Formantinių požymių išskyrimo metodai

Antanas Leonas Lipeika

Matematikos ir informatikos instituto
Atpažinimo procesų skyriaus vyresnysis
mokslo darbuotojas, docentas, daktaras
Institute of Mathematics and Informatics,
Senior Researcher, PhD
Goštauto g. 12, LT-01108 Vilnius

Vilniaus pedagoginio universiteto
Informacinių technologijų katedros profesorius,
docentas, daktaras
Vilnius Pedagogical University,
Department of Information Technology, Professor, PhD
Studentų g. 39, Vilnius
Tel. 266 03 90, faks. 261 99 05
El. paštas: lipeika@ktl.mii.lt

Straipsnyje nagrinėjami formantinių požymių išskyrimo metodai. Formantinių požymių išskyrimas remiasi spektro pikų radimu apskaičiuotame iš tiesinės prognozės modelio parametrų spektre. Formantinių požymių išskyrimo patikimumas priklauso nuo tiesinės prognozės modelio parametrų vertinimo metodo. Anksčiau tiesinės prognozės modelio parametrų vertinimui naudojome autokoreliacinį metodą, kuris neužtikrindavo patikimo formantinių požymių išskyrimo. Todėl, siekiant padidinti formantinių požymių išskyrimo patikimumą, ieškoma geresnio tiesinės prognozės modelio parametrų vertinimo metodo. Autokoreliacinis tiesinės prognozės modelio parametrų vertinimo metodas lyginamas su kovariaciniu, Burg, Marple metodais ir modifikuotu Split Levinson algoritmu. Tyrimais nustatyta, kad pagal formančių trajektorijų išskyrimą kovariacinis, Burg, Marple tiesinės prognozės modelio parametrų vertinimo metodai iš esmės nesiskiria nuo autokoreliacinio, o modifikuotu Split Levinson algoritmu gauname daug patikimesnius formančių trajektorijų įverčius.

Literatūroje (De Wet ir kt., 2004) diskutuojama, ar formantiniai požymiai (balso trakto rezonansiniai dažniai) yra naudingi atpažįstant kalbą. Pripažįstama, kad formantiniai požymiai gali būti naudojami kalbos atpažinimui, tačiau jie iki šiol nepopuliarūs dėl jų išskyrimo problemų. Balso trakto rezonansiniai dažniai dažniausiai randami iš tiesinės prognozės modelio parametrų (Huang ir kt., 2001). Įmanomi du problemos sprendimo būdai. Galima iš tiesinės prognozės modelio parametrų apskaičiuoti amplitudinį spektrą ir surasti spektro pikus, kurie ir yra traktuojami kaip balso trakto rezonansiniai dažniai – formantės. Alternatyvus būdas yra rasti charakteringo tiesinės prognozės polinomo šaknis. Kompleksinės šaknys, kurių kampai su realia ašimi yra intervale $0 < \omega_i < \pi$, atitinka kalbos signalo *formantes*. Tokiu būdu suradę charakteringo polinomo šaknis, mes atmetame realias šaknis ir šaknis, atitinkančias neigiamus dažnius. Likusios šaknys yra surūšiuojamos dažnio didėjimo tvarka, ir tada šaknies numeris atitinka formantės numerį, o kampas su realia ašimi – formantės dažnį (Lipeika, 2005).

Pastarajame darbe buvo naudojamas autokoreliacinis tiesinės prognozės (LPC) modelio parametrų vertinimo metodas (Rabiner ir kt., 1993) ir atliekant formančių dažnių išskyrimo eksperimentus buvo pastebėta, kad formančių dažniai randami ne visada patikimai. Kai kalbos signale yra triukšmo dedamoji, spektro pikai neaštrūs, susilieja ir dažnai nepavyksta patikimai rasti formančių trajektorijų. Kartais gretimos formantės susilieja ir iškyla jų numeracijos problemų. Autokoreliacinis metodas geras tuo, kad tiesinės prognozės modelio parametrams rasti naudojama autokoreliacinė matrica ir dėl jos savybių nežinomiems parametrams rasti galima taikyti skaičiavimų apimties požiūriu efektyvų, rekurentinį Durbino algoritmą (Makhoul, 1975). Tačiau autokoreliacinis metodas turi vieną trūkumą: prognozės klaida yra skaičiuojama laikant, kad už analizės kadro ribų $0 \le m \le N-1$ kalbos signalas yra lygus 0. Taigi prognozės klaidą gauname

