

## UTICAJ MP3 KOMPRESIJE U FORENZIČKOJ ANALIZI GOVORA

Miodrag M. Đorđević  
mijadjo@drenik.net

**Sadržaj** – U radu su prezentirani rezultati analize uticaja mp3 algoritma audio kompresije govornog signala na akustičke parametre koji su relevantni za forenzičku primenu. Govorni materijal koji je korišćen u eksperimentu je podkorpus (diskurs) iz korpusa emotivnog govora za srpski jezik (GEES). Izvorni snimci 6 spikera su konvertovani u mp3 format sa 6 različitih nivoa kompresije. Na svim snimcima je vršeno merenje određenih akustičkih parametara i poređenje sa parametrima dobijenim iz izvornih snimaka.

### 1. UVOD

U savremenoj akustičkoj forenzičkoj praksi se susrećemo sa govornim materijalom snimljenim u komprimovanoj formi i to najčešće u mp3 formatu. Sve češća upotreba ovog načina arhiviranja govornog materijala je uzrokovana činjenicom da je stepen kompresije ovde vrlo visok (praktično 10 pa i više puta), što omogućava uštedu memorijskih kapaciteta. Takođe, tako komprimovan govorni signal je moguće i da se prenosi putem računarskih mreža do udaljenih lokacija (centara). Cilj ovog rada je da se ustanovi da li postoji i koliki je uticaj ove vrste kompresije na parametre govora koji su relevantni u forenzičkim primenama.

### 2. ANALIZA PARAMETARA GOVORA

Za potrebe ove analize iskorišćen je govorni diskurs (podkorpus) iz korpusa GEES [1], na kome je izvršeno merenje izabranih parametara govora, koji se redovno koriste u forenzičkoj analizi. Diskurs predstavlja snimljeni govorni materijal u trajanju od 20 - 30sec, koji ima 79 reči i izgovaran je neutralnim izgovorom. Ovaj diskurs je izgovaralo 6 glumaca sa FDU, 3 muška i 3 ženska govornika.

U originalnom obliku govorni signali u bazi su snimljeni u wav audio formatu sa učestanošću odmeravanja od 22050Hz. U cilju analize snimci svih 6 spikera su konvertovani (i komprimovani) u mp3 format sa šest različitih faktora kompresije, odnosno, šest različitih brzina protoka podataka. Naime, konvertovanje je izvršeno sa frekvencijom semplovanja od 11025Hz sa protokom podataka od 8, 16 i 24 kb/s, kao i 32, 56 i 64 kb/s pri semplovanju od 22050Hz. Takodje, izvršeno je podsemplovanje izvornog wav signala sa 11025Hz. Sve ove konverzije signala izvršene su pomoću programa CoolEditPro. Na ovako dobijenim signalima u mp3 formatu (6 različitih brzina protoka podataka) i dva signala u wav formatu, je izvršeno merenje određenih akustičkih parametara govora za svih 6 spikera.

Upoređivani su dugovremeni i kratkovremeni spektri, kao i sonogrami, a mereni su i spektralni parametri (prva tri formanta i njihove standardne devijacije,  $F1-sr$ ,  $F1-st.dev$   $F2-sr$ ,  $F2-st.dev$   $F3-sr$ ,  $F3-st.dev$ ) kao i osnovna frekvencija

govora i njihove standardne devijacije,  $F0-sr$  i  $F0-st.dev$ . Za ove potrebe merenja korišćen je softver Praat [2].

#### 2.1. Određivanje $f_0$

U tabelama 1 i 2 su date izmerene srednje vrednosti osnovnih frekvencija  $f_0$  glasova sva tri ženska (tabela 1) i sva tri muška spikera (tabela 2), kao i odgovarajućih standardnih devijacija ( $st-dev$ ), za svih 6 brzina protoka komprimovanih snimaka mp3 algoritmom, kao i za podsemplovani snimak wav (11025Hz). Zasenčena vrsta se odnosi na izvorni snimak u wav formatu sa ušestanošću semplovanja od 22050Hz. Poslednje dve vrste predstavljaju usrednjene vrednosti po kolonama za svih 8 načina kodovanja jednog istog snimka, tj., preposlednja se odnosi na  $f_0-sr$  a poslednja vrsta na standardnu devijaciju vrednosti u odgovarajućoj koloni.

