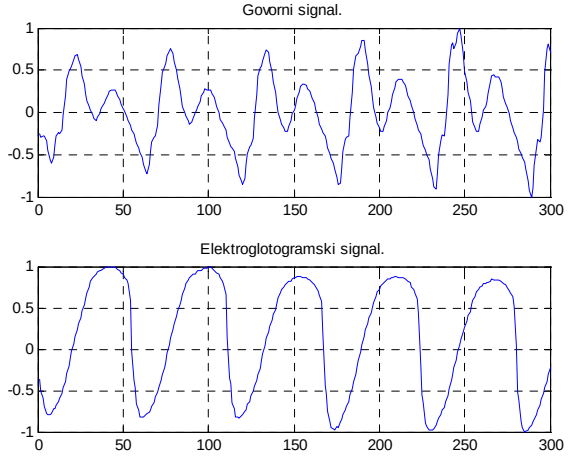
POBOLJŠANJE PREDIKCIJE GOVORA KORIŠĆENJEM ELEKTROGLOTOGRAMSKOG SIGNALA: UPOREDNA ANALIZA

Danijela Protić, *Institut za primenjenu matematiku i elektroniku, Beograd*
Milan Milosavljević, *Elektrotehnički fakultet, Beograd*

Sadržaj *- U radu je prikazan metod predikcije govornog signala, pri čemu je uz govorni signal, elelktroglotogramski signal korišćen kao dodatni ulaz. Primenjeni su linearni AR, ARX i ARMAX modeli i feedforward neuronska mreža za NNARMAX modelovanje. Izvedena je uporedna analiza prediktovanih govornih signala i grešaka predikcije. Rezultati su upoređeni sa rezultatima koji su dobijeni WRLS-VFF algoritmom,* [1].

1. UVOD

Oralni, vokalni, nazalni, glotalni trakt i pobudni signal definišu sistem za proizvođenje govora kod čoveka. Obrasci modelovanja ovog sistema na osnovu poznatog govornog signala poznati su a nameću ih pragovi zahteva za model, u zavisnosti od upotrebe. Snimak elektroglotogramskog signala (egg) omogućio je da se izvedu tačniji obrasci ovih modela. Snimci egg signala, govornog signala, baza podataka i Toolbox za MATLAB izvedeni su u [2] i korišćeni su u ovom radu. Na Sl.1. Prikazani su deo vokala i odgovarajućeg egg signala ženskog govornika iz internacionalnog fonetskog alfabeta, koji su normalizovani u granice [-1,1].



Sl.1. *Govorni signal i elektroglotogramski signal*

Sa Sl.1 se uočava jedna od glavnih karakteristika vokala - stacionarnost u relativno dugom vremenskom periodu, što omogućuje dobru procenu modela. Pobudni signal vokala je (kvazi)periodični niz impulsa, velike snage, obzirom da vazduh iz pluća nailazi na mali prečnik otvora glasnih žica. Kod ostalih fonema vreme trajanja je bitno kraće, procena modela je samim tim otežana. Pobuda kod ovih signala je šum ili kombinacija šuma i niza impulsa, ali je signal manje snage jer je otvor između glasnih žica veći. It tog razloga u ovom radu je za procenu modela korišćen digitalizovan, kvazistacionarni deo ženskog vokala od 600 odbiraka (diskretizacija analognog signala izvedena je sa učestanošću od f_s=10kHz) i odgovarajući deo egg signala. Modeli i metode koje su primenjene u ovom radu opisane su u drugom poglavlju. Treće poglavlje je prikaz eksperimentalnih rezultata. Poslednje poglavlje je zaključak.

2. MODELI

U radu su korišćeni linearni AR (Auto Regressive), ARX (Auto Regressive with eXtra input) i ARMAX (Auto Regressive Moving Average with eXtra input) modeli sistema za proizvođenje govornog signala koji su određeni izrazom (1):

AR:

$$y(n)+a_1y(n-1)+\ldots+a_{n_a}y(n-n_a)=e(n), \qquad (1a)$$

ARX:

