

DFE-FBC algoritam za popravku kvaliteta govornog signala

Zoran N. Milivojević, Dragiša Ž. Balanesković

Sadržaj — U ovom radu prikazan je algoritam za popravku kvaliteta govornog signala u uslovima superponiranih šumova. Algoritam se bazira na filtriranju disonantnih frekvencija F#, B i C# u odnosu na frekvenciju osnovnog tona C. Posredstvom subjektivnog MOS testa analiziran je efekat popravljivanja kvaliteta govornog signala. Analiza rezultata MOS testa, koja je prikazana je u drugom delu rada, ukazuje na povećanje kvaliteta govornog signala u prisustvu superponiranih šumova.

Ključne reči — Disonantne frekvencije, Fundamentalna frekvencija, Kvaliteta govora.

I. UVOD

SUPERPONIRANJE akustičkih ambijentalnih smetnji dovodi do smanjenja kvaliteta govornog signala koje se, pored ostalog, manifestuje kao smanjenje razumljivosti (*engl. intelligibility*) ([1-3]). Govorni signal sa superponiranim akustičkim smetnjama se prenosi posredstvom komunikacionih linija, tako da se na prijemnoj strani može javiti degradacija u tolikoj meri da reprodukovani govor bude nerazumljiv ili vrlo neprijatan.

Kod procesiranja govornog signala u cilju kompresije signala, kao i kod sistema za prepoznavanje govora, superponirani akustički šum u velikoj meri degradira performanse algoritama za procesiranje. U cilju popravke performansa algoritama procesiranja nameće se potreba za predprocesiranjem govornog signala radi smanjenja komponente šuma. Povećanje kvaliteta govornog signala je aktuelan problem i razvijen je veliki broj algoritama za njegovu realizaciju [4-6]. U radu [7] opisan je novi algoritam za popravku govornog signala koji obavlja filtriranje disonantnih frekvencija. Algoritam se bazira na procesiranju govornog signala u spektralnom domenu: a) određivanju fundamentalne frekvencije i b) filtriranju disonantnih frekvencija u odnosu na fundamentalnu frekvenciju u svim oktavama. Filtrirane su disonantne frekvencije koje u odnosu na fundamentalnu frekvenciju stoje kao ton F# u odnosu na ton C. F# je poznat kao the Devil's interval u muzici [8]. Rezultati, koji su bazirani na primeni MOS testa pokazuju da predloženi algoritam dovodi do blagog povećanja kvaliteta govora kod Gauss-ovog šuma. U radu [9] autor proširuje opseg delovanja algoritma opisanog u [7] time što povećava broj filtriranih disonantnih frekvencija (B i F# u odnosu na C). Subjektivni MOS test je pokazao povećanje kvaliteta govornog signala u odnosu na rezultate iz [7]. Efikasnost algoritama opisanih u [7,9] zavisi od preciznosti procene

fundamentalne frekvencije govornog signala. U cilju preciznije procene fundamentalne frekvencije primenjen je PCC (*engl. Parametric Cubic Convolution*) algoritam koji je opisan u [10].

U ovom radu prikazan je algoritam za popravku kvaliteta govornog signala koji se zasniva na filtriranju disonantnih frekvencija (F#, B i C# u odnosu na C). Predloženi algoritam se bazira na algoritmima opisanim u [7] i [9] i predstavlja proširenje opsega delovanja. Efikasnost algoritma testiran je procesiranjem govornih signala kojima je superponiran beli Gauss-ov šum (*engl. White Gaussian Noise, WGN*).

Organizacija rada je sledeća. U sekciji II prikazana je muzikološka definicija disonantnih frekvencija. U sekciji III prikazani su algoritmi za filtriranje disonantnih frekvencija. Rezultati MOS testa obavljenim nad filtriranim govornim signalom sa superponiranim šumovima, prikazani su u sekciji IV. Analiza rezultata MOS testa izvršena je u sekciji V.

II. MUZIČKA DEFINICIJA DISONANTNIH FREKVENCIJA

Teorijom muzike definisane su osnovne osobine zvuka: a) trajanje, b) jačina i c) boja. Izraz *boja* se na zvuk primenjuje u prenosnom smislu, što ukazuje na složenost ove osobine zvuka. Zvučni izvor generiše zvuk sa osnovnom frekvencijom (osnovni ton) kao i prizvuke (aliquotni tonovi u odnosu na osnovni ton). Različiti broj prisutnih alikvota (*lat. aliquoties*, nekoliko puta) i njihova različita relativna jačina u sklopu ukupnog zvučanja, određuje boju zvuka.

