

# Softverska implementacija jednog pitch shift algoritma

Tijana Sutara, Boris Kapor

**Sadržaj** — U radu je prikazan jedan popularan vid obrade audio signala - pitch shift. Opisan je pojam pitch shift-a, njegova implementacija i primena u muzičkoj industriji. Detaljno je prikazan jedan pitch shift algoritam. Izloženi su detalji softverske implementacije u programskom paketu MATLAB®. Prikazani su rezultati nekoliko primera izvršavanja napisanog programa sa različitim tipovima ulaznog signala i za različite zadate parametre obrade.

**Ključne reči** — fazni vokoder, pitch shift, STFT, visina tona.

## I. UVOD

Pitch shift-er je uređaj ili algoritam kojim se „razvlači“ ili „skuplja“ frekvencijski spektar audio signala. Kao posledica toga, svi zvuci u audio signalu pomereni su na više ili niže tonove bez promene brzine signala, što se dešava ako se signal samo resampljuje. Ovakva promena spektra postiže se skaliranjem svih visina tonova, ali tako da su sačuvani harmonični odnosi, drugim rečima, treći harmonik je i dalje na tri puta većoj frekvenciji posle operacije pitch shift-ovanja. Jedan od najpoznatijih primera pitch shift-ing efekta je glas veverica iz crtanog filma „Alvin i Veverice“.

Sve pojave koje zapažamo čulom sluha spadaju u zvuk [1]. Zvuk je mehanički talas uzrokovan vibracijama nekog objekta, izvora zvuka, koji se prostire kroz medijum prenoseći energiju. Telo koje osciluje naizmenično sabija i razređuje vazduh oko sebe, stvarajući tako zvučne talase. Zvuk se može okarakterisati visinom tona, jačinom i kvalitetom. Visina tona (ton) predstavlja [2] način na koji spoznajemo fundamentalnu frekvenciju zvuka. Pored fundamentalne frekvencije, visina tona je određena amplitudom, bojom i trajanjem. Tonovi mogu da se označavaju korišćenjem slova, kao u Helmholtz-ovoj notaciji tonova, zatim korišćenjem kombinacije brojeva i slova, što predstavlja naučnu notaciju tonova, ili brojem koji predstavlja frekvenciju. Način na koji ljudi percipiraju odnos dva tona je približno logaritamskog tipa, što omogućava predstavljanje tona korišćenjem numeričke skale zasnovne na logaritmu fundamentalne frekvencije. Jedan od standarda mapiranja fundamentalne frekvencije  $f$  (data u Hercima) u realan broj  $p$  je MIDI (Musical

Instrument Digital Interface) standard, predstavljen jednačinom (1):

$$p = 69 + 12 \cdot \log_2 \left( \frac{f}{440 \text{ Hz}} \right). \quad (1)$$

Izraz (1) stvara linearni tonski prostor gde oktave imaju vrednost 12, polutonovi vrednost 1, a tonu A440 je dodeljena vrednost 69. Ton A440 ili A4 ima fundamentalnu frekvenciju na 440 Hz i predstavlja referencu za određivanje fundamentalnih frekvencija ostalih tonova.

Sa muzičkog stanovišta [3], pitch shift se sastoji iz menjanja visine tona za jedan ili više polutonova kao što je prikazano u jednačini (2), gde  $ch\_initial$  označava početni poluton,  $s$  predstavlja broj polutonova za koliko želimo da promenimo početni poluton, a  $ch\_shifted$  dobijeni poluton:

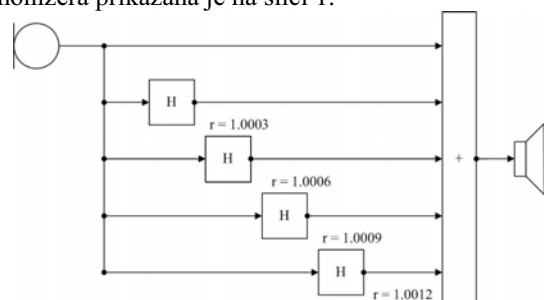
$$ch\_shifted = ch\_initial + s. \quad (2)$$

