

# Garsų trukmės modeliavimas naudojant klasifikavimo ir regresijos medžius

Giedrius Norkevičius, Asta Kazlauskienė, Gailius Raškinis

Vytauto Didžiojo universitetas, K. Donelaičio g. 58, 3000 Kaunas

Šiuolaikinių šnekos sintezės sistemų signalo kokybė pakankama, kad jas būtų galima praktiškai naudoti, tačiau signalui trūksta natūralumo. To priežastys paprastai esti dvi: netikslūs sintezuotos šnekos garsų trukmės santykiai ir netinkama intonacijos moduliacija ar jos nebuvimas apskritai. Šios problemos mažai tyrinėtos ir lietuvių kalboje.

Straipsnyje aprašomi keli garsų trukmės modeliavimo metodai bei plačiau analizuojamos klasifikavimo ir regresijos medžių panaudojimo lietuvių kalboje galimybės. Žvalgomojo pobūdžio tyrimui pasirinkta anotuota 60 tūkst. rišlaus teksto garsų pavyzdžių duomenų bazė. Regresijos medis leido sumažinti balsių ir priebalsių trukmių dispersiją atitinkamai 28 % ir 25 %.

## 1. Įvadas

Šiuolaikinių šnekos sintezės sistemų signalo kokybė pakankama, kad jas būtų galima praktiškai naudoti, tačiau signalui trūksta natūralumo. To priežastys paprastai esti dvi: netikslūs sintezuotos šnekos garsų trukmės santykiai ir netinkama intonacijos moduliacija ar jos nebuvimas apskritai. Šios problemos mažai tyrinėtos ir lietuvių kalboje.

Reikia pasakyti, kad lietuvių bendrinės kalbos balsyno tos pačios pozicijos ilgųjų ir trumpųjų balsių trukmės santykiai, kokybės, priegaidės ir kirčio įtaka balsių kiekybei gana išsamiai aprašyta [1, 2, 3, 4, 5, 6, 7]. Ypač plačiai balsių kiekybė analizuota Parkerio [4]. Visi šitie darbai yra gera atrama modeliuojant balsių trukmės santykius. Tačiau bendrinės kalbos priebalsių kiekybė visiškai neanalizuota. Reikiamo lietuvių kalbininkų dėmesio vis dar nesulaukia ir fonetiniai intonacijos požymiai, nors techninės tokių tyrimų galimybės dabar tikrai geros. Todėl šiuo atveju paprastai minimi tik senesnės kartos mokslininkų gana fragmentiški darbai [8, 9, 10, 11, 12].

Šio straipsnio tikslas – apžvelgti populiariausius garsų trukmės modelius; išsiaiškinti vieno iš garsų trukmės modeliavimo metodų – klasifikavimo ir regresijos medžių – panaudojimo lietuvių kalboje galimybės.

## 2. Garsų trukmės modeliavimo metodai

### 2.1. Modeliavimas naudojant taisykles

Garsų trukmė gali būti modeliuojama įvairiais metodais: sudarant kiekybės santykius nusakančias taisykles [15], naudojant sandaugų sumų modelį [13,14], sukuriant sprendimo medžius [18,19], taikant dirbtinius neuroninius tinklus [16] ar Bajeso tikimybinis tinklus [17]. Pastarieji du metodai – neuroniniai ir tikimybiniai tinklai, – nors ir gali būti gana patikimi modeliuojant trukmę, tačiau sunku (jeigu apskritai įmanoma) interpretuoti jų rezultatus, be to, tai nėra itin populiarūs metodai trukmei modeliuoti.

Vienas iš populiariausių yra Klatt taisyklių modelis [15]. Kiekviena taisyklė nusako kiekybės mažėjimą ar didėjimą procentais, o klasifikuotini segmentai negali būti trumpesni nei tam tikras minimumas. Modeliuoti pradama nuo savaiminės garsų trukmės ir pridėjamas kiekvieną garso požymį atitinkantis trukmės santykis.

$$DUR = ((INHDUR - MINDUR) * PRCNT) / 100 + MINDUR$$

Kur  $INHDUR$  yra segmento savaiminė trukmė ms,  $MINDUR$  segmento minimali trukmė ir  $PRCNT$  trukmės padidėjimas/sumažėjimas procentais.

