

Lietuvių šnekos balsių aprašymo autoregresijos modeliu adekvatumo tyrimas

Jonas Kaukėnas, Gintautas Tamulevičius

Vilniaus universitetas, Matematikos ir informatikos institutas

Akademijos g. 4, LT-08663 Vilnius

E. paštas: jonas.kaukenas@mii.vu.lt, gintautas.tamulevicius@mii.vu.lt

Santrauka. Šnekos signalo aprašymas autoregresijos (AR) modeliu remiasi žiniomis apie signalo susidarymą balso trakte. Mums nepavyko rasti darbų, kuriuose būtų ištirtas AR modelio eilės parinkimas lietuvių šnekai. Atliktas balsių aprašymo AR modeliu tyrimas rodo, kad adekvačiam lietuvių šnekos balsių aprašymui būtina žymiai didesnė AR modelio eilė negu naudojama kitų autorių darbuose. Nustatyta, kad tik aukštos eilės AR modelis leidžia atkurti lietuvių šnekos balsių dažnines charakteristikas.

Raktiniai žodžiai: autoregresijos modelis, šnekos signalo tyrimas.

1 Įvadas

Šnekos signalui aprašyti naudojami įvairūs modeliai, kaip antai: autoregresijos-slenkančio vidurkio [5], autoregresijos (AR) [9, 7, 8], sinusinis [10] ir kt. AR modelio taikymas šnekos signalui aprašyti remiasi žiniomis apie signalo susidarymą balso trakte bei statistinėmis signalo charakteristikomis [9, 7, 8]. Todėl AR modelio panaudojimas šnekos signalui aprašyti diskusijų nekelia. Lietuvių šnekos signalo analizei taip pat yra naudojamas AR modelis [2, 7]. Mums nepavyko rasti darbų, kuriuose būtų ištirtas bei pagrįstas AR modelio eilės parinkimas lietuvių šnekai. Be to, darbe [6] yra parodyta, kad lietuvių šnekai naudojami AR eilės bei parametru įvertinimo metodai [1, 3] duoda paslinktus AR modelio eilės įverčius. Todėl būtina ištirti lietuvių šnekos aprašymo autoregresijos modeliu adekvatumą.

2 Tyrimo metodas

Tegul procesas $\{Z_t, t = 1, 2, \dots\}$ su nuliniu vidurkiu yra aprašomas AR modeliu

$$\sum_{i=0}^M a_i Z_{t-i} = bV_t, \quad a_0 = 1, \quad t = 1, 2, \dots, \quad (1)$$

kur $\{V_t, t = 1, 2, \dots\}$ yra procesas tarpusavyje nepriklausomų atsitiktinių dydžių, aprašomų standartiniu normaliuoju pasiskirstymo dėsnium, t. y. $V_t \sim N(0, 1)$. Tarkime, kad stebime proceso Z_t realizaciją $\{z_1, z_2, \dots, z_N\}$ ir mums reikia įvertinti AR modelio eilę M bei parametrus $\{a_1, a_2, \dots, a_M\}$ ir b .

Uždavinys susideda iš dviejų dalių. Pirmiausia sukonstruojamas parametų $\{a_1, a_2, \dots, a_M\}$ ir b įvertinimo metodas, teigiant, kad eilė M yra žinoma. Po to sprendžiamas AR modelio eilės įvertinimo uždavinys.

Nagrinsime du, dažniausiai naudojamus, AR modelio parametų vertinimo metodus: maksimalaus tikėtinumo ir mažiausių kvadratų [1, 3, 5, 6].

Tikėtinumo funkcija [5]

$$f(z_{1,N}) = \frac{1}{(2\pi)^{N/2} b^{N-M} |K|^{1/2}} \times \exp \left\{ -\frac{1}{2} (z_{1,M})^T K^{-1} (z_{1,M}) - \frac{1}{2b^2} \sum_{t=M+1}^N \left(\sum_{v=0}^M a_v z_{t-v} \right)^2 \right\}, \quad (2)$$

čia K yra M ilgio proceso realizacijos AR modelio kovariacijos matrica.