Tabela 1. Vrednosti osnovne frekvencije ( $f_0-sr$ ) i standardne devijacije ( $st-dev$ ) za tri ženska spikera izmerene za svih 8 načina kodovanja signala.

Vrsta kodovanja	BM		OK		SZ	
	$f_0-sr$ [Hz]	$st-dev$ [Hz]	$f_0-sr$ [Hz]	$st-dev$ [Hz]	$f_0-sr$ [Hz]	$st-dev$ [Hz]
wav (11025)	245.77	49.79	175.97	31.28	203.14	47.15
mp3 (8-11025)	245.79	50.21	176.20	31.88	203.49	43.64
mp3 (16-11025)	244.79	45.28	176.14	31.67	201.32	39.47
mp3 (24-11025)	244.48	50.77	176.30	31.32	203.09	47.81
mp3 (32-22050)	246.91	52.97	176.37	31.08	202.65	46.42
mp3 (56-22050)	246.51	51.80	176.70	31.17	202.24	46.82
mp3 (64-22050)	246.35	51.58	176.52	31.18	202.44	46.78
wav (22050)	245.43	49.12	176.27	31.31	203.20	47.06
srednje $f_0-sr$	245.75	50.19	176.31	31.36	202.70	45.64
srednja $st-dev$	0.84	2.33	0.23	0.27	0.70	2.79

Tabela 2. Vrednosti osnovne frekvencije ( $f_0-sr$ ) i standardne devijacije ( $st-dev$ ) za tri muška spikera izmerene za svih 8 načina kodovanja signala.

Vrsta kodovanja	MM		MV		SK	
	$f_0-sr$ [Hz]	$st-dev$ [Hz]	$f_0-sr$ [Hz]	$st-dev$ [Hz]	$f_0-sr$ [Hz]	$st-dev$ [Hz]
wav (11025)	103.67	22.23	125.30	29.15	96.88	12.78
mp3 (8-11025)	105.90	29.95	125.28	30.29	97.30	12.38
mp3 (16-11025)	105.37	27.40	124.66	27.12	98.21	20.78
mp3 (24-11025)	104.64	23.49	124.60	27.18	97.38	12.50
mp3 (32-22050)	103.81	19.79	124.81	27.71	98.48	28.52
mp3 (56-22050)	103.86	20.30	125.11	27.55	98.62	28.43
mp3 (64-22050)	103.82	20.10	125.03	27.58	97.79	22.00
wav (22050)	103.69	22.32	125.38	29.14	96.93	12.68
srednje $f_0-sr$	104.35	23.20	125.02	28.21	97.70	18.76
srednja $st-dev$	0.86	3.68	0.30	1.16	0.68	7.13

Iz tabela se vidi da su vrednosti odgovarajućih parametara dobijenih iz signala različito kodovanih vrlo ujednačene. Nešto veće odstupanje se javlja za  $st-dev$  kod muškog spikera SK. Ovo potvrđuje i srednja  $st-dev$  (poslednje vrste u tabelama) čija je vrednost u svim kolonama vrlo mala.

Odstupanje  $f_0-sr$  u odnosu na vrednost izmerenu na izvornom wav signalu iznosi od -1% do +1%, za ženske spikere, dok je za muške u opsegu -1% do +2.5%.

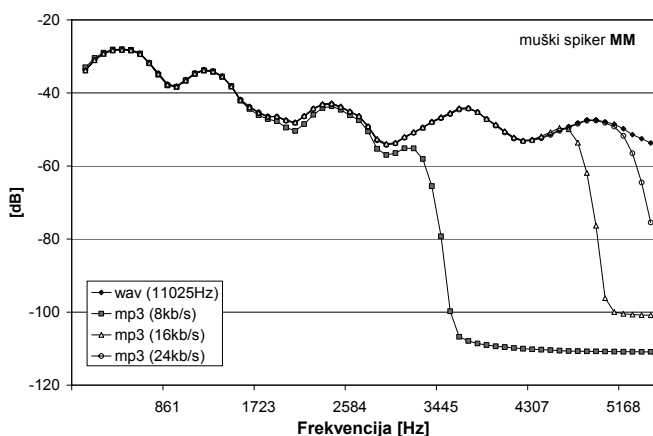
#### 2.2. Spektralna analiza

Na slici 1 i slici 2 su prikazani dugovremeni usrednjeni spektri kompletnih signala za sve vrste kodovanja signala za jednog muškog spikera (MM). Slika 1 se odnosi na kodovanja bazirana na semplovanju signala učestanošću od