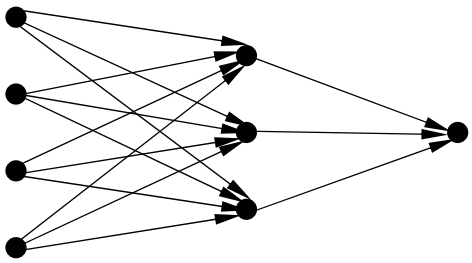
$$\begin{aligned} &y(n)+a_1y(n-1)+\ldots+a_{n_a}y(n-n_a)=\ldots \\ &\ldots=b_1u(n-1)+\ldots+b_{n_b}u(n-n_b)+e(n) \end{aligned}, \qquad (1b)$$

ARMAX:

$$\begin{aligned} &y(n)+a_1y(n-1)+\ldots+a_{n_a}y(n-n_a)=\ldots \\ &\ldots=b_1u(n-1)+\ldots+b_{n_b}u(n-n_b)+\ldots \\ &\ldots+e(n)+c_1e(n-1)+\ldots+c_{n_c}(n-n_c) \end{aligned}, \qquad (1c)$$

pri čemu je $y(n)$ odbirak signala govora, $e(n)$ je greška, dok su a_i (i=1…n_a) AR parametri, b_i (i=1…n_b) su parametri extra input - X dela, c_i (i=1…n_c) su MA parametri modela.

Nealinearni model datog sistema koji je korišćen u ovom radu izveden je feedforvard neuronskom mrežom sa jednim skrivenim slojem (sa tri neurona) i jednim izlaznim neuronom. Prenosna funkcija neurona je tangenshiperbolična za celu mrežu, dok je broj ulaza promenljiv. Sl.2. prikazuje izgled neuronske mreže feedforward tipa.



Sl.2. *Feedforward neuronska mreža sa jednim skrivenim slojem*

Prenosna funkcija ovakve strukture može se opisati izrazom (2):

$$g(y(n),\delta(n),n)=\varepsilon(n), \qquad (2)$$

$\delta^T(n)$ je vektor ulaza u neuronsku mrežu, $\varepsilon(n)$ je greška a g je nelinearna parametarska prenosna funkcija. Za slučaj neuronske mreže sa Sl.2. izlaz ovakve strukture ima oblik koji je definisan izrazom (3):

$$y_i(\mathbf{w},\mathbf{W}) = F_i\left(\sum_{j=1}^{q} W_{ij} f_j\left(\sum_{l=1}^{m} w_{ij} z_l + w_{j0}\right) + W_{f0}\right), \quad (3)$$

pri čemu je y_i odziv, **w** i **W** su matrice parametara neuronske mreže, f_j i F_i su (tangenshiperbolične) prenosne funkcije skrivenog i izlaznog sloja, respektivno, q je broj elemenata skrivenog a m je broj elemenata ulaznog sloja. Poznato je da se ovakvom strukturom može modelovati bilo koja nelinearna funkcija, sa proizvoljnom tačnošću, ukoliko postoji sloboda u izboru matrica parametara i prenosnih funkcija po slojevima, [3, 4, 5, 6]. Početne vrednosti matrica parametara korišćenih neuronskih modela izabrane su tako da su početne vrednosti parametara realni slučajni brojevi u granicama [-1,1]. Primenjen je gradijentni metod (gradient descent) u proceni greške obučavanja i BPA (Back Propagation Alogorithm) algoritam obučavanja. Skup metoda obučavanja i testiranja ovakvih neuronskih mreža definiše NNSYSID Toolbox za MATLAB 7.0. i korišćen je u ovom radu. Modelovane su NNAR, NNARX i NNARMAX feedforvard neuronske mreže (NN je oznaka neuronske mreže a ostale oznake odgovaraju oznakama u linearnim modelima).

Pored rezultantnih signala predikcije, dobijene greške predikcije poređene su sa rezultatom WRLS-VFF (Weighted Recursive Least Square algorithm with a Variable Forgetting Factor) algoritma, [1]. Ovaj algoritam podrazumeva da je govorni signal opisan ARMA modelom, pri čemu je AR deo određen prethodnim odbircima govornog signala dok MA deo određuje tip pobude, koji se procenjuje na osnovu procene karakteristike signala, [7] i postavlja da bude rezidualna greška ili greška predikcije. Poređenje grešaka predikcije modela u ovom radu izvedeno je sa rezultatom WRLS-VFF algoritma jer je ovaj algoritam pokazao veću tačnost od niza standardnih metoda. Na drugoj strani, u WRLS-VFF algoritmu nije primenjen dodatni ulaz u vidu egg signala, što je omogućilo da se direktno proceni uticaj egg signala na smanjenje ukupne greške.