Sa stanovišta obrade signala pitch shift se sastoji iz skaliranja fundamentalne frekvencije i harmonika za specifičan faktor, kao što je pokazano u jednačini (3), gde  $f\_initial$  predstavlja frekvenciju početnog tona,  $s$  je broj polutonova za koliko se vrši pitch shift i  $f\_shifted$  je frekvencija dobijene note:

$$f\_shifted = 2^{\frac{s}{12}} \cdot f\_initial. \quad (3)$$

## II. IMPLEMENTACIJA PITCH SHIFT-A

Primenu pitch shift-ing efekta moguće je ostvariti hardverski ili softverski. Istorijski gledano prvi primeri pitch shift-ing efekta su bili hardverskog tipa i u analognom domenu, tako što se audio kasetna puštala na uređaju koji je mogao da menja brzinu puštanja, ili da usporava ili ubrzava puštanje audio kasete. Prvi muzičari koji su eksperimentisali sa zvukom u ovom smislu bili su Džon Lenon, Džimi Hendriks i Dejvid Bouvi [4]. Sredinom sedamdesetih godina dvadesetog veka pojavili su se i prvi digitalni uređaji. Među prvima koji je pronašao primenu je Harmonizer koji predstavlja uređaj koji elektronski duplicira signal na drugačijoj visini tona. Šema Harmonizera prikazana je na slici 1.



Sl. 1. Harmonizer.

Tijana Sutara, student Elektrotehničkog fakulteta u Beogradu, Srbija (telefon: 381-60-3018084, e-mail: [tijana.sutara@gmail.com](mailto:tijana.sutara@gmail.com))

Boris Kapor, student Elektrotehničkog fakulteta u Beogradu, Srbija (telefon: 381-63-7058645, e-mail: [bkapor@gmail.com](mailto:bkapor@gmail.com))

### A. Softverska implementacija u vremenskom domenu

Mnogi algoritmi pitch shift-a u vremenskom domenu su razvijeni, pri čemu su najjednostavniji oni koji uključuju modifikaciju osnovne frekvencije signala ili samo reprodukciju signala brže ili sporije u odnosu na originalni signal. Uprkos čujnim promenama u visini tona, te metode ne predstavljaju pravi pitch shift, jer trajanje signala nije očuvano. Najpoznatiji vremenski algoritam je PSOLA (Pitch Synchronous OverLap/Add) algoritam [4]. Prednost ovakvog algoritma je u tome da se ukupna spektralna anvelopa signala ne menja, samim tim postizemo visoko kvalitetni izlazni signal.

Vremenski algoritmi pitch shift-a su doživeli veliku komercijalnu eksploataciju krajem dvadesetog veka [5]. Danas su oni u upotrebi u nizu uređaja, kao što su telefonski sistemi za primanje poziva, procesori za muzičke efekte, profesionalni CD plejeri, hard-disk rekorderi...

Algoritam implementacije pitch shift-a u vremenskom domenu sastoji se iz više etapa. Ulazna komponenta je detektor osnovne frekvencije ulaznog signala, gde je tipična vrednost za muški glas između 80 Hz i 200 Hz, dok je ženski glas između 150 Hz i 350 Hz. Primenom pravougaone prozorske funkcije ulazni signal se deli u blokove dužine  $k*fo$ , gde je  $k$  pozitivan ceo broj. U cilju da se snizi visina tona, odbacuje se  $p$  perioda ( $p$  je pozitivan ceo broj,  $p \leq k$ ) iz svakog bloka pre nego se obrade na izlazu. Pre obrade signala na izlazu dodaje se dodatni blok podataka, zakašnjena replika obrađenog bloka iz ulaznog signala. Opisana procedura rezultira proširenjem signala dobijenog nakon odvajanja  $p$  perioda. Da bi se povratila prvobitna dužina signala vrši se resemplovanje. U ovom slučaju je potreban resampling sa koeficijentom  $\alpha$  koji je u obliku razlomka pa se proces resamplovanja sastoji od interpolacije i decimacije.

