

# Uticaj načina upotrebe mobilnog telefona na dugovremeni usrednjeni spektar govora

Nikola Jovanović

**Sadržaj** — Cilj ovog istraživanja je pronalaženje intra-spikerskih i inter-spikerskih razlika kod govornika koji su koristili mobilni telefon na 12 načina. Analizirane su promene dugovremenog usrednjenog spektra govora (*Long-Term Average Speech Spectrum, LTASS*). U istraživanju su učestvovala tri muška i tri ženska govornika. U proseku, spektri LTASS muških govornika su oslabljeni za bar 5 dB između približno 1900 i 2900 Hz, dok su kod ženskih govornika pojačani za bar 5 dB u tom frekvencijskom opsegu.

**Gljučne reči** — dugovremeni usrednjeni spektar govora, LTASS, prenos govora mobilnim telefonom, prepoznavanje govora i govornika.

## I. UVOD

GOVORNA komunikacija služi ljudima za razmenu informacija i predstavlja najlakši, najefikasniji i najrobusniji oblik komuniciranja. Ljudi koriste mobilni telefon za razgovor u situacijama kada govorna komunikacija nema svoja uobičajena svojstva. Prema [1], pod uobičajenom govornom komunikacijom podrazumeva se komunikacija dve zdrave osobe normalnim intenzitetom glasa i izraženim emocijama, bez napora i stresa, i u ambijentu čija buka ne predstavlja smetnju u sporazumevanju. Neuobičajena govorna komunikacija podrazumeva šaputanje, govor povišenog glasa (viku), patološki govor i slično.

Promene u izgovoru istih reči/rečenica od strane različitih govornika nazivaju se inter-spikerske promene, a promene u izgovoru iste reči/rečenice od strane istog govornika se nazivaju intra-spikerske promene. U ovom radu analizirana je zavisnost intra-spikerskih i inter-spikerskih razlika od načina upotrebe mobilnog telefona analiziranjem promena dugovremenog usrednjenog spektra govora (*Long-Term Average Speech Spectrum, LTASS*). Poznavanje osobina spektara LTASS može doprineti boljem prepoznavanju govora i govornika.

Rad je organizovan tako da posle upoznavanja sa dugovremenim usrednjenim spektrom govora u drugom odeljku, sledi opis uticaja upotrebe mobilnog telefona na govor u trećem odeljku. U četvrtom je opisan sadržaj istraživanja, u petom formiranje govorne baze i u šestom su analizirani eksperimentalni rezultati. U zaključku su sumirani dobijeni rezultati.

Nikola Jovanović, student-diplomac, Elektrotehnički fakultet u Beogradu, Bulevar kralja Aleksandra 73, 11120 Beograd, Srbija (telefon: 065-985-12-09; e-mail: etfovac@yahoo.com).

## II. DUGOVREMENI USREDNJNI SPEKTAR GOVORA

Usrednjavanjem više kratkovremenih spektara izračunatih na segmentima govornog signala dužine od 10 do 20 ms dobija se dugovremeni usrednjeni spektar govora (LTASS). Prema [2], u praktičnim situacijama za validnu estimaciju LTASS minimalno trajanje govornog signala treba da bude 30 ili više sekundi.

Osnovne karakteristike LTASS, prema [3], su: maksimum u okolini frekvencije 500 Hz i pad visokofrekvencijskih komponenti od 9 dB/oktavi. Maksimum na oko 500 Hz je u vezi sa prosečnom dužinom vokalnog trakta odraslog muškarca koja iznosi približno 17 cm. Kada se vokalni trakt modelira pomoću cevi dužine 17 cm koja je na jednom kraju zatvorena a na drugom otvorena onda ona ima prvu rezonancu, tj. pol prenosne funkcije, na oko 500 Hz.

LTASS se koristi kada se posmatra spektralni sastav govornog signala bez obzira na njegov fonetski sadržaj. Odnosno, usrednjavanjem tokom dužeg govornog iskaza akustičke karakteristike fonema se gube i ostaje samo opšti profil spektralne gustine snage govora.

