinės teksto analizės metu nustatoma kokiai kalbos daliai priklauso žodžiai bei kokios gramatinės formos jie yra. Pagal analizės rezultatus žodžiai sukirčiuojami [Kas00] [Kas01]. Kirčiuotas tekstas transkribuojamas – tai yra raidės verčiamos į fonemas [Kas99]. Fonema yra nedalomas garsinis vienetas, žymimas specialiu simboliu (arba keliais simboliais). Pavyzdžiui, žodžio „parlame`ntinė“ fonetinė transkripcija atrodo taip: „p a r l a m‘ E N‘ t‘ i n‘ ee“.

Čia fonemos atskirtos tarpais. Mažosiomis raidėmis žymimos fonemos atitinka trumpus garsus. Iš pavyzdžio matyti, kad pirmos penkios žodžio raidės sutampa su fonemų žymėjimais. Toliau raidė *m* pakeista fonema /m‘/. Fonema su viengubos kabutės simboliu reiškia garso minkštumą. Žodžio skiemuo *en* kirčiuotas, todėl transkripcijoje tai pažymėta fonemomis /E/ /N‘/. Didžiosios raidės reiškia kirčiuotas fonemas. Ne visų raidžių ir fonemų simboliai sutampa: štai paskutinę raidę *ė* atitinka ilgo garso fonema /ee/. Taip pat yra specialios fonemos ilgiems kirčiuotiems garsams žymėti, pvz.: priklausomai nuo kirčio raidė ‚a‘ gali būti transkribuojama skirtingai: fonema /Aa/ - atitinka ilgą garsą su tvirtaprade priegaide (pvz.: „na^mas“→“n **Aa** m a s“), o /aA/ - žymi tvirtagalę priegaidę (pvz.: „skanda~las“→„s k a n d **aA** l a s“).

2 etapas: Prozodijos modeliavimas

Kalba, gauta jungiant trumpus garsų segmentus turi gerą suprantamumą (intelligibility) tačiau skamba nenatūraliai. Natūralumas yra svarbus faktorius *for user acceptance*. Kalbai natūralumo suteikia prozodiniai komponentai, tai garsų trukmė, tonas. Šiame etape fonemoms priskiriama prozodinė informacija.

Garsų trukmių parinkimui dažniausiai naudojami dviejų tipų modeliai: vienas pagrįstas taisyklėmis, kitas – duomenimis.

Vienas iš paprastų ir populiarių taisyklėmis pagrįstų modelių yra dar 1979m. pasiūlytas D. Klatt modelis [Kla87]. Šiame modelyje garsų trukmė paskaičiuojama pagal formulę

dur = (inhdur - mindur) * prct / 100 + mindur

parametrai:

- inhdur – prigimtinė garso trukmė
- mindur – minimali garso trukmė
- prct – faktorius, nustatomas pagal konkretaus garso kirtį, jo poziciją žodyje/frazėje/sakinyje bei pagal garsų kontekstą.

Parametrų reikšmės gaunamos eksperimentiniu būdu. Kuriant šį modelį buvo padarytos kelios prielaidos: 1) kiekvienas garsas turi savo prigimtinę trukmę 2) kiekvienas efektas nusakomas kaip trukmės padidėjimas arba sumažėjimas procentais 3) garsas negali būti trumpesnis už jam būdingą minimalų ilgį.

3 etapas: Kalbos signalo formavimas

Šio etapo įvesties duomenys – fonemų eilutė su garsų trukmėmis bei tono duomenimis. Akustinė sintezė atlieka garso vienetų jungimą. Tam, kad sintezuojama kalba skambėtų natūraliai šiuolaikiniai sintezatoriai naudoja balso duomenų bazes sudarytas iš trumpų garsų segmentų. Populiariausi garsų vienetai yra difonai, taip pat kai kurių kalbų sintezei gali būti naudojami skiemenys. Segmentai apjungiami vadovaujantis fonemų kontroline informacija. Balso duomenų

bazės gamyba yra kritiškai svarbi kalbos sintezės rezultatams. Duomenų bazės dažnai sudaromos iš difonų.

1.2. Kalbos intonacija

Šiuo metu viena labiausiai nagrinėjamų problemų kalbos sintezės sistemose yra balso tono valdymas. Kalbos sintezė jau pakankamai pažengusi, kad klausytojas galėtų nesunkiai suprasti kas yra sakoma, tačiau neįmanoma nepastebėti, kad sintezuotos kalbos intonacija kol kas neprilygsta natūraliai. Aišku viena: tam, kad dirbtinė kalba skambėtų natūraliai jai reikia suteikti natūralią intonaciją. Tuomet iškyla klausimas: kas yra intonacija? Intonacija – tai balso tono varijavimas kalbant.

Tono informacijų lygiai:

- Lingvistinis - kirtis, sakinio tipas
- Para-lingvistinis - kalbančiojo laikysena(nuostata), intencija, dialektas
- Ne-lingvistinis: - sveikatos būklė, emocinė būsena

Tam, kad sugeneruoti intonaciją kalbos sintezei, reikia:

- Formalaus intonacijos apibrėžimo tam skirtos teorijos pagrindu
- Būdo, kaip intonacijos apibrėžimą konvertuoti į atitinkamą tono kontūrą arba kontrolinių taškų aibę, iš kurios būtų atkuriamas tonas.

Intonacijos charakteristikoms aprašyti naudojami intonacijų modeliai, modeliuojantys toną pagal apsibrėžtas taisykles. Pagrindiniai reikalavimai keliami intonacijų modeliui:

- Ribotumas. Modelio generuojamas tono transkripcija turėtų būti kiek įmanoma trumpesnė bei neperteklinė - be galimybės iš vienos transkripcijos išvesti kitą.
- Platus padengimas. Generuojama transkripcija turėtų apimti kiek galima daugiau intonacinių fenomenų bei būti pakankamai lanksti, kad pajėgtų išreikšti skirtumus tarp skirtingai suvokiamų kalbos elementų.
- Lingvistinis prasmingumas. Generuojama transkripcija turėtų būti tokios formos, kad ją pilnai galėtų generuoti ir interpretuoti aukštesnio abstrakcijos lygio kalbos sintezės komponentai.

Pagal tai kiek modeliai tenkina lingvistinio prasmingumo reikalavimą gali būti skirstomi į tris lygius: aukšto abstrakcijos lygio (labiausiai tenkinantis reikalavimą), vidutinio bei žemiausio abstrakcijos lygio (mažiausiai tenkinantis reikalavimą).

Taip pat, modeliai pagal transkribavimo metodikas klasifikuojami į tris kategorijas: žymėjimo (angl. labeling), stilizavimo (angl. stylization) ir modeliavimo (angl. modeling).

Žymėjimo kategorijai priklausančios schemos pagrįstos svarbių tono fragmentų pažymėjimu specialiomis žymėmis. Ši kategorija laikoma aukšto abstrakcijos lygio tono apibrėžimo kategorija. Vienas populiariausių kategorijos modeliu laikomas ToBI (žr. sk. 2.2.) bei INTSINT (žr. sk. 2.3.).

Stilizavimo kategorijos schemoms būdingas aproksimacinis tono formavimas iš elementarių segmentų, pvz. tiesių linijų, splainų arba parabolinių funkcijų. Natūralaus balso tono segmentavimas pagrįstas subjektyviu sprendimu, tačiau po to sekantis aproksimacinis segmentų suliejimas išlaiko kiekybines originalaus tono savybes.

Modeliavimo kategorija priskiriama žemiausiam abstrakcijos lygiui. Šiai kategorijai priskiriami metodai, generuojantys toną pagal tam tikrą modelį, pagrįstą komandomis, turinčiomis lingvistinę informaciją. Šios kategorijos žinomiausias modelis yra Fujisaki.

1.3. Kalbos signalo formavimo metodai

Vienos technologijos geriau išgauna kalbos natūralumą, kitos – suprantamumą. Sintezavimo sistemos paskirtis dažnai nulemia technologijų pasirinkimą. Pagrindinės naudojamos sintezavimo technologijos: konkatenacinė sintezė ir formantinė sintezė [Kas05].

Konkatenacinė sintezė

Konkatenacinė sintezė paremta įrašytų garsų segmentų jungimu. Palyginti su kitomis technologijomis garsų jungimas išgauna natūraliausiai skambančią kalbą. Yra dvi populiariausios konkatenacinės sintezės technikos: Fonetinių vienetų išrinkimo (Unit selection) sintezė bei difonų sintezė.

Fonetinių vienetų išrinkimo sintezė naudoja dideles įrašyto balso duomenų bazes (daugiau nei valandos balso įrašas). Įrašyta kalba suskaidoma į: fonemas, skiemenis, morfemas, žodžius, frazes ir sakinius. Dažniausiai kalbos skaidymas į vienetus atliekamas automatiškai, papildomai gautus rezultatus pakoreguojant rankiniu būdu. Po to, sukuriamas indeksas, remiantis gautų vienetų akustiniais parametrais tokiais kaip tonas, trukmė, pozicija skiemenyje bei kaimyniniai garsai. Sintezavimo metu norimas kalbos fragmentas sukuriamas iš duomenų bazės išrenkant tinkamiausią kandidatų grandinę. Šis procesas dažniausiai naudoja specialų sprendimų medį. Kaip jau buvo minėta, ši sintezavimo technologija leidžia sudaryti bazes iš ištisų žodžių. Technologijos privalumas, kad tokiu būdu siaurai kalbos sričiai (pvz.: autobusų išvykimo/atvykimo laiko pranešinėjimas) galima sukurti nedidelę ir visiškai natūraliai skambančią duomenų bazę. Tačiau ši technika turi trūkumų - norint realizuoti visos kalbos sintezę, tektų sukurti duomenų bazę iš visų egzistuojančių žodžių, be to, tie patys žodžiai turėtų būti įrašyti keliomis intonacijomis, kad būtų

galima sintezuoti ne vieno tono balsą. Tokia duomenų bazė užimtų labai daug vietos - gigabaitus atminties.

Difonų sintezės technika paremta dvigarsių jungimu. Iš visų balso duomenų bazių mažiausiai atminties užima duomenų bazė iš difonų. Įrašomi tik tie difonai, kurie pasitaiko duotoje kalboje. Pavyzdžiui, lietuvių kalboje yra – apie 5000 difonų, ispanų - apie 800, vokiečių – apie 2500. Difonų duomenų bazėje visų difonų užtenka tik po vieną pavyzdį (nereikia saugoti to paties difono skirtingomis intonacijomis). Sintezavimo metu, difonai jungiami į reikalingą seką ir suteikiama reikalinga prozodija. Prozodija bazės vienetams suteikiama panaudojant skaitmeninio signalo apdorojimo techniką, tai gali būti: tiesinė prognozė (Linear predictive coding), PSOLA, MBROLA ir kiti.

Sintezuojamo garso kokybė prastesnė nei Unit selection metodo tačiau skamba natūraliau nei formantinė sintezė.

Formantinė sintezė

Formantinėje kalbos sintezėje visai nenaudojama įrašyto balso duomenų. Vietoj to, sintezuota kalba išgaunama naudojant akustinį modelį. Dirbtinė kalba sukuriama varijuojant formantinius dažnius bei jų amplitudes. Reikiamas spektras sukuriamas sužadinant rezonatorių rinkinį garso šaltiniu arba triukšmo generatoriumi, priklausomai nuo to, ar imituojamas skardus, ar duslus. Dauguma formantine sinteze pagrįstų sistemų generuoja robotiškai skambančią kalbą, tačiau ji būna gerai suprantama. Formantinė sintezė neturi garsų jungimo problemos, būdingos konkatenacinei sintezei.

2. Balso tono modeliavimo metodai

2.1. Fujisaki modelis

Fujisaki modelis (Fujisaki 1983) [FOW99], dar dažnai vadinamas *Command-Response* modeliu. Iš pradžių modelis buvo sukurtas japonų kalbai, o vėliau buvo sėkmingai adaptuotas ir kitoms kalboms. Šis modelis priskiriamas superpozocinių modelių grupei. Superpozicinis reiškia, jog sudėtingas tono kontūras traktuojamas kaip kelių lygių signalas, sudarytas iš smulkesnių individualių komponentų, kurie pozicionuojami vienas virš kito.

Fujisaki modelyje pagrindinis tonas sudaromas iš dviejų lygių komandų: frazių komandų ir kirčių komandų. Kiekviena komanda tai tam tikrų parametrų rinkinys. Komandų sekos paduodamos į frazės ir kirčio tonų generavimo mechanizmus atitinkamai. Gauti atskirų lygių tonai pridedami prie bazinio tono. Ir taip gaunamas pagrindinis tonas. Fujisaki pagrindinis tonas generuojamas pagal šią formulę:

$$\ln F_0(t) = \ln Fb + \sum_{i=1}^{I} Ap_i Gp(t - T_{0i}) + \sum_{j=1}^{J} Aa_j [Ga(t - T_{1j}) - Ga(t - T_{2j})],$$

$$Gp(t) = \begin{cases} \alpha^2 t e^{-\alpha t}, & t \geq 0 \\ 0, & t < 0 \end{cases},$$

$$Ga(t) = \begin{cases} 1 + \cos(\pi t), & t \in [-1,1] \\ 0, & t < -1, t > 1 \end{cases}$$

Čia $Gp(t)$ frazės tono funkcija, o $Ga(t)$ - kirčio tono funkcija. Kitų formulės simbolių paaiškinimas:

Fb : pagrindinio tono bazinis dažnis

I : frazės komandų skaičius

J : kirčio komandų skaičius

Ap_i : i-osios frazės komandos dažnio faktorius

Aa_j : j-osios kirčio komandos dažnio faktorius

T_{0i} : i-osios frazės komandos pradžia

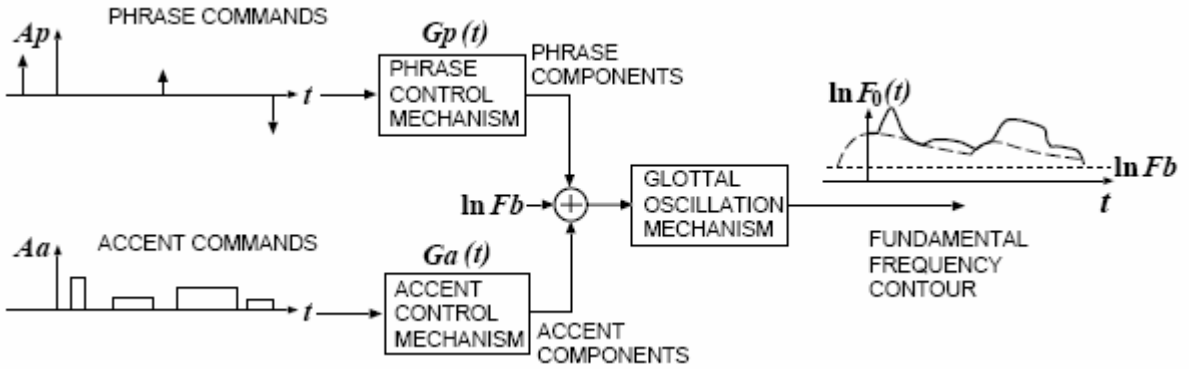
T_{1j} : j-osios kirčio komandos pradžia

T_{2j} : j-osios kirčio komandos pabaiga

α : kertinis frazės tono dažnis

β : kertinis kirčio tono dažnis

γ : santykinis kirčio komponentų ribinis aukštis



2 pav. Pagrindinio tono generavimo procesas pagal Fujisaki modelį.

2.2. ToBI modelis

ToBI (Tones and Break Indices) modelis pristatytas 1992 metais [BE97]. Pagrindiniai siekti tikslai kuriant modelį buvo: sukurti bendrą anglų kalbos intonacijos transkribavimo standartą, kuris

būtų aukšto abstrakcijos lygio, paprastas, lengvai išmokstamas bei apimantis geriausias tuometinių modelių savybes. Vėliau modelis adaptuotas vokiečių (GToBI), japonų (J_ToBI) ir kitoms kalboms.

ToBI modelis priskiriamas žymėjimo (labeling) modelių kategorijai, nes specialiais simboliais pažymimos tik tos vietos, kur vyksta prozodiškai reikšmingi pokyčiai, tarkim kurioje vietoje tonas krenta arba kyla, praleidžiant vietas kur jis pastovus.

