

Uticaj klipovanja govornog signala na osnovnu frekvenciju glasa

Milan Vojnović

Sadržaj — U radu je analiziran uticaj klipovanja na tačnost estimacije osnovne frekvencije glasa. Analizirana je osnovna frekvencija glasa osam govornika (četiri ženska i četiri muška) za različite nivoe klipovanja: od 0 do 30 dB. Rezultati analize pokazuju da klipovanje govornog signala praktično nema nikakav uticaj na estimaciju osnovne frekvencije glasa. Ova činjenica favorizuje njenu upotrebu u postupku forenzičke identifikacije govornika.

Gljučne reči — forenzika, klipovanje, identifikacija govornika, klipovanje, osnovna frekvencija glasa.

I. UVOD

Snimci govora koji se koriste u forenzičkoj identifikaciji govornika su najčešće telefonski razgovori dobijeni u toku istražnog postupka. Prenos govora kroz komunikacione kanale prate različite vrste izobličenja i smetnji. Sve ovo ima uticaja na razumljivost govora. Sa stanovišta identifikacije govornika, izobličenja govornog signala imaju uticaja na tačnost estimacije parametara govora koji se analiziraju. Kao krajnja konsekvencija, izobličenja govornog signala se manifestuju na validnost procesa identifikacije govornika. Zbog svega toga, nameće se potreba za sistematskim izučavanjem uticaja pojedinih vrsta izobličenja na validnost i pouzdanost procesa identifikacije govornika. Ovim istraživanjima treba da se odrede granice dozvoljenih izobličenja signala za validnu estimaciju parametara govora. Drugim rečima, utvrđuje se robustnost pojedinih parametara govora na različite vrste izobličenja. Ova saznanja su neophodna u trijažnom delu analize govornih signala gde se određuje koji parametri govora će biti analizirani u konkretnom slučaju identifikacije govornika. Izbor parametara, pre svega, zavisi od kvaliteta snimaka govora i vrste izobličenja koja su oni pretrpeli. Ceo postupak se može posmatrati iz ugla ponderisane analize govora: parametrima govora se daje različit značaj (ponderacija) u zavisnosti od tipa izobličenja koji se nalazi u snimku govora.

U forenzičkoj identifikaciji govornika najčešće se analiziraju sledeći parametri govora:

- osnovna frekvencija glasa,
- formantne frekvencije vokala,
- kepralni koeficijenti,

Ovaj rad je nastao u okviru projekta "E-medicine sistem za procenu kvaliteta sluha" broj 13011 koji je finansiran od strane ministarstva za nauku i zaštitu životne sredine Republike Srbije.

Milan Vojnović, Centar za unapređenje životnih aktivnosti – Inovacioni centar, Gospodar Jovanova 35, Beograd, Srbija (e-mail: vojnovicmilan@yahoo.com).

- dugovremeni spektar govora,
- raspodela energije govora i dr.

Osnovna frekvencija govora je jedan od elementarnih parametara analize glasa. To je parametar koji se, relativno, jednostavno izračunava. Verovatno je to parametar koji se najčešće analizira i to za najrazličitije namene: pevani glas, prepoznavanje emocija, prepoznavanje psiho-fizičkog stanja, poremećaji govornog aparata i sl. Postoje stotine razrađenih algoritama za njegovu ekstrakciju iz govornog signala, što implicira na njegov značaj. Sa druge strane, praksa je pokazala veliku robustnost ovog parametra na najrazličitije smetnje.

