

Preciznost procene razumljivosti govora na bazi akustičkog impulsnog odziva

Violeta Stojanović, Zoran Milivojević, Zoran Veličković,
SP Komunikacione tehnologije
Visoka tehnička škola strukovnih studija
Niš, Srbija

violeta.stojanovic@vtsnis.edu.rs zoran.milivojevic@vtsnis.edu.rs zoran.velickovic@vtsnis.edu.rs

Sažetak—U prvom delu rada opisan je eksperiment u okviru koga je na osnovu objektivnog parametra *STI* akustičkog impulsnog odziva procenjena razumljivost govora u analiziranom akustičkom okruženju. Sproveden je MOS test i određena je ocena subjektivne razumljivosti govora. Rezultati su prikazani tabelarno i grafički. U drugom delu rada izvršena je komparativna analiza rezultata i određena greška procene razumljivosti govora na osnovu objektivnog parametra *STI*.

Ključne riječi—*Impulsni odziv prostorije, objektivni parametar STI, MOS test, subjektivna razumljivost govora.*

I. UVOD

Impulsni odziv prostorije *RIR* (**engl.** *Room Impulse Response*), predstavlja osnovni izvor informacija o akustičkim osobinama prostorije [1]. On se dobija kao odziv prostorije na specijalnu vrstu zvučne pobude (pucanj, aplauz, sweep signal i dr.) i u potpunosti karakteriše ponašanje i osobine prostorije kao linearnog i vremenski invarijantnog akustičkog sistema. Ako se analizira subjektivni stav slušaoca, kao merilo akustičkog kvaliteta, vremenska struktura impulsnog odziva se može razložiti na: direktan zvuk, rane refleksije i reverberaciju. Deo odziva, reverberacija se opisuje objektivnim parametrom koji se zove vreme reverberacije *RT*, (**engl.** *Reverberation Time*). Od vremena reverberacije zavisi razumljivost i kvalitet zvuka u prostoriji. Što je veće vreme reverberacije lošija je razumljivost govora u prostoriji.

Razumljivost govora se definiše tačnošću kojom slušalac prima sadržaj govora [2]. Na razumljivost govora utiču: a) vreme reverberacije, b) izobličenja u prenosnom sistemu, c) jačina govora, d) odnos SNR (**eng.** *Signal Noise Ratio*), e) direktan i reflektovani zvuk u prostoru,... Za ocenjivanje razumljivosti govora mogu se primeniti: a) metoda objektivne procene razumljivosti govora koja se bazira na fizičkim parametrima prenosnog sistema. Ocena razumljivosti govora se vrši pomoću objektivnih parametara AL_{cons} (**engl.** *Articulation Loss of Consonants*), *STI* (**engl.** *Speech Transmission Index*) i *RASTI* (**engl.** *Rapid Speech Transmission Index*) koji karakterišu subjektivni osećaj razumljivosti govora [1]; i b) metoda subjektivne procene razumljivosti govora koja se bazira na primeni testova razumljivosti sa logatomima tipa CVC, VCV, CV, VC, CCVC, CVCC, (C—konsonant, V—vokal)), sa slogovima ili sa test rečenicama. Rezultati testova razumljivosti zavise od: sadržine govora, preciznosti izgovora

glasova i dikcije govornika, koncentracije slušaoca na govor, godine starosti i td.

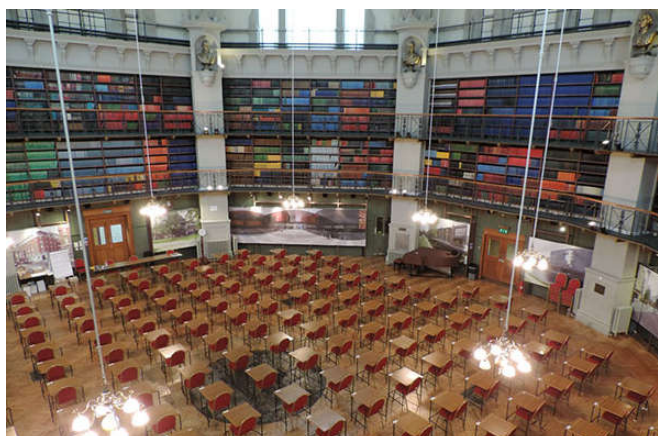
Prvo merenje vremena reverberacije je izvršio Sabine 1900. god., a 1922. je uspostavio vezu između vremena reverberacije, zapremine prostorije i apsorpcionih osobina materijala upotrebljenih za unutrašnju akustičku obradu prostorija [1]. French, Steinberg i Beranek 1947. god. su prvi istakli problem prenosa i razumljivosti govora [3]. Kryter je 1962. god. uveo indeks artikulacije, *AI*. Peutz je 1971. god. razvio algoritam za predviđanje razumljivosti govora u auditorijumima i predložio parametar koji se odnosi na gubitak artikulacije suglasnika, AL_{cons} [4]. Houtgast i Steeneken su 1980. god. potvrdili objektivnu metodu za merenje kvaliteta prenosa govora u prostoriji [3], i predložili akustički parametar, indeks prenosa govora *STI* [5]. Subjektivna ocenjivanja sistema za prenos govora pomoću testova razumljivosti otvorenog i zatvorenog tipa su vršili, najpre, Fletcher i Steinberg, 1929. god. [5], a zatim Egan 1944., Miller i Nicely, 1955., House 1965., Voiers 1977. i dr. Pregled testova ocenjivanja za razumljivost govora su dali Pols 1991. i Steeneken 1992.