3. EKSPERIMENTALNI REZULTATI

U eksperimentima su korišćeni linearni modeli (4):

ARX:

$$n_a = 14,\ n_b = 4,$$

ARMAX:

$$n_a = 14, n_b = 4,\ n_c = 1, \quad (4)$$

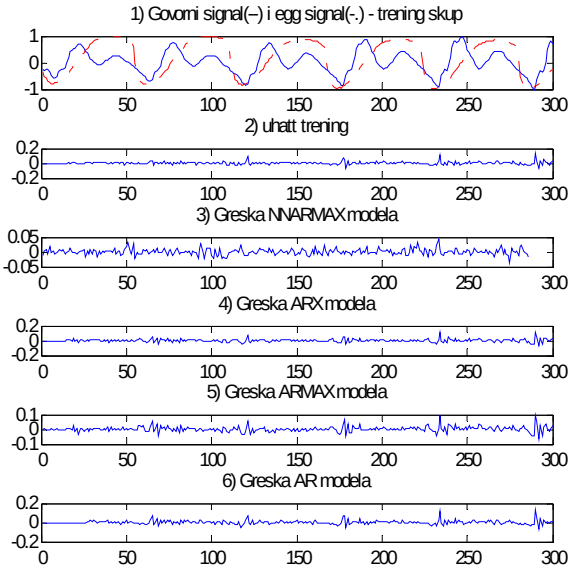
i AR model visokog reda (5):

AR:

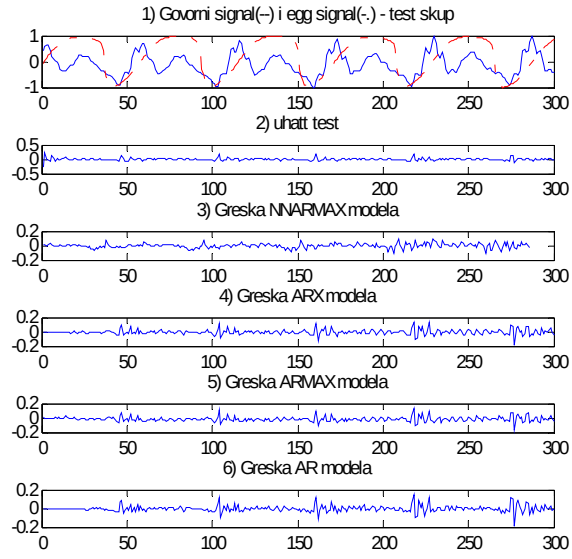
$$N = 25. \quad (5)$$

Veliki broj eksperimenata u kojima se koristi model vokalno-nazalnog trakta koristi AR model 10-tog do 16-tog reda, u zavisnosti od frekvencijskog opsega posmatranog govornog signala (dva pola za, približno, svakih [(2*n+1)*500]Hz, n = ..., -1, 0, 1, ...) [1, 3]. U ovom slučaju procena greške predikcije izvedena je na AR modelu višeg reda zbog adekvatnog poređenja sa greškama ostalih modela, obzirom na nešto veći red ovih modela. Takođe, obzirom na jednostavnost strukture ovih modela, ukoliko bi se pokazalo da AR model visokog reda daje grešku čija bi vrednost bila u okvirima drugih procenjenih grešaka, bilo bi nepotrebno uzimati elektroglotogramski signal u razmatranje, obzirom na činjenicu da je govorni signal dostupniji i široko primenjen u praksi.

Kod nelinearnih modela korišćeni su isti redovi modela kao i za odgovarajuće linearne modele i odgovarajući signali su dovedeni na ulaz feedforward neuronske mreže sa tri parametra jednog skrivenog sloja i jednim elementom izlaznog sloja (3). U eksperimentima je korišćen skup od 600 odbiraka normalizovanog govornog signala, i odgovarajući skup egg signala. Za obučavajući skup korišćeno je prvih 300 odbiraka signala, za test skup korišćeno je narednih 300 odbiraka. Signal izlaza WRLS-VFF algoritma (obeležen je sa u_hatt, što je standardna oznaka u [1]) takođe se sastoji od odgovarajućih trening i test signala, u ukupnoj dužini 600 odbiraka. Sl.3. daje prikaz obučavajućeg skupa govornog signala i egg signala, potom u_hatt signala, i grešaka NNARMAX, ARX, ARMAX i AR modela, respektivno. Na Sl.4 je prikazan odgovarajući niz signala za testirajući skup.



