

Implementacija IIR filtara u aritmetici fiksne tačke

Nevena I. Todorović

Sadržaj – Ovaj rad se bavi problematikom softverske implementacije IIR filtara višeg reda, realizovanih preko kaskadne veze sekcija drugog reda. U radu su upoređeni rezultati različitih realizacija sekcija drugog reda, direktne i kanonične. Pri softverskoj implementaciji ispoljavaju se efekti konačne dužine reči koji unose razlike u rezultatima za različite sekcije drugog reda.

Ključne reči – Efekti konačne dužine reči, IIR filtri

I. UVOD

Digitalni filtri beskonačnog impulsnog odziva (IIR, *Infinite Impulse Response*) predstavljaju značajnu klasu digitalnih filtara. Funkcija prenosa IIR filtara je racionalna funkcija kompleksne promenljive z :

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^N b_k z^{-k}}{1 + \sum_{k=0}^N a_k z^{-k}} \quad (1)$$

gde su b_k, a_k koeficijenti filtra, a N red filtra.

Realizacija digitalnih filtara može biti hardverska ili softverska. U ovom radu biće izloženi rezultati softverske realizacije IIR filtara, implementirane u razvojnom okruženju Visual DSP++ 4.0, namenjenom za Analog Devices procesore signala.

Rad je organizovan na sledeći način. U drugom poglavlju predstavljena je realizacija IIR filtara višeg reda, a u trećem različite realizacije sekcija drugog reda koje se povezuju pri formiranju filtra višeg reda. Četvrto poglavlje se bavi problemom prekoračenja i kako ga eliminisati, a peto opisuje efekte konačne dužine reči koji se ispoljavaju. U šestom je opisan način realizacije praktičnog dela rada i analizirani rezultati rada.

II. REALIZACIJA IIR FILTARA VIŠEG REDA

Direktna realizacija IIR filtra podrazumeva da su konstante množenja koeficijenti funkcije prenosa.

Konstante množenja se, pri implementaciji, predstavljaju određenim brojem bita [1] (konačna dužina

reči), što unosi odstupanja realizovane od projektovane karakteristike filtra. Kao što se iz funkcije prenosa IIR filtra (1) vidi, IIR filter ima nule i polove u kompleksnoj z ravni. Uslov stabilnosti filtra je da se svi polovi nalaze unutar jediničnog kruga u kompleksnoj z ravni. Usled kvantizacije koeficijenata dolazi do pomeranja nula i polova. Pomeranje polova može dovesti do nestabilnosti filtra ukoliko se oni nađu izvan jediničnog kruga.

Ako se funkcija prenosa filtra $H(z)$ predstavi preko nula i polova $z_k, p_k, k=1, \dots, N$ [1]

$$H(z) = b_0 \frac{\prod_{k=1}^N (z - z_k)}{\prod_{k=1}^N (z - p_k)} \quad (2)$$

iz ovog izraza može se izvesti izraz za osetljivost nula i polova. Izraz za osetljivost pola p_i u odnosu na koeficijent filtra a_k glasi [1]

$$S_{a_k}^{p_i} = \frac{\partial p_i}{\partial a_k} = - \frac{p_i^{N-k}}{\prod_{i=1, j \neq i}^N (p_i - p_j)} \quad (3)$$

Sledi da je osetljivost velika ukoliko je razlika $p_i - p_j$ mala, što je slučaj kod filtara višeg reda. Filtri drugog reda imaju najmanju razliku $p_i - p_j$. Iz ovog razloga filtri višeg reda se dekomponuju na sekcije prvog i drugog reda, koje se vezuju paralelno ili kaskadno.

Ovde je razmatrana kaskadna realizaciona struktura koja se dobija faktorizacijom funkcije prenosa [1]

$$H(z) = \prod_{i=1}^K H_i(z) \quad (4)$$

gde je K broj sekcija, a $H_i(z)$ funkcija prenosa i -te sekcije nižeg reda. U praktičnim realizacijama koriste se sekcije prvog i drugog reda.