U ovom radu izvršena je analiza preciznosti procene razumljivosti govora na osnovu objektivnog akustičkog parametra *STI*. Parametar *STI* izračunava se iz akustičkog impulsnog odziva. Primenom dijagrama iz Standarda IEC 60268 – 16 [6] vrši se prognoza razumljivosti govora, izražena procentualno. U cilju verifikacija preciznosti procene, na osnovu objektivnog parametra *STI*, sproveden je eksperiment gde je testna grupa ocenjivala razumljivost govornog signala koji je dobijen konvolucijom čistog govornog signala iz SMST (**engl.** *Serbian Matrix Sentence Test*) baze i snimljenog impulsnog odziva (subjektivna ocena). Impulsni odziv odnosi se na prostoriju „Octagon“ sa „Queen Mary“ Univerziteta u Londonu, a snimljen je od strane „Centre for Digital Music“ [7] i dostupan je na: <http://c4dm.eecs.qmul.ac.uk/rdr/handle/123456789/10>. Eksperiment je sproveden na Visokoj tehničkoj školi strukovnih studija u Nišu, a testnu grupu je činila studentska populacija.

Organizacija rada je sledeća: U Sekciji II je opisan eksperiment i prikazani su rezultati tabelarno i grafički. U Sekciji III prikazana je analiza rezultata. Sekcija IV je zaključak.

II. EKSPERIMENT

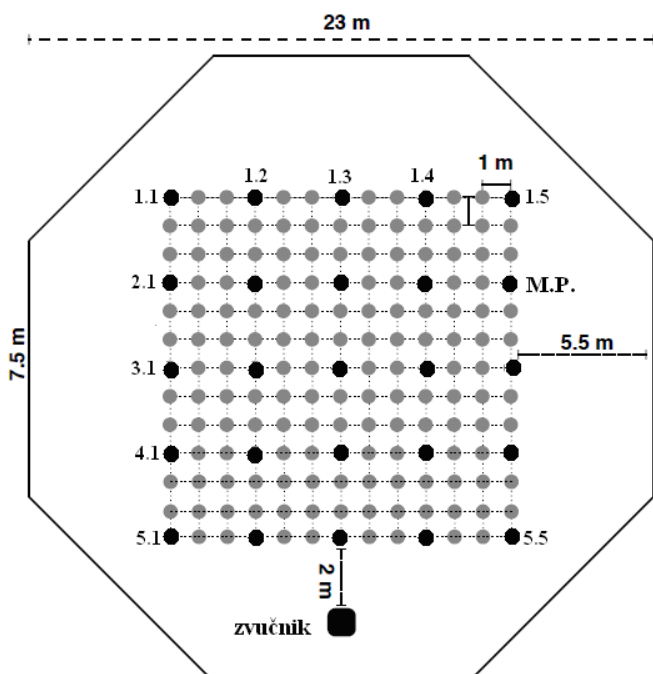
Sproveden je eksperiment sa ciljem određivanja razumljivosti govora u „Octagon at the Mile End campus of Queen Mary“, (Sl. 1.), na Univerzitetu u Londonu.



Slika 1. Prikaz „Octagon“ prostorije.

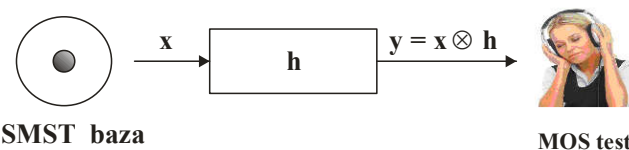
Na osnovu impulsnih odziva određena je vrednost objektivnih akustičkih parametara RT_{30} i STI . Subjektivna razumljivost govora određena je pomoću MOS testa. Prognoza razumljivosti govora u prostoriji se bazira na upotrebi Standarda IEC 60268 – 16 i klasifikaciji odnosa procentualne zavisnosti subjektivne razumljivosti govora i objektivnog akustičkog parametra STI [6, 8].