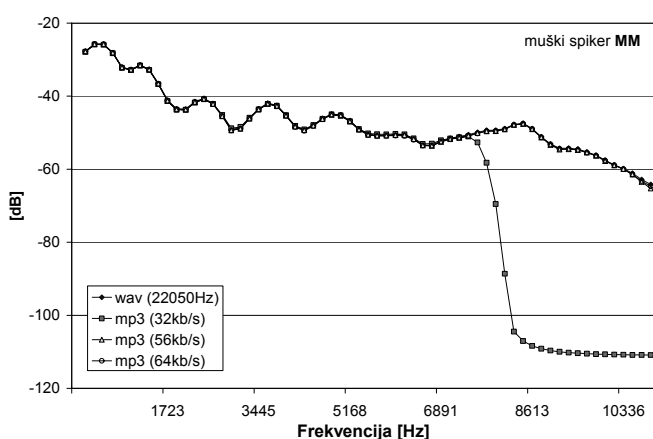
11025Hz, pri čemu su tri kodovana *mp3* algoritmom sa brzinama protoka od 8, 16 i 24 kb/s, dok je četvrti u *wav* formatu. Druga slika se odnosi na kodovanja bazirana na semplovanju signala učestanošću od 22050Hz, pri čemu su tri kodovana *mp3* algoritmom sa brzinama protoka od 32, 56 i 64 kb/s, dok četvrti predstavlja izvorni signal u *wav* formatu semplovan učestanošću od 22050Hz.

Sa slike 1 se može zapaziti da, u odnosu na dugovremeni spektar *wav* signala (11025Hz), primena algoritma kompresije ima za posledicu sužavanje frekvencijskog opsega u gornjem delu spektra i to proporcionalno stepenu kompresije. Pored toga, za algoritam kompresije *mp3*(8kb/s) se pokazuje i malo odstupanje (slabljenje) u prenosnoj karakteristici, već od oko 1700Hz.

Na slici 2 se uočava potpuno drugačija situacija. Naime, ovde je perfektno poklapanje dugovremenih spektara dobijenih za signale *mp3*(56kb/s) i *mp3*(64kb/s) sa onim dobijenim za izvorni signal *wav*(22050Hz). Spektar trećeg signala komprimovanog *mp3* algoritmom sa 32kb/s je skraćen na oko 7700Hz, ali se u ovom opsegu veoma dobro poklapa sa ostalim spektrima.



Sl.1. Uporedni prikaz usrednjenih spektara (za kompletan snimak) za tri stepena kompresije *mp3* algoritma u odnosu na *wav* snimak ( $f_s=11025\text{Hz}$ ).

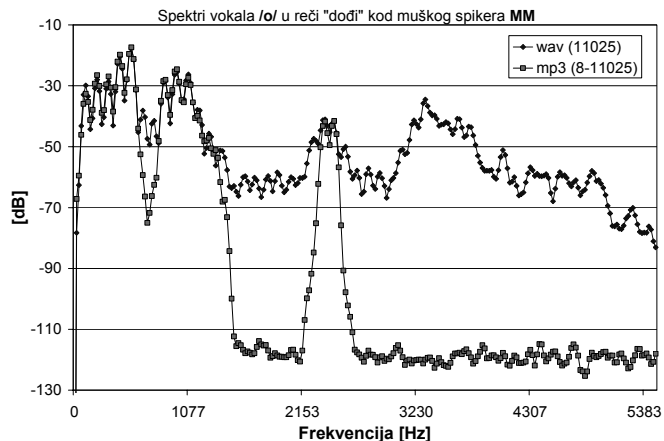


Sl.2. Uporedni prikaz usrednjenih spektara (za kompletan snimak) za navedena tri stepena kompresije *mp3* algoritma u odnosu na izvorni snimak ( $f_s=22050\text{Hz}$ ).