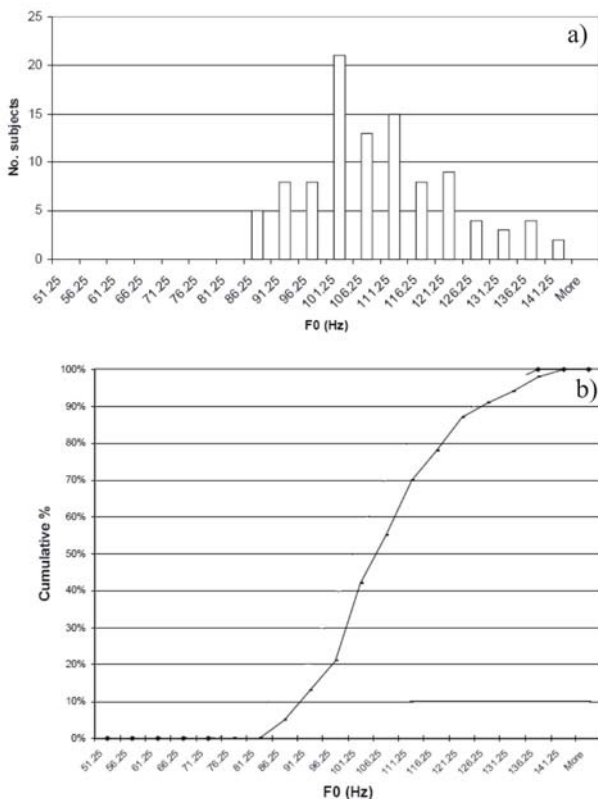
Za osnovnu frekvenciju glasa se ne može reći da je "čvrst" intra-govornički parametar, tj. ne može se reći da svaka osoba ima egzaktanu, fiksnu vrednost osnovne frekvencije glasa. Osnovna frekvencija glasa se menja u toku govora što znači da se pod ovim terminom podrazumeva neka srednja vrednost, najčešće statistička srednja vrednost. Zbog nestacionarne prirode osnovne frekvencije glasa moraju se uzeti u obzir i drugi parametri vezani za nju:

- opseg promena,
- standardna devijacija,
- intonacioni tok,
- statistička raspodela i dr.

Velika intra-govornička promena osnovne frekvencije glasa je njena najveća "mana" u postupku forenzičke identifikacije govornika. Ako se uzmu u obzir i različita psiho-fizička stanja govornika [1] onda je opseg promena još širi.

Na Sl. 1 prikazana je histogramska i kumulativna raspodela osnovne frekvencije glasa za engleski jezik [2]. Raspodela se odnosi na 100 ispitanika čiji je glas snimljen u simuliranom policijskom intervjuu. Svi ispitanici su bili muškog pola starosti između 18 i 25 godina. Srednja vrednost osnovne frekvencije glasa za ovu populaciju iznosi 106 Hz, a opseg promena je od 76 do 141 Hz. Interesantno je primetiti da 60% govornika ima osnovnu frekvenciju glasa između 94 i 113 Hz (opseg od 19 Hz). Osnovnu frekvenciju glasa nižu od 94 Hz ima 20% ispitanika. Isto toliko ispitanika ima osnovnu frekvenciju glasa veću od 113 Hz. Ovi rezultati se dobro poklapaju sa rezultatima drugih istraživača. U [3] je analiziran nemački jezik. Ušestvovalo je 105 muških govornika mlađe starostne dobi. Srednja vrednost osnovne frekvencije glasa za ovu populaciju iznosi oko 116 Hz. Dobijeno je da 60% govornika ima osnovnu frekvenciju glasa između 99 i

120 Hz (opseg od 27 Hz). U oba ova slučaja analiziran je spontani govor. Kod glume i čitanja dobijaju se drugačije vrednosti osnovne frekvencije glasa; uglavnom više vrednosti.



Sl. 1. Histogram i kumulativna raspodela osnovne frekvencije glasa za 100 ispitanika [2].

Krajevi kumulativne raspodele (Sl. 1b)) su veoma bitni jer govornici sa relativno niskom i relativno visokom osnovnom frekvencijom glasa su retki. Ako se u postupku forenzičke identifikacije govornika konstatuje da osumnjičeni ima osnovnu frekvenciju glasa iz ovih regiona onda postoje dobri uslovi za validniju identifikaciju govornika.