$$E_n = \sum_{m=0}^{N-1+p} e_n^2(m),$$

čia N yra analizės kadro ilgis, p – tiesinės prognozės modelio eilė, n – analizės kadro numeris. Vadinasi, kai m = 0,1,..., p–1, padaugintas iš lango funkcijos kalbos signalas $s_n(m)$ yra prognozuojamas pagal ankstesnes imtis ir kai kurios iš jų yra lygios nuliui. Dėl to dažnai atsiranda didelių prognozės klaidų šioje srityje, ypač vokalizuotiems garsams. Be to, m = N – 1,..., N –1 + p srityje irgi dažnai pasitaiko didelių prognozės klaidų dėl to, kad kalbos signalas yra prognozuojamas tik pagal keletą (mažiau negu p) ankstesnių kalbos signalo reikšmių. Dėl šios prielaidos gauname ne visai tikslius tiesinės prognozės modelio parametrų įverčius. Todėl mes tyrėme alternatyvius tiesinės prognozės modelio parametrų vertinimo būdus.

Tiesinės prognozės modelio parametrų vertinimo metodai

Nagrinėsime alternatyvius tiesinės prognozės modelio parametrų vertinimo metodus ir parinksime tą, kuris geriausiai tinka formančių trajektorijoms išskirti. Iš pradžių apžvelgsime išskirtines nagrinėjamų metodų savybes.

Kovariacinis metodas. Kovariacinis metodas (Rabiner ir kt., 1993) taiko kitą lango funkcijos naudojimo būdą – fiksuojamas intervalas (kadras) $0 \le m \le N-1$, kuriame kvadratinė klaida yra skaičiuojama kaip

$$E_n = \sum_{m=0}^{N-1} e_n^2(m)$$

ir kalbos signalas nedauginamas iš lango funkcijos. Naudojant šį metodą gauta kovariacinė matrica yra simetrinė ($\Phi_n(i,k) = \Phi_n(k,i)$), bet ne Tioplico. Skirtingai nuo autokoreliacinio metodo, kovariacinės matricos elementai ant įstrižainių nėra vienodi. Ši lygčių sistema gali būti išspręsta naudojant Cholesky dekompozicijos metodą. Kovariacinis metodas skaičiavimo požiūriu yra daug imlesnis ir todėl rečiau naudojamas. Šis metodas yra panašus į Proni metodą (McDonough ir kt., 1968), kur signalas yra aproksimuojamas gęstančiomis eksponentėmis.

Burg metodas. Burg metodas (Kay ir kt., 1981) įvertina atspindžio koeficientus, ir tiesinės prognozės modelio parametrų įverčiams gauti naudojama Levinsono rekursija. Atspindžio koeficientų įverčiai gaunami rekursyviai minimizuojant prognozės klaidos energiją, esant skirtingoms prognozės eilėms. Skaičiuojant m-ojo atspindžio koeficiento įvertį laikoma, kad jau yra įvertinti (m – 1) eilės prognozės klaidos filtro koeficientai $\{\hat{a}_1^{(m-1)}, \hat{a}_2^{(m-1)}, \ldots, \hat{a}_{m-1}^{(m-1)}\}$, kurie jau buvo gauti minimizuojant $(m-1)$ eilės prognozės klaidos energiją. Burg pasiūlė įvertinti k_m minimizuojant tiesioginės ir atgalinės prognozės klaidų energijas. Vadinasi, kad gautume k_m įvertį, minimizuojame

$$J_m = \sum_{n=m}^{N-1} \left\{ (e^{(m)}[n])^2 + (r^{(m)}[n])^2 \right\} \tag{1}$$

ir tada

$$a_i^{(m)} = \begin{cases} a_i^{(m-1)} + k_m a_{m-i}^{(m-1)}, & i = 1,\ldots,m-1 \\ k_m, & i = m. \end{cases}$$