III. RAZLIČITE REALIZACIJE SEKCIJA DRUGOG REDA

Iako imaju istu funkciju prenosa

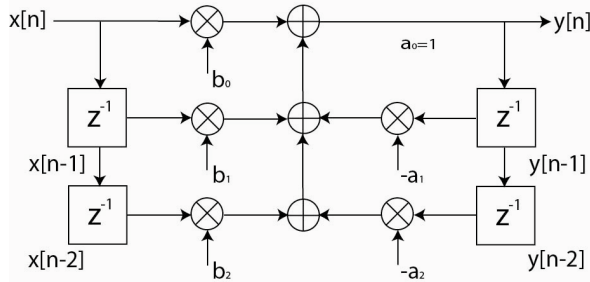
$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \quad (5)$$

i identične su u uslovima beskonačne preciznosti (vrednost odbirka i koeficijenta funkcije prenosa može biti bilo koja vrednost od $-\infty$ do $+\infty$), u uslovima konačne preciznosti različite realizacije sekcija drugog reda daju različite

N. I. Todorović, Elektrotehnički fakultet, Bulevar Kralja Aleksandra 73, 11000 Beograd, Srbija (tel: +381646356557, e-mail: nevena.todorovic@gmail.com)

Mentor: Prof. Miroslav Lutovac, Elektrotehnički fakultet, Bulevar Kralja Aleksandra 73, 11000 Beograd, Srbija (tel: +381113218348, e-mail: lutovac@etf.bg.ac.yu)

rezultate što ovaj rad ima za cilj da pokaže. Direktna realizaciona struktura prikazana je na Sl. 1 i poznata je pod nazivom direktna struktura forma I.

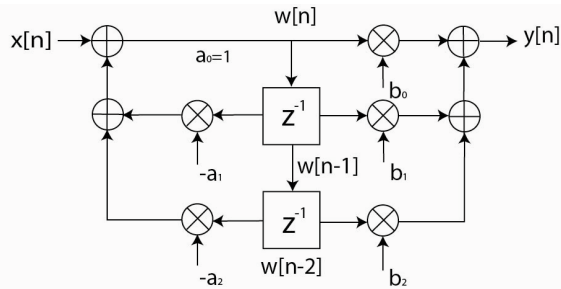


Sl. 1. Direktna realizacija sekcije drugog reda

Direktna struktura forma I može se softverski realizovati prema odgovarajućoj diferencnoj jednačini [2]:

$$y[n] = \sum_{k=0}^2 b_k x[n-k] - \sum_{k=0}^2 a_k y[n-k] \quad (6)$$

Ova struktura ima dvostruko više elemenata za kašnjenje nego što je potrebno pa se naziva i nekanonična. Na Sl. 2 je prikazana direktna kanonična struktura poznata i kao direktna struktura forma II.



Sl. 2. Direktna kanonična realizacija sekcije drugog reda

Direktna struktura forma II može se softverski realizovati prema izrazima:

$$w[n] = x[n] - \sum_{k=1}^2 a_k w[n-k] \quad (7a)$$

$$y[n] = \sum_{k=0}^2 b_k w[n-k] \quad (7b)$$

gde je $w[n]$ pomoćni niz čiji se elementi čuvaju u liniji za kašnjenje. Iz izraza (7a) i (7b) jasno je da se elementi niza $w[n]$ skraćuju na dužinu kodne reči što unosi odstupanje u odnosu na teorijski rezultat.

IV. PROBLEM PREKORAČENJA I KAKO GA IZBEĆI

Prekoračenje pri realizaciji IIR filtera nastupa kad je rezultat veći od sistemom definisanog dinamičkog opsega. Najjednostavniji korekcionni metod za izbegavanje prekoračenja jeste skaliranje [3]. Skaliranje se vrši ispred sekcije tj. skalira se ulaz, a izlaz (tj. koeficijenti u numeratoru) se množi sa faktorom kojim je skaliran ulaz da bi celokupno pojačanje filtera ostalo isto.

Kod kanonične realizacije, Sl. 2. prekoračenje se može javiti u tačkama $w[n]$ i $y[n]$. Postoji više metoda za

izračunavanje faktora skaliranja. Ovde je odabran metod poznat pod nazivom "II norma" [4] jer je faktor koji se dobija ovim postupkom veći od faktora koji se dobijaju drugim postupcima i pouzdano obezbeđuje da ne dođe do prekoračenja. Faktor skaliranja se određuje iz formule

$$s = \max \left\{ \sum_{k=0}^{\infty} |h_w(k)|, \sum_{k=0}^{\infty} |h_y(k)| \right\} \quad (8)$$

gde je $h_w(k)$ impulsni odziv filtera do tačke $w[n]$, a $h_y(k)$ do $y[n]$. Ulaz prve sekcije drugog reda se skalira sa s , a koeficijenti u imeniocu se množe sa istim brojem. Postupak se ponavlja za sve naredne sekcije, za svaku se računa njen korekcionni faktor s .