Jedan od najčešćih izobličenja telefonskog govornog signala je klipovanje (odsecanje vršnih vrednosti) i različite šumne i impulsne smetnje. Klipovanje je posebno interesantna vrsta izobličenja obzirom na njeno destruktivno dejstvo. Naime, ne postoji način da se rekonstruiše klipovani govorni signal. Kod ostalih vrsta izobličenja postoje različite tehnike za kondicioniranje govornog signala.

U ovom radu analizirano je kako klipovanje utiče na tačnost estimacije osnovne frekvencije glasa.

II. PROCEDURA ISPITIVANJA

Govorni materijal koji je korišćen za ovu analizu predstavlja snimke dnevnika televizijske stanice B-92. Snimanje je napravljeno tokom 2005. godine i predstavlja internu bazu autora ovog rada. Svi snimci su očišćeni u smislu da su izbrisani neželjeni glasovi i zvuci tako da je ostao samo glas voditelja dnevnika. Ukupno je analizirano osam govornika: četiri ženska i četiri muška. Od ovih osam

govornika pet su profesionalni voditelji (tri ženska i dva muška), a preostala tri govornika su bili gosti emisije (jedan ženski i dva muška).

Trajanje snimaka govora kreće se u rasponu od 39 do 307 s (prosečno trajanje je oko 216 s). Početni govorni signali su normalizovani na vršnu vrednost, tj. signali su maksimalno pojačani ali nema klipovanja. Svaki od ovih osam normalizovanih snimaka govora je pojačan za 5, 10, 15, 20, 25 i 30 dB tako da je dobijeno još 48 klipovanih snimaka govora. Nivo klipovanja se definiše kao decibelski nivo pojačanja vršno normalizovanog signala [4]. Dakle, ukupno je analizirano 56 snimaka govora: 8 govornika \times 7 nivoa klipovanja.

Estimacija osnovne frekvencije govora je urađena programom PRAAT [5]. Parametri analize su bili sledeći:

- frekvencija odmeravanja : 22050 Hz
- metod analize : kros-korelacioni
- donja granica analize : 50 Hz
- gornja granica analize : 600 Hz
- korak analize : 20 ms
- gornji nivo "tišine" : 0,03
- donji nivo govora : 0,45

Za 56 snimaka govora estimirane su osnovne frekvencije glasa. Tačnije, estimirane su statističke raspodele (histogrami). Na osnovu ove raspodele određene su srednje vrednosti i standardne devijacije. Komparativnom analizom je utvrđen uticaj klipovanja na tačnost estimacije osnovne frekvencije glasa. U suštini, upoređeni su oblici histogramске raspodele, srednje vrednosti i standardne devijacije osnovne frekvencije glasa za slučaj klipovanog i neklipovanog govornog signala.

III. REZULTATI ANALIZE

Sumarni rezultati analize prikazani su u tabeli 1. Date su srednje vrednosti i standardne devijacije osnovnih frekvencija govora za osam ispitanika i za sedam nivoa klipovanja. Na osnovu podataka o osnovnoj frekvenciji glasa koja je u funkciji vremena (intonacioni tok), srednja vrednost i standardna devijacija se mogu odrediti prema sledećim relacijama:

$$\overline{F_0} = \frac{F_{0,1} + F_{0,2} + F_{0,3} + \dots + F_{0,N}}{N} \quad (1)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (F_{0,i} - \overline{F_0})^2}{N-1}} \quad (2)$$

$F_{0,i}$ je pojedinačno estimirana vrednost osnovne frekvencije glasa, a N njihov ukupni broj u govornom signalu.

Drugi način je da se histogramске raspodele osnovne frekvencije glasa aproksimiraju (fituju) normalnom, odnosno Gauss-ovom raspodelom. Ovo je opravdano obzirom da osnovna frekvencija glasa ima Gauss-ovu raspodelu. Rezultati prikazani u tabeli 1 odnose se na fitovanje histogramске raspodele Gauss-ovom krivom. U

poređenju sa rezultatima koji se dobijaju kada se koriste relacije (1) i (2) postoje male razlike, ali su one male i nemaju neki praktični značaj na konačne zaključke o uticaju klipovanja na osnovnu frekvenciju glasa.