$(e^{(m)}[n])^2$ ir $(r^{(m)}[n])^2$ priklauso tiktai nuo k_m, nes (m – 1) eilės prognozės koeficientai jau yra įvertinti minimizuojant J_{m-1}. Įvertintos tiesioginės ir atgalinės prognozės klaidos yra

$$e^{(m)}[n] = x[n] + \sum_{i=1}^{m} a_i^{(m)} x[n-i];$$

$$r^{(m)}[n] = x[n-m] + \sum_{i=1}^{m} a_i^{(m)} x[n-m+i].$$

Grotelinio filtro priklausomybės, kurios nusako modelio eilės atnaujinimą tiesioginės ir atgalinės prognozės klaidoms, yra

$$e^{(m)}[n] = e^{(m-1)}[n] + k_m r^{(m-1)}[n-1],$$
$$n = m, m+1, \ldots, N-1;$$
$$r^{(m)}[n] = r^{(m-1)}[n-1] + k_m e^{(m-1)}[n],$$
$$n = m, m = 1, \ldots, N-1;$$

čia $e^{(0)}[n] = r^{(0)}[n] = x[n]$. Įrašę šias priklausomybes į (1), gauname vidutinę įvertintą prognozės energiją. Negana to, diferencijuodami J_m atžvilgiu k_m, išvestinę prilygindami nuliui ir spręsdami k_m atžvilgiu, gauname

$$k_m = \frac{-2\sum_{n=m}^{N-1} e^{(m-1)}[n] r^{(m-1)}[n-1]}{\sum_{n=m}^{N-1}\left\{(e^{(m-1)}[n])^2 + (r^{(m-1)}[n-1])^2\right\}}.$$

Tai ir yra *m*-ojo atspindžio koeficiento įvertinimas Burg metodu.

Marple metodas. Šis metodas (Marple, 1980), kaip ir Burg metodas, minimizuoja tiesioginės ir atgalinės prognozės klaidų energiją

$$e_M = \sum_{k=1}^{N-M} | f_{M,k} |^2 + \sum_{k=1}^{N-M} | b_{M,k} |^2 , \qquad (2)$$

kur tiesioginės prognozės klaida yra

$$f_{M,k} = x_{k+M} + \sum_{i=1}^{M} a_{M,i} x_{k+M-i} = \sum_{i=0}^{M} a_{M,i} x_{k+M-i} , \qquad (3)$$

atgalinės prognozės klaida

$$b_{M,k} = x_k + \sum_{i=1}^{M} a_{M,i} x_{k+i} = \sum_{i=0}^{M} a_{M,i} x_{k+i} , \qquad (4)$$

čia M yra prognozės eilė, N – analizės intervalo ilgis.

Ne taip kaip Burg metodas, šis metodas netaiko prielaidos, kad tiesioginės ir atgalinės prognozės klaidos tenkina Levinsono rekursijos apribojimus

$$f_{M,k} = f_{M-1,k+1} + a_{M,M} b_{M-1,k} ;$$
$$b_{M,k} = b_{M-1,k} + a_{M,M} f_{M-1,k+1} .$$

Naudojant Marple metodą, įrašius į (2) išraišką (3) ir (4), ieškoma e_M išvestinių $a_{M,1}, \ldots, a_{M,M}$ atžvilgiu ir jos prilyginamos nuliui. Gauname

$$\frac{\partial e_M}{\partial a_{M,i}} = 2\sum_{j=0}^{M} a_{M,j} r_M(i,j) = 0, \quad i = 1, \ldots, M; (a_{M,0} = 1),$$