Sl.3. *Govorni i egg signal (1), u_hatt (2), greške NNARMAX, ARX, ARMAX i AR modela (3),(4),(5),(6), obučavajući skup*



Sl.4. *Govorni i egg signal (1), u_hatt (2), greške NNARMAX, ARX, ARMAX i AR modela (3),(4),(5),(6), test skup*

Signali grešaka ukazuju na momenat otvaranja odn zatvaranja glasnih žica, što odgovara promeni govornog i egg

signala sa Sl3 i Sl4. i ukazuje na periodičnu povorku impulsa pobude. Tabelom 1. prikazane su minimalne odn maxsimalne vrednosti u_hatt signala i grešaka svih modela za obučavajući i test skup. Obzirom da je signal govora normalizovan u granice [-1,1] rezultati u tabeli istovremeno prikazuju procentualnu vrednost signala greške u odnosu na odgovarajući govorni signal.

Tabela 1. *Minimalne/maxskimalne vrednosti grešaka*

	Obučavajući skup		**Test skup**	
Signal greške	*min*	*max*	*min*	*max*
u_hatt	-0.0646	0.1285	-0.2646	0.2305
ARMAX model	-0.0614	0.0996	-0.1777	0.1213
AR model	-0.0737	0.1251	-0.1956	0.1411
ARX model	-0.0738	0.1084	-0.1893	0.1308
NNARMAX	-0.0366	0.0440	-0.1161	0.0927

Rezultati ukazuju na činjenicu da greška ARMAX modela najviše odgovara u_hatt signalu, što se obzirom na ARMA model WRLS-VFF algoritma moglo očekivati. Vrednosti ovih signala gotovo su identične, međutim greška na testirajućem skupu ARMAX modela je bitno manja. Na drugoj strani AR model visokog reda i ARX model na oba skupa pokazuju gotovo iste rezultate, uz činjenicu da je greška ARX modela neznatno niža. Linearni modeli pokazuju poznate osobine koje diktira njihova struktura, odn greška je najmanja kod ARMAX modela, nakon čega sledi nešto veća greška ARX i AR modela, respektivno. U slučaju neuronske mreže greška na obučavajućem skupu, kao i na test skupu je bitno manja od grešaka linearnih modela. Međutim, može se primetiti da je kod linearnih modela greška kod test skupova oko dva puta veća nego kod grešaka trening skupova. Greška test skupa kod neuronske mreže je 2.6 puta veća od greške obučavanja. Rezultat upućuje na robusnost linearnih modela u odnosu na nelinearne modele, iako je egg signal korišćen kao dodatni ulaz u procesu obučavanja neuronskih mreža.

Gradijentni metod koji je korišćen u eksperimentima podrazumeva procenu greške signala oblika (6):

$$e(n)=\frac{\Delta y}{\Delta n}=\frac{y(n)-y(n-1)}{\Delta n}\Big|_{n=2,3,\ldots\,/\,\Delta n=1,\forall n}=\cdots$$
$$\cdots=\frac{y(n)-y(n-1)}{1}=y(n)-y(n-1)=\Delta y(n)\big|_{\forall n} \quad (6)$$

Greška $e(n)$ ima oblik reziduala trenutne vrednosti signala $y(n)$ i vrednosti signala koji je procenjen u prethodnoj mernoj instanci $y(n\text{-}1)$. Iz tog razloga izveden je niz eksperimenata u kojima se iz poznate vrednosti greške računa nepoznata vrednost signala, obzirom da je u određenim situacijama dostupna greška signala, dok sam signal nije dostupan. Potrebno je voditi računa da je kod linearnih/nelinearnih modela greška n-tog odbirka određena procenom prethodnih vrednosti odbiraka, koje su određene karakteristikama (strukturom) modela, i da ova činjenica može izazvati izvesne netačnosti rezultata. Ukoliko je greška definisana kao (6) onda važi (7):

$$\begin{aligned} &y(n)=y(n-1)+\Delta y(n),\\ &y(n-1)=y(n-2)+\Delta y(n-1),\\ &\vdots\\ &y(2)=y(1)+\Delta y(1),\\ &\Rightarrow\\ &y(n)=y(1)+\Delta y(1)+\cdots+\Delta y(n) \end{aligned} \quad , \quad (7)$$