Frekvencija muzički definisanih tonova u odnosu na osnovni ton u intervalu od jedne oktave određena je formulom:

$$F_k = F_0 \times 2^{\left(\frac{k}{12}\right)}, \quad k = 0, 1, \dots, 12. \quad (1)$$

gde je F_0 frekvencija osnovnog tona a F_k frekvencija k-tog polutona. Prema osnovnom tonu polutonovi formiraju intervale. Interval je definisan odnosom frekvencija polutona i osnovnog tona. Razlomci F_k/F_0 , za $k=0,1,\dots,12$, koji predstavljaju pojedine intervale (1/1, 135/128, 9/8, 6/5, 5/4, 4/3, 45/32, 3/2, 8/5, 27/16, 9/5, 15/8, 2/1) predstavljaju aproksimaciju prave vrednosti (1). Klasifikacija intervala po zvučnosti ostvaruje se na osnovu razlomka koji ga opisuje. Kada je razlomak jednostavniji interval je, kao sazvučje tonova, stabilniji, odnosno, konsonantniji. Kada je razlomak složeniji stabilnost intervala je manja, tako da je disonantnost veća.

Zoran N. Milivojević, Visoka tehnička škola, Aleksandra Medvedeva 20, Niš; e-mail: milivojevic@vtsnis.edu.yu.

Dragiša Ž. Balanesković, Fakultet umetnosti, Niš e-mail: balanesko@bankerinter.net.

Konsonanca i disonanca nisu oštro razgraničene već zajedno čine jednu izdiferenciranu skalu međustepena, od potpune stabilnosti na jednom, do krajnje labilnosti na drugom kraju skale. U skali se razlikuju: a) savršene (potpune) konsonance (prima (1/1), oktava (2/1), kvinta (3/2) i kvarta (4/3)), b) nesavršene (nepotpune) konsonance (velika terca (5/4), velika seksta (5/3), mala terca (6/5) i mala seksta (8/5)), c) nesavršene (nepotpune) disonance (mala septima (9/5) i velika sekunda (9/8)) i d) savršene (potpune) disonance (mala sekunda (135/138), tritonus ili prekomerna kvarta (45/32) i velika septima (15/8)). Sa stanovišta doživljaja, odnosno percepcije zvuka, muzički interval je definisan kao konsonantan ako je zvuk prijatan (**engl. pleasant**) ili miran (**engl. restful**). Muzički interval je disonantan u velikoj meri ako je zvuk neprijatan ili grub (**engl. rough - hrapav**).

U odnosu na osnovni ton definišu se frekvencije polutonova koji sa osnovnim tonom čine konsonance u svim oktavama iz čujnog područja:

$$F_d = F_0 \times 2^{(n+k/12)}, \quad n = 0, 1, \dots, 7; \quad k = \{1, 6, 11\}, \quad (2)$$

gde je F_0 frekvencija osnovnog tona, n broj oktave i k broj polutona u pojedinim oktavama. Uzimajući ton C kao referentni, odnosno osnovni ton, tada su njemu disonantni polutonovi B, F# i C# kao i njihovi harmonici u svim oktavama.

III. FILTRIRANJE DISONANTNIH FREKVENCIJA

Govor nastaje eksitacijom vokalnog trakta čoveka [11]. Uvodeći analogiju govornog signala sa muzikološkom definicijom zvuka, može se uspostaviti odgovarajuća korespodencija osnovnog tona i njemu odgovarajućih polutonova i alikvota, sa fundamentalnom frekvencijom (F_0) i pratećih frekvencija govornog signala. Samim tim, moguće je definisati disonantne frekvencije u odnosu na F_0 .