Prema rezultatima iz tabele 1, može se zaključiti da klipovanje nema nikakav praktični uticaj na tačnost estimacije srednje vrednosti i standardne devijacije osnovne frekvencije glasa. Razlike koje izaziva klipovanje su reda 1-2 Hz kako za srednju vrednost tako i za standardnu devijaciju. Jedino je malo veća razlika dobijena kod prvog ispitanika (GD), oko 5 Hz, što je isto tako zanemarljivo.

Na Sl. 2 prikazane su histogramske raspodele osnovne frekvencije glasa za svih osam ispitanika, ali samo za dva krajnja slučaja klipovanja:

- nivo klipovanja 0 dB (nema klipovanja) i
- nivo klipovanja 30 dB.

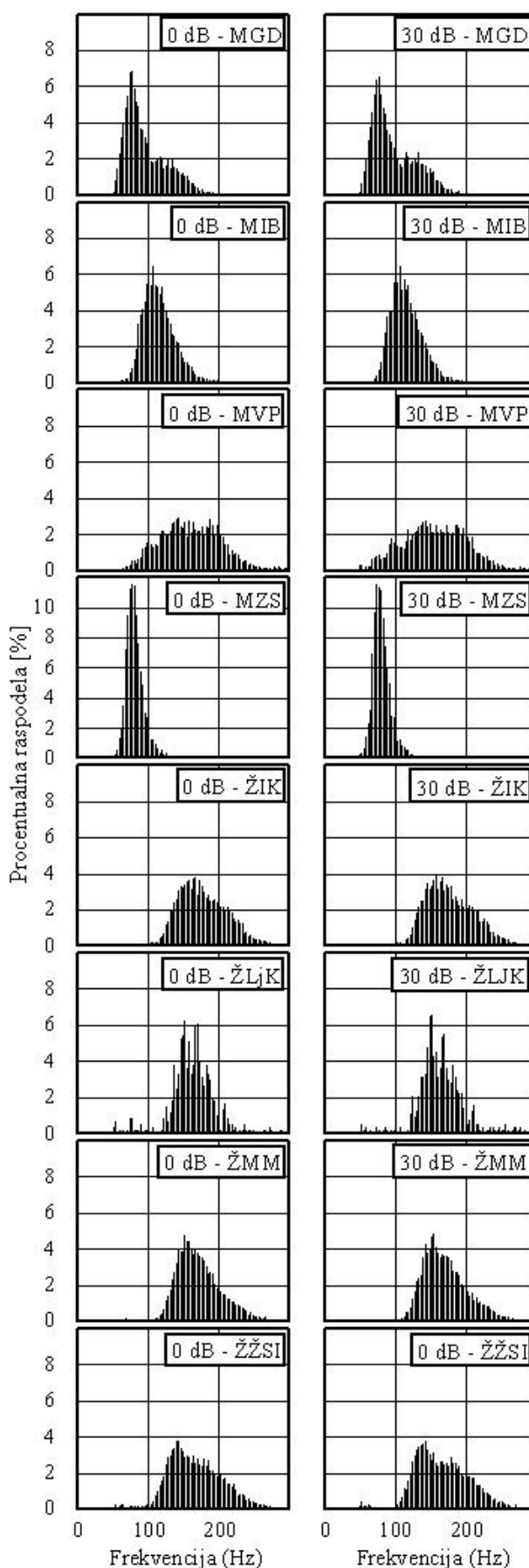
Ispred inicijala ispitanika dodato je slovo za označavanje pola ispitanika: M-muški, Ž-ženski.

Ostali nivoi klipovanja nisu prikazani jer nema dodatnih promena u raspodeli. Ono što se sa sigurnošću može zaključiti je da klipovanje govora nema nikakav praktični uticaj ni na statističku raspodelu osnovne frekvencije glasa.