Visos taisyklės sudaromos rankomis, o tam reikia išsamių tyrimų. Kaip jau minėta, tik lietuvių kalbos balsių trukmė nemažai tirta, vadinasi, tik jų kiekybės santykius, išanalizavus tempo įtaką balsių trukmei, galima būtų modeliuoti šiuo metodu.

Pasirinkus šį metodą, būtina atkreipti dėmesį į tai, kad kurtinų taisyklių gali būti labai daug, jos gali turėti išimčių, kurioms aprašyti vėl reikės naujų taisyklių. Be to, sudarydamas taisykles žmogus dažnai pasikliauja išankstine nuostata, kuri ne visada yra teisinga.

### 2.2. Sumų sandaugų modelis

Sandaugų sumų modelis, sukurtas Van Santeno [13, 14], apibendrinamas formule:

$$DUR(d) = \sum_{i \in K} \prod_{j \in I_i} S_{i,j}(d_j),$$

Čia  $d$  yra parametrų vektorius, nusakantis prognozuojamą segmentą,  $K$  – indeksų, atitinkančių kiekvieną sandaugą, aibė,  $I_i$  – aibė parametrų, įeinančių į  $i$ -tąją sandaugą. Parametrai  $S_{i,j}$  yra vadinami parametrų svoriais (*factor scales*).

Modeliavimas šiuo metodu vyksta trimis etapais:

- pagal jau žinomus garsų kiekybės santykius kalbininkai sudaro kategorijų medžius – vienas medžio lapas atspindi garsų grupę, kuriai turi įtakos tam tikri faktoriai/parametrai ar jų sąveikos,
- kiekvienam lapui/kategorijai sudaromas atskiras modelis,
- skaičiuojami modelių parametrai.

Daugelio nurodoma, kad tai vienas patikimiausių metodų, t. y. geriausiai prognozuojantis bei didžiausia koreliacija tarp prognozuojamos ir tikrosios reikšmės pasižymintis modelis. Literatūroje taip pat minima, jog šis modelis geriausiai atsiskleidžia krypties invariantiškumą. Krypties invariantiškumą geriausia pailustruoti pavyzdžiu: pavyzdžiui, kirčiuotas  $u$  yra ilgesnis už nekirčiuotą  $u$ , toks pat šių balsių kiekybės santykis išliks ir tuo atveju, jeigu  $u$  papriešakės (taigi ankstesnio priebalsio minkštumas neturės įtakos kiekybės santykiams, t. y. kirčiuotas  $u$  išliks ilgesnis).

Šio metodo, kaip ir taisyklių, kol kas negalima taikyti lietuvių kalbai, nes nėra pakankamai išsamių garsų trukmės tyrimų duomenų, kurie būtini norint sudaryti kategorijų medį.

### 2.3. Modeliavimas klasifikavimo ir regresijos medžiais

Šiuo metu lietuvių kalbos garsų trukmei modeliuoti galima pritaikyti vieną iš mašininio mokymo metodų – sprendimo medžio metodą.

Sprendimo medžių vienas iš variantų – klasifikavimo ir regresijos medžiai – yra statistinio modeliavimo metodas, naudojamas prognozuoti kintamojo  $y$  reikšmei, atitinkančiai parametru vektoriu  $f$ . Modeliavimas susideda iš trijų etapų:

- a) medžio konstravimas,
- b) jo paprastinimas (genėjimas),
- c) optimalaus medžio parinkimas.

Kaip ir kiekvienam mašininio mokymo algoritmui, taip ir pastarajam reikalinga  $\{f_n, y_n\}$  pavidalo mokymo imtis  $L$ , kur  $y_n$  – nuo parametru vektoriaus  $f_n$  priklausomo objekto reikšmė. Iš pradžių medis susideda iš vieno, vadinamojo šakninio, mazgo  $t_1$ , kurį sudaro visi aibės  $L$  mokymo pavyzdžiai. Užduotis yra surasti optimalų aibės  $L$  padalinimą į dvi dalis. Šiuo atveju optimalumo kriterijus yra vidutinė kvadratinė klaida:

$$\sum_{f \in t_L} (d(f) - \bar{y}_L)^2 + \sum_{f \in t_R} (d(f) - \bar{y}_R)^2$$

Realaus tipo parametru nustatomi visi  $f_n^i < \tau$  padalinimai. Išvardijamojo tipo parametru padalinimo pavidalas yra:  $f^i \in \Theta$ , kur  $\Theta$  gali būti bet koks aibės, sudarytos iš  $i$ -tojo požymio reikšmių, poaibis. Tokiu būdu yra išrenkamas geriausių padalinimų atitinkantis parametras ir visi šakninio mazgo pavyzdžiai padalinami į mazgus  $T_L, T_R$ . Su gautaisiais mazgais kartojama tokia pati procedūra tol, kol įvykdoma sustojimo sąlyga (paprastai dalinama tol, kol pasiekiamas tam tikras iš anksto apibrėžtas klaidos mažėjimas).