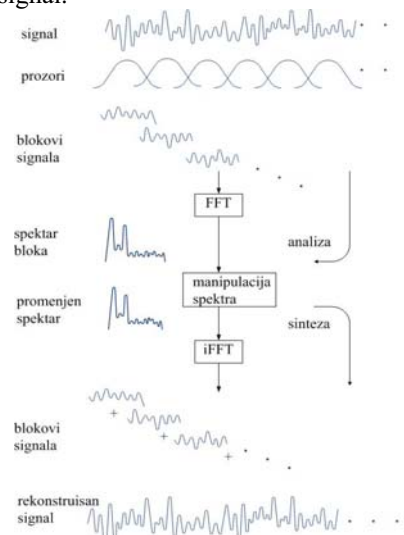
Ovaj metod implementacije pitch shift-a je popularan zbog jednostavnog proračuna koji ne opterećuje memoriju računara i daje odlične rezultate u obradi monofonog zvuka [5]. Međutim, ovakva tehnika obrade signala će uvesti neke visoko frekvencijske komponente koje su posledica pre svega diskontinuiteta do kojih dolazi usled neslaganja faza blokova koji se preklapaju i izbora pravougaonog prozora. Takođe daje slabe rezultate kada se primeni na polifone signale. Detaljniji pregled implementacije u vremenskom domenu može se naći u radu Laroche-a i Moulines-a [6] i [7].

### B. Softverska implementacija u frekvencijskom domenu

Uvođenje koncepta frekvencijske obrade signala predstavlja povoljno rešenje za neke specijalizovane aplikacije, kao što je pitch shift audio signala. Ova aplikacija ima veliku praktičnu primenu u današnjim audio-sistemima za obradu. Jedan proces koji ga primenjuje je "fazni vokoder". Metod faznog vokodera su uveli Flanagan i Golden 1966. godine, a digitalno ga sprovodi Portnoff deset godina kasnije. On je u svom radu [8] opisao implementaciju faznog vokodera pomoću FFT-a (Fast Fourier Transform) i njegovu efikasnost. U daljem razmatranju ćemo objasniti osnovni tip faznog vokodera. Za razliku od implementacije u vremenskom domenu fazni

vokoder daje visoko kvalitetne rezultate za polifone signale, ali zahteva komplikovaniji proračun.

Kako su govor i muzika kvazistacionarni, jer su gotovo stacionarni u roku od nekoliko milisekundi [9], nije dobro da se vrši Furijeova transformacija celog signala, jer će se izgubiti promene u spektru signala, a time individualne karakteristike neće biti lako uočljive. Ako se, s druge strane, signal deli na manje "blokove", prilikom obrade imaćemo stacionaran signal u svakom bloku. Ovaj način posmatranja ulaznog signala segmentiranog na kratke blokove, pri čemu za svaki odredimo vrednost njegove FFT, zove se Short Time Fourier Transform (STFT) signala. Zato se i sam algoritam frekvencijske implementacije pitch shift-a zasniva na primeni STFT-a, a ceo proces obrade je prikazan na slici 2. Nakon primene STFT-a na svaki blok, vrši se njihova obrada pomoću koje se tonovi snižavaju ili povisuju. Zatim se računa vrednost njegove iFFT (inverse Fast Fourier Transform) da bi signal iz frekvencijskog vratili u vremenski domen. Potom sledi preklapanje vremenskog oblika blokova da bi dobili pitch shift-ovan signal.



Sl. 2. Algoritam obrade signala u faznom vokoderu.

Algoritmi faznog vokodera su se uglavnom koristili za naučne i obrazovne softverske proizvode da bi se pokazala upotreba i ograničenja Fourierove transformacije, ali su stekli popularnost tokom poslednjih nekoliko godina zbog poboljšanja koja omogućavaju da se u velikoj meri smanje nedostaci originalnog faznog vokodera, koji na izlazu daje signal sa izobličenjima koja se čuju kao "zamazivanje" i "reverberacija". Objašnjenje i načini na koji se značajno smanjuje reverberacija opisani su u radu Laroche-a i Dolson-a [5]. Iako kvalitet modifikovanog faznog vokodera prikazanog u radu daje bolji kvalitet obrađenog signala, on je i dalje inferioran u odnosu na implementaciju pitch shift-a u vremenskom domenu za monofone signale kao što je govor.