Modelis intonaciją transkribuoja keliais lygiagrečiais lygiais. Pirmas lygis – toninis, antras – trūkių, trečias – įvairiarūšis bei ketvirtas - ortografinis. Kiekvienas lygis sudarytas iš specialių simbolių, aprašančių sintezuojamo fragmento prozodinius įvykius laiko skalėje.

Toninis lygis žymi pagrindinius kontrastinius intonacijos elementus. Žymimi trijų tipų įvykiai:

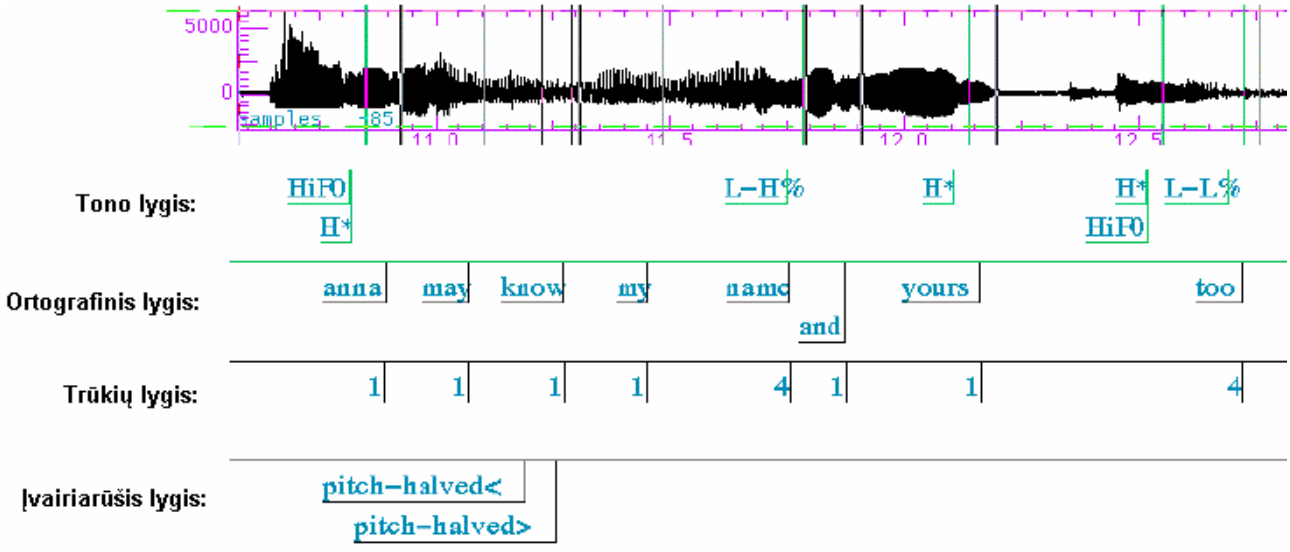
- Frazės kraštinis tonas, žymintis frazės pabaigos toną. Naudojami žymėjimai: H% arba L%.
- Frazės kirtis, esantis tarp paskutinio tono kirčio bei frazės kratinio tono. Įvykis žymimas H- arba L-.
- Kirčiuotų skiemenų tonai. Išskirti penki kirčio tono tipai:
    - H* - paprastas aukštas,
    - L* - paprastas žemas ,
    - L+H* - iš žemo kylantis į aukštą ,
    - L*+H - pavėluotas kilimas,
    - H+!H* - kritimas ant kirčio.

Trūkių lygis žymi ryšio stiprumą tarp gretimų žodžių. Apibrėžti keturi ryšio tipai, žymimi skaičiais nuo 0 iki 4:

- 0 – ryšys tarp fonetiškai artimai sugrupuotų žodžių, pvz. greitai šnekamoje kalboje,
- 1 – ryšys tarp dviejų skirtingų prozodinių žodžių
- 2 – žymi tarpinės frazės ribą
- 3 – žymi pilną intonacinę frazę

Įvairiarūšis lygis skirtas žymėti įvairiems spontaniškiems kalbos efektams, pvz.: juokas, dvejojimas ir pan.

Ortografinis lygis skirtas kalbos ortografiniams simboliams žymėti.



3 pav. ToBI modelio pavyzdys. Schemoje pavaizduoti modelio naudojami tono transkribavimo lygiai.

2.3. INTSINT modelis

INTSINT (**IN**ternational **T**ranscription **S**ystem for **INT**onation) modelį 1987m. pasiūlė D. Hirst [Wik06c]. Tai dar vienas populiarus žymėjimais paremtas intonacijos generavimo modelis. INTSINT toną koduoja aštuoniais simboliais:

T (aukščiausias), **H** (aukštesnis), **U** (žingsniu aukštesnis), **S** (nepakitęs), **M** (vidutinis), **D** (žingsniu žemesnis), **L** (žemesnis), **B** (žemiausias).

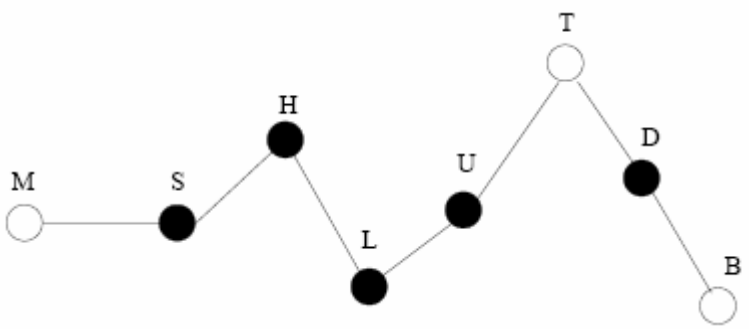
Tonų pozicionavimas kalbos fragmentuose žymimas specialiais simboliais:

[(pradinis), < (pirma pusė), **:** (vidurinis), > (antra pusė), **]** (finalinis)

Pavyzdyje pateikiama frazės „It‘s time to go“ tono transkripcija kartu su frazės fonetiniais IPA (International Phonetic Alphabet) simboliais:

M: /ɪts/ T: /taɪmtə/ D<B] /gəʊ/

Pavyzdyje esantis M: žymėjimas reiškia vidutinį toną ties fragmento „It‘s“ viduriu. Toliau seka žymėjimas T:, reiškiantis aukščiausią toną fragmento „time to“ viduryje. Tolesnis D< žymėjimas reiškia žingsniu žemesnį toną „go“ fragmento pirmoje dalyje bei B] - žemiausią toną to pačio fragmento pabaigoje.



4 pav. INTSINT modelio žymėjimai

Fonetinė INTSINT modelio interpretacija gaunama įvedant du nuo kalbėtojo (arba netgi nuo sintezuojamo fragmento) priklausomus parametrus:

- **key**: nurodo absoliutų tono tašką hercais
- **range**: dydis, nusakantis tono minimumo ir maksimumo intervalą

Tada tono segmentai gali būti išreikšti absoliutaus tono taškais pagal žemiau aprašytas išraiškas. T, M ir B žymėjimai reiškiami absoliučiais tono taškais neatsižvelgiant į prieš juos esančius tono taškus:

- T : P(i) := key + range/2
- M : P(i) := key
- B : P(i) := key – range/2

Kiti žymėjimai apibrėžiami atsižvelgiant į prieš juos esančius tono taškus:

- H : P(i) := (P(i-1) + T) / 2
- U : P(i) := (3*P(i-1) + T) / 4
- S : P(i) := P(i-1)
- D : P(i) := (3*P(i-1) + B) / 4
- L : P(i) := (P(i-1) + B) / 2

Čia P(i) – nagrinėjamas tono taškas, P(i-1) – ankstesnis tono taškas.

3. Įrankiai

Šiame skyriuje aprašomi darbe naudoti įrankiai.

3.1. MBROLA

MBROLA – tai kalbos sintezatorius, naudojantis MBR-PSOLA sintezavimo algoritmą, pagrįstas difonų konkatenacija [Mon96]. Programos įvesties duomenys – fonemų sąrašas kartu su jų prozodijos informacija (fonemos trukmė bei tono duomenys). Programos darbo rezultatas – 16-os bitų balso įrašas .wav formatu.

Tai nėra TTS (Text-To-Speech) sistema, tai reiškia, kad balsas nesintezuojamas iš neapdoroto teksto. Norint gauti pilną TTS sistemą reikalingas teksto procesorius, apdorojantis tekstą ir paverčiantis į fonetinių bei prozodinių komandų rinkinį.

Programa laisvai platinama, galima parsisiųsti iš MBROLA projekto namų svetainės: http://tcts.fpms.ac.be/synthesis .

3.1.1. Įvesties duomenų formatas

Failo pvz.pho pirmos aštuonios eilutės:

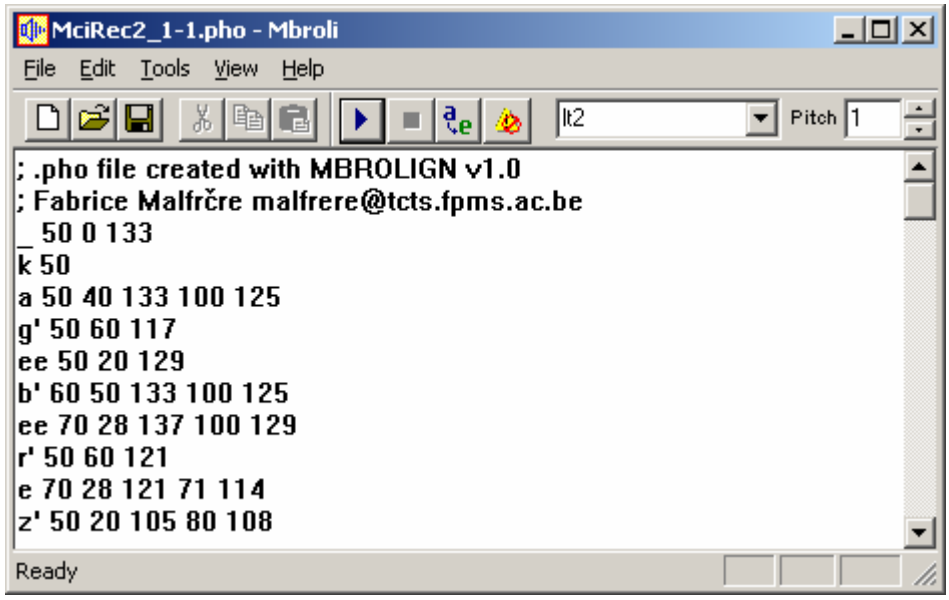
```
; Komentaras
_ 50 0 133
k 50
a 50 40 133 100 125
g' 50 60 117
ee 50 20 129
b' 60 50 133 100 125
ee 70 28 137 100 129
_ 50 0 133
```

Pavyzdyje parodytas įvesties duomenų formatas, reikalingas MBROLA sintezatoriui. Kiekviena eilutė prasideda fonemos simboliu, po jo seka fonemos trukmė (milisekundėmis), o po trukmės – tono taškai, susidedantys iš reikšmių porų: taško pozicijos fonemoje (procentais) bei tono aukščio (hercais). Taipogi, fonema gali visai neturėti tono taškų. Duomenys eilutėse skiriami tarpais. Eilutė, prasidedantis kabliataškiu, reiškia komentarą.

Antroji pavyzdžio eilutė:

```
_ 50 0 133
```

Sako sintezatoriui generuoti 50 ms trukmės tylą bei nustatyti 133Hz toną fonemos pradžioje (50ms fonemos ilgio 0% taške). Pavienių fonemų tono taškai formuoja visos frazės tiesinę tono kreivę. Tono kreivė generuojama be trūkių, t.y. generuojant frazės toną fonemos, neturinčios tono taškų, tiesiog praleidžiamos.



5 pav. Mbroli - MBROLA sintezatoriaus vartotojo sąsaja

3.1.2. Balso duomenų bazė

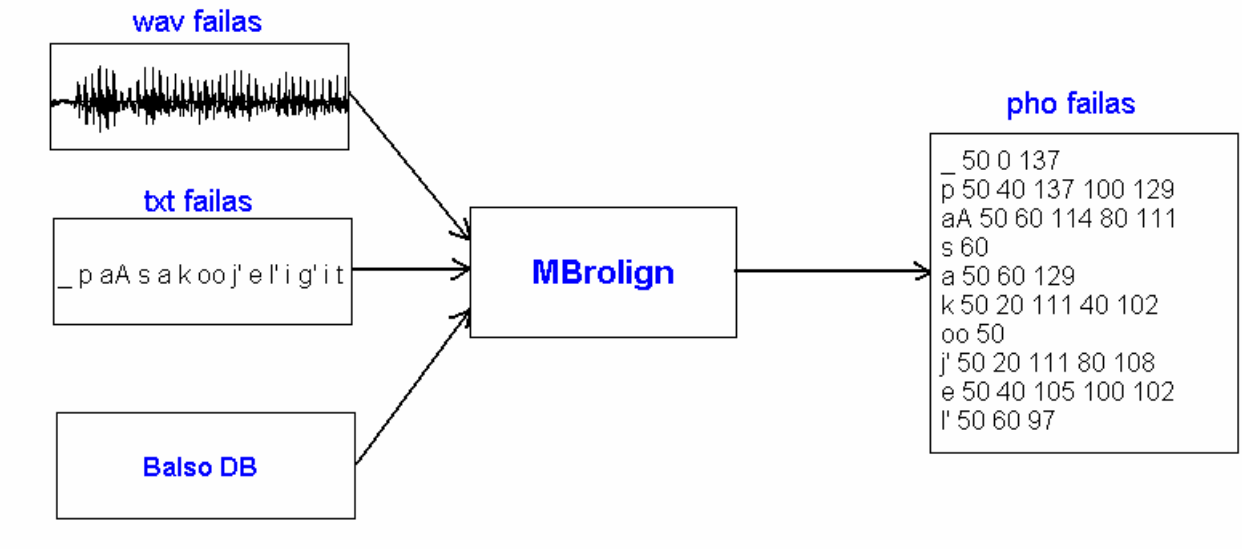
Balso sintezei iš .pho failų MBROLA programa naudoja balso duomenų bazes, sudarytas iš dvigarsių (difonų) aibės, įrašytos vieno žmogaus balsu, neutraliu tonu. Šiuo metu nemokamai

platinama 35 kalbų nemokamos MBROLA duomenų bazės. Tarp jų yra dvi lietuvių kalbos duomenų bazės *lt1* ir *lt2*. *lt1* įrašyta profesoriaus hab. Dr. A. S. Girdenio balsu, *lt2* – G. Deksnio balsu (galima parsisiųsti iš: http://tcts.fpms.ac.be/synthesis/mbrola/dba/lt2/lt2.zip). Abiejose bazėse sukaupta po 5003 difonus.

3.2. Mbrolign

Mbrolign – tai prozodijos perkėlimo įrankis, veikiantis MBROLA sintezatoriaus pagrindu [Mon99]. Programos įvesties duomenys: .wav 16kHz (16 bitų) garso failas bei .txt jo fonetinės transkripcijos failas (SAMPA abėcėlė). Programa altieka laikinį fonetinės transkripcijos ir garso signalo lygiavimą (temporal alignment) rezultate sugeneruodama .pho failą, kuris po to gali būti naudojamas MBROLA sintezatoriuje natūraliai skambančio balso išgavimui.

Prozodijos transplantacijai atlikti be minėtų įvesties duomenų taip pat būtina nurodyti MBROLA duomenų bazę, kuri bus naudojamas procedūros eigoje.