čia $r_M(i,j) = \sum_{k=1}^{N-m} (x_{k+M-j} x_{k+M-i} + x_{k+j} x_{k+i}),$

$0 \le i, j \le M.$

Minimali prognozės klaidos energija yra

$$e_M = \sum_{j=0}^{M} a_{M,j} r_M(0, j) .$$

Šias išraiškas galima užrašyti matricine forma

$$\mathbf{R_M A_M} = \mathbf{E_M} ;$$

čia

$$\mathbf{A_M} = \begin{bmatrix} 1 \\ a_{M,1} \\ \vdots \\ a_{M,M} \end{bmatrix}, \quad \mathbf{E_M} = \begin{bmatrix} e_M \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$\mathbf{R_M} = \begin{bmatrix} r_M(0,0) & \cdots & r_M(0,M) \\ \vdots & & \vdots \\ r_M(M,0) & \cdots & r_M(M,M) \end{bmatrix}.$$

Šią lygčių sistemą galima spręsti naudojant klasikinius metodus, tačiau tuo atveju skaičiavimo operacijų skaičius yra proporcingas M^3. Jai spręsti Marple pasiūlė rekurentinį metodą, kurį naudojant operacijų skaičius yra proporcingas NM.

Split Levinson algoritmas. Split Levinson algoritmas (Delsarte ir kt., 1986) tiesinės prognozės modelio parametrams įvertinti buvo sukurtas siekiant sumažinti autokoreliacinio LPC parametrų vertinimo metodo skaičiavimo operacijų skaičių. Naudojant Split Levinson algoritmą, LPC parametrų įverčiai gaunami tie patys, kaip ir naudojant Levinson algoritmą, tačiau Split Levinson algoritmas reikalauja dvigubai mažesnio sandaugų skaičiaus negu Levinson algoritmas ir to paties sumų skaičiaus. Esminis skirtumas, kad Split Levinson algoritmas remiasi išsigimusių prognozės polinomų skaičiavimu. Jeigu turime tiesinės prognozės polinomų aibę

$A_k(z) = 1 + a_k(1)z^{-1} + a_k(2)z^{-2} + \cdots + a_k(k)z^{-k}$,

$k = 1, \cdots, p$

jie yra susieti priklausomybe

$$A_{k+1}(z) = A_k(z) + \rho_{k+1} z^{-(k+1)} A_k(z^{-1}); \quad (5)$$

čia $\rho_1, \rho_2, \ldots, \rho_p$ yra atspindžio koeficientai. Jeigu atspindžio koeficientą ρ_{k+1} prilyginsime 1 arba -1, iš (5) gausime du išsigimusius prognozės polinomus:

$$P_{k+1}(z) = A_k(z) + z^{-(k+1)} A_k(z^{-1}) =$$
$$= 1 + (a_k(1) + a_k(k))z^{-1} + (a_k(2) + a_k(k-1))z^{-2} +$$
$$+ \cdots + (a_k(k) + a_k(1))z^{-k} + z^{-(k+1)}$$

ir

$$Q_{k+1}(z) = A_k(z) - z^{-(k+1)} A_k(z^{-1}) =$$
$$= 1 + (a_k(1) - a_k(k))z^{-1} + (a_k(2) - a_k(k-1))z^{-2} +$$
$$+ \cdots + (a_k(k) - a_k(1))z^{-k} - z^{-(k+1)}.$$

Polinomas $P_{k+1}(z)$ yra simetrinis, $Q_{k+1}(z)$ – antisimetrinis. Todėl

$$A_{k+1}(z) = 1/2[P_{k+1}(z) + Q_{k+1}(z)].$$

Didinant modelio eilę rekurentiškai skaičiuojami išsigimę prognozės polinomai ir p-ajai eilei LPC modelio parametrai randami iš $p+1$ ir žemesnės eilės išsigimusių prognozės polinomų.