### C. Hardverska implementacija

Hardverska primena pitch shift efekta je najčešća u profesionalnim studijima za snimanje zvuka ili na koncertima i njena prednost je u tome što radi u realnom vremenu. Koristi se pri ispravljanju intonacije instrumenata ili pevača. Moguće je napraviti i horski efekat, a da pri tom peva samo jedan pevač tako što se glas

pevača miksuje sa njegovim kopijama koje su neznatno pomerenе u vremenu.

### III. IMPLEMENTACIJA U MATLAB® -U

U programskom paketu MATLAB® napravili smo program koji implementira fazni vokoder. Signale koje smo koristili su govorni signal, sinusoida na 440 Hz i muzički signali. Sam algoritam našeg MATLAB® koda biće opisan dalje u tekstu.

Program pre početka obrade zahteva odabir željenog signala, ime izlaznog fajla i za koliko polutonova je potrebno sniziti ili povisiti ton. Frekvenciju odabiranja, broj bita po odbirku i učitavanje odabranog signala smo ostvarili upotrebom MATLAB® funkcije *wavread*. Koeficijent pitch shifta  $\alpha$  dat je jednačinom (4), gde je  $s$  zadati broj polutonova za koji je potrebno promeniti signal

$$\alpha = 2^{\frac{s}{12}}. \quad (4)$$

Za prozorsku funkciju uzet je Hanov prozor dužine 1024 i koeficijenta preklapanja 0.1. Sledeći korak obrade je inicijalizacija radnih parametara: dužine ulaznog signala, broj blokova u ulaznom signalu i normalizovane učestanosti.

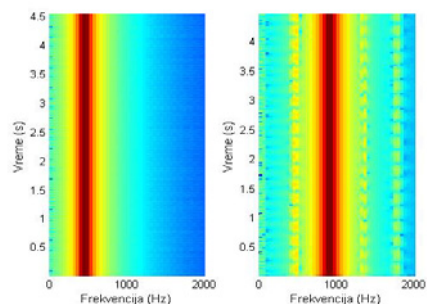
Prvo je potrebno izvršiti inicijalizaciju izlaznog signala, radi rezervisanja memorije, čime se postiže ubrzanje programa. Određuju se amplitude i faze Furijeove transformacije prvog bloka ulaznog signala pre ulaska u for-petlju. Algoritam faznog vokodera, koji je opisan u drugom (II) poglavlju i prikazan na slici 2., simulirali smo unutar for-petlje gde se vrši obrada amplitude i faze Furijeove transformacije blokova ulaznog signala primenom STFT. Amplituda bloka izlaznog signala iz for-petlje dobija se tako što se računa amplituda FFT proizvoda prozorske funkcije i bloka ulaznog signala, dok se faza izlaznog bloka računa kao razlika faza trenutnog bloka, predhodnog bloka i prozorske funkcije. Da bi se dobio pitch shift-ovan signal izvršili smo preklapanje blokova obrađenih u for-petlji kao i resemplovanje pomoću MATLAB® funkcije *resample*. Poslednji korak je slušanje dobijenog signala pomoću MATLAB® funkcije *wavplay* i iscrtavanje spektrograma signala.

Na nekoliko primera biće prikazani rezultati našeg programa. Za prikaz rezultata korišćen je spektrogram jer je uobičajeni prikaz u obradi audio signala, a mogu se jasno uočiti promene u spektru.

Pored naše simulacije i sam MATLAB® ima demo verziju faznog vokodera koja se može pokrenuti naredbom *dspitchtime* iz komandnog prozora (Command Window), koji smo koristili da uporedimo s našim rezultatima.