6 pav. Prozodijos išgavimas iš balso įrašo panaudojant *Mbrolign* programą

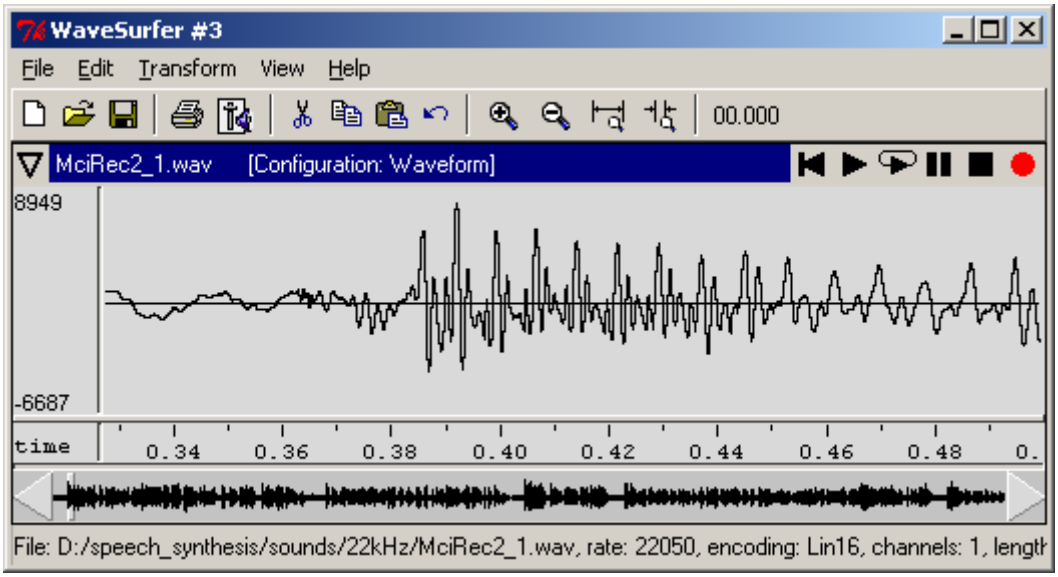
3.3. AAA real recorder

AAA real recorder – tai garsų įrašymo programa. Pagrindinė programos savybė - gebėjimas nuskaityti garsų srautą tiesiogiai iš garso kortos ir įrašyti į .wav failą. Su šia programa buvo įrašinėjamas natūralus balsas, kuris vėliau apdorotas ir rezultate gauta balso tono duomenų bazė.

3.4. WaveSurfer

WaveSurfer – tai garso vizualizavimo bei manipuliavimo programa. Programa lengvai ir intuityviai valdoma. Puikiai tinka .wav failų apdorojimui. Turi daugybę garso manipuliavimo funkcijų, viena iš daugelio – garso dažnio keitimas. Taip pat turi garso grafiko didinimo įrankį, kuriuo garsą galima padidinti iki milisekundės fragmento. Su šiuo įrankiu galima labai tiksliai karpyti garsus.

WaveSurfer programos pagalba balso įrašai suskaidyti į frazes. Taip pat su šia programa garsai konvertuoti iš 22kHz į 16kHz garsus.



7 pav. WaveSurfer – garso vizualizavimo bei manipuliavimo programa

3.5. Maple

Maple – tai plataus funkcionalumo kompiuterinė sistema pažangiai matematikai. Ji turi reikiamas funkcijas algebrai, diksrečiai matematikai bei daugybei kitų sričių. Taip pat numatytas nepamainomas grafiko piešimo funkcionalumas. Su Maple galima generuoti įvairiausių funkcijų grafikus. Vizuali funkcijų išraiška padeda greičiau ir lengviau palyginti kelių funkcijų elgesį, jų skirtumus ir panašumus.

3.6. MS Excel

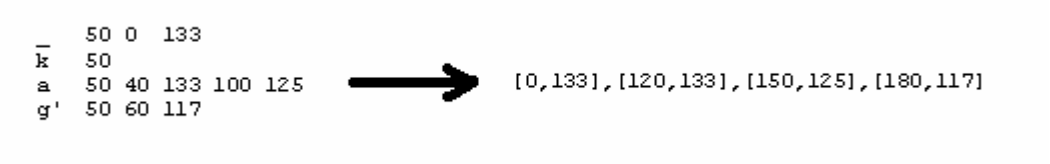
MS Excel – tai dar viena galinga skaičiavimo programa. Gardelės principu paremta darbo lapų struktūra leidžia patogiai atvaizduoti nagrinėjamus duomenis bei lengvai taikyti matematines, statistines priemones duomenims analizuoti.

3.7. Specialiai šiam darbui sukurtos JAVA programos

Viena iš tiriamojo darbo dalių buvo trūkstamų įrankių kūrimas nagrinėjamų duomenų apdorojimui, manipuliacijai bei analizei. Visi sukurti įrankiai skirti darbui su tonu transkripcinių .pho failų duomenų formate (MBROLA sintezatoriaus įvesties duomenys).

3.7.1. Frazės tono konvertavimo programa

Frazės tono konvertavimo programa konvertuoja prozodijos transkripciją į *Maple* programai suprantamą formatą. Įvesties duomenys: frazės .pho failas. Programos išvestis: frazės tono taškai *Maple* formatu. Tono konvertavimo iš .pho į Maple formatą pavyzdys:

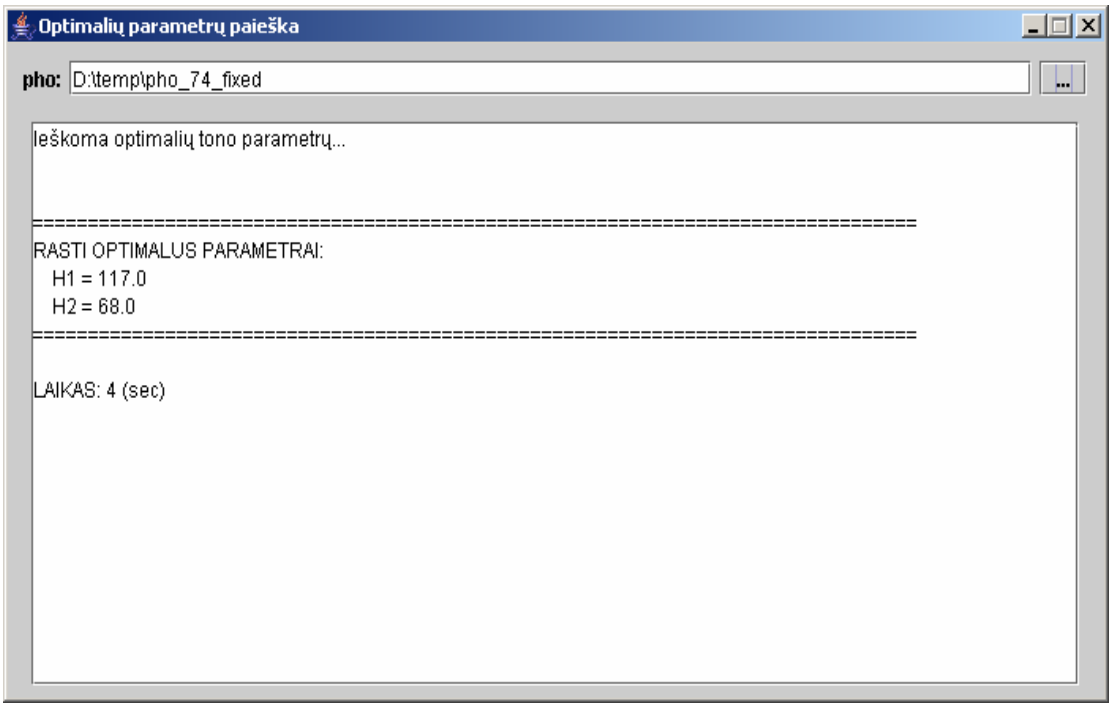


Iš pavyzdžio matyti, kad .pho formato fonemų tono procentiškai nurodytos pozicijos keičiamos į absoliučius laiko taškus.

3.7.2. Optimalaus frazės tono paieškos programa

Frazės tono paieškos programa atlieka optimalaus frazės tono paiešką. Programa remiasi idėja, kad frazės tonui vienareikšmiškai apibrėžti pakanka dviejų parametrų: tono H1 frazės pradžioje bei tono H2 frazės pabaigoje.

Programos įvesties duomenys – frazių tonų prozodijos transkripcinių .pho failų rinkinys. Vienas failas traktuojamas kaip viena frazė. Programa kiekvienai frazei randa H1 ir H2 parametrų porą. Iš gautų porų išskaičiuojama nauja – optimalių parametrų pora.

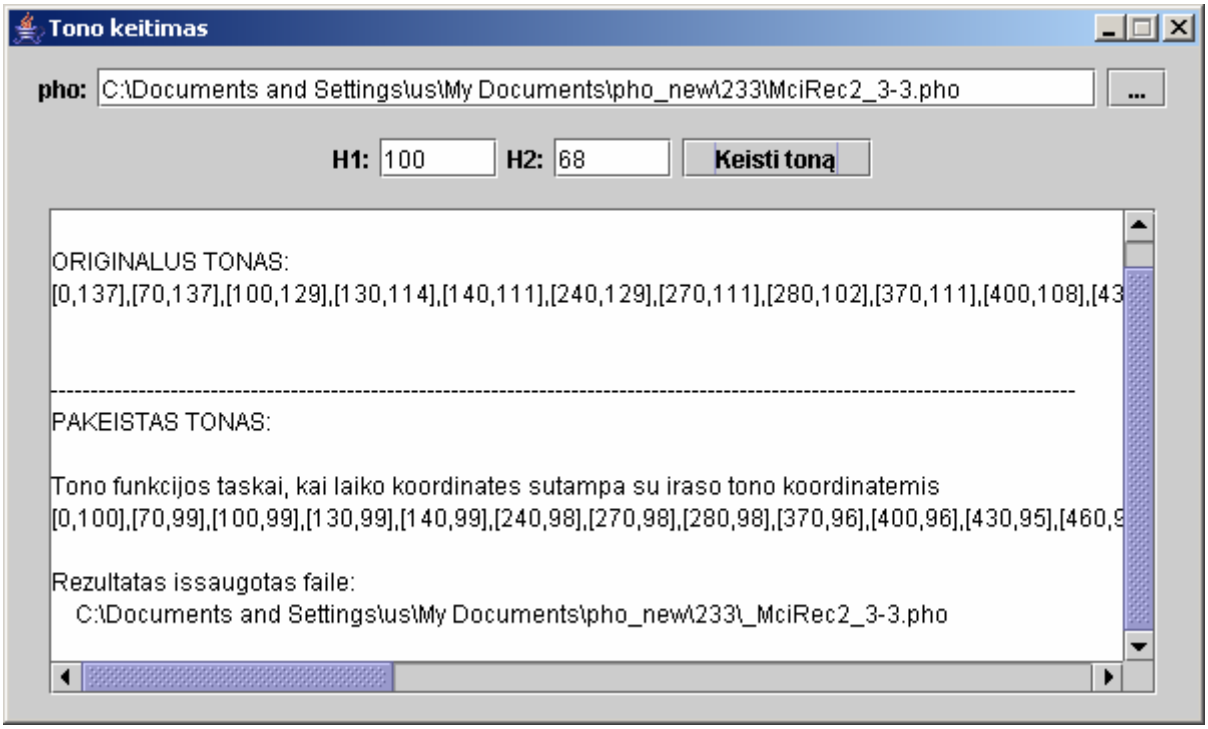


8 pav. Java programa, iš frazių tonų duomenų bazės išrenkanti (*H1*,*H2*) poras bei paskaičiuojanti optimalią (*H1*,*H2*) porą

3.7.3. Frazės tono keitimo programa

Frazės tono keitimo programa skirta pasirinktai frazei (čia žodžiu „frazė“ vadinamas frazės pagrindinio tono transkripcinis failas .pho) suteikti naują toną. Suteikiamas tonas – yra tik frazės tonas, tai nėra pagrindinis tonas, kuris susideda iš frazės tono bei kirčių tonų. Programa pasirinktos frazės toną konvertuoja į naują toną, kuris vienareikšmiškai aprašomas frazės tono pradžios

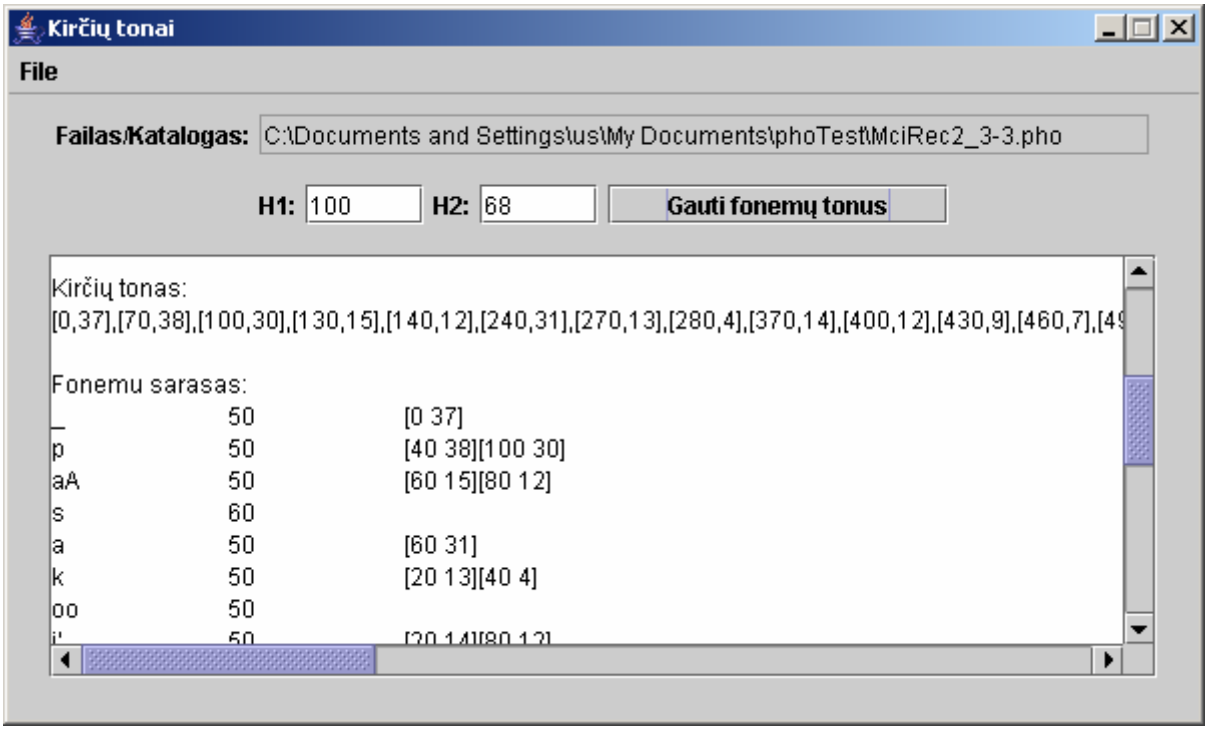
parametru (H1) bei frazės tono pabaigos parametru (H2). Gautas naujas tonas įrašomas į .pho failą tuo pačiu MBROLA formatu.



9 pav. Java programa, generuojanti frazės toną pasirinktai frazei panaudodama užduotas *H1 ir H2* reikšmes

3.7.4. Kirčių tonų išskyrimo programa

Kirčių tonų išskyrimo programa iš frazės pagrindinio tono išskiria kirčių tonus. Kirčių tonų išskyrimas atliekamas iš pagrindinio tono „atimant“ frazės toną. Gauti kirčių tonai įrašomi į naują .pho failą tuo pačiu MBROLA formatu.