TABELA 1: SREDNJE VREDNOSTI I STANDARDNE DEVIJACIJE OSNOVNE FREKVENCIJE GLASA U FUNKCIJI KLIPOVANJA.

	Isp.	GD	IB	VP	ZS	IK	LjK	MM	ŽSI								
										Pol	M	M	M	Ž	Ž	Ž	Ž
										T [s]	307	261	156	187	238	39	292
Nivo klipovanja	0 dB	$\overline{F_0}$	81	110	158	78	172	161	164	160							
		σ	20	20	46	10	35	22	27	37							
	5 dB	$\overline{F_0}$	81	110	158	78	172	161	164	160							
		σ	20	21	47	10	35	22	28	38							
	10 dB	$\overline{F_0}$	82	110	157	78	172	161	163	161							
		σ	20	21	48	10	35	23	28	38							
	15 dB	$\overline{F_0}$	82	110	156	78	172	161	163	161							
		σ	20	21	48	10	35	23	28	38							
	20 dB	$\overline{F_0}$	85	110	156	78	172	161	163	161							
		σ	23	21	49	10	35	23	28	38							
	25 dB	$\overline{F_0}$	86	110	156	78	171	161	163	161							
		σ	24	21	49	10	35	23	28	38							
	30 dB	$\overline{F_0}$	85	111	156	78	171	160	163	161							
		σ	25	21	49	10	35	23	28	38							
	∞ dB	$\overline{F_0}$	94	112	159	78	172	163	164	162							
		σ	31	20	47	9	35	22	28	38							

Dakle, konačno se može zaključiti da je osnovna frekvencija glasa potpuno imuna na klipovanje govornog signala jer nema promena ni srednje vrednosti, ni standardne devijacije ni oblika histogramske raspodele. Pri tome ne treba ispustiti iz vida da je analizom obuhvaćeno



Sl. 2. Histogramska raspodela osnovne frekvencije glasa za slučaj neklipovanog govora (leva strana) i za slučaj klipovanja nivoom 30 dB (desna strana).

klipovanje nivoa 30 dB koje uzrokuje prilično veliku degradaciju govornog signala. Veći nivoi klipovanja od ovog nemaju praktični značaj u forenzičkoj identifikaciji govornika jer je onemogućena ekstrakcija ostalih parametara glasa.

Kao kuriozitet, analiziran je i krajnji slučaj klipovanja, tzv. beskonačno klipovanje koje se realizuje po sledećem algoritmu:

$$a(t) = \begin{cases} +A & \text{ako je } a(t) > 0 \\ -A & \text{ako je } a(t) < 0 \end{cases} \quad (3)$$

Sve amplitude signala veće od nule zamenjuju se maksimalnom pozitivnom vrednošću, a ako su manje od nule maksimalnom negativnom vrednošću. Čak i za ovaj krajnji slučaj nema većih promena u estimaciji osnovne frekvencije glasa. Pri tome se misli na promene srednje vrednosti, standardne devijacije i histogramске raspodele.

U ovom radu je prikazana izuzetno velika imunost osnovne frekvencije glasa na klipovanje. U nekim ranijim radovima [6] je konstatovan delimičan uticaj klipovanja na dugovremeni usrednjeni spektar govora, odnosno u [4] veliki uticaj klipovanja na formantne frekvencije vokala.

Izbor parametara analize u postupku identifikacije govornika svakako treba prilagoditi ovim činjenicama. U tom smislu, ako je došlo do klipovanja govornog signala analizirane parametre glasa treba vrednovati (ponderisati) na sledeći način:

- najveći značaj dati osnovnoj frekvenciji glasa,
- nešto manji značaj dati dugovremenom spektru i
- najmanji značaj dati formantnim frekvencijama vokala.