#### A. Sinusoida na 440 Hz

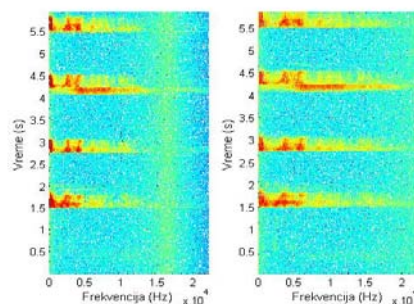
Ulazni signal predstavlja sinusni signal na frekvenciji od 440 Hz i trajanja četiri sekunde. Ovaj primer plastično prikazuje promenu tona, jer sinusni signal na 440 Hz predstavlja A4 notu, a nakon pitch shifta za celu oktavu dobije se nota A5 koja predstavlja sinusni signal na 880 Hz. Sa spektrograma na slici 3. se jasno vidi ostvarena promena frekvencije.



Sl. 3. Spektrogram sinusoida na 440 Hz dat je sa leve strane, dok je sinusoida na 880 Hz dat sa desne strane slike.

#### B. Signal govora I

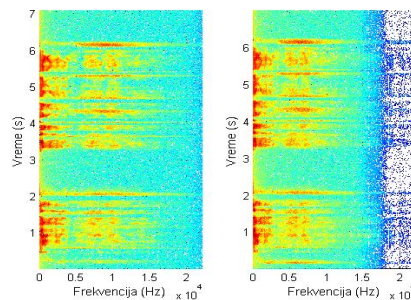
Ovaj primer predstavlja snimljen muški glas koji izražajno izgovara prva četiri slova engleskog alfabeta. Kako je ovaj audio zapis snimljen u “gluvoj sobi” odlikuje ga odsustvo šuma na spektrogramu. Ulazni signal smo povisili za pola oktave što se može uočiti analiziranjem spektrograma izlaznog signala gde su vidno izražene više frekvencije.



Sl. 4. Spektrogram originalnog govornog signala dat je na levoj strani, dok je spektrogram pitch shift-ovanog signala prikazan na desnoj strani slike.

#### C. Signal govora II

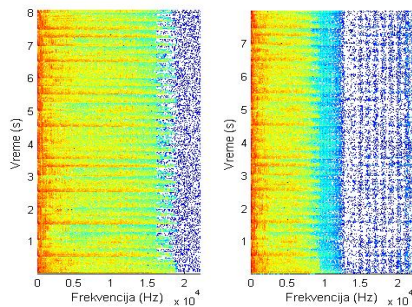
Drugi primer predstavlja popularnu primenu pitch shifta, a to je imitacija glasa “Velikog brata”. Ovaj primer predstavlja originalni snimljeni signal govora snižen za pola oktave. Na slici 5. se vidi da su izraženije niže komponente spektra što dokazuje da je glas dublji nakon pitch shift-a signala. Kako ovaj signal govora nije snimljen u “gluvoj sobi” jasno se uočava postojanje šuma u ulaznom signalu. Pored tog šuma u izlaznom signalu se primećuje i postojanje šuma generisanog u samom programu.



Sl. 5. Spektrogram originalnog govornog signala dat je na levoj strani, dok je spektrogram pitch shift-ovanog signala prikazan na desnoj strani slike.

#### D. Muzički signal I

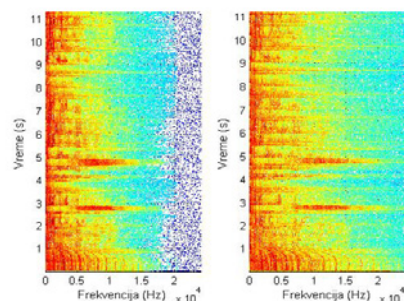
Muzički signal koji smo u ovom primeru koristili je deo pesme "Lady Gaga feat. Beyonce: Telephone", trajanja osam sekundi. Nakon pitch shift-a dobili smo sniženje ženskog glasa za celu oktavu, što se analizom spektrograma na slici 6. može zaključiti po izraženijim nižim frekvencijama obrađenog signala. U samom signalu samo je snižen ton izvođača, dok je melodija ostala očuvana.