Paprastai sudaromas gana didelis medis  $T_{\max}$ . Genėdami šakas sukonstruojame medžių seką  $T_{\max} \supseteq \dots \supseteq T_k \supseteq \dots \supseteq T_1 = t_1$ . Iš šios sekos, panaudodami nuo mokymo duomenų nepriklausomą validavimo aibę, išrenkame geriausių (mažiausiai klaidų generuojanti) medį.

Segmento trukmę lemiantys parametrai (paties segmento identifikacija, kirtis ir priegaidė, gretimų segmentų nustatymas, segmento pozicija skiemenyje, žodyje, sakinyje ir kt.) [18, 19] yra pasirenkami. Medžius gali sudaryti išvardijimo ir skaitmeniniai parametrai, tačiau jie turi būti nustatomi iš gryno, įprastine rašyba parašyto teksto. Einant medžio šakomis yra tenkinamos įvairios sąlygos, susijusios su parametrais, pvz.: kairiąją šaką reikia rinktis tuomet, kai segmento dešinys kontekstas lygus  $X$ , o dešiniąją šaką, kai dešinys kontekstas nelygus  $X$ . Kai pasiekiamas lapas, turime prognozuojamą trukmę – lape esančių segmentų trukmių vidurkį. Taip sudaryti regresijos medžiai lengvai interpretuojami ir gali būti koreguojami (to negalima daryti naudojant neuroninius ar tikimybinus tinklus). Šis metodas labai parankus tada, kai kalba mažai ištirta, ir būtent jo duomenys gali būti atspirties taškas atliekant išsamius kiekybės tyrimus, ruošiant taisyklių ar sandaugų sumų modelius, nustatant, kurie parametrai turi didžiausią įtaką trukmei. Dėl šių priežasčių pradiniais trukmės tyrimams pasirinktas klasifikacijos ir regresijos medžio metodas.



- 1) ilgiausi yra pučiamieji priebalsiai (visi jie dešiniajame medžio krašte), o sprogstamųjų ir pusbalsių trukmė įvairuoja;
- 2) priebalsių minkštumas neturi akivaizdžios įtakos trukmei, bent jau šie duomenys to nerodo;
- 3) dvigarsyje esantys nekirčiuoti pusbalsiai irgi ne visada trumpesni už tokius pačius ne dvigarsio priebalsius, kirčiuoti tvirtagaliai šie dėmenys visada ilgesni už atitinkamus nekirčiuotus;
- 4) gana akivaizdžiai skardieji priebalsiai ilgesni už atitinkamus dusliuosius.

Jeigu įtraukiame ir konteksto iš kairės bei dešinės požymį, tada sunku aprėpti visą vienetų gausą, nustatyti patikimus dėsningumas ar akivaizdesnes tendencijas. Todėl, norint išsiaiškinti gretimų garsų įtaka, reikia mažinti tiriamų vienetų bazę ir didinti pasirinktinių požymių kiekį.

#### Anotation

##### Decision trees in phoneme's duration modelling

Currently, intelligibility of the best TTS systems is extremely good, and certainly good enough for many real applications. However, it definitely lacks naturalness. It is commonly assumed that lack of natural prosody is the main reason for this. It is generally accepted that, next to intonation, timing plays a crucial role for encoding and decoding speech. The prerequisite for appropriate timing in speech synthesis is a high quality model for duration prediction. Research on Text-to-Speech conversion for Lithuanian is a much younger enterprise in comparison with the Text-to-Speech research for English and other European languages. Unfortunately there are no any investigations on duration modeling for Lithuanian. Therefore the purpose of this paper is to review existing models of duration prediction imposing more attention to Decision trees, in particular CART (classification and regression trees) like decision trees and to do some preliminary experiments on modeling phonemes duration for Lithuanian language.