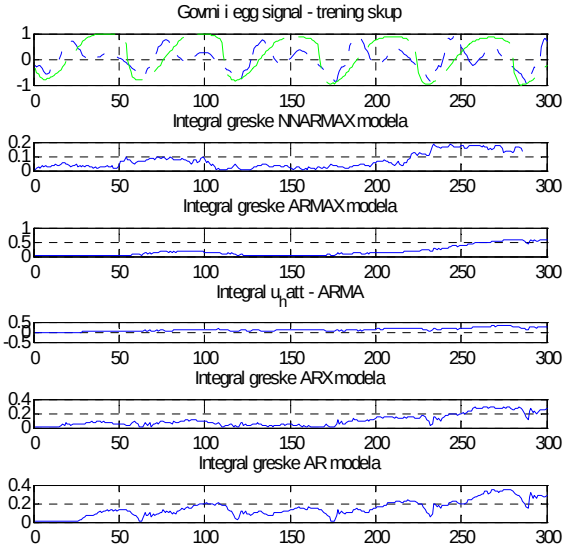
Obzirom da iz (6) važi:

$$e(n)=\Delta y(n)\big|_{\forall n} \quad (8)$$

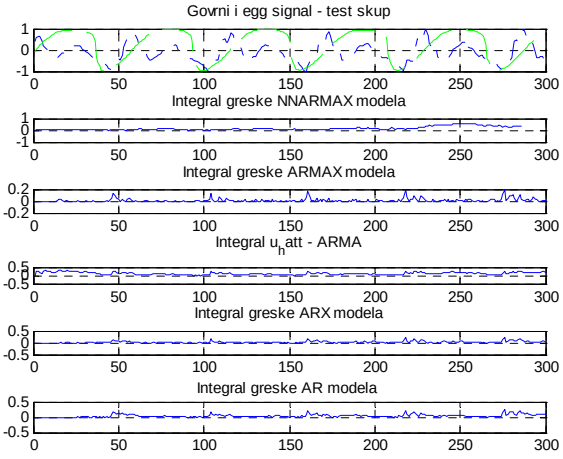
onda je procenjena vrednost signala, na osnovu (6-8) određena izrazom (9):

$$y(n)=y(1)+e(1)+\cdots+e(n)\Big|_{y(1)=\frac{e(1)+\cdots+e(n)}{n}} . \quad (9)$$

Problem koji se kod nepoznatog signala javlja je kako odrediti njegovu početnu vrednost $y(1)$. U eksperimentima izvedenim na normalizovanom govornom signalu srednja vrednost signala greške uzeta je kao početna vrednost $y(1)$. Ukoliko se ovaj uslov ne ispuni, vrednosti procenjenog signala imaju pomerene amplitude, ali je oblik procenjenog signala isti. Na osnovu izraza (6-9) eksperiment koji je izveden na signalu greške predikcije razmatranih modela dao je rezultat sa Sl.5., za trening signale i greške, dok je Sl.6. prikaz istovetnog eksperimenta izvedenog na test signalima.



Sl.5. Govor i egg(1), integrali grešaka za NNARMAX(2), ARMAX(3), ARMA(4), ARX(5), AR(6) model, trening skup



Sl.6. Govor i egg(1), integrali grešaka za NNARMAX(2), ARMAX(3), ARMA(4), ARX(5), AR(6) model, test skup

Rezultati ovih eksperimenata zanimljivi su jer ukazuju na trade-off između greške modela i izvođenja signala integraljenjem na osnovu greške modela. Kada je trening skup u pitanju, AR model pokazuje najveću grešku, međutim procena nepoznatog izlaznog signala izvodi se na osnovu