Sl. 6. Spektrogram originalnog muzičkog signala prikazan je na levoj strani, dok je spektrogram pitch shift-ovanog signala dat na desnoj strani slike.

#### E. Muzički signal II

Kao drugi muzički signal iskorišćen je deo pesme "Frank Sinatra: Fly Me To The Moon", trajanja jedanaest sekundi, da bi prikazali povišenje tona muškog glasa za pola oktave. Rezultat je prikazan na slici 7. u vidu spektrograma obrađenog signala na kome su izraženije više frekvencije. Kada se sluša izlazni signal, kao i u prethodnom primeru, promena u visini tona se jasno čuje samo u glasu izvođača, dok je melodija očuvana.



Sl. 7. Spektrogram originalnog muzičkog signala dat je na levoj strani, dok je spektrogram pitch shift-ovanog signala prikazan na desnoj strani slike.

### IV. ZAKLJUČAK

Na osnovu dobijenih rezultata u simulaciji pitch shift-a u programskom okruženju MATLAB<sup>®</sup> može se zaključiti da je moguća uspešna implementacija ovog efekta na osnovu algoritma faznog vokodera. Rezultati dobijeni

simulacijom su zadovoljavajućeg kvaliteta i ukazuju na moguću primenu algoritma pre svega u stvaranju efekata poput glasa „Velokog Brata“ ili piskutavih glasova. Manipulacijom faktora  $\alpha$  moguće je i proizvesti različite varijacije glasa koje imaju primenu u tzv. elektronskoj muzici. Kako je proces jako spor učitavanje cele pesme i njena obrada je računarski zahtevna i, u zavisnosti od raspoloživih resursa, može dovesti do preopterećenja sistema. Takođe opisan program podrazumeva da se obrađuje prethodno snimljen signal što onemogućava primenu u realnom vremenu. Zbog toga smatramo da će se dalji razvoj i primena pitch shift-a naći u konvergenciji hardverske i softverske implementacije, pre svega zbog mogućnosti za rad u realnom vremenu upotrebom hardverske implementacije i lakoj izmeni performansi pri softverskoj implementaciji.

#### ZAHVALNICA

Zahvaljujemo se asistentkinji Jeleni Čertić na pomoći pri izradi ovog projekta u okviru predmeta Obrada signala II na katedri za telekomunikacije i informacione tehnologije Elektrotehničkog fakulteta u Beogradu.

#### LITERATURA

- [1] <http://ljljar.spaces.live.com/default.aspx>
- [2] [http://en.wikipedia.org/wiki/Pitch\\_%28music%29](http://en.wikipedia.org/wiki/Pitch_%28music%29)
- [3] <http://www.guitarpitchshifter.com>
- [4] U. Zölzer, "DAFX – Digital Audio Effects", John Wiley & Sons, Ltd.
- [5] J. Laroche, M. Dolson, "Improved phase vocoder time-scale modification of audio," *Speech and Audio Processing, IEEE Transactions on*, vol.7, no.3, pp.323-332, May 1999.
- [6] J.Laroche, "Time and pitch scale modification of audio signals" *Applications of Digital Signal Processing to Audio and Acoustics*, M.Kahrs and K. Brandenburg, Eds. Boston, MA: Kluwer, 1998.
- [7] E. Moulines and J.Laroche, "Non parametric techniques for pitch-scale and time-scale modification of speech", *Speech Commun.*, vol.16, pp. 175-205, Feb. 1995.
- [8] M. Portnoff, "Implementation of the digital phase vocoder using the fast Fourier transform," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol.24, no.3, pp. 243- 248, Jun 1976.
- [9] <http://www.dspdimension.com/admin/pitch-shifting-using-the-ft/>.

#### ABSTRACT

This paper presents a popular form of audio signal processing that is called pitch shift. It describes the concept of pitch shift, its implementation and application in the music industry. The paper presents detailed survey of a pitch shift algorithm implemented in MATLAB<sup>®</sup> with the results obtained after several program execution with different types of input signal and parameters.

### SOFTWARE IMPLEMENTATION OF A PITCH SHIFT ALGORITHM

Tijana Sutara, Boris Kapor