U slučaju direktne realizacije, pošto filter ima jedan akumulator, prekoračenja unutar sekcije drugog reda nisu problem tako da skaliranje ulaza nije neophodno. To je još jedna prednost ove realizacije. Jedino treba voditi računa da li dolazi do prekoračenja na izlazu iz svake sekcije, $y[n]$.

V. EFEKTI KONAČNE DUŽINE REČI

U ovom radu je korišćena fiksna dužina reči od 16 bita sa aritmetikom sa fiksnom tačkom [5]. Q format je zapis koji pokazuje poziciju tačke tj. koliko je bita levo, a koliko desno od tačke. Q1.15 ili uobičajniji zapis Q.15 govori da je 1 bit za znak, a 15 iza tačke predstavljaju vrednost.

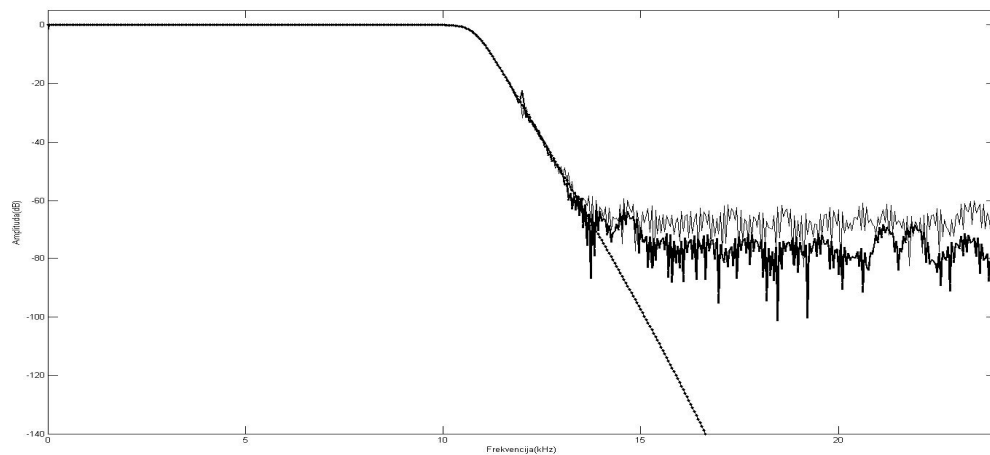
Množenjem dva Q.15 broja kao rezultat se dobija Q.30 broj. DSP (*Digital Signal Processor*) ima akumulator veći od dužine duple reči tako da se u njega mogu smeštati međurezultati predstavljeni sa više od 32 bita, bez odsecanja.

Ova mogućnost koristi direktnoj realizaciji kod koje se u svakoj iteraciji, u akumulatoru, na prethodnu vrednost dodaje naredna i tek se finalni rezultat predstavlja sa 16 bita najveće težine tako što se niži biti odbacuju zaokruživanjem ili odsecanjem rezultata. Time što se tek finalni rezultat zaokružuje pravi se manja greška i manje odstupanje od projektovane karakteristike.

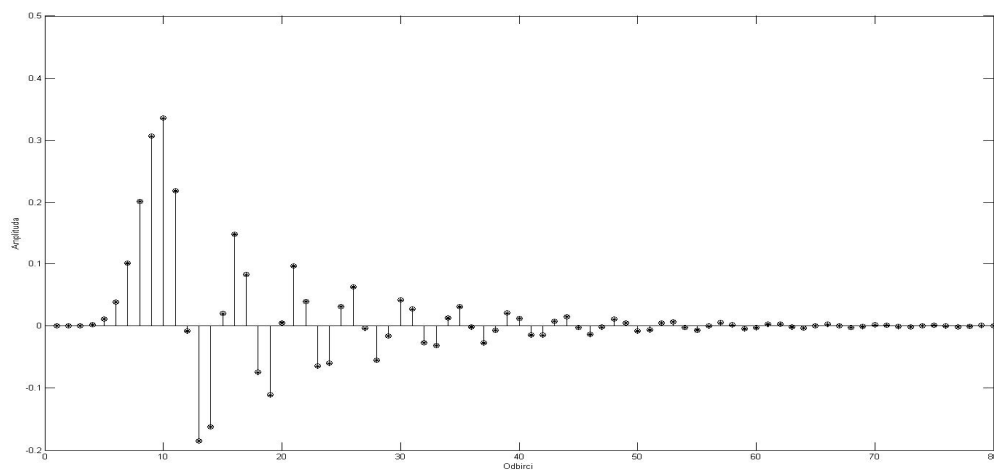
Kod kanonične strukture realizacija je takva da je neophodno smeštati međurezultate, koji su često predstavljeni većim brojem bita nego što je dužina reči, u memoriju. Pri tome se svaki međurezultat mora svesti na dužinu reči i tako se greška unosi u svaki međurezultat. Finalni rezultat zavisi od međurezultata i greška se akumulira. U ovoj realizaciji, pri softverskoj implementaciji, greška je veća nego što je to slučaj sa direktnom. Takođe, greška raste se porastom reda filtera.