Statistička ispitivanja spektara LTASS različitih govornika su pokazala da se oni razlikuju po specifičnostima kao što su lokalni minimumi i maksimumi. Dugovremeni usrednjeni spektar govora je vrlo pouzdan i robusan parametar u procesu identifikacije govornika jer je imun na različite vrste smetnji i izobličenja [3].

## III. UTICAJ UPOTREBE MOBILNOG TELEFONA NA GOVOR

Rezultati istraživanja uticaja telefona na govorni signal, sumirani u [4], otkrili su tri grupe efekata koje zajedno izobličavaju govor prenet telefonom: efekte okoline, efekte govornika i efekte tehnologije.

Efekte okoline uzrokovani su fizičkim okruženjem u kojem je obavljen telefonski razgovor. Na primer, telefonski poziv može biti upućen iz okruženja sa visokim nivoom ambijentalnog šuma ili iz prostorije sa izraženom reverberacijom.

Efekte govornika nastaju kada čovek svesno ili nesvesno promeni svoje ponašanje pri pričanju preko telefona. Jako izražen efekat koji se javlja kod nekih govornika je usvajanje drugačijeg načina izražavanja koji se naziva „telefonski glas“. Drugačije izražavanje podrazumeva promene u kvalitetu glasa, brzini pričanja i izgovoru segmenata rečenica. Već na prvi pogled je očigledno da korisnici mobilnih telefona imaju tendenciju

da pričaju glasnije, možda glasnije i od korisnika fiksnih telefona [4]. Zbog toga se efekat povećanja osnovne frekvencije glasa, primećen kod upotrebe fiksnog telefona, pogoršava pri upotrebi mobilnog telefona.

Efekti tehnologije se odnose na pojave koje su rezultat prenosa govornog signala putem mobilne i putem fiksne telefonije. Izobličenja govornog signala zavise od: tipa telefonskog aparata, prenosnog puta do i od telefonske centrale, ali i od opreme za snimanje koja je korišćena pri analizi govornog signala. Najznačajniji efekat tehnologije je svakako selektivan prenos frekvencija govornog signala jer se tako smanjuje njegova razumljivost. Govorni signal prenet telefonskim putem je ograničen u frekvencijskom domenu na opseg od 300 Hz do 3400 Hz. Prema [1], kod uskopojasnog govora sa smanjenjem frekvencijskog opsega smanjuje se razumljivost, ali se ona u određenim granicama može održati konstantnom sa povećanjem nivoa govora. Ovim se takođe može objasniti glasnije pričanje preko telefona i povećanje osnovne frekvencije glasa.

Prema [4], u proseku su frekvencije prvog formanta u uslovima mobilnog prenosa oko 29 % više od onih izmerenih u direktnom snimku, vrednosti drugih i trećih formantnih frekvencija nisu izmenjene, osim kod govornika sa relativno visokim frekvencijama trećeg formanta koje se smanjuju.

#### IV. SADRŽAJ ISTRAŽIVANJA

##### A. Cilj istraživanja

Cilj ovog istraživanja je pronalaženje intra-spikerskih i inter-spikerskih razlika kod govornika koji su koristili mobilni telefon na 12 uobičajenih ali različitih načina. U tu svrhu su analizirane promene LTASS spektara govora u međusobnom relativnom odnosu.

##### B. Postupak istraživanja

Dugovremeni spektri su razvrstani po tri kriterijuma: polu govornika, načinu držanja mobilnog telefona i načinu/okolnostima izgovora diskursa. Tri muške osobe i tri ženske osobe su izgovorile isti tekst diskursa koristeći mobilni telefon na 12 načina, odnosno u 12 situacija čiji je pregled dat u tabeli 1. Snimanje signala i estimacija spektra LTASS je obavljena pomoću programa *Adobe Audition*.

Govorni signali su segmentirani na segmente dužine 11,61 ms korišćenjem Hamingovog prozora sa 512 odbiraka. Izračunavanje kratkovremenih spektara je izvršeno pomoću FFT u 512 ekvidistantnih tačaka sa rezolucijom od 86,132 Hz.