#### Literatūros sąrašas

1. **Anusienė L.** Kirčiuotų ilgųjų balsių trukmė lietuvių bendrinės kalbos frazėse, *Kalbotyra*, 34 (1), 1983, 5–13
2. **Dambrauskaitė-Urbelienė J.** K voprosy o nekotoryx osobenostex dolgix litovskix glasnix a i e. *Kalbotyra*, 1967, T. 17, 17–25
3. **Pakerys A., Plakunova, J. Urbelienė,** Otnositelnaja dlitelnost glasnyx litovskogo jazyka, *Kalbos garsai ir intonacija*, 1970 – P. 30–53.
4. **Pakerys A.** *Lietuvių bendrinės kalbos prozodija*, 1982
5. **Svecevičius B.,** Nauji lietuvių literatūrinės kalbos paprastųjų balsių eksperimentiniai duomenys, *Eksperimentinės fonetikos ir kalbos psichologijos kolokviumo medžiaga*, 1964, T. 1, 14–32.
6. **Vaitkevičiūtė V.,** 1960, Lietuvių kalbos balsių ir dvibalsių ilgumas arba kiekybė, *Lietuvių kalbotyros klausimai*, T. 3, 207–217.
7. **Vaitkevičiūtė V.,** 1961, Lietuvių literatūrinės kalbos balsinės ir dvibalsinės fonemos, *Lietuvių kalbotyros klausimai*, T. 3, 19–39.
8. **Bikulčienė P.** Skatinimo ir konstatavimo intonacijų gretinimas, *Kalbos garsai ir melodika*, 1978, 3–11
9. **Bikulčienė P.** Liepimo intonacijos, *Kalbos garsai ir prozodija*, 1982, 3–16
10. **Pukelis V.** Kai kurie fiziniai pagrindinio tono požymiai lietuvių kalbos patikrinamuosiuose klausimuose, *Garsai, priegaidė, intonacija*, 1972, 161–164
11. **Pukelis V.** Frazės kirčiu pabrėžto žodžio ir jo kirčiuoto skiemens akustiniai požymiai lietuvių kalbos patikrinamuosiuose klausimuose, *Eksperimentinė ir praktinė fonetika*, 1974, 199–217
12. **Statkevičienė J.** Vienarūšių ir nevienarūšių pažymiųjų pagrindinis tonas, *Eksperimentinė ir praktinė fonetika*, 1974, 218–223
13. **Jan P. H. van Santen,** Prosodic modeling in Text-To-Speech Synthesis ,Lucent Technologies – Bell Labs, 600 Mountain Ave., Murray Hill, NJ 07974, U.S.A.
14. **Jan P. H. van Santen,** Quantitative modeling of segmental duration , Bell Labs, 600 Mountain Ave., Murray Hill, NJ 07974, U.S.A.
15. **D H Klatt,** Synthesis by rule of Segmental Durations in English Sentences, in *Frontiers of Speech Communication Research* edited by Lindblom & Ohman, Academic Press 1979 (pp 287-299)
16. **Martti Vainio<sup>1</sup> & Toomas<sup>2</sup> Altsaar,** Pitch, loudness, and segmental duration correlates : towards a model for the phonetic aspects of finnish prosody, [Department of phonetics, University of Helsinki, Finland]<sup>1</sup>, [Acoustics Laboratory, Helsinki University of Technologie, Finland]<sup>2</sup>
17. **Olga Gaubanova,** Using Bayesian belief networks for model duration in text-to-speech systems, Centre for Speech Technology Research, University of Edinburgh
18. **Robert Batušek,** A Duration Model for Czech Text-To-Speech Synthesis , Laboratory of Speech and Dialogue, Faculty of Informatics, Masaryk University, Brno, Czech Republic
19. **Sridhar Krishna & Hema A. Murthy,** Duration modelilng of Indian languages Hindi and Telugu, Indian Institute of Technology, Madras, Chennai – 60003
20. **Raškinis A., G. Raškinis, A. Kazlauskienė.** SAMPA (Speech Assessment Methods Phonetic Alphabet) for Encoding Transcriptions of Lithuanian Speech Corpora. *Information technology and control*. **Kaunas: Technologija**, 2003, No. 4(29), p. 52–55.