10 pav. Programa, iš frazės pagrindinio tono išgaunanti kirčių tonus

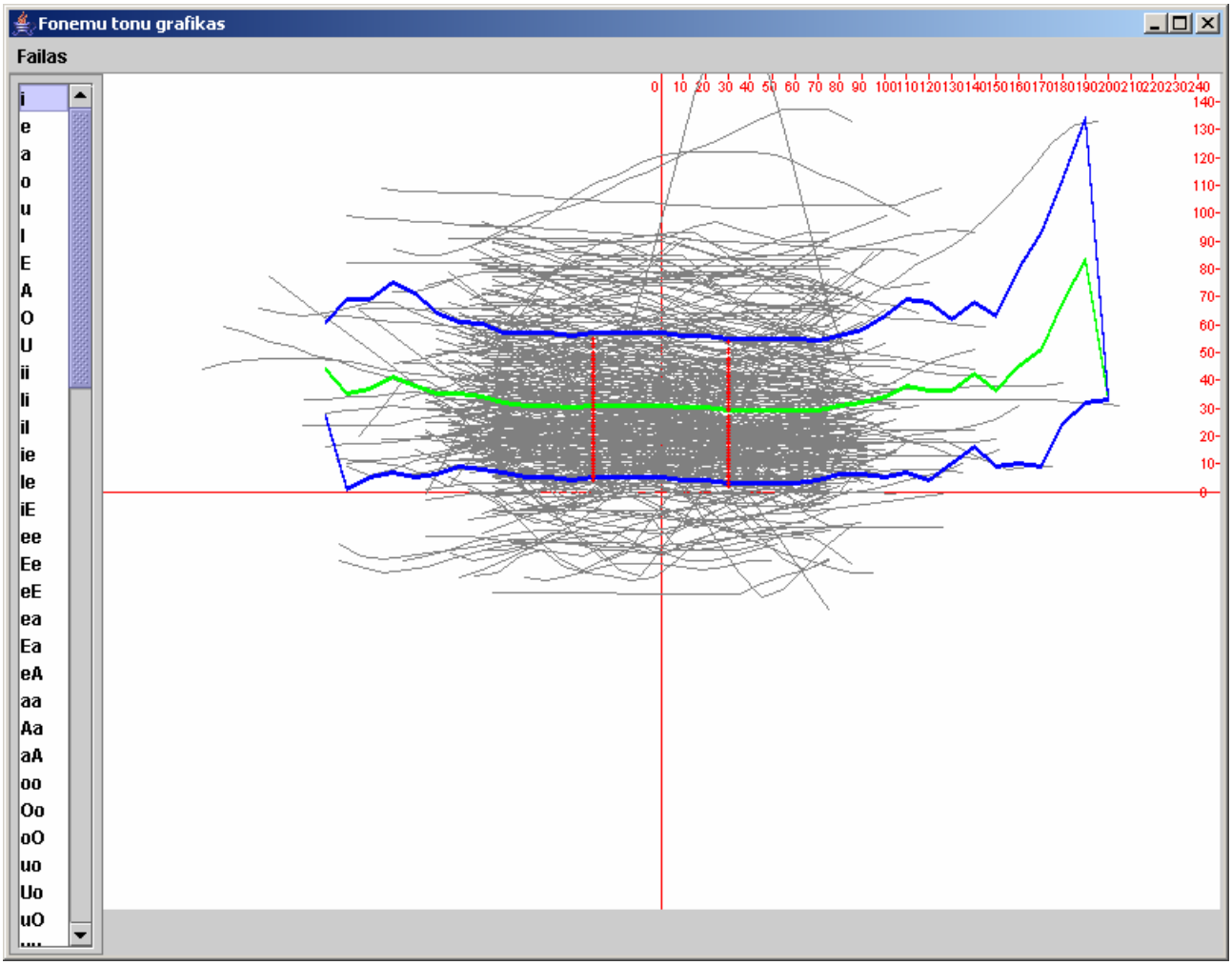
3.7.5. Fonemų tonų analizės programa

Fonemų tonų analizės programa skirta fonemų tonų vizualizavimui, analizei bei išrinkimui. Programos įvesties duomenys – tai frazių fonemų tonų failų rinkinys. Iš nurodytų failų išrenkami fonemų tonai ir sugrupuojami pagal fonemas, rezultate gaunama fonemų abėcėlė su kiekvienai fonemai priskirtu pasitaikiusių tonų sąrašu.

Užkrovus duomenis ir iš fonemų sąrašo pasirinkus fonemą parodomas jos visų pasitaikiusių tonų grafikas, taip pat pavaizduojamas tonų vidurkis bei standartinis nuokrypis nuo vidurkio.

Visos grafike vaizduojamos kreivės centruotos pagal fonemos vidurio tašką, fonemos vidurio taškas sutampa su x (laiko) ašies nuline verte.

Kiekvienos fonemos kraštinės tonų reikšmės, patenkančios į standartinio nuokrypio rėžius, išsaugomos faile. Kiekvienos fonemos duomenys išsaugomi atskirame faile. Išsaugoti duomenys naudojami nuodugnesnei tonų analizei.

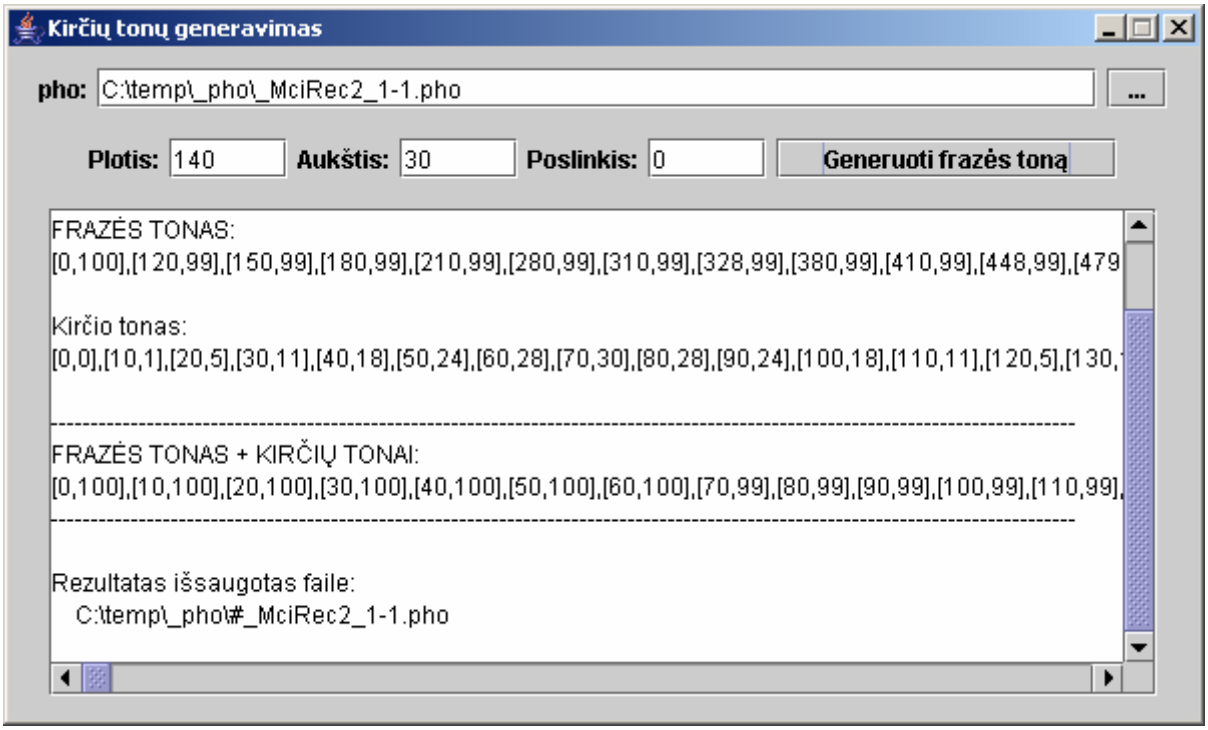


11 pav. Fonemų tonų grafikas. Iš kairės – fonemų sąrašas. Centre – iš fonemų sąrašo pasirinktos fonemos grafikas, x ašis – laikas (ms), y ašis – tono aukštis (Hz). Pilkos spalvos kreivės – užkrautose frazėse pasitaikiusios fonemos tonai. Žalia kreivė – visų fonemos tono kreivių vidurkis. Mėlynos kreivės – standartinis nuokrypis. Dvi vertikalios raudonos atkarpos (iš kairės į dešinę): fonemos pradžia, fonemos pabaiga.

3.7.6. Kirčių tonų modeliavimo programa

Kirčių tonų modeliavimo programa generuoja kirčių tonus frazės kirčiuotų skiemenų srityse. Įvesties parametrai:

- Frazės tonas (.pho failas) – tik frazės tonas (be fonemų tonų)
- Plotis – kirčio tono plotis
- Aukštis – kirčio tono aukštis
- Poslinkis – kirčio tono poslinkis nuo kirčiuotos fonemos centro. Jei poslinkis < 0 – tonas paslenkamas į kairę nuo fonemos centro, jei poslinkis > 0 – į dešinę.



12 pav. Kirčių tono modeliavimo programa

4. Balso tono duomenų bazė

4.1. Įvadas

Tam, kad sumodeliuoti natūralią žmogaus kalbą būtina turėti kalbos duomenų bazę. Šiame darbe taip pat iš pradžių buvo įrašyta natūrali lietuvių kalba tam, kad iš jos būtų sudaryta balso tono duomenų bazė. Sukaupta duomenų bazė buvo panaudota natūralaus balso tono analizei bei frazės ir kirčių tono modeliavimui.

Balso tono duomenų bazės sudarymas apima šiuos etapus:

- Balso įrašymas
- Balso įrašų redagavimas
- Balso įrašų transkribavimas
- Tono kopijavimas iš balso įrašų

4.2. Balso įrašymas

Įrašyta 16Mb balso įrašų, t.y. ~85 sakiniai (>85 frazės). Įrašinėtas LNK žinių vedėjo Gintaro Deksnio balsas. Balsas įrašytas sujungus televizorių su kompiuteriu per garso išvesties ir įvesties lizdus ir panaudojus programą AAA real recorder (žr. sk. 3.3.), kuri garso signalo srautą iš garso kortos įrašo į .wav failą.

Pasirinktas būtent G. Deksnio balsas, nes buvo reikalingas maksimalus suderinamumas su turima *Mbrolos* difonų duomenų baze, kuri buvo sukurta G. Deksnio balso pagrindu. Balsų suderinamumas reikalingas frazių tonų generavimui iš balso įrašų panaudojant MBrolign programą, plačiau apie tai aprašyta frazių tonų generavimo skyriuje.

4.3. Balso įrašų redagavimas

Balso įrašų redagavimas atliktas su garsų redagavimo įrankiu *WaveSurfer* (žr. sk. 3.4.).

Įrašyto balso failai konvertuojami į Mbrolign programai tinkamą 16kHz kokybę (garso kokybė turi neviršyti 16kHz, priešingu atveju MBrolign programa neteisingai atlieka tono kopijavimą iš įrašų).

Balso įrašai sukarpomi į frazės ilgio fragmentus. Kiekvienas naujas fragmentas įrašomas į atskirą failą. Iš visų fragmentų išrenkamos tik konstatuojamojo tono frazės.

4.4. Balso įrašų transkribavimas

Vienas iš būtinų tono kopijavimo proceso įvesties duomenų yra balso įrašo transkripcija. Įrašytų frazių transkribavimas susideda iš šių etapų:

- Išklausoma frazė ir užrašoma jos stenograma
- Stenograma sukirčiuojama (stenogramos tekstas papildomas kirčių simboliais)
- Speciali programa kirčiuotą stenogramos tekstą konvertuoja į transkripcijos tekstą.

Antrame bei trečiame etapuose man talkino P. Kasparaitis.

4.5. Tono kopijavimas iš balso įrašų

Natūralaus balso prozodijos informacijai išgauti panaudota programa *Mbrolign* (žr. sk. 3.2.), kuri iš balso įrašo ir jo stenogramos transkripcijos sugeneruoja frazės tono transkripcijos .pho failą.

Užduočiai atlikti *MBrolign* programa taip pat naudoja balso dvigarsių duomenų bazę. Šiame darbe buvo panaudota G.Deksnio balso duomenų bazė (žr. sk. 3.1.2.). Iš viso su *Mbrolign* nukopijuotas natūralus tonas iš 74 konstatuojamojo tono frazių.

Sugeneruotuose .pho failuose saugomos fonemos su jų trukmėmis bei tono aukščiu. Fonetinių transkripcijų .pho failai naudojami *Mbrola* programoje (žr. sk. 3.1.) balso sintezei. Šiame darbe *Mbrola* panaudota visų transkripcinių .pho failų sintezavimui, tiek *Mbrolign* sukurtų, tiek sugeneruotų su tono modifikavimo programomis.

Perklausius sugeneruotus transkripcinius failus pastebėta keletas *Mbrolign* programos darbo rezultatų trūkumų:

- kuo ilgesnė frazė tuo jos balsas dirbtinesnis ir tuo daugiau pasitaiko tono perkėlimo klaidų. Tono perkėlimo klaida yra tuose tono taškuose, kuriuose tonas lygus 400Hz, taip *Mbrolign* programa rezultatų faile pažymi taškus, kuriuose nepavyko atkurti tono.

- trumpiausios fonemos ilgis niekada nemažesnis nei 50ms, o iš tikrųjų natūralioje kalboje pvz. fonemos /t/ trukmė gali siekti vos 20ms!

Pastebėti trūkumai gali būti tarpusavyje susiję, t.y. dėl neteisingo fonemų trukmės nustatymo ilgėjant frazei įvyksta daugiau nusimušimų balso įrašo ir jo stenogramos transkripcijų palyginime, ko pasekoje generuojama neatpažinto tono klaida (400Hz).

5. Frazės konstatuojamojo tono modeliavimas

5.1. Įvadas

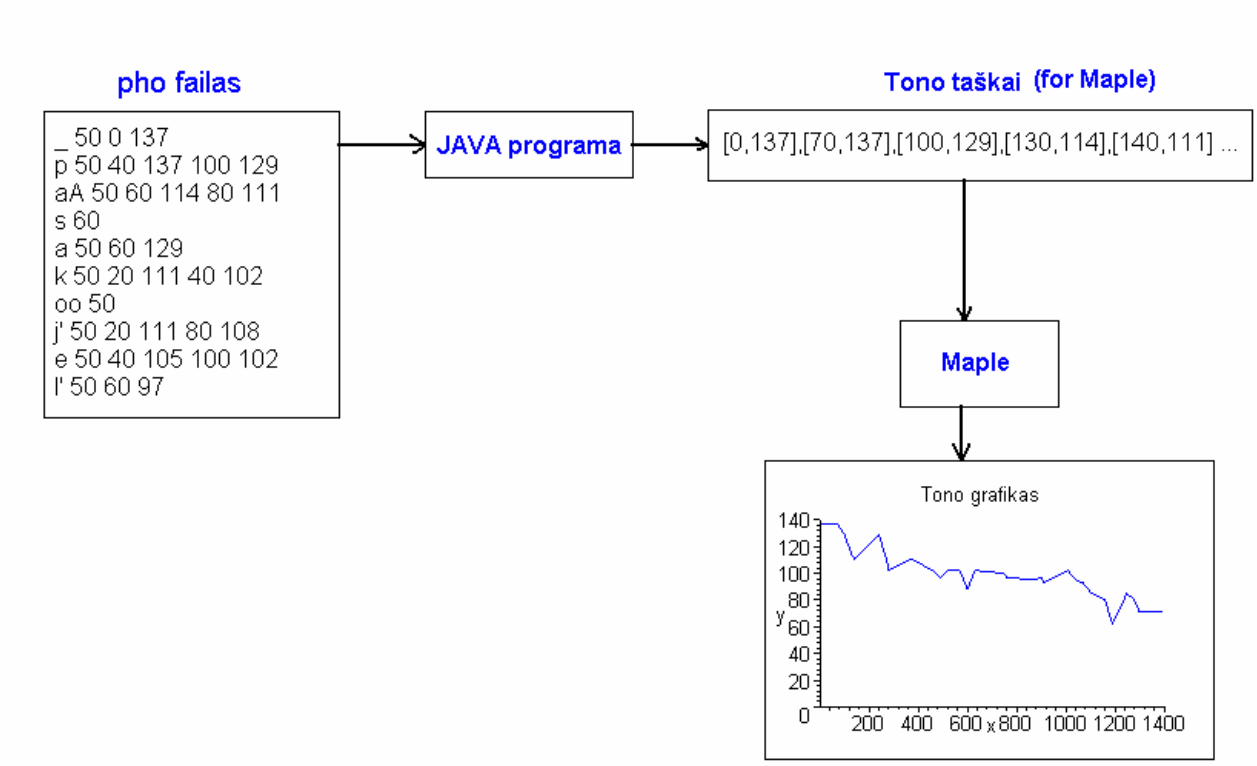
Šiame skyriuje aprašomo darbo tikslas – parinkti funkciją, kuri modeliuotų lietuvių kalbos frazės konstatuojamąjį toną. Parinktoji funkcija turėtų turėti kiek galima mažiau parametrų ir būti lengvai valdoma.