Pri analiziranju promena spektara LTASS posmatrane su spektralne razlike. Usvojeno je da vrednost svakog spektra LTASS na frekvenciji 517 Hz iznosi 0 dB, tj. spektri su normalizovani (frekvencija od 517 Hz je odabrana jer je najbliža frekvenciji od 500 Hz). Spektar LTASS pri normalnom držanju mobilnog telefona je usvojen za referentni i on je oduziman od ostalih spektara, tako da su analizirane promene dugovremenih spektara u 11 načina upotrebe mobilnog telefona. Opseg frekvencija

u kojima su analizirane spektralne razlike je 258÷3531 Hz.

TABELA 1: DVANAEST ANALIZIRANIH NAČINA UPOTREBE MOBILNOG TELEFONA.

Sit.	Držanje mob. telefona	Izgovor diskursa
1	Normalno	Normalan
2	Normalno	Sa bombonom u ustima
3	Normalno	Sa cigaretom u ustima
4	Normalno	Ljutito
5	Normalno	Šapatom
6	Normalno	U <i>babble</i> buci
7	Normalno	U automobilskoj buci
8	Upravno na uvo	Normalan
9	Sa maramicom preko mobilnog telefona	Normalan
10	Sa rukom preko usta i telefona	Normalan
11	Upravno na uvo kada je govornik savijen na bok	Normalan
12	Između ramena i obraza	Normalan

#### V. FORMIRANJE GOVORNE BAZE

##### A. Govorni materijal

Govorni materijal čine 72 govorna snimka dobijena izgovaranjem unapred pripremljenog neinteraktivnog govornog diskursa. Snimanje je obavljeno u prostoriji Instituta za eksperimentalnu fonetiku i patologiju govora u Beogradu početkom jula 2009. godine. Svi govornici su koristili mobilni telefon proizvođača *Nokia* model *6600 slide* čiji pretplatnički broj pripada mobilnoj telefonskoj mreži 064.

Govorni materijal je sačuvan u obliku standardnih Microsoft Wave (.wav) mono audio fajlova u nekompresovanom PCM formatu sa frekvencijom odabiranja 44,1 kHz i sa rezolucijom od 16 bita po odbirku.

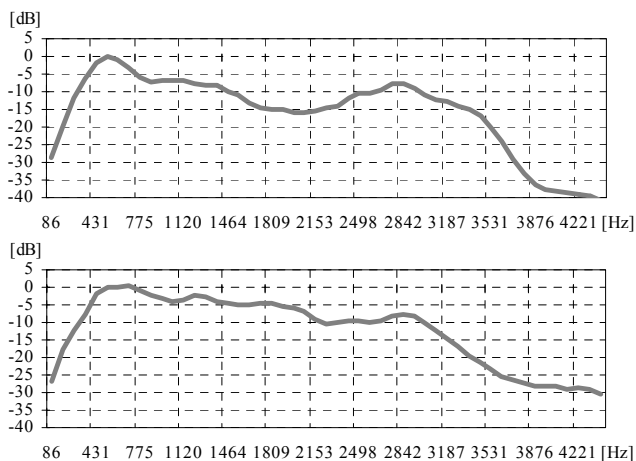
##### B. Organizacija snimanja

Svaki učesnik je svoja snimanja obavio tokom istog dana. Govornik bi pozvao broj fiksnog telefona koji se nalazi u istoj prostoriji. Kabl fiksnog telefona je preko razdelnika povezan na zvučnu kartu računara. Kada se veza uspostavi, govornik bi seo na stolicu i počeo sa čitanjem teksta trudeći se da na najbolji način dočara svaku od 12 situacija iz tabele 1. Prosečno trajanje govornih signala je 31,65 s.

Snimanje nije obavljeno u zvučno izolovanoj („gluvoj“) prostoriji, tako da su govorni signali izloženi spoljašnjim uticajima, poput reverberacije, baš kao u realnom slučaju. Ipak, prostorija u kojoj je snimano bila je tiha i zaklonjena od spoljašnje buke.