U [7] opisan je algoritam za poboljšanje kvaliteta govornog signala filtriranjem disonantnih frekvencija. Algoritam se sastoji iz sledećih koraka: a) podele govornog signala na sekvence dužine N i izračunavanja FFT-a svake sekvence, b) određivanja fundamentalne frekvencije F_0 , c) određivanja disonantne frekvencije F_d u odnosu na fundamentalnu frekvenciju F_0 (saglasno odnosu C i F#), d) filtriranja disonantnih frekvencija i e) generisanja sekvence govornog signala primenom IFFT. U [9] opisan je algoritam popravke kvaliteta govornog signala koji je dobijen proširenjem opsega delovanja algoritma iz [7]. Proširenje algoritma odnosi se na eliminisanje disonantnih tonova F# i B u odnosu na C.

U nastavku ovog rada prikazan je algoritam (sl.1) za eliminaciju disonantnih frekvencija (**engl. Dissonant Frequency Filtering - DFF**) F#, B i C# u odnosu na frekvenciju osnovnog tona C, koji je baziran na algoritmima iz [7,9]. DFF-FBC algoritam se sastoji iz sledećih koraka:

1: podele govornog signala na sekvence dužine N i određivanja FFT-a svake sekvence,

2: Procene fundamentalne frekvencije F_0 korišćenjem PCC,

3: Određivanja disonantne frekvencije F_{d1} , F_{d2} i F_{d3} u odnosu na fundamentalnu frekvenciju F_0 (saglasno odnosu tonova C prema F#, B i C#) kao:

$$F_{d1} = F_0 \times 2^{(n+6/12)}, \quad n = 0, 1, \dots, 7, \quad (3)$$

$$F_{d2} = F_0 \times 2^{(n+11/12)}, \quad n = 0, 1, \dots, 7. \quad (4)$$

$$F_{d3} = F_0 \times 2^{(n+1/12)}, \quad n = 0, 1, \dots, 7. \quad (5)$$

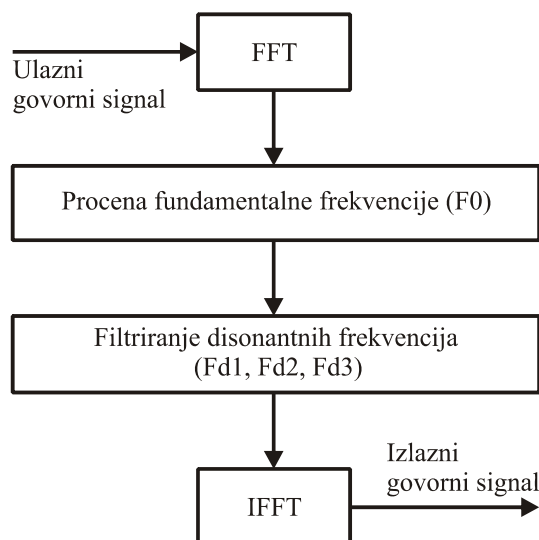
4: Filtriranja disonantnih frekvencija iz opsega (koji odgovara polutonu)

$$F_0 \times 2^{(n+1/24)} < F_{d1} < F_0 \times 2^{(n+13/24)}, \quad n = 1, 3, \dots, 7, \quad (6)$$

$$F_0 \times 2^{(n+21/24)} < F_{d2} < F_0 \times 2^{(n+23/24)}, \quad n = 2, 3, \dots, 7, \quad (7)$$

$$F_0 \times 2^{(n+1/24)} < F_{d3} < F_0 \times 2^{(n+13/24)}, \quad n = 2, 3, \dots, 7. \quad (8)$$

5: generisanja vremenske sekvence govornog signala primenom IFFT.



Sl.1. Algoritam procesiranja govornog signala filtriranjem disonantnih frekvencija.

IV. EKSPERIMENTALNI REZULTATI

U cilju testiranja efekta DFF-FBC algoritma na perceptualne karakteristike govornog signala i komparativne analize sa rezultatima procesiranja govornog signala algoritmima iz [7,9] formirane su baze fajlova govornih signala saglasno bazama iz [9] i izvršen subjective MOS test.

A. Baza govornih signala

Baza govornih signala predstavlja kolekciju od 14 fajlova sa arhiviranim govornim signalima 7 govornika (5 muških i 2 ženska), pri čemu je svaki govornik izgovorio 2 rečenice na srpskom jeziku. Proces formiranja baze obavljen je sa frekvencijom smplovanja 8 kHz sa 8-bit rezolucijom. Primenjen je Hanning-ov prozor dužine 256. U eksperimentima je korišćen beli Gauss-ov šum i

formirani govorni signali u opsegu SNR=0,5, 10,15,20 dB.