Histogramske raspodele prikazane na Sl. 2 imaju Gauss-ovu raspodelu. Jedino postoji mala deformacija ove raspodele kod prvog ispitanika (GD) koja je bimodalna. Zbog ovog odstupanja od Gauss-ove raspodele javile su se veće greške u estimaciji srednje vrednosti i standardne devijacije osnovne frekvencije glasa (tabela 1). Dakle, veće razlike u osnovnoj frekvenciji glasa kod ispitanika GD nisu posledica klipovanja već su rezultat greške u procesu fitovanja Gauss-ovom raspodelom. Ako se srednje vrednosti i standardne devijacije za ispitanika GD izračunaju relacijama (1) i (2), respektivno, onda će promene biti manje i iznosiće 2 Hz što je istog reda kao i kod ostalih ispitanika.

Treba obratiti pažnju na još jednu činjenicu: histogramske raspodele se razlikuju za pojedine govornike. Ova činjenica bi se mogla koristiti u postupku forenzičke identifikacije govornika. I ne samo to, treba koristiti i druge kvazi-prozodijske parametre kao što su prvi i drugi izvod intonacionog toka. Isto kao i osnovna frekvencija glasa, i ovi parametri pokazuju veliku robustnost na klipovanje pa ih treba češće koristiti u postupku identifikacije govornika.

IV. ZAKLJUČAK

Rezultati analize pokazuju da je osnovna frekvencija glasa izuzetno robustan parametar kada je u pitanju

klipovanje. Čak i za velike nivoe klipovanja (reda 30 dB) praktično nema nikakvih promena srednje vrednosti, standardne devijacije i histogramске raspodele.

Kada se u postupku forenzičke identifikacije govornika koriste snimci govora degradirani klipovanjem onda najveći značaj treba dati osnovnoj frekvenciji glasa, zatim dugovremenom usrednjenom spektru i na kraju formantnim frekvencijama vokala.

Jedan od problema koji ostaje za naredna istraživanja je definisanje nivoa klipovanja u govornom signalu. Treba videti da li statistička raspodela trenutnih vrednosti govornog signala može da bude pouzdan pokazatelj nivoa klipovanja.

Histogramske raspodele sa Sl. 2 pokazuju razlike u obliku za pojedine govornike. Da li se ova konstatacija može koristiti kao parametar u forenzičkoj identifikaciji govornika? Ovo treba proveriti u nekim budućim istraživanjima, stim da broj ispitanika bude veći, najmanje 30-tak.

LITERATURA

- [1] Murray I.R., Arnnot, J.L., "Towards the simulation of emotion in synthetic speech: A review of the literature of human vocal emotion", *J. Acoust. Soc. Am.*, 93(2), pp.1097-1108, 1993.
- [2] Hudson T., de Jong G., McDougall K., Harrison P., Nolan F., "F0 statistics for 100 young male speakers of southern british english", *16th International Congress of Phonetic Sciences ICPhS 2007*, pp.1809-1812, Saarbrücken, 2007, Germany.
- [3] Künzel H.J., "How well does average fundamental frequency correlate with speaker height and weight?", *Phonetica*, 46, pp.117-125, 1989.
- [4] Vojnović M., "Uticaj klipovanja govornog signala na formantne frekvencije vokala", *XVI Telekomunikacioni forum TELFOR*, 2008, Beograd, biće objavljen.
- [5] Boersma P., Weenink D., "PRAAT: A system for doing phonetics by computer", 1992-2005, <http://www.praat.org/>.
- [6] Vojnović M., "Prepoznavanje govornika pomoću dugovremenog usrednjenog spektra", *Nauka Tehnika Bezbednost*, br. 2, str. 53-66, Beograd, 2005.

ABSTRACT

The influence of speech clipping on the voice fundamental frequency estimation had been analyzed. Eight speakers (four female and four male) take part in experiment. The speech signal was clipped with different levels from zero to 30 dB. The analyze results show that speech signal clipping have no any practical influence on voice fundamental frequency estimation. This fact favors usage of voice fundamental frequency in forensic speaker identification procedure.

THE INFLUENCE OF THE CLIPPING ON THE VOICE FUNDAMENTAL FREQUENCY

Milan Vojnović