Ištyrus populiariausius balso tono modeliavimo metodus buvo nuspręsta darbe remtis superpozicinu Fujisaki modeliu (žr. sk. 2.1.). Pasirinktas modelis pasižymi paprastumu ir praktiškumu. Modelis pagrįstas idėja, kad pagrindinis tonas gali būti traktuojamas tarsi sudarytas iš dviejų nepriklausomų lygių, kur pirmas lygis reiškia frazės toną, o antras – kirčių toną. Abu lygiai sudaryti iš komandų: pirmas lygis – iš frazės komandų, antras – iš kirčio komandų. Komandos sąvoką atitinka parametrizuota funkcija. Varijuojant parametrais gaunami norimos trukmės bei amplitudės tono fragmentai.

Parinkus frazės konstatuojamojo tono modeliuojančią funkciją atliekama balso tono duomenų bazės analizė ir parenkami funkcijos optimalūs parametrai. Optimalūs parametrai, tai tokie parametrai, su kuriais funkcijos modeliuojamas tonas atitinka statistiškai dažniausiai pasitaikantį frazės toną natūralioje kalboje.

5.2. Konstatuojamąjį toną modeliuojančios funkcijos parinkimas

Iš pradžių atlikta balso tono duomenų bazės analizė. Tam tikslui sukurta frazės tono konvertavimo programa (žr. sk. 3.7.1.), kuri iš .pho formato konvertuoja tono taškus į koordinačių eilutę, kurią su *Maple* (žr. sk. 3.5.) programa galima pavaizduoti grafiškai (13 pav.).



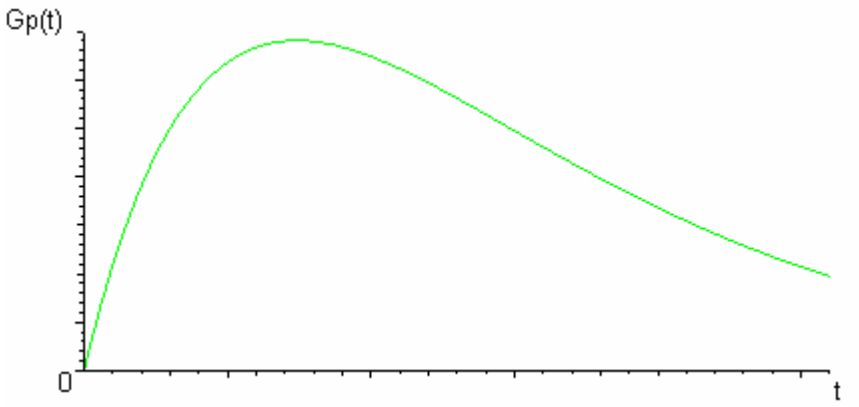
13 pav. Balso tono vizualizacijos procesas

Atsitiktinai iš balso tono parinktų frazių tonų grafikų matyti, kad visų frazių tonas frazės pradžioje būna aukščiausias ir frazės metu leidžiasi frazės pabaigoje įgydamas mažiausią reikšmę (Šis pastebėjimas vėliau pagrindžiamas statistiškai (17 pav.)).

Kaip jau buvo minėta, parenkant frazės toną modeliuojančią funkciją remtasi Fujisaki modeliu, kuriame frazės tonas modeliuojamas tokia funkcija:

$$Gp(t) = \begin{cases} \alpha^2 t e^{-\alpha t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \qquad (1)$$

kurios grafikas toks:



Fujisaki pasiūlytos frazės toną modeliuojančios funkcijos grafikas iš pradžių kyla iki aukščiausios reikšmės, o po to leidžiasi. Tačiau, kaip jau buvo minėta, ištyrus turimą balso tono duomenų bazę nustatyta, kad frazės tonas jau pačioje frazės pradžioje įgyja aukščiausią reikšmę. Tam, kad Fujisaki frazės tono funkcija atitiktų duomenis, ji buvo modifikuota:

$$f(t) = h \cdot e^{-st^2} \qquad (2)$$

, kur h – Pradinio tono aukščio parametras (Hz)

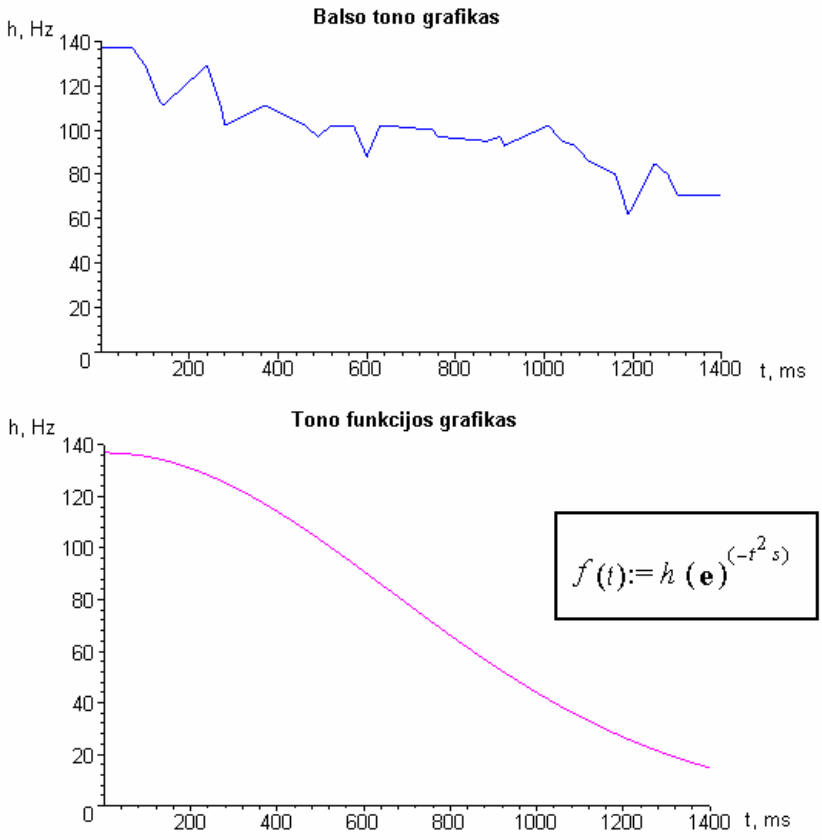
s – Tono "gesimo" greičio parametras

t – Laiko argumentas (ms)

f(t) – Tonas (Hz) laiko momentu t

$s \in R$, h, t, f $\in$ N

Modifikuotos funkcijos grafikas pateiktas 14 paveikslėlyje. Funkcijos grafikas kaip ir frazės teigiamojo tono grafikas pradžioje (kai t=0) įgyja maksimalią reikšmę, kuri frazės bėgyje leidžiasi iki minimalios reikšmės frazės pabaigoje (t=N, kur N – frazės trukmė). Funkcijos privalumas tame, kad parametrai h ir s nepriklausomi vienas nuo kito, t.y. nustačius pradinį tono aukštį (parametras *h*) galima varijuoti tono „gesimo" greitį (parametras *s*) tačiau pradinis tono aukštis visada išliks nepakitęs ir atvirkščiai.



14 pav. Balso tono ir parinktos tono funkcijos grafikai

5.3. Optimaliausių tono funkcijos parametrų parinkimas

Ieškant optimaliausių tono funkcijos parametrų buvo išmėgintos dvi strategijos:

- Balso tono ir tono funkcijos reikšmių palyginimas fiksuoto dydžio žingsniu
- Kraštinių tono funkcijos reikšmių fiksavimas

5.3.1. Balso tono ir tono funkcijos reikšmių palyginimas

Ši strategija paremta balso tono ir tono funkcijos reikšmių palyginimu pagal formulę:

$$\min_{s}\left[\sum_{t=0}^{\max T}\left|f(t)-Tonas(t)\right|\right] \qquad (3)$$

, f – pasirinktoji tono funkcija

, $Tonas$ – frazės įrašo tonas

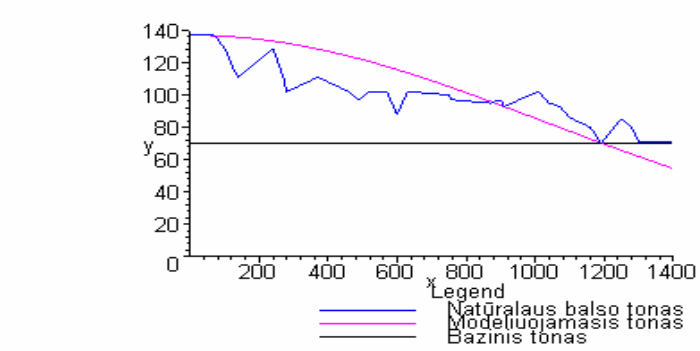
, $maxT$ – frazės ilgis

, s – optimizuojamasis f f-jos parametras

Iš pradžių parenkamas parametras h. Kadangi jis fiksuoja pradinį tono aukštį, kuris, kaip žinoma, yra aukščiausia tono reikšmė visoje frazėje, todėl tiesiog iš įrašo tono taškų išrenkama didžiausia reikšmė ir priskiriama parametrui h. Čia nekreipiama dėmesio į kirčių tonus, todėl daroma prielaida, kad aukščiausias pagrindinio tono taškas yra aukščiausias frazės tono taškas.

Optimaliausia parametro s reikšmė randama perrinkimo būdu panaudojant (3) formulę. Parametro reikšmės perrenkamos iš intervalo $[10^{-9}, 10^{-5}]$. Perrinkimas vyksta dviem etapais: 1) Intervalas pereinamas stambiu žingsniu (10^{-7}) (greitas perrinkimas) tam, kad rasti mažesnę sritį, kurioje randasi optimalioji parametro reikšmė; 2) Rasta siauresnioji sritis perrenkama smulkiu žingsniu (10^{-9}) ir randama tiksli optimalioji parametro s reikšmė. Naudojamas dviejų etapų reikšmių perrinkimas žymiai pagreitina parametro reikšmės paiešką, nes nereikia viso s reikšmių intervalo perrinkinėti smulkiu žingsniu.

Nors atrodytų, kad toks algoritmas turėtų parinkti pačius optimaliausius parametrus su kuriais modeliuojamas balso tonas turėtų skambėti labai panašiai į natūralų tačiau taip nebuvo. Pastebėta, kad šio algoritmo modeliuojamas tonas frazės pabaigoje būna šiek tiek žemesnis už lyginamojo balso toną ir dažnai būna žemiau bazinio tono (Bazinis tonas – žemiausias žmogaus balso tonas: 70Hz), ko pasekoje frazės pabaiga skamba nenatūraliai.



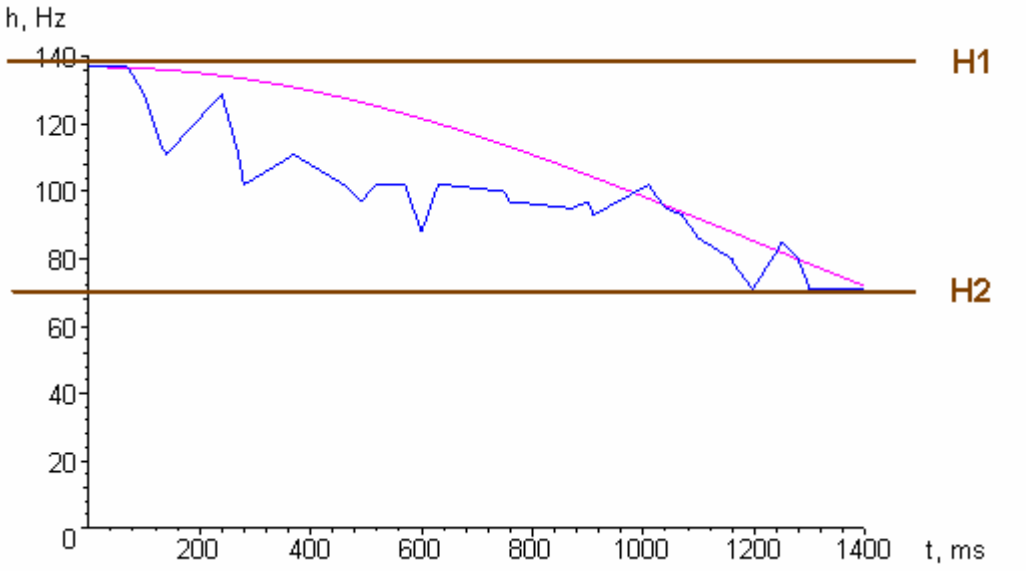
15 pav. Modeliuojamasis tonas, su parinktais optimaliais parametrais, frazės pabaigoje nusileidžia žemiau bazinio tono tokiu būdu prarasdamas natūralų skambėjimą

5.3.2. Frazės pradžios bei pabaigos tono reikšmių fiksavimas

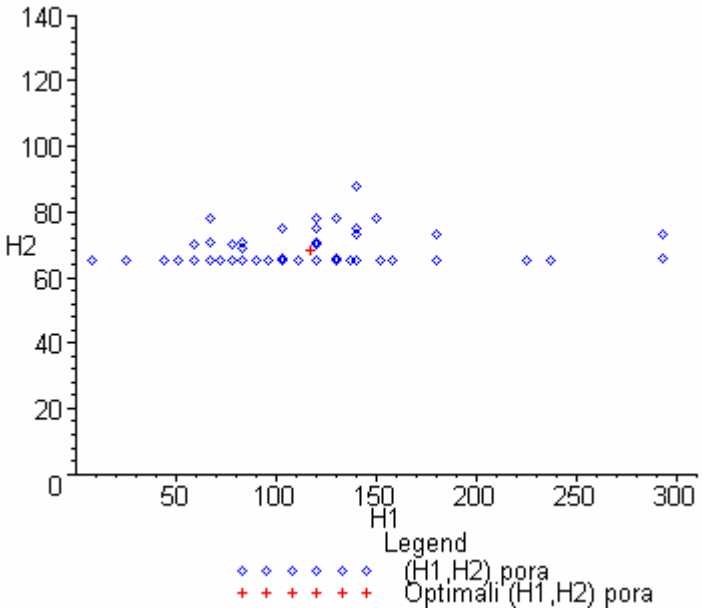
Atliekant bandymus buvo pastebėta, kad geresni rezultatai (natūralesnis balsas) gaunami tono funkcijos reikšmes frazės galuose sutapatinant su balso tono reikšmėmis (16 pav.). Šios paprastos sąlygos pakanka gerų rezultatų užtikrinimui bei sutaupoma daug skaičiavimų lyginant su pirmuoju metodu. Kadangi tono funkcijos reikšmė frazės pradžioje visuomet būna didžiausia, o frazės pabaigoje mažiausia todėl skaitome, kad tono reikšmė frazės pradžioje lygi $H1 = \max_t[Tonas(t)]$, o tono pabaigoje - $H2 = \min_t[Tonas(t)]$ ($Tonas(t)$ - natūralaus balso įrašo frazės tonas).

Iš balso įrašų tonų duomenų bazės buvo išrinktos frazių (*H1, H2*) poros su tam tikslui sukurta java programa (žr. sk. 3.7.2.). (17-ame paveikslėlyje pavaizduotas (*H1,H2*) porų pasiskirstymas). Optimali *H1* reikšmė gauta paskaičiavus visų gautų *H1* vidurkį, optimali *H2* reikšmė gauta tokiu pačiu principu. Paskaičiuota optimalių parametrų pora: (117, 68).

Žinant frazės pradinę tono reikšmę *H1* bei frazės galinę tono reikšmę *H2* nesunku paskaičiuoti tono funkcijos *f(t)* parametrus *h* bei *s*. Tam, kad parametrai būtų nepriklausomi nuo frazės ilgio skaičiuojant parametrus frazės ilgis prilyginimas 1-ui, tokiu būdu apibrėžiamas laiko argumentas: $t \in [0,1]$. Kadangi parametras *h* nusako pradinį tono aukštį tuomet jis lygus *H1* (h := H1). Parametras *s* randamas jį išsireiškus iš tono funkcijos $f(t) = h \cdot e^{-st^2}$ => $s = \ln\frac{h}{f(1)}$, kai *t=1*. Gautoje išraiškoje vietoj *h* įsistatome *H1*, o vietoj *f(1)* – *H2*.