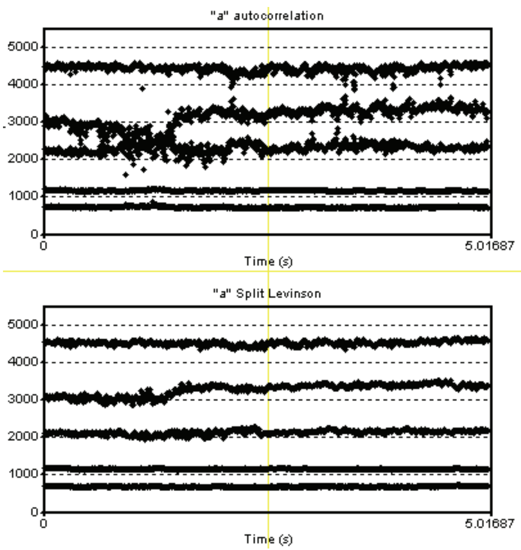
Modifikuotas Split Levinson algoritmas. Split Levinson algoritmas buvo modifikuotas (Willems, 1987) formančių trajektorijoms įvertinti. Šis algoritmas naudoja Split Levinson algoritmą kiekvienam kalbos signalo kadrui nustatyti fiksuotą spektro maksimumų skaičių. Vietoje šaknų ieškojimo standartiniame LPC polinome, kad nustatytume spektro maksimumų vietas, yra konstruojamas vadinamasis išsigimęs prognozės polinomas, kurio nuliai yra randami naudojant iteratyvią procedūrą. Visi šio išsigimusio prognozės polinomo nuliai yra ant vienetinio apskritimo ir rastų spektro maksimumų skaičius visais atvejais yra pusė LPC modelio eilės. Spektro maksimumų vietos yra traktuojamos kaip formantės, rastos pasitelkiant šį algoritmą.

Formantinių požymių išskyrimo metodų palyginimas

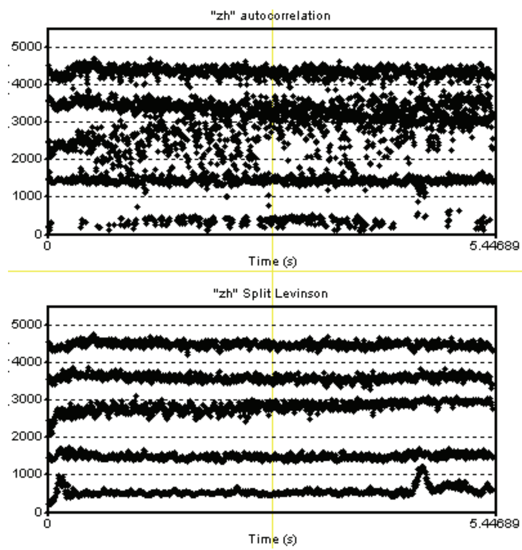
Eksperimentiškai buvo lyginami nagrinėti LPC parametrų vertinimo metodai: autokoreliacinis, kovariacinis, Burg, Marple ir modifikuotas Split Levinson algoritmas. Metodams palyginti buvo naudojama kalbos signalų analizės programinė įranga „Praat“ (Boersma ir kt., 2005).

Kadangi mūsų sukurtas formantinių požymių išskyrimo būdas (Lipeika, 2005) naudoja autokoreliacinį tiesinės prognozės modelio parametrų vertinimo metodą, šį metodą lyginome su kitais mūsų aprašytais tiesinės prognozės modelio parametrų vertinimo metodais. Lygindami autokoreliacinį tiesinės prognozės parametrų vertinimo metodą su kovariaciniu, Burg ir Marple metodais, esminių skirtumų tarp išskirtų formančių trajektorijų nepastebėjome. Pagal formančių dažnių išskyrimą šie metodai yra lygiaverčiai. Lygindami autokoreliacinį tiesinės prognozės parametrų vertinimo metodą su modifikuotu Split Levinson algoritmu pastebėjome, kad modifikuotu Split Levinson algoritmu gaunamos daug stabilesnės formančių trajektorijos.