B. MOS test

Kvalitet reprodukovanog govora testiran je subjektivnim MOS (**engl. Mean Opinion Score**) testom. U testu je učestvovalo dvadeset slušaoca. Slušaoci sede udobno u tihoj prostoriji i slušaju rečenice koje se emituju sa slučajnom redosledu. Slušaoci na osnovu individualnog doživljaja kvaliteta donose ocenu u opsegu 1 (vrlo loš kvalitet) to 5 (odličan kvalitet).

U tabeli 1 su prikazani rezultati MOS testova za: a) neprocesiran govorni signal (MOS_{UP}), b) govorni signal sa eliminisanim F# intervalima ($MOS_{F\#}$) (algoritam iz [7]), c) govorni signal sa eliminisanim F# i B intervalima ($MOS_{F\#B}$) (algoritam iz [9]) i d) govorni signal sa eliminisanim F#, B, C# intervalima ($MOS_{F\#BC\#}$) (DFF-FBC algoritam iz ovog rada).

TABELA 1: REZULTATI MOS TESTA KOD WGN.

SNR	MOS_{UP}	$MOS_{F\#}$	$MOS_{F\#B}$	$MOS_{F\#BC\#}$
0 dB	1.85	1.98	2.02	2.03
5 dB	2.64	2.81	2.83	2.84
10 dB	2.88	3.06	3.11	3.13
15 dB	3.05	3.24	3.31	3.33
20 dB	3.89	4.03	4.07	4.08
Original	4.8	4.85	4.89	4.9

U cilju komparativne analize efekata algoritama na kvalitet govornog signala izračunati su procentualni priraštaji MOS vrednosti procesiranih ($\Delta MOS_{F\#}$, $\Delta MOS_{F\#B}$, $\Delta MOS_{F\#BC\#}$) u odnosu na neprocesirane (MOS_{UP}) govorne signale prema:

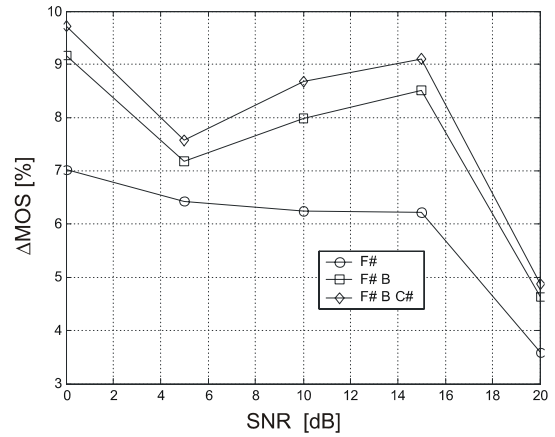
$$\Delta MOS_{F\#} = \frac{100 \cdot MOS_{F\#}}{MOS_{UP}} - 100, \quad (9)$$