Da bi se simulirao ambijent sa žamorom, tj. *babble* bukom, i ambijent sa bukom koja postoji pri vožnji u automobilu preko zvučnika drugog računara su emitovana dva test signala *Pub\_Noise\_Binaural\_V2.wav* i *Midsized\_Car2\_100\_Kmh\_binaural\_(0.00-30.00\_s).wav* preuzeta iz ETSI baze na internet adresi

<http://portal.etsi.org>. U oba slučaja je odnos signal-šum bio približno 18 dB. Na taj način govornici su izgovarali tekst u buci i pod utiskom ambijentalne buke izmenili su svoju uobičajenu govornu komunikaciju. Ove izmene u govoru mobilni telefon je verno preneo dok je buka bila znatno potisnuta, jer savremeni mobilni telefoni imaju efikasne algoritme za potiskivanje buke, i nije uticala na spektre LTASS. Na Sl. 1 prikazane su usrednjene vrednosti normalizovanih referentnih spektara LTASS za muške i ženske govornike.



Sl. 1. Prosečne vrednosti normalizovanih referentnih LTASS za muške (gore) i ženske (dole) govornike.

## VI. ANALIZA REZULTATA

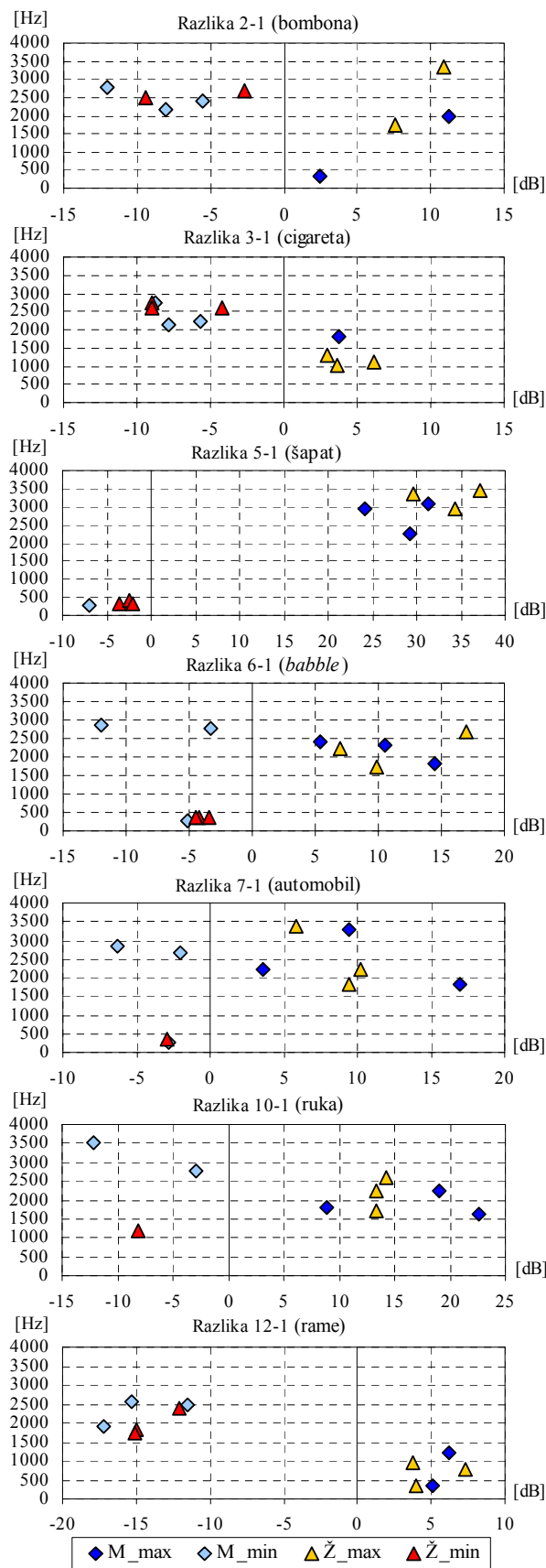
Na osnovu dobijenih rezultata može se zaključiti da su posebno interesantni slučajevi kada govornici drže bombonu ili cigaretu u ustima, kada se nalaze u ambijentu sa bukom, kada šapuću, kada drže ruku preko usta i telefona i kada pridržavaju mobilni telefon između obraza i ramena. U ovim situacijama apsolutne vrednosti minimuma i maksimuma spektralnih razlika dostižu i 15 dB, a kod šapata i preko 30 dB (videti Sl. 2). Tumačenje ovih razlika je sledeće.