VI. REALIZACIJA I REZULTATI

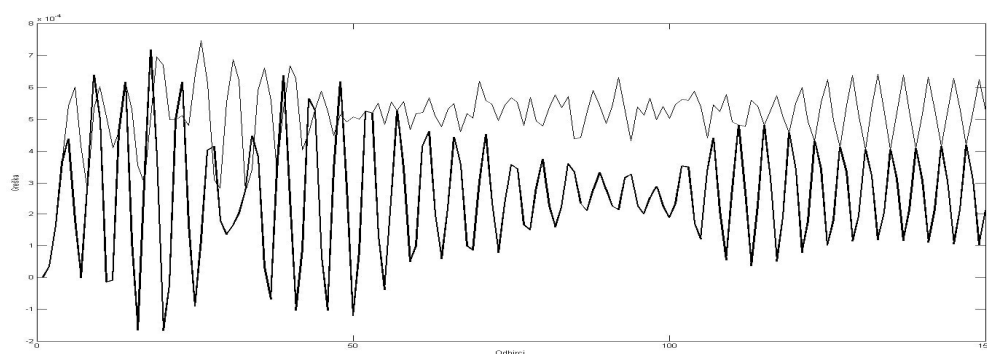
U programskom paketu Matlab 2007 napravljen je IIR filter, propusnik niskih frekvencija, koji je upotrebljen da ilustruje razlike u rezultatima direktne i kanonične realizacije sekcija drugog reda odnosno da bude referentni. Filter je napravljen koristeći Matlab FdaTool. U Matlabu je razložen na sekcije drugog reda i vršeno je skaliranje II normom (8) da bi se izbeglo prekoračenje, odnosno da samo skaliranje ne unosi dodatnu grešku. Izvršena je



Sl. 3. Amplituske karakteristike filtera



Sl. 4. Impulsni odziv filtera



Sl. 5. Odstupanja vrednosti odbiraka impulsnog odziva dve realizacije od referentnog impulsnog odziva

kvantizacija. Kvantizovani koeficijenti sekcija drugog reda eksportovani su iz Matlaba u fajl iz koga se učitavaju u Visual DSP++ 4.0.

ADSP-BF533 EZ-KIT Lite je razvojni sistem Analog Devices za Blackfin procesore [6]. Razvojna ploča je dizajnirana da se koristi sa Visual DSP++ razvojnim

okruženjem koje omogućava kreiranje, kompajliranje, linkovanje programa pisanih u C, C++ i ADSP-BF533 assembleru, kao i debugovanje. Pristup ADSP-BF533 procesoru i periferijama razvojne ploče od strane PC-a se obavlja preko USB porta ili opcionog JTAG.

U Visual DSP++ 4.0 su realizovane dve realizacije. U

prvoj su sekcije drugog reda u direktnoj realizaciji, a u drugoj u kanoničnoj. Za obe realizacije napisane su i testirane C++ rutine koje predstavljaju softversku kaskadnu realizaciju IIR filtra prilagođenu ADSP-BF533 procesoru.

Iz arhitekture direktne realizacije, Sl. 1 se vidi da se izlaz sekcije može izračunati u jednoj programskoj petlji, sabiranjem proizvoda 16-bitnih koeficijenata i odgovarajućih 16-bitnih odbiraka ulaznog i izlaznog signala. Izlaz sekcije se zaokružuje sa 32 bita na 16 zbog čega nastaje greška. Međutim, kod kanonične realizacije postoje dva zaokruživanja zbog čega je i greška veća. Zbog njene arhitekture, Sl. 2, izlaz sekcije se ne može izračunati u jednoj petlji. U prvoj iteraciji se sabiraju proizvodi koeficijenata imenioca i zakašnjenih odbiraka pomoćnog niza $w[n]$ i zbir je predstavljen sa 32 bita. Da bi se taj rezultat sabrao sa 16-bitnim aktuelnim odbirkom ulaznog signala, i potom taj rezultat smestio u memoriju, potrebno ga je zaokružiti na 16 bita. Tu dolazi do prvog odstupanja koje dalje propagira. Zatim se taj međurezultat množi sa koeficijentima brojioca i izlaz sekcije je zbir 32-bitnih brojeva. Izlaz se skraćuje na 16 bita i to je drugo zaokruživanje u ovoj realizaciji.