Šios funkcijos maksimizavimas yra sudėtingas uždavinys, todėl praktiniams tikslams yra maksimizuojama apytikslė tikėtinumo funkcija [5]

$$f_{M+1, M+2, \dots, N}(z_{M+1}, z_{M+2}, \dots, z_N | z_1, z_2, \dots, z_M) \times \frac{1}{(\sqrt{2\pi} b^2)^{N-M}} \exp \left\{ -\frac{1}{2b^2} \sum_{t=M+1}^N \left(\sum_{v=0}^M a_v z_{t-v} \right)^2 \right\}. \quad (3)$$

Gaunamoms paklaidoms skaičiuoti išraiškų nėra, bet esant eilei $M \ll N$ manoma, kad paklaidos yra nežymios [1, 3, 5]. Nesunku pastebėti, kad rezultatas naudojant supaprastintą maksimalaus tikėtinumo metodą yra toks pats, kaip ir gautasis mažiausių kvadratų metodu [5]. Todėl dažniausiai AR parametrams vertinti naudojamas mažiausių kvadratų metodas.

Iš išraiškos (1) galime gauti

$$\begin{aligned} -z_{M+1} &= a_1 z_M + a_2 z_{M-1} + \dots + a_M z_1 + b v_{M+1}, \\ -z_{M+2} &= a_1 z_{M+1} + a_2 z_M + \dots + a_M z_2 + b v_{M+2}, \\ &\vdots \\ -z_N &= a_1 z_{N-1} + a_2 z_{N-2} + \dots + a_M z_{N-M} + b v_N. \end{aligned} \quad (4)$$

Pažymėkime

$$\begin{aligned} Y' &= (-z_{M+1}, -z_{M+2}, \dots, -z_N), \\ X'_1 &= (z_M, z_{M+1}, \dots, z_{N-1}), \\ X'_2 &= (z_{M-1}, z_M, \dots, z_{N-2}), \\ &\vdots \\ X'_M &= (z_1, z_2, \dots, z_{N-M}), \\ V' &= (V_{M+1}, V_{M+2}, \dots, V_N), \end{aligned}$$

$$X = (X_1, X_2, \dots, X_M), \quad A' = (a_1, a_2, \dots, a_M), \quad N_0 = N - M.$$

Tuomet

$$Y = X \cdot A + bV. \quad (5)$$

Naudosime darbe [6] pateiktą rekurentinį AR parametrų vertinimo metodą. Suformuojame Efroymson matricą [4]

$$E = \begin{bmatrix} R(M \times M) & T'(M \times 1) & I(M \times M) \\ T(1 \times M) & I(1 \times 1) & O(1 \times M) \\ -I(M \times M) & O(M \times 1) & O(M \times M) \end{bmatrix}. \quad (6)$$

Čia $R(M \times M) - X_i$ ir X_j koreliacijos matrica; $T'(M \times 1)$ ir $T(1 \times M) - Y$ ir X_i koreliacijos vektoriai; $O(1 \times M)$ ir $O(M \times 1) -$ nuliniai vektoriai; $O(M \times M) -$ nulinė matrica; $I(1 \times 1)$, $I(M \times M) -$ vienetinės matricos.

Kintamasis X_i įtraukiamas į modelį naudojant rekurentines išraiškas

$$\begin{aligned} E'(i, j) &= E(i, j)/E(i, i), \quad j = 1, 2, \dots, 2M + 1, \\ E'(k, l) &= E(k, l) - \frac{E(k, i) \cdot E(i, l)}{E(i, i)}, \quad k, l = 1, 2, \dots, 2M + 1, l \neq i, \end{aligned} \quad (7)$$

kur $E(i, j) -$ Efroymson matrica prieš įtraukiant X_i ; $E'(i, j) -$ Efroymson matrica, įtraukus X_i į modelį.