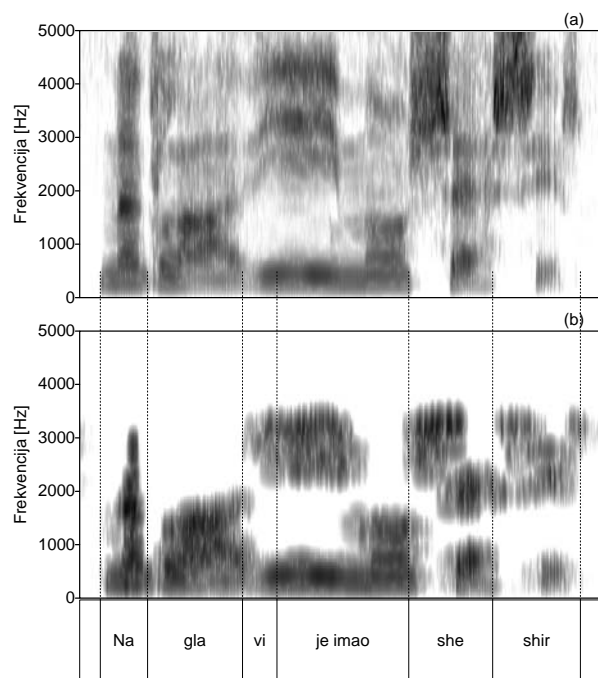
Na slici 3 je dat uporedni prikaz kratkovremenih spektara (za vokal /o/ usrednjen na segmentu trajanja od oko 80ms) za signal komprimovan *mp3* algoritmom sa protokom 8kb/s u odnosu na *wav* snimak pri  $f_s=11025\text{Hz}$ . Uočavaju se veliki spektralni procepi u komprimovanom signalu. Praktično se u cilju uštede prenose samo formanti dok se ostali delovi

ukidaju. Za slučajeve sa manjim faktorom kompresije ovi spektralni procepi su ređi i manji.

Na slici 4 je prikazan izgled sonograma za jednu rečenicu iz diskursa, koju je izgovorio isti spiker. Uporedni prikaz sonograma navedene rečenice za *wav* snimak pri  $f_s=11025\text{Hz}$  (gornja slika) i za signal komprimovan *mp3* algoritmom sa protokom 8kb/s (donja slika). I ovde se uočavaju velike spektralne praznine.



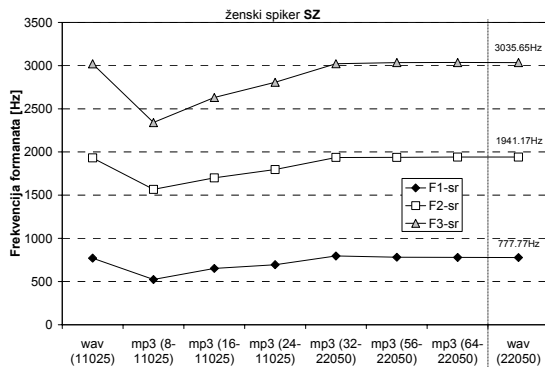
Sl.3. Uporedni prikaz spektara (za vokal /o/ usrednjen na segmentu od 80ms) za signal komprimovan *mp3* algoritmom sa protokom 8kb/s u odnosu na *wav* snimak pri  $f_s=11025\text{Hz}$ .



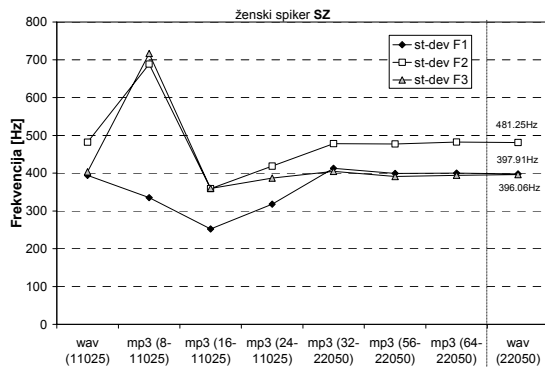
Sl.4. Uporedni prikaz sonograma navedene rečenice za *wav* snimak pri  $f_s=11025\text{Hz}$  (a) i za signal komprimovan *mp3* algoritmom sa protokom 8kb/s (b).