Kad būtų aiškiau, pateikiame keletą autokoreliacinio ir modifikuoto Split Levinson metodų palyginimo pavyzdžių. Iš balsio „a“ išskirti formantiniai požymiai naudojant autokoreliacinį metodą (viršuje) ir modifikuotą Split Levinson algoritmą (apačioje) pavaizduoti 1 pav. Matome, kad naudojant modifikuotą Split Levinson algoritmą formantiniai požymiai išskiriami daug patikimiau negu autokoreliaciniu metodu. Analogiškas priebalsio „ž“ palyginimas pateiktas 2 pav. Rezultatas vėl rodo modifikuoto Split Levinson algoritmo pranašumą. Šio algoritmo pranašumas būdingas visiems lietuvių kalbos garsams.



1 pav. ***Iš balsio „a" išskirti formantiniai požymiai naudojant autokoreliacinį metodą (viršuje) ir modifikuotą Split Levinson algoritmą (apačioje)***



2 pav. ***Iš priebalsio „ž" išskirti formantiniai požymiai naudojant autokoreliacinį metodą (viršuje) ir modifikuotą Split Levinson algoritmą (apačioje)***

Išvados

Straipsnyje formančių trajektorijų išskyrimo kokybės požiūriu tiriami tiesinės prognozės modelio parametrų įvertinimo metodai. Išnagrinėti autokoreliacinio, kovariacinio, Burg, Marple metodų ir modifikuoto Split Levinson algoritmo ypatumai. Šie metodai palyginti pagal formančių trajektorijų išskyrimo kokybę ir nustatyta, kad autokoreliacinis, kovaraicinis, Burg ir Marple metodai pagal formančių trajektorijų išskyrimo kokybę beveik nesiskiria. Naudojant modifikuotą Split Levinson algoritmą gaunamos daug stabilesnės formančių trajektorijos, ir šį algoritmą siūlome taikyti tiesinės prognozės modelio parametrams vertinti iš kalbos signalo išskiriant formantinius požymius kalbai atpažinti.

LITERATŪRA

BOERSMA, P.; WEENINK, D. (2005). Praat: doing phonetics by computer. Prieiga per internetą: www.praat.org

De WET, F.; WEBER, K.; BOVES, L.; CRANEN, B.; BENGIO, S.; BOURLAND H. (2004). Evaluation of formant-like features on an automatic vowel classification task. *The Journal of the Acoustical Society of America*, vol. 116(3), p. 1781–1792.

DELSARTE, P.; GENIN, Y. (1986). The Split Levinson Algorithm. *IEEE Trans. on Acoustics, Speech and Signal Processing*, ASSP-34 (3), p. 470–478.

HUANG, X.; ACERO, A.; HON, H-W. (2001). *Spoken Language Processing*. Prentice Hall.

LIPEIKA, A. L. (2005). Formantiniai požymiai atpažįstant kalbą. *Informacijos mokslai*, t. 34, p. 215–219.

MARPLE, L. (1980). A New Autoregressive Spectrum Analysis Algorithm. *IEEE Trans. On Acoustics, Speech and Signal Processing*, ASSP-28 (4), p. 441–454.

RABINER, L. R.; JUANG, B. H. (1993). *Fundamentals of Speech Recognition*. Prentice-Hall.

WILLEMS, L.F. (1987). Robust Formant Analysis for Speech Synthesis Applications. *Proceedings of the European Conference on Speech Technology,* vol.1, p. 250–253.

FORMANT FEATURE EXTRACTION METHODS

Antanas Leonas Lipeika

S u m m a r y

Formant feature extraction is investigated in the paper. Extraction of formant features is based on calculating frequency positions of spectral peaks. The spectrum has been calculated from parameters of linear prediction model. Reliability of formant feature extraction depends on the method used for linear prediction model parameter estimation. The autocorrelation method previously used for linear prediction model parameter estimation was not reliable enough for formant feature extraction. Therefore we were looking for more reliable method of linear prediction model parameter estimation. The previously used autocorrelation method was compared with covariance, Burg, Marple methods and the modified Split Levinson algorithm. It was concluded, that autocorrelation, covariance, Burg and Marple methods are similar from the point of view of formant feature extraction. The modified Split Levinson algorithm provides the best formant feature estimates.