16 pav. Fiksuojamas tono funkcijos *f(t)* (raudona kreivė) leistinų reikšmių intervalas [*H2, H1*]



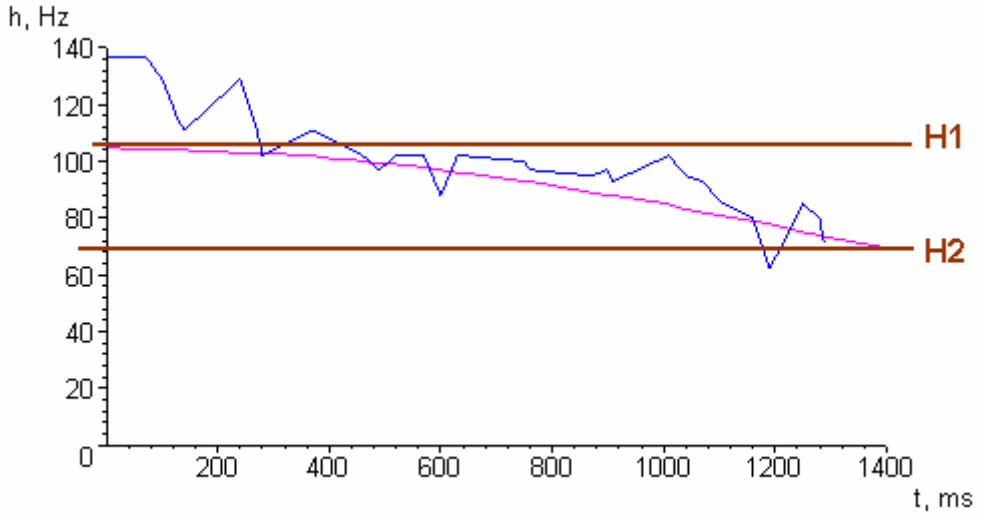
17 pav. (H1, H2) porų pasiskirstymas. H1 – tono aukštis (Hz) frazės pradžioje, H2 – tono aukštis (Hz) frazės pabaigoje. Optimali parametrų pora: (117, 68)

5.4. Išvados

Nors bandomieji įrašai, gauti pritaikius frazės toną su optimaliais parametrais, skamba pakankamai natūraliai, tačiau rezultatai ne visai atitinka natūralaus balso toną, nes be abejo dar trūksta kirčių tonų, be to, gautas frazės tonas yra ne visai korektiškas. Nekorektiškas todėl, kad pagal Fujisaki modelį pagrindinis tonas sudarytas iš frazės tono bei virš jo esančių kirčių tonų, taigi

galima teigti, kad aukščiausi pagrindinio tono taškai priklauso ne frazės tonui, o kirčių tonams. Tuo tarpu, dėl žinių stokos apie kirčių tonų charakteristikas šiame skyriuje jie nebuvo išskiriami iš pagrindinio tono. Todėl frazės tono optimalūs parametrai paskaičiuoti neatsižvelgiant į kirčių tonus. Grįžtant prie Fujisaki modelio, ideologiškai teisingesnis frazės tonas turėtų atrodyti maždaug taip kaip pavaizduota 18-ame paveikslėlyje, čia matyti, kad virš frazės tono (raudona kreivė) yra aukštesnių pagrindinio tono taškų (mėlyna kreivė), kurie signalizuoja apie kirčius. Iš paveikslėlio matyti, kad pakoreguotas tik frazės tono pradžios taškas (H1 parametras), tuo tarpu tonas frazės pabaigoje (H2 parametras) visiškai tenkina sąlygas.

Kirčių tonų modeliavimas apžvelgiamas kitame skyriuje.



18 pav. Prielaida, kaip iš tikrųjų turėtų atrodyti frazės tono grafikas (raudona kreivė), t.y. frazės tonas turėtų būti „po“ kirčių tonu. Numanomi kirčių tonai – pagrindino tono (mėlyna kreivė) „kalneliai“ virš frazės tono. Čia padaryta prielaida, kad optimalūs parametrai (H1,H2) = (103, 68).

6. Kirčių tono modeliavimas

6.1. Įvadas

Šiame skyriuje aprašomo darbo tikslas – parinkti funkciją, kuri modeliuotų lietuvių kalbos kirčių toną. Parinktoji funkcija turėtų turėti kiek galima mažiau parametrų ir būti lengvai valdoma.

Kirčių tono modeliavimas, kaip ir frazės tono modeliavimas (ankstesnis skyrius) remiasi Fujisaki modeliu (žr. sk. 2.1.). Pasirinktas modelis pasižymi paprastumu ir praktiškumu. Modelis pagrįstas idėja, kad pagrindinis tonas gali būti traktuojamas tarsi sudarytas iš dviejų nepriklausomų lygių, kur pirmas lygis reiškia frazės toną, o antras – kirčių toną. Abu lygiai sudaryti iš komandų: pirmas lygis – iš frazės komandų, antras – iš kirčio komandų. Komandos sąvoką atitinka parametrizuota funkcija. Varijuojant parametrais gaunami norimo dažnio bei trukmės tono fragmentai.

Parinkus kirčių tono modeliuojančią funkciją atliekama balso tono duomenų bazės tyrimas bei pateikiamos tyrimų išvados.

6.2. Kirčių tonų išskyrimas

Kaip jau buvo minėta, šiame skyriuje remiamasi Fujisaki pagrindinio tono modeliu. Modelyje daroma prielaida, kad pagrindinis tonas sudarytas iš frazių tono bei kirčių tono.

Pasinaudojus minėta prielaida kirčių tonas iš pagrindinio tono išskirtas kreivių skirtumo principu. Tai yra, žinant pagrindinį toną bei frazės toną, įmanoma kreivių skirtumo principu išskaičiuoti tą pagrindinio tono dalį, kuri nepriklauso frazės tonui arba kitaip sakant yra „virš“ frazės tono. Kirčių tonas išskaičiuojamas pagal formulę:

$$\mathbf{F_{fo}(x) = F_a(x) - F_{fr}(x)}$$

, kur $x \in [0, \mathit{FrazėsIlgis}]$

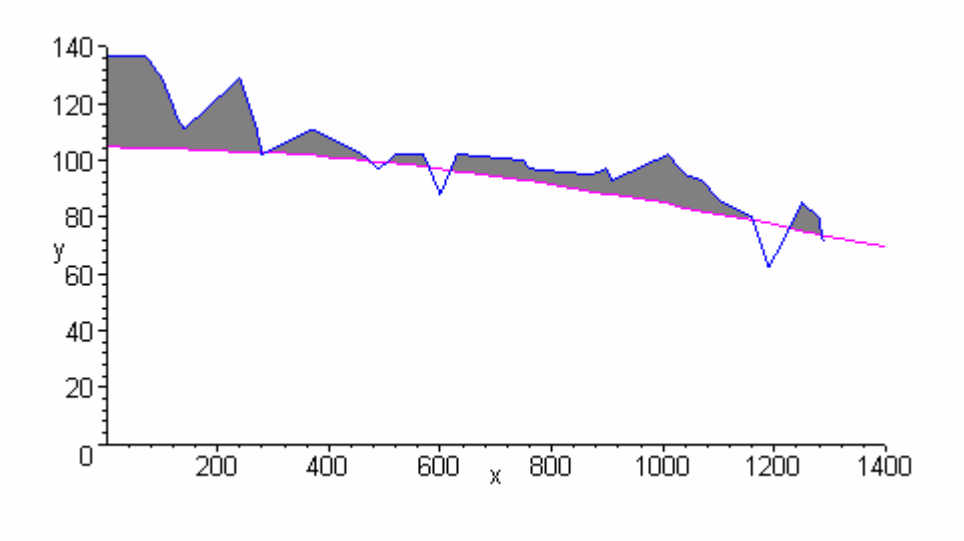
F_{fo} – kirčių tonas, F_a – pagrindinis tonas, F_{fr} – frazės tonas

Tam tikslui sukurta programa (žr. sk. 3.7.4.), išskirianti fonemų toną minėtu principu. Programos darbo rezultato vizualizcija pateikta 19 paveikslėlyje. Fonemų tonas iš pagrindinio išskiriamas nurodžius šiuos parametrus:

- Pagrindinis tonas – nurodomas kelias iki frazės pagrindinio tono transkripcinio failo
- Frazės tonas – nurodomi du dydžiai: H1 - frazės tono aukštis frazės pradžioje bei H2 - frazės tono aukštis frazės pabaigoje. Frazės tono modeliavimas aprašytas ankstesniame skyriuje.

Kirčio tono išskyrimo iš pagrindinio tono bandymai atlikti taikant įvairaus aukščio frazės toną. Iš pradžių bandymai atlikti su praeitame skyriuje parinktu optimaliu frazės tonu, kuris vienareikšmiškai aprašomas parametrų pora (H1=117, H2=68). Reikėtų paminėti, kad optimalus frazės tonas apskaičiuotas neatsižvelgiant į kirčių tonus, tai reiškia, kad optimalios frazės pradžios tono parametras yra per aukštas, tačiau kiek per aukštas įmanoma sužinoti tik eksperimentinių bandymų būdu. Tuo tarpu, parametras H2 yra apytiksliai lygus baziniam balso tonui. Baziniu balso tonu vadinamas žemiausias natūralaus balso dažnis ir yra lygus 70Hz. Atlikus eksperimentinius bandymus mažinant H2 parametro reikšmę žemiau bazinio balso tono buvo patvirtintas teiginys, jog tonas, žemesnis nei 70Hz praranda natūralų skambesį. Taigi, galima teigti, kad H2 parametras iš tiesų parinktas optimaliai ir jo keisti nėra prasmės.

Taigi, kirčių tonų išskyrimo bandymai buvo atliekami varijuojant tik H1 parametro reikšmę. Atlikti tyrimai parodė, kad geriausi rezultatai gaunami kuomet H1 parametro reikšmė apytiksliai lygi 102 hercams. Detalesnis kirčių tono modeliavimas aprašytas tolesniuose poskyriuose.



19 pav. Užtušuotoji sritis tarp pagrindinio tono (viršutinė kreivė) bei frazės tono (apatinė kreivė) žymi kirčių toną.

6.3. Kirčių toną modeliuojanti funkcija

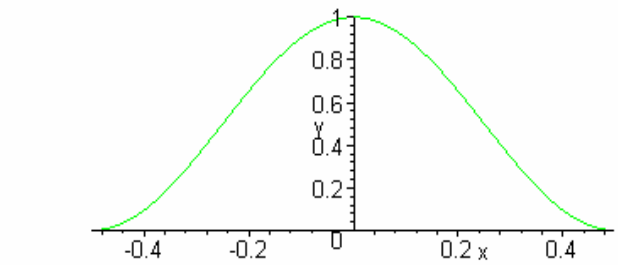
Pagal Fujisaki modelį, kirčių toną modeliuojanti funkcija turi tokią išraišką:

$$Ga(t) = \begin{cases} 1+\cos(\pi t),\ t \in [-1,1] \\ 0, \qquad\quad t < -1, t > 1 \end{cases}$$

Darbe panaudota šiek tiek modifikuota Fujisaki funkcija:

$$f(t) = \frac{h}{2}(1+\cos(t\pi\frac{2}{w}))\ ,\ t \in [-0.5\cdot w, 0.5\cdot w],\ \ f \in [0,h]$$

, kur h - tono aukštis (Hz), w - tono plotis (ms). Taip pat atlikti funkcijos pataisymai, kad šios aukštis bei plotis būtų lygūs vienetui, kuomet h=1 ir w=1. Funkcijos grafikas pateiktas 20 pav.

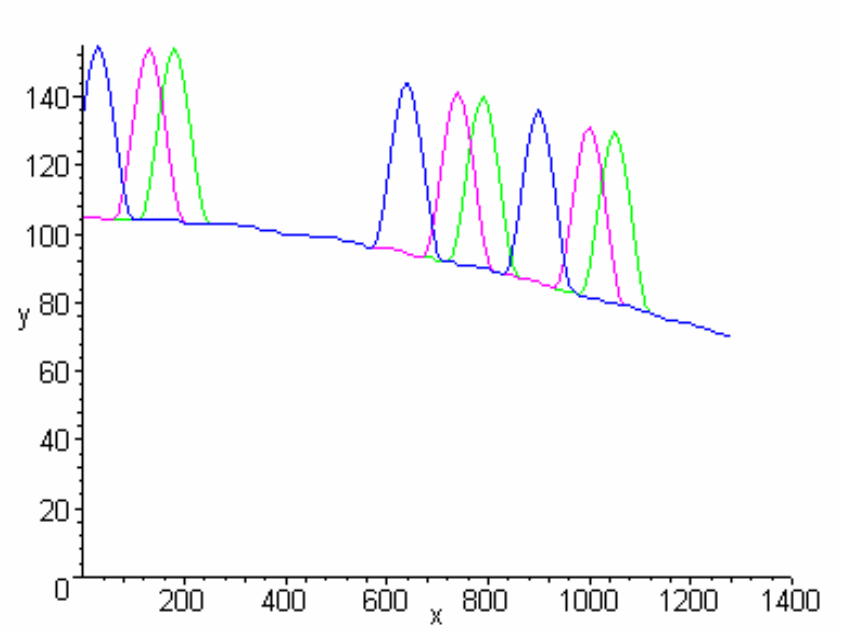


20 pav. Kirčių toną modeliuojančios funkcijos grafikas

6.4. Kirčių tono modeliavimas

Kirčių tonų modeliavimui sukurta programa (žr. sk. 3.7.6.) prie frazės tono kirčiuotų fonemų srityse pridedanti sugeneruotus kirčio tonus. Programai nurodomas .pho failas su frazės tonu bei generuojamo kirčio tono aukštis, plotis bei postūmis nuo kirčiuotos fonemos centro. Čia postūmis

nurodo kiek milisekundžių pastumti kirčio toną į kairę (su neigiamu ženklu) arba į dešinę. Programa automatiškai suranda kirčiuotas fonemas bei pagal nurodytus parametrus prideda kirčio toną prie frazės tono. Gautas rezultatas išsaugomas į failą .pho formatu bei automatiškai su Mbrola susintezuojamas balsas perklausymui.



21 pav. Pavaizduoti trys (skirtingų spalvų) kirčių tonų modeliavimo bandymai kiekviename bandyme keičiant kirčio tono poslinkį nuo kirčiuotos fonemos centro

6.5. Kirčių modeliavimo rezultatų analizė

Atlikus gautų duomenų analizę pastebėta, kad dažniausiai kirčiuotų skiemenų srityse tonas padidėja vidutiniškai 20 – 40 Hz.

Atlikus kirčiuotų skiemenų tonų modeliavimo bandymus, verta paminėti keletą pastebėjimų:

- Skiemenys su ilgomis kirčiuotomis balsėmis, einančiomis po sprogstamosios priebalsės skamba geriau kuomet kirčio tonas paslenkamas link sprogstamosios priebalsės
- Skiemenys su ilgais kirčiuotais garsais turi aukštesnį toną negu tie, kuriuose kirčiuoti garsai yra trumpi
- Darbo metu gautoje kirčių tonų duomenų bazėje pastebėtos padidinto tono sritys apimančios ne vieną skiemenį, o visą žodį. Šiose srityse tonas palyginti su kirčiuotais skiemenimis yra neaukštas. Buvo atliktas bandymas, kurio metu tonas buvo paaukštintas ties atsitiktiniu žodžiu, maksimalus aukštis siekė 50Hz (Bandymo metu specialiai tonas buvo pasirinktas didesnis nei įprasta tam, kad būtų lengviau akustiškai bei grafiškai išskirti). Tokio bandymo metu pastebėtas žodžio pabrėžimas, išskyrimas

iš konteksto, o svarbiausia, toks pabrėžimas skamba pakankamai natūraliai. Manau labiau pasigilinus šioje srityje galima būtų tikėtis apčiuopiamų rezultatų.