Na Sl. 3 prikazane su amplitudske karakteristike filtra sve tri realizacije. Referentni filter, napravljen u FdaTool, je filter propusnik niskih frekvencija, 20. reda, dekomponovan na sekcije drugog reda. Učestanost odabiranja je 48kHz, a granična učestanost (učestanost na kojoj je slabljenje 3dB) je 10,8kHz. Realizovan je u uslovima "beskonačne" preciznosti, a kanonična i direktna realizacija u uslovima konačne preciznosti određene dužinom digitalne reči od 16 bita. Tanjom linijom je prikazana amplitudska karakteristika kanonično realizovanog filtra, a debljom direktnog. Sa slike se vidi da je potiskivanje u nepropusnom opsegu direktne realizacije veće u odnosu na kanoničnu. Potiskivanje referentnog filtra u nepropusnom opsegu je mnogo veće od ove dve realizacije jer su one ograničene mogućnostima tj. preciznošću koju realizacija sa 16 bita pruža.

Na Sl. 4 je dat impulsni odziv sve tri realizacije predstavljen u 80 odbiraka. Sve tri realizacije imaju isti oblik impulsnog odziva, ali postoje odstupanja u vrednostima odbiraka koja su veća kod kanonične nego kod direktne realizacije. Te razlike su grafički prikazane na Sl. 5. Debljom linijom je predstavljeno odstupanje direktne realizacije od referentne, a tanjom kanonične od referentne. Sa grafika se vidi da je greška veća u slučaju kanonične realizacije, odnosno, vrednosti odbiraka impulsnog odziva filtra realizovanog direktno manje odstupaju od vrednosti odbiraka impulsnog odziva referentnog filtra nego što je to slučaj sa kanoničnom realizacijom.

VII. ZAKLJUČAK

Usled različitih realizacija ove dve strukture i efekata konačne dužine reči koji se pri tom ispoljavaju, dobijaju se i različiti rezultati pri softverskoj implementaciji IIR filtera. Upoređujući ove dve realizacije, na osnovu prethodno prikazanih rezultata, zaključuje se da je direktna

realizacija sekcija drugog reda vezanih kaskadno bolja, za softversku implementaciju, od kanonične realizacije.

ZAHVALNICA

Laboratorija za Obradu signala katedre za Telekomunikacije na Elektrotehničkom fakultetu u Beogradu je ADSP-BF533 EZ-KIT Lite razvojni sistem dobila zahvaljujući univerzitetkom programu proizvođača Analog Devices.

LITERATURA

- [1] Lj. Milić, Z. Dobrosavljević, "Uvod u digitalnu obradu signala", Beograd, 1999.
- [2] A. V. Oppenheim, R. W. Schaffer, "Discrete-time signal processing", Prentice Hall, Englewood Cliffs, New Jersey, 1989.
- [3] E. C. Ifeachor, B. W. Jervis, "Digital Signal Processing", Addison-Wesley, 1993.
- [4] *Overflow Avoidance Techniques in Cascaded IIR Filter Implementations on the TMS320 DSP's*, Texas Instruments, 1999.
- [5] R. C. Cofer, B. Harding, "Fixed-Point DSP and Algorithm Implementation"
<http://www.dspdesignline.com/showArticle.jhtml;jsessionid=10M0MDBM3HT2KQSNLPCXHSCJUNN2JVN?articleID=193402220>
- [6] ADSP-BF533 EZ-KIT Lite Evaluation System Manual (Revision 3.1, September 2007)

ABSTRACT

This paper presents differences between two software realization of digital IIR filter, cascaded direct and canonic second order sections. It has been shown some of finite word length effects on these two digital filter implementations. It has also been shown that direct realization is less sensitive to finite word length effects and better than canonic. IIR filter is implemented in fixed-point arithmetic on a ADSP-BF533 digital signal processor using Analog Devices VisualDSP++ 4.0.

IIR FILTER IMPLEMENTATION IN FIXED-POINT ARITHMETIC

Nevena I. Todorović