Na slici 5 su grafički prikazane pozicije prva tri formanta dobijene za 8 različitih vrsta konverzije signala govornog diskursa za ženskog spikera SZ. Krajnje (leva i desna) se odnose na *wav* snimke, pri čemu desni odgovara izvornom snimku ( $f_s=22050\text{Hz}$ , odvojen isprekidanom vertikalnom linijom), dok se srednjih 6 grupacija odnosi na signale komprimovane *mp3* algoritmom. Slika 9 se odnosi na prikaz istih merenja za muškog spikera MM.

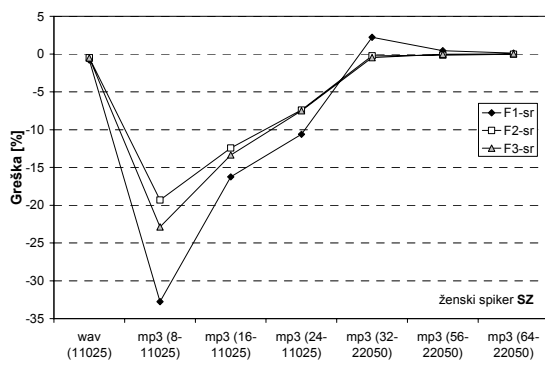
Grafici na slici 6 prikazuju pozicije prva tri formanta dobijene za 8 različitih vrsta konverzije signala govornog diskursa za ženskog spikera SZ.



Sl.5. Vrednosti prva tri formanta za govorni diskurs ženskog spikera SZ izmeren za osam načina digitalizacije signala (šest baziranih na mp3 algoritmu i dva u wav formatu).



Sl.6. Vrednosti standardne devijacije za prva tri formanta za govorni diskurs ženskog spikera SZ izmeren za osam načina digitalizacije signala (šest baziranih na mp3 algoritmu i dva u wav formatu).

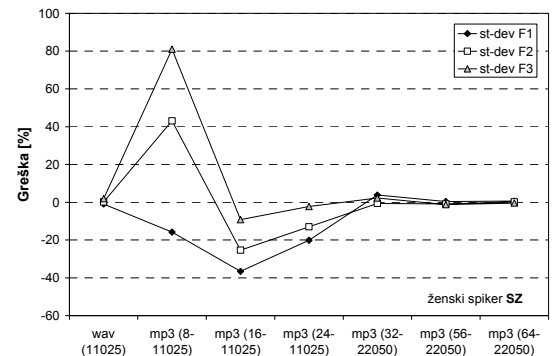


Sl.7. Odstupanje (greška u %) vrednosti formanta izmerenih za 6 načina kompresije izvornog signala i dvostruko podsemplovanom wav signalu u odnosu na vrednosti izmerene za izvornisignalya ženskog spikera SZ.

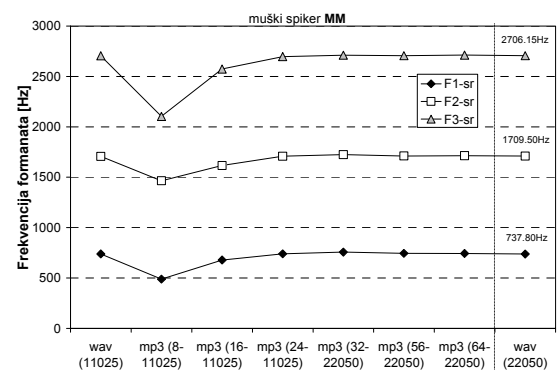
Grafici na slikama 7 i 8 predstavljaju odstupanje (grešku u %) određivanja vrednosti formanta za 7 različitih kompresija u odnosu na vrednosti dobijene za izvorni signal. Slika 7 je bazirana podacima predstavljenim na slici 5, dok je slika 8 u vezi sa slikom 6.