6.6. Išvados

Kirčių kaip ir frazės tonui modeliuoti remtasi Fujisako balso tono modeliu (žr. sk. 2.1.). Kirčių tono modeliavime panaudotas optimalus tonas, paskaičiuotas 5-ame skyriuje. Atlikti bandymai parodė, kad optimalus tonas (aprašomas parametrais H1=117, H2=68) per aukštas frazės pradžioje. Tai nustatyta sumodeliuotą pagrindinį toną (frazės tonas + kirčių tonas) akustiškai lyginant su originalia frazės intonacija. Pastebėta, kad frazės tonas pradžioje aukštesnis nei originale, o kirčiai mažiau išsiskiriantys. Tada, frazės tonas buvo žeminamas, o kirčių tonai aukštinami sumoje išlaikant tą patį aukštį. Eksperimentai buvo tęsiami tol, kol tarp originalios ir sumodeliuotos intonacijų buvo juntamas mažiausias skirtumas. Rezultate gautas optimalus frazės tonas: H1=103, H2=68. Kirčių tono modeliavimo bandymai parodė, kad kirčių tono dažnis svyruoja 20-40Hz. Ryškiausios pastebėtos kirčio tono tendencijos:

- tonas aukštesnis ties ilgomis kirčiuotomis balsėmis;
- kirčiuotos ilgos balsės, einančios po sprogstamosios priebalsės, tonas pasislinkęs link priebalsės.

7. Fonemų tono tyrimas

7.1. Įvadas

Šiame skyriuje aprašomas atskirų fonemų tonų tyrimas. Tyrimo tikslas – sužinoti atskirų fonemų tono tendencijas. Tam reikia atlikti nuodugnią surinktų fonemų tonų analizę. Būtina išanalizuoti kaip kinta tonas ties kiekviena iš fonemų, koks tono aukštis fonemos pradžioje bei fonemos pabaigoje, taip pat reikia sužinoti vidutinį fonemos toną, bei nustatyti ar skirtingų fonemų tonai skiriasi, t.y. ar fonemos turi specifinius tonus.

Užsibrėžtam tikslui pasiekti buvo atlikti šie darbai:

- Išskirti fonemų tonai iš sukauptos balso tono duomenų bazės;
- Tonai sugrupuoti pagal fonemas;
- Fonemų tonų grupėms atlikta statistinė analizė.

7.2. Fonemų tonų išskyrimas

Fonemų tonas išskirtas iš sukauptos natūralios balso tono duomenų bazės (žr. sk. 4). Tonas išskirtas funkcijų skirtumo principu, aprašytu 6-ame skyriuje. Tono atskyrimo procedūroje panaudotas frazės optimalus tonas, paskaičiuotas 5-ame bei patikslintas 6-ame skyriuose.

7.3. Fonemų tonų grafinis atvaizdavimas bei paruošimas statistinei analizei

Fonemų tonų duomenų išgavimui buvo sukurta programa grafiškai vaizduojanti kiekvienos fonemos toną (žr. sk. 3.7.5.). Programa leidžia užkrauti po vieną frazę (t.y. po vieną frazės transkripcinį failą) ir peržiūrėti kiekvienos fonemos visus pasitaikiusius tonus toje frazėje, arba galima nurodyti katalogą ir tokiu būdu užkrauti visų jame esančių frazių fonemų tonus. Pastarasis metodas buvo panaudotas atliekant sukauptos balso tono bazės fonemų tonų analizę. Tam, kad būtų galima tiksliau vizualiai įvertinti fonemų tonų kreivių charakteristikas, vaizduojamas ne vienos nagrinėjamosios fonemos tonas, o tonas sudarytas iš trijų fonemų tonų – tai yra centre yra nagrinėjamosios fonemos tonas, o iš kairės ir iš dešinės po vieną kaimyninės fonemos toną. Taip buvo būtina padaryti ir dėl kitos priežasties – tos pačios fonemos trukmė įvairiose frazėse gali skirtis (pvz.: fonemos *aA* trukmė įvairiose frazėse svyruoja nuo 50 iki 120 Hz), todėl fonemų tonas turi būti pakankamai ilgas. Programos grafike fonemų tonas vaizduojamas centruotas pagal x ašį taip, kad kiekvienos fonemos vidurio taško tonas būtų ties x=0. Tai reikalinga tam, kad būtų galima įvertinti fonemos tonų kreivių elgesį ties lygiaverčiais taškais.

Užkrovus visus fonemų tonų duomenis programa paskaičiuoja tonų kreivių vidurkį ir pavaizduoja grafike (žalia kreivė) (11 pav.). Taip pat paskaičiuojamas ir pavaizduojamas standartinis tono nukrypimas nuo vidurkio (mėlynos kreivės). Standartinis nuokrypis (angl.: Standard Deviation) [Wik06b] toliau žymimas SD, randamas pagal formulę:

$$\partial_j = \sqrt{\frac{\sum_{i=1}^{N} (f_i^{(j)} - f_j^{-})^2}{N}}$$

, kur j $\in$ [0, *TonoIlgis*], $f_i^{(j)}$ - i-asis fonemos tonas j-ajame taške, f_j^{-} - fonemos tonų vidurkio kreivė j-ame taške,

N – fonemos tonų kreivių skaičius

SD kreivės gaunamos prie vidutinio fonemos tono pridėjus/atėmus SD reikšmes: $f^{-} + \partial$ ir $f^{-} - \partial$. Standartinis nuokrypis parodo vidutinį skirtumą tarp aibės elementų. Jei visi elementai panašūs (artimi vidurkiui) tuomet SD artimas nuliui. Jei elementai smarkiai skiriasi tuomet SD didelis (tolsta nuo nulio). SD visada tik teigiamas bei visada matuojamas tais pačiais dydžiais, kokiais matuojami tiriamieji duomenys. Šiuo atveju SD, kaip ir tonas, matuojamas hercais.

Fonemų tonų grafinis vaizdas yra vertingas - jis leidžia geriau įsivaizduoti kaip tonai pasiskirstę laiko ir dažnio erdvėje. Tačiau kuomet fonemai tenka dešimtys tonų ir jų vizualus pasiskirstymas atrodo chaotiškai, tuomet vien tik iš grafiko sunku pastebėti bendras tendencijas ir tokiu atveju praverčia duomenų statistinė analizė. Tam tikslui, fonemų tonų grafiką vaizduojanti programa lygiagrečiai su atvaizdavimu, atlieka vaizduojamų tonų duomenų įrašymą į failus, kurie po to gali

būti analizuojami statistinės analizės priemonėmis. Į failus įrašoma fonemų pradžios bei pabaigos tonų reikšmės. Tonų reikšmės atrenkamos vadovaujantis trimis taisyklėmis:

- fonemos pradžios tonų reikšmės $f^{(j)}(x_1)$ turi tenkinti sąlygą

  $(f^{-}(x_1) - \partial) \le f^{(j)}(x_1) \le (f^{-}(x_1) + \partial)$,

  čia f^{-} - vidutinis fonemos tonų tonas, ∂ - Standartinis nukrypimas , *j* – tono indeksas fonemos tonų aibėje
- fonemos pabaigos tonų reikšmės $f^{(j)}(x_2)$ turi tenkinti sąlygą

  $(f^{-}(x_2) - \partial) \le f^{(j)}(x_2) \le (f^{-}(x_2) + \partial)$
- Į pirmą aibę patekusios reikšmės būtinai turi sudaryti porą su viena ir tik viena reikšme iš antros aibės. Porą sudarančios reikšmės privalo priklausyti tai pačiai tono kreivei!

7.4. Fonemų tonų statistinė analizė

Fonemų tonų statistinė analizė buvo atlikta su MS Excel programa. Iš pradžių, kiekvienai fonemai buvo sukurta po naują lapą (angl.: *sheet*) ir užpildyta duomenimis, gautais iš fonemų tonų analizės programos (žr. sk. 3.7.5.).

Šio darbo kalbos sintezėje naudota balso duomenų bazė iš 91 fonemos. Apdorojant fonemų tonus dėl duomenų trūkumo atmesta 11 fonemų: /o/, /O/, /p/, /ts/, /dz/, /dz‘/, /tS/, /x/, /x‘/, /h‘/, /M‘/. Likusių 80 fonemų kraštiniai tonai buvo užkrauti į MS Excel programą (23 pav.). Kiekvienos fonemos duomenims atliktas duomenų patikimumo testas panaudojant MS Excel statistinę funkciją TTEST, kuri žinoma *Student‘s t-Test* vardu.

7.4.1. Student‘s t-Test metodas

Student‘s t-Test funkcija grąžina tikimybę ar du mėginiai priklauso dviems aibėms, kurios turi tą patį vidurkį [Tro02]. Tai statistinis testas, skirtas nustatyti ar dvi grupės žymiai skiriasi remiantis jų vidurkiais. T-Test funkcija atsako į klausimą „Ar rezultatai statistiškai patikimi?“. T.y. šio fonemų tonų tyrimo atveju patikrinama ar gautas fonemos tonų vidurkis patikimas, ar tai nėra atsitiktinumas. Kritinė t-Test funkcijos reikšmė yra p=0.05. Rezultatai, kurių tikimybė $p \le 0.05$ laikomi statistiškai patikimais, tai reiškia $\ge 95\%$ tikimybę, kad dviejų lyginamų aibių vidurkiai yra skirtingi. T-Test funkcija turi kelias atmainas, viena iš jų skirta nepriklausomų aibių palyginimui, o kita - poromis priklausomų elementų aibėms. Fonemų tonų atveju panaudota pastaroji, nes fonemos tono pradžios ir pabaigos taškai turi būti poruojami, kadangi jie reiškia tą pačią tono kreivę ir mums svarbu nustatyti kaip kraštuose elgiasi kreivė, o ne kraštinių taškų aibė. Dar vienas t-Test funkcijos parametras nusako ar tiriamieji duomenys prognozuojami esantys kryptiniai ar nekryptiniai. Kryptiniai duomenys – kuomet vienos iš tiriamų aibių vidurkis prognozuojamas didesnis už kitos

aibės vidurkį. Jei hipotezė kryptinė, naudojamas *one-tailed t-Test*, jei nekryptinė – *two-tailed t-Test*. Fonemų tonų tyrimo atveju nebuvo tikslo įrodyti, jog duomenys yra kryptingi, todėl kryptingumo parametras nustatytas į vadinamą *two-tailed* reikšmę.

Fonemų tonų elgesiui tirti panaudotas t-Test metodas, nes tai populiariausias metodas, aibių skirtingumui nustatyti.

7.5. Rezultatų analizė

Atlikus t-Test testą iš 80 tirtų fonemų, statistiškai patikimi pasirodė esantys tik 31 fonemų duomenys, t.y. 39% tirtų fonemų. Iš viso 32 balsės (40% tirtų fonemų) ir 48 priebalsės (60% tirtų fonemų). Statistiškai patikimos pasirodė esančios 53% balsių (17 iš 32), bei 33% priebalsių (16 iš 48). 22 paveikslėlyje pateiktos fonemos su testo rezultatais. Fonemos sugrupuotos į septynias fonetines grupes, kiekviena fonemų grupė pateikta lentelės pavidalu.

Pirma grupė: Trumpos nekirčiuotos balsės. Šioje vienintelėje grupėje iš visų septynių grupių pastebima visai grupei būdinga vieninga tendencija – tono mažėjimo tendencija (tamsiai mėlynos spalvos fonemos). Taip pat, visos tyrime dalyvaujančios fonemos yra statistiškai patikimos (t-Test reikšmė ≤ 0.05 (žalios spalvos)). Šioje grupėje dėl duomenų trūkumo praleista tik viena /o/ fonema. Taigi, tyrime dalyvauja 4 iš 5 (80%) trumpos nekirčiuotos balsės ir visos jos su ta pačia - tono žemėjimo tendencija!

Antra grupė: Trumpos kirčiuotos balsės. Šioje grupėje mažiau ryški, bet visgi pastebima tono aukštėjimo tendencija (šviesiai mėlyna spalva). Tyrime dalyvauja 4 iš 5 fonemų, iš kurių 3 yra statistiškai patikimos, taigi iš viso 60% fonemų turi aukštėjančio tono tendenciją.

Trečia grupė: Ilgos nekirčiuotos balsės. Matyti, kad tyrime dalyvauja visos galimos ilgos nekirčiuotos balsės (8 iš 8). Statistinis patikimumas šios grupės taip pat pakankamai aukštas – 75%. Statistiškai patikimai rezultatai teigia, kad 4 iš 8 (50%) tonų turi tendenciją žemėti ir 2 iš 8 (25%) - aukštėti.

Ketvirta grupė: Ilgos balsės su tvirtaprade priegaide. Šios grupės statistinis patikimumas nepakankamas – tik 38%, todėl nėra pagrindo teigti apie kokią nors tendenciją.

Penkta grupė: Ilgos balsės su tvirtagale priegaide. Statistiškai patikima tik 1 iš 8 (13%) fonemų – tendencijos nėra.

Šešta grupė: Kirčiuotos priebalsės. 6 iš 16 (38%) fonemos statistiškai patikimos. Nepakankama patikimų rezultatų dalis.

Septinta grupė: Nekirčiuotos priebalsės. 9 fonemų tonai statistiškai patikimi iš 32, tai sudaro 28%. Bendros tendencijos taip pat nėra.

7.6. Išvados

Tyrimas parodė, kad statistiškai patikimi rezultatai gauti tik 31-ai fonemai. Iš viso tirta 80 fonemų, taigi bendras duomenų patikimumas yra 39% - mažiau nei pusė visų fonemų tonai yra statistiškai patikimi. Statistiškai patikimesni balsių tonai (53%), nei priebalsių (33%). Tai reiškia, kad tonas ties šiek tiek daugiau nei puse visų balsių gali būti prognozuojamas, t.y. turi pastovią kryptingumo (aukštėjimo arba žemėjimo) tendenciją. Tuo tarpu tik trečdalis priebalsių pasižymi tokia savybe. Tai galima paaiškinti tuo, kad balsių trukmės yra žymiai ilgesnės nei priebalsių, taip pat kirčiuojant kirtis dedamas ties balsėmis, vadinasi tono kaita vyksta ties balsėmis, taigi, priebalsių fonemų tonai įtakojami jas supančių balsių tonų, todėl dėl skirtingos įvairių balsių tono įtakos priebalsėms bendri tyrimo rezultatai rodo priebalsių didesnį nepastovumą.

Nors visumoje rezultatai nėra patikimi, tačiau sugrupavus fonemas pagal trukmės ir tono charakteristikas (22 pav.) keliose fonemų grupėse pastebima ryški tono tendencija. Ypač aiškiai matyti tono kryptis pirmoje bei antroje fonemų grupėse. Abiejose grupėse trumpos balsės. Balsės grupėse skiriasi tuo, kad pirmoje jos nekirčiuotos, o antroje – kirčiuotos. Tai puikiai atsispindi tono tendencijoje: pirmoje grupėje tonas statistiškai patikimai žemėja, tuo tarpu kirčiuotoms balsėms (antroji grupė) tonas aukštėja. Kadangi kirtis sąlygoja tono paaukštėjimą todėl tyrimo rezultatai visiškai atitinka balsių fonemų tonų tendencijas – kirčiuotos balsės turi aukštėjantį toną. Mažiau ryški (50%), bet vis dėlto pastebima tono žemėjimo tendencija ilgų nekirčiuotų balsių grupėje (trečia lentelė).

Ilgų kirčiuotų balsių bei priebalsių grupėse statistiškai patikimų duomenų santykis nėra pakankamas (<50%), taigi, bendrų tendencijų nėra.

Apibendrinant fonemų grupių rezultatus galima teigti, kad šio tyrimo sąlygomis tendencingiausias tonas yra trumpų nekirčiuotų balsių bei trumpų kirčiuotų balsių grupėse. Fonemų tonas grupėse vienodo kryptingumo, o duomenų statistinis patikimumas rodo, kad balsių tonai yra pastovūs, taigi – šių fonemų tonai gali būti prognozuojami.