$$\Delta MOS_{F\#B} = \frac{100 \cdot MOS_{F\#B}}{MOS_{UP}} - 100, \quad (10)$$

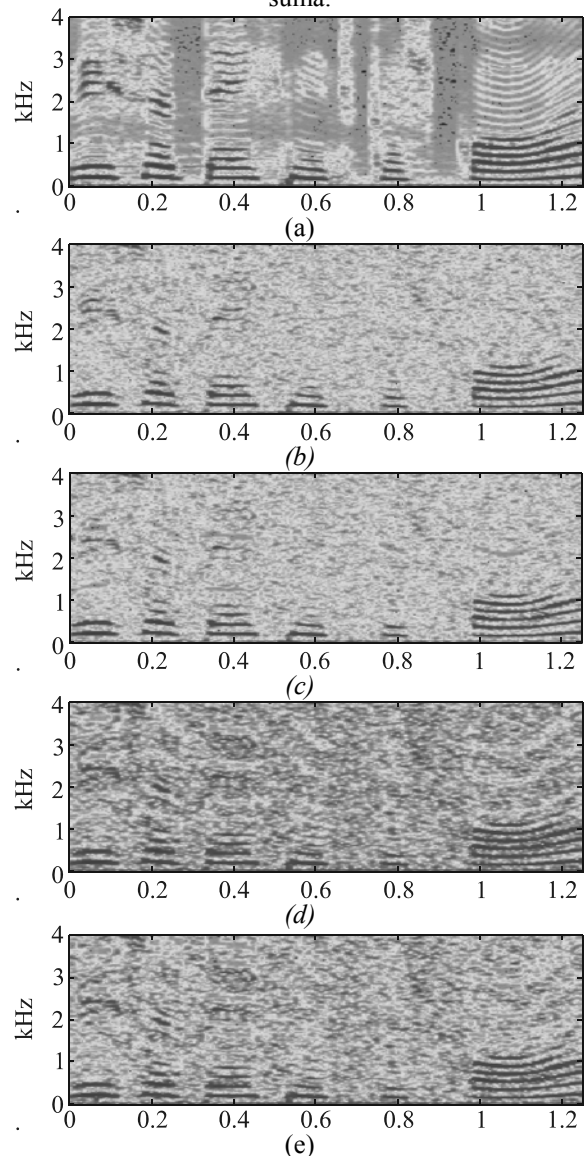
$$\Delta MOS_{F\#BC\#} = \frac{100 \cdot MOS_{F\#BC\#}}{MOS_{UP}} - 100, \quad (11)$$

gde su $\Delta MOS_{F\#}$, $\Delta MOS_{F\#B}$ i $\Delta MOS_{F\#BC\#}$ procentualni priraštaji MOS vrednosti za algoritme iz [7,9] i DFF-FBC algoritam respektivno. Procentualni priraštaji MOS vrednosti grafički su prikazani na sl.2.

Spektrogrami govornog signala (rečenica 'Visoka tehnička škola' izgovorena na srpskom jeziku) prikazana je na sl.3.: a) čist govorni signal, b) čist govorni signal sa superponiranim WGN SNR=5 dB, c) filtrirane F#, d) F#B i e) F#BC# disonance.



Sl.2. Procentualni priraštaj MOS vrednosti kod filtriranja pojedinih disonantnih frekvencija kod belog Gausovog šuma.



Sl. 3. Spektrogram govornog signala sa 5 dB WGN: a) čist govorni signal, b) čist govorni signal sa superponiranim WGN SNR=5 dB, c) filtrirane F#, d) F#B i e) F#BC# disonance.

V. ANALIZA REZULTATA

U cilju analize efekta DFF-FBC algoritma na subjektivni kvalitet govornog signala, na osnovu podataka (procentualno povećanje rezultata MOS testa u odnosu na rezultat neprocesiranog signala (jed.9-11)) koji su grafički prikazani na sl.2 i sl.3, određene su srednje vrednosti procentualnog priraštaja vrednosti MOS testa:

$$\overline{\Delta MOS_d} = \frac{\sum_{SNR} \Delta MOS(SNR)}{len(SNR)}, \quad (12)$$

gde je $d=\{F\#, F\#B, F\#BC\# \}$, $SNR=\{0,5,10,15,20\}$ [dB] a $len(SNR)$ predstavlja broj elemenata niza SNR, i prikazane u tabeli 2.

TABELA 2: PROCENTUALNO POVEĆANJE REZULTATA MOS TESTA.

SNR	$\overline{\Delta MOS_{F\#}}$ [%]	$\overline{\Delta MOS_{F\#,B}}$ [%]	$\overline{\Delta MOS_{F\#,B,C\#}}$ [%]
WGN	5.91	7.49	7.99

Na osnovu rezultata MOS testova grafički prikazanih na sl.2 i u tabelama 1 i 2 može se zaključiti da:

a) algoritam za eliminisanje disonance F# kod govornog signala sa superponiranim belim Gauss-ovim šumom generiše govorni signal čiji je rezultat MOS testa 5.91 % veći od neprocesiranog signala. Ovaj rezultat se slaže sa rezultatom algoritma iz [7];

b) algoritam za eliminisanje disonance F# i B kod govornog signala sa superponiranim šumom generiše govorni signal čiji je rezultat MOS testa 7.49% veći od neprocesiranog signala. Ovaj rezultat se slaže sa rezultatom algoritma iz [9];

c) algoritam za eliminisanje disonance F#, B i C# (DFF-FBC algoritam predložen u ovom radu) kod govornog signala sa superponiranim šumom generiše govorni signal čiji je rezultat MOS testa 7.99% veći od neprocesiranog signala;

Poseban kvalitet DFF-FBC algoritma predstavlja povećanje efekta subjektivnog kvaliteta zvuka sa smanjenjem SNR (sl.2).