Parametro a_i įvertis gaunamas pagal formulę

$$\hat{a}_i = E(i, M + 1) \sqrt{(Y'Y)/(X'_i X_i)}. \quad (8)$$

Parametro b įvertis gaunamas pagal formulę

$$\hat{b}^2 = E(M + 1, M + 1)(Y'Y)/(N_0 - M). \quad (9)$$

AR modelio įvertinimo uždavinys yra sprendžiamas lyginant eilės M_1 modelį su eilės M_2 modeliu. Dažniausiai yra lyginami modelių prognozės klaidos dispersijų įverčiai $\hat{b}_{M_1}^2$ ir $\hat{b}_{M_2}^2$, gauti pagal (9). Darbe [6] yra pateikta AR modelio eilės parinkimo procedūra, paremta nuosekliu AR modelio eilės didinimu nuo 0 iki M_{\max} , kur $M_{\max} -$ maksimali galima eilė, parinkta naudojantis empirinėmis žiniomis apie tiriamą procesą.

$$\widehat{M}_i = \begin{cases} i, & \text{jeigu } \left(\frac{\hat{b}_i^2 - \hat{b}_{i-1}^2}{\hat{b}_i^2} \right) (N_0 - i) > F_{kr}(1, N - i), \\ 0, & \text{jeigu } \left(\frac{\hat{b}_i^2 - \hat{b}_{i-1}^2}{\hat{b}_i^2} \right) (N_0 - i) \leq F_{kr}(1, N - i), \end{cases} \quad i = 1, \dots, M_{\max},$$

$$\widehat{M} = \max_i \{\widehat{M}_i\}; \quad i = 1, \dots, M_{\max}, \quad (10)$$

čia $F_{kr}(1, N - i)$ yra kvantilis Fišerio pasiskirstymo su 1 ir $(N - i)$ laisvės laipsniais. \hat{b}_0^2 yra realizacijos dispersijos įvertis, t. y. $\hat{b}_0^2 = \widehat{D}$.

3 Tyrimo rezultatai

Atlikome trumpųjų ir ilgųjų lietuvių šnekos balsių analizę, siekdami išsiaiškinti ar literatūroje pateikiami AR modelio eilės ir parametrų vertinimo metodai adekvačiai aprašo konkrečius lietuvių šnekos balsius, išstartus skirtingų kalbėtojų. Nagrinėjome trumpuosius $[a]$, $[e]$, $[i]$, $[o]$, $[u]$ ir ilguosius $[a:]$, $[e:]$, $[è:]$, $[i:]$, $[o:]$, $[u:]$ balsius, iškirptus iš natūralios šnekos. Įvertindami pagrindinio tono dažnio skirtumus, vyrų ir moterų šnekos įrašus nagrinėjome atskirai. Tyrime panaudoti 16 moterų ir 17 vyrų natūralios šnekos įrašai. Iš jų atsitiktine tvarka iškirpome po 3 atkarpas vyrų ir 3 atkarpas moterų išstartų aukščiau paminėtų balsių. Kadangi natūralioje lietuvių šnekoje balsių trukmė beveik visada yra didesnė nei 50 ms, vienodumo dėlei naudojome tokios pat trukmės (50 ms) balsių įrašų atkarpas. Šneka buvo įrašyta 11025 Hz diskretizavimo dažniu. Todėl įrašų ilgiai $N = 551$, kas labai artima įrašų ilgiams, naudojamiems kitų autorių darbuose [7, 8].

Naudodami pasirinktąjį metodą tiriamiesiems balsiams aprašyti įvertinome AR modelio eilę bei parametrus. Objektiviam AR modelio adekvatumo vertinimui naudojome prognozės klaidos kvadrato ir signalo dispersijos įverčių santykį $\widehat{b}^2/\widehat{D}$. AR modelio adekvatumą dažnių srityje tikrinome lygindami signalo Furjė transformaciją su spektrinio tankio funkcija, gauta iš AR modelio parametrų įverčių.