Fonema	t-Test
i	0
e	0
a	0
u	0,004

Lentelė 1

Fonema	t-Test
I	0,0251
E	0
A	0,0364
U	0,6329

Lentelė 2

Fonema	t-Test
ii	0,9566
ie	0,0015
ee	0,0004
ea	0,0028
aa	0,0051
oo	0
uo	0,8205
uu	0,0004

Lentelė 3

Fonema	t-Test
Ii	0,0005
Ie	0,8695
Ee	0,0014
Ea	0,0668
Aa	0,046
Oo	0,8069
Uo	0,2123
Uu	0,9568

Lentelė 4

Fonema	t-Test
iI	0,6495
iE	0,5216
eE	0,5914
eA	1
aA	0,5602
oO	0,8647
uO	0,6926
uU	0,0364

Lentelė 5

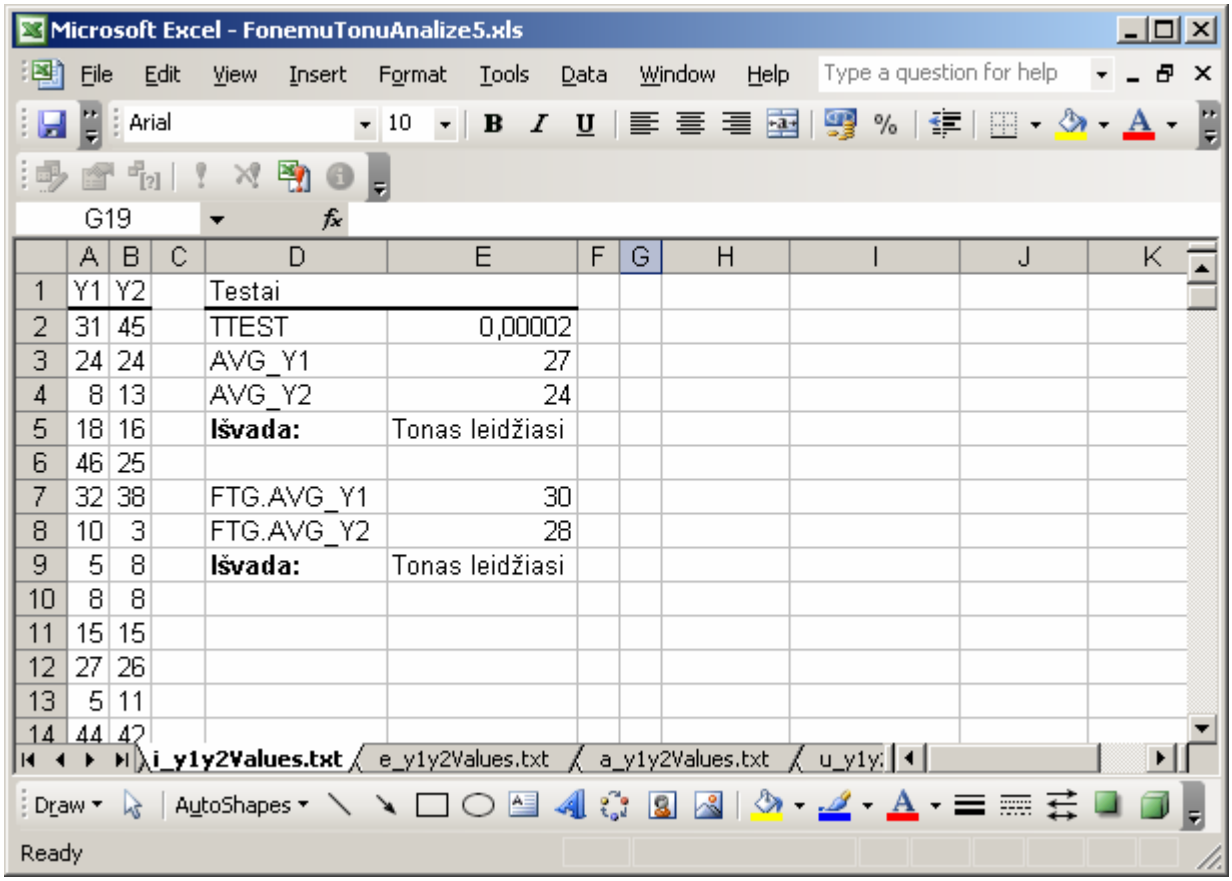
Fonema	t-Test
tS'	0,2134
dZ	0,5
dZ'	0,8834
S	0,0061
S'	0,0008
Z	0,0174
Z'	0,1091
J	0,0357
W	0,2608
L	0,0903
L'	0,0401
R	0,0863
R'	0,1002
M	0,7308
N	0,9823
N'	0,0177

Lentelė 6

Fonema	t-Test
p'	0,3313
b	0,1364
b'	0,1101
t	0,7298
t'	0,4341
d	0,0517
d'	0,854
k	0,374
k'	0,0061
g	0,1108
g'	0,2383
ts'	0,8941
s	0,3026
s'	0,0288
z	0,7578
z'	0,0059
h	0,5277
f	0,6257
f'	0,0329
j	0
j'	0,0171
v	0,8292
v'	0,1641
w	0,1964
l	0,7618
l'	0,4108
r	0,0032
r'	0,05
m	0,2675
m'	0,0528
n	0,7076
n'	0,5962

Lentelė 7

22 pav. Septynios lentelės su sugrupuotomis fonemomis pagal trukmės ir tono charakteristikas. Šviesiai mėlyna fonema reiškia, kad jos tonas aukštėja, tamsiai mėlyna – tonas žemėja, balta spalva – tonas nekinta. Žalios spalvos t-Test reikšmė reiškia, kad fonemos tonas statistiškai patikimas, baltos spalvos reikšmė – tonas statistiškai nepatikimas. Lentelė 1: Trumpos nekirčiuotos balsės; Lentelė 2: Trumpos kirčiuotos balsės; Lentelė 3: Ilgos nekirčiuotos balsės; Lentelė 4: Ilgos balsės su tvirtaprade priegaide; Lentelė 5: Ilgos balsės su tvirtagale priegaide; Lentelė 6: Kirčiuotos priebalsės; Lentelė 7: Nekirčiuotos priebalsės;



23 pav. MS Excel programos panaudijimo pavyzdyz. Pavaizduotas lapas su fonemos '*i*' duomenimis.

Skaičių seka stulpelyje A – tai fonemos pradžios tonų reikšmės $y_1^{(j)}(x_1)$ intervale

$$(f^{-}(x_1) - \partial) \le y_1^{(j)}(x_1) \le (f^{-}(x_1) + \partial),$$

Skaičių seka stulpelyje B – tai fonemos pabaigos tonų reikšmės $y_2^{(j)}(x_2)$ intervale

$$(f^{-}(x_2) - \partial) \le y_2^{(j)}(x_2) \le (f^{-}(x_2) + \partial),$$

, čia x_1 - fonemos pradžios taškas, x_2 - fonemos pabaigos taškas, f^{-} - vidutinis fonemos tonas, ∂ - Standartinis nukrypimas

Stulpelyje E pateiktos reikšmės išskaičiuotos iš A ir B stulpeliuose esančių reikšmių:

TTEST – statistinės funkcijos ttest() reikšmė

AVG_Y1 – vidutinė A stulpelio reikšmių reikšmė

AVG_Y2 – vidutinė B stulpelio reikšmių reikšmė

FTG.AVG_Y1 – vidutinė fonemos tonų reikšmė taške x_1

FTG.AVG_Y2 – vidutinė fonemos tonų reikšmė taške x_2

Išvados

Šiuo metu viena labiausiai negrinėjamų problemų kalbos sintezės sistemoje yra balso tono valdymas. Kalbos sintezė jau pakankamai pažengusi, kad klausytojas galėtų nesunkiai suprasti kas yra sakoma, tačiau neįmanoma nepastebėti, kad sintezuotos kalbos intonacija kol kas neprilygsta natūraliai. Aišku viena: tam, kad dirbtinė kalba skambėtų natūraliai, jai reikia suteikti natūralią intonaciją.

Šiame darbe atliktas natūralaus balso frazės konstatuojamojo tono bei kirčių tono modeliavimas. Taip pat atliktas fonemų tonų tyrimas.

Tono modeliavimui panaudota natūralaus balso tono duomenų bazė, kurioje iš viso sukaupta virš 70 konstatuojamojo tono frazių. Bazė sukurta iš G. Deksnio balso įrašų. Diktoriaus balsas pasirinktas neatsitiktinai – taip siekta maksimalaus suderinamumo su balso sintezei naudota G. Deksnio balso MBROLA difonų baze. Sukauptoji duomenų bazė turi trūkumų: sukaupti tonai nėra identiški natūraliems, iš kurių jie buvo išgauti. To priežastis – Mbrolign programa, kurios pagalba iš natūralių balso įrašų išgautas balso tonas. Mbrolign programa balso toną išgauna apytiksliai, todėl sudaryta tono duomenų bazė apytikslė. Tai reiškia, kad darbe modeliuotas tonas tėra apytikslis lyginant su natūraliu balso tonu, t.y. akustiškai juntamas skirtumas.

Modeliuojant balso toną remtasi superpoziciniu Fujisaki balso tono modeliu, kuris pagrįstas prielaida, kad pagrindinis tonas F_0 sudarytas iš dviejų tono lygių, esančių vienas virš kito. Pirmas lygis – tai frazės tonas, antras – kirčių tonas. Kiekvieno lygio tonui modeliuoti apibrėžta tam tikra, parametrizuota funkcija. Šiame darbe Fujisaki toną modeliuojančios funkcijos buvo šiek tiek modifikuotos tam, kad padaryti patogesnį jų valdymą. Modifikavimo esmė – funkcijų parametrų pakeitimas naujais, paprasčiau valdomais.

Balso tono modeliavimas atliktas dviem etapais. Iš pradžių modeliuotas konstatuojamosios frazės tonas, po to – kirčių tonas. Ištyrus balso tono duomenų bazę, parinkti optimalūs konstatuojamąjį toną modeliuojančios funkcijos parametrai. Optimaliais vadinami parametrai su kuriais funkcijos modeliuojamas tonas artimiausias natūraliam. Atlikus klausymo testus nustatyta, kad sumodeliuotas tonas artimas natūraliam konstatuojamajam tonui.

Optimalių kirčių tonų radimas kur kas sudėtingesnis dėl daugybės įvairių faktorių. Kaip žinome, kalboje kirčiai dedami ant balsių. Lietuvių kalboje naudojami trijų tipų kirčiai: kairinis, tvirtapradis bei tvirtagalis. Kiekvieno iš šių kirčių tonas skirtingo aukščio bei amplitudės. Be to, kirčių tono aukštį bei poziciją skiemenyje įtakoja pati kirčiuota fonema bei tai, kokias fonema stovi prieš ją, žodžiu svarbūs fonemų junginiai. Pavyzdžiui, kirčiuotuose skiemenyse kirčio tonas dažniausiai

pasislinkęs link sprogstamojo garso, jei jis stovi prieš kirčiuotą fonemą. Įvertinus kirčių tonų modeliavimo problemos sudėtingumą, nuspręsta pasigilinti į atskirų fonemų tonų charakteristikas.

Fonemų tonų tyrimas atliktas panaudojus sukurtą balso tono duomenų bazę. Iš viso lietuvių kalbos sintezėje naudojama 91 fonema. Dėl duomenų trūkumo, tyrime panaudoti 80-ies fonemų (32 balsės ir 48 priebalsės) tonai. Fonemų tonams tirti naudoti standartinio nuokrypio bei t-Test metodai. Standartinio nuokrypio metodu išrinktiems fonemų tonams taikytas t-Test metodas. Tokiu būdu nustatytas fonemų tonų statistinis patikimumas. Statistiškai patikimi duomenys reiškia, kad stebimas fonemos tono aukštėjimas ar žemėjimas yra tendencingas, t.y. būdingas fonemai ir tai nėra tik atsitiktinumas.

53% tyrime panaudotų balsių tonai pasirodė esantys statistiškai patikimi. Bendras priebalsių patikimumas kur kas mažesnis - tik 33%. Mažas priebalsių duomenų patikimumas gali būti paaiškintas tuo, kad priebalsių fonemų trukmė yra žymiai mažesnė nei balsių todėl jos labiau įtakojamos jas supančių balsių tono, nes kaip žinoma, kalboje kirčiuojamos balsės, o ne priebalsės.

Sugrupavus fonemas pagal tono bei trukmės charakteristikas, statistiškai patikimiausios pasirodė esančios dvi grupės: trumpų nekirčiuotų balsių bei trumpų kirčiuotų balsių. Pirmoje grupėje fonemų tonas statistiškai patikimai žemėja, o antroje – aukštėja. Kadangi kirtis sąlygoja tono paaukštėjimą todėl tyrimo rezultatai visiškai atitinka balsių fonemų tonų tendencijas – kirčiuotos balsės turi aukštėjantį toną.

Žodynas

Kalbos tonas	- dažnai vadinamas pagrindiniu kalbos tonu ir žymimas F_0.
Fonema	- mažiausias kalbos vienetas (garsas), kuri šnekamojoje kalboje realizuojama tam tikru garsu.
Transkripcija	- kalbos garsų užrašymas specialiais simboliais.
Ortografija	- kalbos simbolių aibė, bei taisyklių rinkinys, aprašantis kaip kalba turi būti užrašyta naudojant kalbos simbolius. Ortografija taip pat apima kalbos rašybą ir skyrybą.
Grafema	- atominis rašomosios kalbos vienetas (raidė)
Prozodija	- tai kalbos intonacija, kirčiai, ritmas bei garsų trukmės
Intonacija	- kalbos tono varijavimas

Literatūros sąrašas

[Mon96] Faculte polythenique de Mons. The MBROLA Project. URL: http://tcts.fpms.ac.be/synthesis , 1996

[Mon99] Faculte polythenique de Mons. The MBROLIGN Project. URL: http://tcts.fpms.ac.be/synthesis/mbrolign , 1999

[Kas05] P. Kasparaitis. Kompiuterinė lingvistika. URL: http://www.mif.vu.lt/~pijus/CL/cl.htm , 2005

[Kas99] P. Kasparaitis. Transcribing of the Lithuanian text using formal rules: 367-376p. URL: http://www.informatik.uni-trier.de/~ley/db/journals/informaticaLT/informaticaLT10.html#Kasparaitis99 , 1999

[Kas00] P. Kasparaitis. Automatic stressing of the lithuanian text on the basis of a dictionary: 19-40p. URL: http://www.informatik.uni-trier.de/~ley/db/journals/informaticaLT/informaticaLT11.html#Kasparaitis00 , 2000

[Kas01] P. Kasparaitis. Automatic stressing of the lithuanian nouns and adjectives on the basis of rules: 315-336p. URL: http://www.informatik.uni-trier.de/~ley/db/journals/informaticaLT/informaticaLT12.html#Kasparaitis01 , 2001

[Kla87] Denis H. Klatt. Review of text-to-speech conversion in English. URL: http://www.mindspring.com/~ssshp/ssshp_cd/dk_760.htm , 1987

[FOW99] H. Fujisaki, S. Ohno, Ch. Wang. A command-response model for F_0 contour generation in multilingual speech synthesis. URL: http://www.slt.atr.co.jp/cocosda/jenolan/Proc/r51/r51.pdf , 1999

[Dob96] A. Dobnikar. Modeling segment intonation for Sloven TTS system. URL: http://www.asel.udel.edu/icslp/cdrom/vol3/299/a299.pdf , 1996

[Dre06] Technische Universitat Dresden. Elements of TTS system. URL: http://www.ias.et.tu-dresden.de/sprache/lehre/multimedia/tutorial/rahmen.htm, 2006

[Mix03] H. Mixdorff. Speech technology, ToBI, and making sence of prosody. URL: http://www.lpl.univ-aix.fr/sp2002/pdf/mixdorff.pdf , 2003

[Wik06a] Wikipedia encyclopedia. Speech synthesis.
URL: http://en.wikipedia.org/wiki/Speech_synthesis, 2006

[BE97] Mary E. Beckman, Gayle A. Elam. Guidelines for ToBI labelling.
URL: http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf , 1997

[LB02] J.A. Louw, E. Barnard. Automatic intonation modeling with INTSINT.
URL: http://www.meraka.org.za/pubs/louwja04intsint.pdf , 2002

[Tro02] William M. K. Trochim. Student‘s t-Test.
URL: http://www.socialresearchmethods.net/kb/stat_t.htm , 2002

[Wik06b] Wikipedia encyclopedia. Standard deviation.
URL: http://en.wikipedia.org/wiki/Standard_deviation , 2006

[Wik06c] Wikipedia encyclopedia. INTSINT.
URL: http://en.wikipedia.org/wiki/INTSINT , 2006