VI. ZAKLJUČAK

U ovom radu prikazan je DFF-FBC algoritam za eliminisanje disonantnih opsega u spektru govornog signala. Algoritam je formiran na osnovu algoritama koji su opisani u [7] i [9] proširivanjem opsega delovanja. Algoritam se zasniva na iznalaženju fundamentalne frekvencije govornog signala i određivanju frekvencijskih opsega gde se nalaze disonantni tonovi i njihovi harmonici u svim oktavama audio opsega. Analizirani su disonance koje formiraju skalu savršenih (potpunih) disonanca. U odnosu na fundamentalnu frekvenciju F_0 savršene disonance su u relaciji kao ton C u odnosu na tonove F#, B i C#. Efekat dejstva algoritma na govorni signal analiziran je na osnovu rezultata subjektivnog MOS testa. Analize su sprovedene nad govornim signalima kojima je superponiran beli Gauss-ov šum. Rezultati MOS testa pokazali su subjektivnu popravku govora u opsegu 5.71-7.99%. Efekat subjektivnog povećanja kvaliteta se povećava sa smanjenjem SNR-a, što predstavlja poseban

kvalitet algoritma. Rezultati prikazani u ovom radu govore u prilog implementiranju DFF-FBC algoritama za predprocesiranje govornog signala kod algoritama za kompresiju, prepoznavanje govora i dr.

LITERATURA

- [1] J. Festen, R. Plomp, "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing", *J. Acoust. Soc. Am.* 88, (1990), 1725-1736.
- [2] Gerhard Schmidt, Tim Haulick, "Signal processing for in-car communication systems", *Signal Processing* 86 (2006) 1307-1326.
- [3] R. Udrea, N. Vizireanu, S. Ciochina, S. Halunga, "Nonlinear spectral subtraction method for colored noise reduction using multi-band Bark scale", *Signal Processing* 88 (2008) 1299-1303.
- [4] H. Lev-Ari, Y. Ephraim, "Extension of the signal subspace speech enhancement approach to colored noise", *IEEE Signal Process. Lett.* 10 (4) (2003) 104-106.
- [5] L.M. Arslan, "Modified Wiener filtering", *Signal Processing* 86 (2006) 267-272.
- [6] P. Lanchantin, J. Lapuyade, W. Pieczynski, "Unsupervised segmentation of triplet Markov chains hidden with long-memory noise", *Signal Processing* 88 (2008) 1134-1151.
- [7] B. Joen, S. Kang, S.J. Baek, K.M. Subg, "Filtering of a Dissonant Frequency Based on Improved Fundamental Frequency Estimation for Speech Enhancement", *IEICE Trans. Fundamentals*, Vol. E86-A, No.8, August 2003, 2063-2064.
- [8] J. Tennenbaum, "The foundations of scientific musical tuning", *FIDELIO Magazine*, 1(1) (Winter 1991-92).
- [9] S. Kang, "Dissonant frequency filtering technique for improving perceptual quality of noisy speech and husky voice", *Signal Processing*, Vol. 84 (2004), 431-433.
- [10] H.S. Pang, S.J. Baek, Sung KM, "Improved Fundamental Frequency Estimation Using Parametric Cubic Convolution", *IEICE Trans. Fundamentals* E83-A (2000) 2747-50.
- [11] P. Mokhtari, H. Takemoto, T. Kitamura, "Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches", *Speech Communication* 50 (2008), 179-190.

ABSTRACT

This paper presents an algorithm for enhancement of the noisy speech signal quality. This algorithm is based on the dissonant frequency filtering- DFF, F#, B and C# in relation to the frequency of the primary tone C (DFF-FBC algorithm). By means of the subjective MOS test, the effect of the enhancement of the speech signal quality was analysed. The analysis of the MOS test results, presented in the second part of this paper, points out to the enhancement of the speech signal quality in the presence of superimposed noises.

DFF-FBC ALGORITHM FOR ENHANCEMENT OF SPEECH SIGNAL

Zoran N. Milivojević, Dragiša Ž. Balanesković.