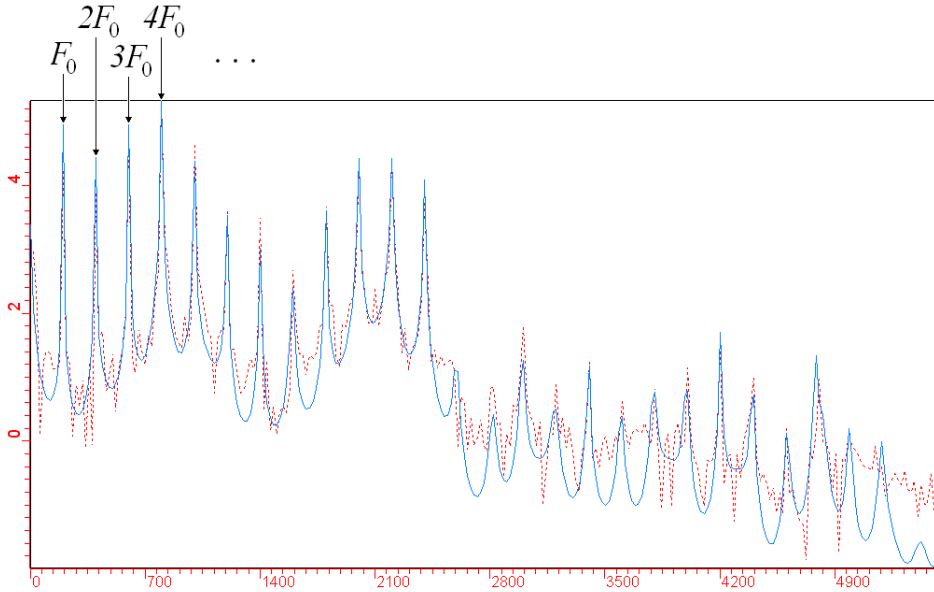
Atlikus lietuvių šnekos balsių aprašymo AR modeliu adekvatumo tyrimus, gauti šie rezultatai:

- Kadangi moterų pagrindinio tono dažnis yra didesnis, tai jų šnekos signalui aprašyti reikalinga mažesnė AR modelio eilė negu vyrų. AR modelio eilės įvertis moterų tariamų balsių atveju svyravo nuo 64 iki 100, vidutiniškai buvo 85. Tuo tarpu vyrų šnekos atveju modelio eilė kito nuo 61 iki 100, vidutiniškai buvo 92.
- AR modelio prognozės klaidos ir signalo dispersijų įverčių santykio $\widehat{b}^2/\widehat{D}$ reikšmė kito nuo 0,0004 iki 0,0408, o vidurkis buvo 0,0093. Tai parodo taikomo AR modelio tikslumą, aprašant lietuvių šnekos balsius.
- Lyginant signalo Furjė transformaciją $S_F(f)$ su spektrinio tankio funkcija $S_{AR}(f)$, gauta iš AR modelio parametrų įverčių, pasiektas didelis atitikimas. Tai iliustruojama paveiksle 1, kur AR modelis atkuria tiek pagrindinį toną, tiek visus virštonius.
- Mūsų gautoji AR modelio eilė yra žymiai didesnė už rekomenduojamą literatūroje [9, 7, 8]. Akivaizdu, kad žemesnės eilės AR modelis aprašo šnekos signalą neadekvačiai (2 pav.), nes neatkuria nei pagrindinio tono, nei virštonių.

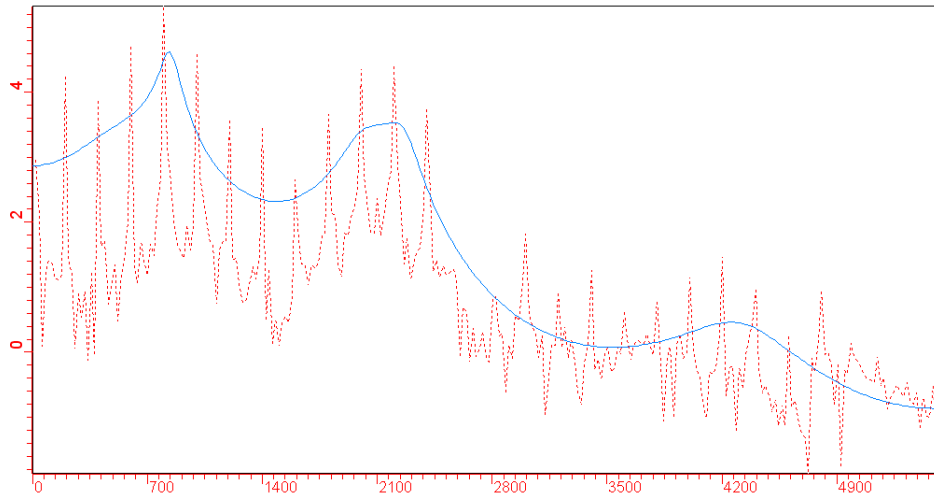
Atliekant tyrimus buvo keletas atvejų, kai metodas parinko žymiai mažesnę eilę nei vidutinę (pvz., $\widehat{M} = 12$). Tais atvejais stebėjome didelį spektrinio tankio funkcijų neatitikimą. Taip pat kai kuriais atvejais buvo gauta žymiai didesnė AR modelio eilė (pvz., $\widehat{M} = 151$).