njegovih prethodnih n_a odbiraka, nema dodatnog ulaza u model, i na taj način je vrednost integrala greške najsličnija govornom signalu. Kako se povećava red modela ili ukoliko se modelu dodaju X i MA komponente, tako je izlazni signal "otežana" suma većeg broja različitih (nepoznatih) komponenti signala, koji se u principu teško mogu razdvojiti. Kod nelinearnih struktura, koje proizvode najmanje greške predikcije, integral ovih grešaka nema realnog smisla ukoliko se izvodi iz izraza (6-9). Ova pojava rezultat je gradijentnog metoda za proračun greške na velikom broju parametara i sa nelinearnim prenosnim funkcijama [4,6]. Kada se posmatra test skup linearni modeli pokazuju sličnost rezultata. Greške ovih modela računate na test skupu imaju gotovo iste vrednosti, što je prikazano Tabelom1., i gotovo identične oblike signala - Sl.4. Rezultat integraljenja ovih grešaka ukazuje na momente nagle promene signala na oblik rezultantnog (zbirnog) signala koji se formira od test signala govora i odgovarajućeg egg signala, dok se kod AR modela može uočiti i pojava da integral greške prati promene govornog test signala. Neuronska mreža i dalje ne pokazuje rezultat kojim bi se mogao odrediti oblik ulaznih signala. Na osnovu svega navedenog dolazi se do činjenice da je kod linearnog/nelinearnog modelovanja u procesu obrade govornog signala dobro koristiti složenije modele koji daju manje greške, međutim ukoliko je potrebno izvesti procenu signala na osnovu greške modela, integraljanjem, najbolje je iskoristiti rezultate koje daju najjednostavniji modeli.

4. ZAKLJUČAK

Rad prikazuje rezultate eksperimenata na linearnim odn nelinearnim modela u kojima je, uz govorni signal, korišćen i elektroglotogramski signal. Ovi signali korišćeni su za trening i testiranje AR, ARX, ARMAX i NNARMAX modela. Izlaz WRLS-VFF algoritma (ARMA struktra) korišćen je za poređenje dobijenih rezultata, obzirom na tačnije rezultate procene, od rezultata klasičnih metoda. Rezultati na AR, ARX, ARMAX i NNARMAX modelima pokazuju da u procesu obučavanja neuronska mreža feedforward tipa daje najbolje rezultate, dok greška raste kako se struktura modela uprošćava. Pretpostavka da bi primena elektroglotogramskog signala kao dodatnog ulaza trebalo da poboljša rezultate procene, odn manju grešku, potvrđena je slikama Sl.3. i Sl.4. dok se rezultati minimalnih/maksimalnih vrednosti grešaka nalaze u Tabeli1.

Pored navedenih rezultata u radu je primenjen metod procene nepoznatog signala ukoliko je poznata greška procene. Polazni signal određuje se na osnovu zbira svih prethodnih grešaka i pretpostavljene početne vrednosti signala.

Obzirom na nepoznavanje početne vrednosti, i karakteristike ispitivanih signala, predloženo je da ta vrednost bude srednja vrednost signala greške. Ova pretpostavka izvedena je iz činjenice da signal greške "prati" oblikom govorni signal, samo je njegova amplituda manja. Rezultat procena signala integraljenjem greške pokazuje da najjednostavnij modeli, kod kojih je greška najveća, daju dobre rezultate u proceni nepoznatog signala. Kod složenih modela, pogotovo kod nelinearnih struktura, ova metoda je teško primenjiva u praksi, obzirom na kombinovane ulazne vrednosti, nelinearne prenosne funkcije i sl.

LITERATURA

[1] Childers, D.G., Principe, J.C. and Thing, Z.T., *"Adaptive WRLS_VFF for Speech Analysis"*, *IEEE Transactions on Speech and Audio Processing,* 209-213, 1995.

[2] Childers, Donald G., *Speech processing and syntesis toolboxes,* Wiley, 2002.

[3] Ljung L., *System Identification: Theory for the User*, Prentice Hall Inc., 1987.

[4] *Neural Networks for Signal Processing*, Claus Svarer, Technical University of Denmark, 1995.

[5] *"Final Prediction Error of Serbian Vowel Neural Network Model"*, Arsenijević D., Milosavljević M., TELFOR, 1998.

[6] *Fitting Autoregressive Models for Prediction,* Akaike H., Ann. Ins. Stat. Mat.,1969.

[7] Childers, D.G., Holm, M., and Larar, J.N. *"Silent and Voiced/Unvoiced/Mixed Exitation (Four Way) Classification of Speech" 37*, *1771-1774*, 1989.

Abstract - This paper represents comparative analisys of different speech signal prediction methods. Improvement is made by extra input electroglottographic signal usage. Linear AR, ARX, and ARMAX models as well as feedforward neural network have been used for experiments. Prediction errors of distincit models were compared. Results were paralleled with WRLS-VFF algorithm ouput signal, [1].

SPEECH SIGNAL PREDICTION IMPROVEMENT BY ELECTROGLOTTOGRAPHIC SIGNAL: COMPARATIVE ANALISYS

Danijela Protić, Milan Milosavljević