Odstupanje (greška u %) vrednosti formanta izmerenih za 6 načina mp3 kompresije izvornog signala i dvostruko podsemplovanom wav signalu u odnosu na vrednosti

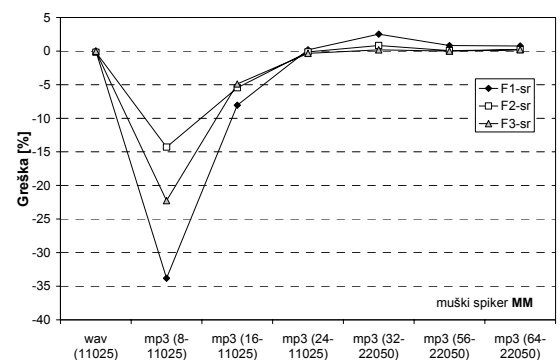
izmerene za izvorni signal za muškog spikera MM, grafički su prikazani na slici 9.



Sl.8. Odstupanje (greška u %) vrednosti standardnih devijacija formanta izmerenih za 6 načina kompresije izvornog signala i dvostruko podsemplovanom wav signalu u odnosu na vrednosti izmerene za izvorni signal za ženskog spikera SZ.



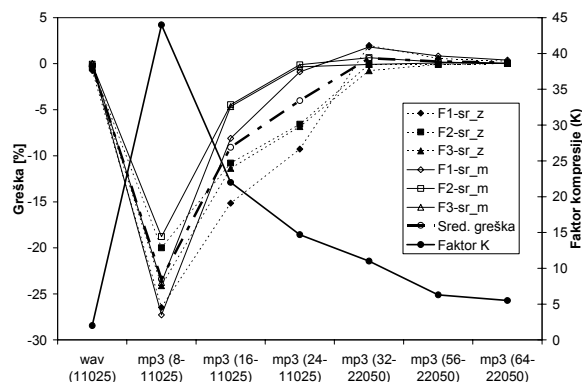
Sl.9. Vrednosti prva tri formanta za govorni diskurs muškog spikera MM izmeren za osam načina digitalizacije signala (šest baziranih na mp3 algoritmu i dva u wav formatu).



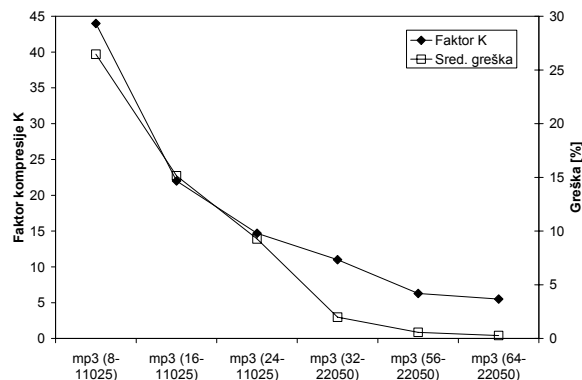
Sl.10. Odstupanje (greška u %) vrednosti formanta izmerenih za 6 načina kompresije izvornog signala i dvostruko podsemplovanom wav signalu u odnosu na vrednosti izmerene za izvornisignalya muškog spikera MM.

Na slici 11 su prikazani grafici Srednje odstupanje (greška u %) vrednosti formanta izmerenih za 6 načina kompresije izvornog signala i dvostruko podsemplovanom wav signalu u odnosu na vrednosti izmerene za izvorni signal za sva tri ženska, kao i za sva tri muška spikera. Na slici su, takođe prikazani i grafici ukupne srednje greške za svih 6 spikera za sva tri formanta, kao i grafik faktora kompresije za svaki od sedam načina konverzije izvornog signala.

Na slici 12 su posebno prikazana poslednja dva grafika i to samo za signale komprimovane mp3 algoritmom. Ovde je uzet apsolutni iznos greške. Interesantno je zapaziti da se ove dve konture imaju vrlo podudaran oblik.



Slika 11. Srednje odstupanje (greška u %) vrednosti formantata izmerenih za 6 načina kompresije izvornog signala i dvostruko podsemplovanom wav signalu u odnosu na vrednosti izmerene za izvorni signal za ženske i za muške spikera, kao i ukupno srednje odstupanje za svih 6 spikera i 3 formanta.