U ambijentu sa bukom svi govornici su povećali nivo glasa posebno na višim frekvencijama (slučajevi 6-1 i 7-1). Ova pojava se naziva Lombardov efekat. U slučaju držanja bombone i cigarete u ustima govornici su promenili svoj uobičajeni način artikulacije tako što su smanjili pokretljivost donje vilice i otvor usana (slučajevi 2-1 i 3-1).

Zbog zamene zvučnog dela pobude vokalnog trakta šumnom pobudom intenzitet šapata na niskim frekvencijama je niži u odnosu na normalan (uobičajen) govor za 25÷35 dB (slučaj 5-1). Prostor ograničen rukom (dlanom), otvorom usta i mobilnim telefonom se ponaša kao akustički rezonator pojačavajući određene komponente spektra, u ovom slučaju na učestanostima između 1464 Hz i 2756 Hz (slučaj 10-1).

Držanje telefona između ramena i obraza znači savijanje vrata prema ramenu. Ovakav položaj govornika je očigledno izmenio konfiguraciju vokalnog trakta, pre svega farinksa i usne šupljine, a ograničio je i pokrete donje vilice. Spektralne razlike u proseku imaju niže

vrednosti (za oko 10 dB) u opsegu 1809÷2842 Hz kod muških, odnosno 1464÷2498 Hz kod ženskih govornika (slučaj 12-1).



Sl. 2. Frekvencijska raspodela ekstremuma kod najinteresantnijih spektralnih razlika.

### A. Maksimumi i minimumi spektralnih razlika

Ekstremumi čija je apsolutna vrednost manja od 2 dB nisu uzeti u obzir. Minimumi razlika 2-1 (bombona) i 3-1 (cigareta) su grupisani između 2000 i 3000 Hz i imaju vrednosti između -5 i -10 dB. Slično je i sa minimumima razlike 12-1 (rame) čiji su minimumi na nižim frekvencijama (1500÷2500 Hz) i imaju najveću apsolutnu vrednost (oko -15 dB).

Najmanje apsolutne vrednosti minimuma i ujedno minimume na najnižim frekvencijama imaju razlike 5-1 (šapat), 6-1 (*babble*) i 7-1 (automobil). Šaputanje se izdvaja po tome što ima najveće maksimume spektralnih razlika koji približno imaju vrednosti od 25 do čak 40 dB što znači da LTASS kod šaputanja ima relativno velike vrednosti na višim učestanostima (iznad 2000 Hz) u odnosu na komponente na niskim frekvencijama. Osim pri šaputanju, veliko pojačanje (reda 15 dB) LTASS ima i pri držanju ruke preko usta i pri pričanju u ambijentu sa reprodukovanom *babble* bukom.

Ako se izuzme šaputanje, spektralne razlike retko premašuju 15 dB. Zapravo, kod većine razlika, tačnije kod 7 od 11, je primećeno da su maksimalne vrednosti grupisane između vrednosti 5 i 10 dB. Takođe, veoma malo minimuma se spušta ispod -15 dB. Prosečne vrednosti ekstremuma dobijene usrednjavanjem za sve govornike iznose -7,4 dB i 10,3 dB. Minimum spektralnih razlika se u proseku nalazi na 2065 Hz, a maksimum na 2129 Hz. Prema tome, ekstremume spektralnih razlika u proseku treba tražiti oko frekvencije 2100 Hz.

### B. Karakteristični opsezi spektralnih razlika

U tabeli 2 dat je pregled najširih frekvencijskih opsega u okviru kojih su odstupanja spektralnih razlika veća od 5 dB. Povećanje spektralnih razlika je obeleženo znakom (+), a smanjenje znakom (-). Znak (/) je upisan ako je broj pluseva i minusa jednak, što se desilo kod razlike 2-1.