4 Išvados

1. Iširtas lietuvių šnekos balsių aprašymo AR modeliu adekvatumas, naudojant metodą [6].
2. Adekvataus AR modelio eilė yra žymiai didesnė negu rekomenduojama literatūroje [9, 7, 8]. Vyrų tariamiems balsiams aprašyti reikalinga šiek tiek didesnė AR modelio eilė negu moterų.



1 pav. Garso [e:], iškirpto iš žodžio „žemė“, ištarto moters, logaritmuoti spektrai: Furjė (raudona punktyrinė linija) ir 82 eilės AR modelio (mėlyna linija). Paveiksle pažymėtos šnekos signalo spektro komponentės: pagrindinis tonas F_0 bei virštoniai kF_0 ($k = 2, 3, \dots$).



2 pav. Garso [e:], iškirpto iš žodžio „žemė“, ištarto moters, logaritmuoti spektrai: Furjė (raudona punktyrinė linija) ir 12 eilės AR modelio (mėlyna linija).

3. Tik naudojant aukštos eilės AR modelį, sugeneruota spektrinio tankio funkcija atkartoja Furjė analizės spektrinio tankio funkciją bei aprašo lietuvių šnekos balsių pagrindinį toną ir virštonius.

Literatūra

- [1] H. Akaike. A new look at the statistical model identification. In *IEEE Trans. Aut. Control*, volume AC-19, pp. 716–723, 1974.
- [2] A. Bastys, A. Kisel and B. Šalna. The use of group delay features of linear prediction model for speaker recognition. *Informatica*, **21**(1):1–12, 2010.
- [3] J. Durbin. The fitting of time series models. *Rev. Inst. Int. Stat.*, **28**(3):233–244, 1960.
- [4] M.A. Efron. Multiple regression analysis. In A. Ralston and H.S. Wilf(Eds.), *Math. Meth. Dig. Comp.*, volume 1, pp. 191–203. John Wiley and Sons, Inc., 1965.
- [5] G.M. Jenkins and D.G. Watts. *Spectral Analysis and Its Applications*. Holden-Day, 1968.
- [6] J. Kaukėnas. On estimation of ar model order and parameters. In *Stat. Prob. Control*, volume 61, pp. 46–60, 1983. [rusų k.]
- [7] A. Lipeika. Optimization of formant feature based speech recognition. *Informatica*, **21**(3):361–374, 2010.
- [8] S. K. Katti M. A. Anusuya. Front end analysis of speech recognition: a review. *Int. J. Speech Tech.*, **14**(2):99–145, 2011.
- [9] J.D. Markel and A.H. Gray. *Linear Prediction of Speech*. Springer-Verlag, 1976.
- [10] R.J. McAulay and T.F. Quatieri. Speech processing based on a sinusoidal model. *Lincoln Lab. J.*, **1**(2):153–168, 1988.

SUMMARY

Analysis of adequacy of autoregressive model for Lithuanian vowels

J. Kaukėnas, G. Tamulevičius

Autoregressive (AR) model is widely used for modeling of speech signals. Nevertheless, the problem of adequate modelling of Lithuanian speech is still open. The results of this study indicate the need of much higher order models for Lithuanian vowel description. Only high order AR models enable us to model frequential properties of Lithuanian vowels adequately.

Keywords: autoregressive model, speech signal analysis.