Slika 12. Srednja vrednost apsolutne greške (u %) vrednosti formantata izmerenih za 6 nivoa mp3 kompresije izvornog signala u odnosu na vrednosti izmerene za izvorni signal i odgovarajuće vrednosti faktora kompresije.

#### 4. DISKUSIJA REZULTATA

Uticao mp3 kompresije bilo kog stepena na osnovnu frekvenciju  $f_0$  govora je zanemarljiv. Slican zaključak se može doneti i za standardnu devijaciju  $f_0$ , osim u slučaju muškog spikera SK, kod koga postoji nešto veće odstupanje.

U pogledu dugovremenog spektra govornog signala algoritam mp3 kompresije uglavnom utiče na skraćivanje frekventijskog opsega i to proporcionalno stepenu kompresije. Poklapanje spektara u propusnom opsegu je odlično, osim u slučaju vrlo velikog stepena kompresije (8kb/s), gde se već od 1700Hz javlja mali pad (slabljenje signala) ali se oblik spektra zadržava. Kod kodovanja sa 56 i 64 kb/s spektri se potpuno, u celom frekventijskom opsegu, poklapaju sa spektrom izvornog signala. Što se tiče kratkovremenog spektra, tu je situacije nešto drugačija. Za vrlo veliki stepen kompresije javljaju se velike spektralne šupljine, koje se smanjuju za slučaj korišćenja mp3 algoritma sa većim brzinama protoka podataka, odnosno manjim faktorom kompresije signala.

Analiziranjem slika 7, 9 i 11 zapaža se mala razlika između rezultata dobijenih za ženske i za muške spikere. Naime, mp3 audio kompresija ima veći uticaj na tačnost merenja formantata kod ženskih glasova, nego kod muških. Najkarakterističniji je slučaj za 24kb/s, za koju muški glasovi ne pokazuju skoro nikakvo odstupanje u odnosu na izvorni signal dok je kod ženskih to odstupanje dosta značajno (od 5 do 10%).

#### 5. ZAKLJUČAK

Uzimajući u obzir napred iznete rezultate i zaključke, generalno se može reći da je uticaj mp3 algoritma audio kompresije govornog signala na tačnost merenja akustičkih parametara, koji se redovno koriste u forenzičkim analizama, zanemarljiv za faktore kompresije manje od 11 za žene i 14 za muškarce.

Kod kratkovremene spektralne analize treba biti oprezan pri njihovom tumačenju zbog unošenja spektralnih procepa.

Treba imati u vidu činjenicu da su ovi zaključci izvedeni na bazi rezultata analize govornih uzoraka snimljenih u studijskim uslovima, tako da bi sve ovo trebalo proveriti sa realnim signalima - dobijenim sa realnih sistema prenosa.

#### LITERATURA

- [1] S.T. Jovičić, Z. Kašić, M.M. Đorđević, M. Vojnović, M. Rajković, "Serbian emotional speech database: design, processing and evaluation", accepted for 9<sup>th</sup> International Conference SPEECH AND COMPUTER - SPECOM'2004, St. Petersburg, Russia, 2004.
- [2] Praat, <http://www.praat.org>.
- [3] A. Braun and C. Grasmück, "The Effect of Standard Audio Compression Algorithms on Forensically Relevant Measurements", Helsinki, Finland, 28th July - 31st July 2004
- [4] V. Kezić, "Kompresija audio signala primenom MPEG/audio standarda" NTB, No.2 1997, Beograd.
- [5] M. Đurđev, "Karakteristike komercijalnih mp3 enkodera" ETRAN'05, Budva, jun 2005.

**Abstract** – This paper presents results of analyses of effect of mp3 audio compression algorithm of speech signals on acoustic parameter, with respect on forensic application. For speech materials are used discurs (sub corpus) from corpus of emotional Serbian speech (GEES). Origin records of six speakers are converted in mp3 audio format with six different bit rates. On all record are performed measurement selected acoustic speech parameters, and compared them with that extracted from original records.

#### EFFECT MP3 COMPRESSION TO FORENSIC ANALYSIS OF SPEECH

Miodrag M. Đorđević