TABELA 2: POJAČANI I OSLABLJENI FREKVENCIJSKI OPSEZI SPEKTRALNIH RAZLIKA KOD SVIH GOVORNIKA.

Spektralna razlika	Znak promene	Srednja vr. donje gran. frekv. [Hz]	Srednja vr. gornje gran. frekv. [Hz]
2-1 bombona	/	1929	2343
3-1 cigareta	-	1929	2274
4-1 ljutito	+	2257	2601
5-1 šapat	+	723	3531
6-1 <i>babble</i>	+	1998	2601
7-1 automobil	+	2084	2687
8-1 upravno	-	2584	3083
9-1 maramica	-	2735	3187
10-1 ruka	+	1378	2722
11-1 bok	+	2084	2463
12-1 rame	-	1723	2980

Karakteristični opseg najveću širinu ima pri šaputanju (2808 Hz), a najmanju kada se drži cigareta u ustima (344 Hz). Prosečna širina svih opsega je približno 820 Hz. Kod

muških govornika je LTASS u proseku oslabljen u opsegu 1921÷2764 Hz, dok je kod ženskih govornika pojačan na višim frekvencijama između 2069 Hz i 2855 Hz.

## VII. ZAKLJUČAK

I analiza karakterističnih frekvencijskih opsega i analiza ekstremuma 11 spektralnih razlika vodi zaključku da su promene posmatranih dugovremenih usrednjenih spektara govora veće od 5 dB uglavnom locirane u opsegu širokom oko 1000 Hz između frekvencija 1900 Hz i 2900 Hz. U proseku, spektri LTASS muških govornika su u tom opsegu oslabljeni, dok su kod ženskih pojačani.

Analiza efekata koje unose razni modaliteti upotrebe mobilnih telefona je značajna u govornim tehnologijama, posebno kod prepoznavanje govora i govornika.

Kao sledeći korak u budućem istraživanju može se sprovesti analiza uticaja položaja mobilnog telefona na pomeranje formanata pojedinih vokala. Takođe, rezultate analize treba proveriti za veći broj govornika i za još situacija kod upotrebe mobilnih telefona.

## ZAHVALNICA

Zahvaljujem se prof. Slobodanu Jovičiću na mentorstvu u ovom istraživanju koje je organizovano u cilju izrade mog diplomskog rada a koje predstavlja deo istraživačkog projekta 148028G koji finansira Ministarstvo za nauku i tehnološki razvoj Republike Srbije i koji se odvija pod rukovodstvom prof. Jovičića. Zahvaljujem se takođe saradnicima Instituta za eksperimentalnu fonetiku i patologiju govora u Beogradu koji su mi omogućili da u njihovim prostorijama izvršim snimanje govorne baze.

## LITERATURA

- [1] Slobodan T. Jovičić, *Govorna komunikacija: fiziologija, psihoakustika i percepcija*, Nauka, Beograd, 1999.
- [2] Milan Vojnović, „Zavisnost dugovremenog usrednjenog spektra govora od dužine govornog signala“, 13. *Telekomunikacioni forum TELFOR*, Beograd, 2005.
- [3] Milan Vojnović, „Identifikacija govornika“, *Bilten Okružnog suda u Beogradu*, br. 69, pp. 27-48, 2006.
- [4] Catherine Byrne and Paul Foulkes, „The ‘Mobile Phone Effect’ on Vowel Formants“, *Speech, Language and the Law*, 11(1), pp. 83-102, 2004.

## ABSTRACT

The goal of this research is finding intra-speaker and inter-speaker differences among speakers who used a mobile phone in 12 ways. Changes of long-term average speech spectrum (LTASS) were analyzed. Three male and three female speakers took part in the research. On the average, LTASSs of males are attenuated between approximately 1900 and 2900 Hz by at least 5 dB, while those of females are amplified by at least 5 dB in that frequency range.

## THE INFLUENCE OF THE WAY OF USE OF THE MOBILE PHONE ON THE LONG-TERM AVERAGE SPEECH SPECTRUM

Nikola Jovanović