Za akustičku analizu objektivnih parametara RT_{30} i STI uzeto je 25 mernih tačaka MP (**engl. Measurement Point**) (od 169) čiji su položaji prikazani na Sl. 2. Analiza je izvršena pomoću softverskih paketa EASERA i Matlab.



Slika 2. Lokacije mernih tačaka u prostoriji.

Subjektivna razumljivost govora je izvršena za mernu tačku MP (3, 4). Ova merna tačka je izabrana zato što je njena vrednost parametra STI medijana svih vrednosti u skupu mernih tačaka. MP(3, 4) je udaljena od zvučnog izvora 8.54 m. Nakon toga izvršena je subjektivna ocena razumljivosti govora eksperimentom čiji je model prikazan na Sl. 3.



Slika 3. Model eksperimenta.

U eksperimentu se programski, prema slučajnom zakonu, formira jedna od mogućih 100000 različitih rečenica na osnovu izabranih reči iz baze SMST. Tako formirana rečenica, odnosno govorni signal x (na Sl.3.) se propušta kroz akustički model analizirane prostorije predstavljene impulsnim odzivom h . Rezultujući govorni signal $y = x \otimes h$, predstavlja ekvivalent govornom signalu snimljenom u samoj prostoriji.

U procesu MOS testiranja slušalac se nalazi u akustički izolovanom prostoru i govorni signal sluša preko slušalice. Nakon emitovanja rečenice, koja je semantički neizvesna kako za ispitivača tako i za slušaoca, slušalac glasno ponavlja rečenicu onako kako je razumeo. Ispitivač beleži rezultat kao tačno ponovljena rečenica ili kao tačno ponovljene određene reči u rečenici. Svaki slušalac iz testne grupe, koja se sastoji od 20 slušalaca, sluša po 10 rečenica (svaka rečenica ima 5 reči). Na osnovu statističke analize rezultata svih slušalaca određena je subjektivna razumljivost govora (u procentima) za a) celu test grupu b) za slušaoca muškog i c) slušaoca ženskog pola.

U Tabeli I prikazana je korelacija između vrednosti parametra STI i kvaliteta razumljivosti govora, razumljivosti rečenica i reči koja je definisana Standardom IEC 60268-16 [6].

TABELA I. KORELACIJA IZMEĐU VREDNOSTI PARAMETRA STI , KVALITETA RAZUMLJIVOSTI GOVORA I RAZUMLJIVOSTI REČENICA I REČI (STANDARD IEC 60268-16).

STI	Kvalitet razumljivosti govora	Razumljivost rečenica (%)	Razumljivost reči (%)
0 ÷ 3	Loša	0 ÷ 89	0 ÷ 67
0.3 ÷ 0.45	Slaba	89 ÷ 92	67 ÷ 78
0.45 ÷ 0.6	Prihvatljiva	92 ÷ 95	78 ÷ 87
0.6 ÷ 0.75	Dobra	95 ÷ 96	87 ÷ 94
0.75 ÷ 1	Odlična	96 ÷ 100	94 ÷ 96

Procenjena razumljivost govora, dobijena analizom akustičkih impulsnih odziva, upoređuje se sa subjektivnom razumljivošću, dobijenom MOS testom, i donosi se zaključak o grešci procene. Greška procene razumljivosti govora svih slušalaca dobijena na osnovu razumljivosti: a) celih rečenica Δ_{sent} i b) pojedinih reči (fonetski balansirane) Δ_{PBW} , definisana je kao:

$$\Delta_{sent} = \left| SI_{P_{sent}} - SI_{MOS_{sent}} \right| \quad (1)$$

$$\Delta PBW = \left| SI_{PPBw} - SI_{MOSPBw} \right| \quad (2)$$

gde su: $SI_{P_{sent}}$ i SI_{PPBw} predviđene razumljivosti govora dobijene na osnovu razumljivosti celih rečenica i na osnovu razumljivosti pojedinih reči, respektivno. Ove vrednosti se određuju pomoću objektivnog parametra STI (dijagram E.1 u Standardu 60268 – 16 IEC: 2003 (E)) [6]. $SI_{MOS_{sent}}$ je subjektivna razumljivost govora određena MOS testom razumljivosti celih rečenica, a $SI_{MOS_{PBw}}$ je subjektivna razumljivost govora određena MOS testom razumljivosti pojedinih reči. Greška procene razumljivosti govora posebno muških i posebno ženskih slušalaca, izračunava se sa dijagrama iz [8] i prema jednačini (2). SI_{PPBw} predstavlja predviđenu srednju vrednost razumljivosti govora određenu razumljivošću pojedinih reči i definiše se sa:

$$SI_{PPBw} = \frac{1}{4} \sum_{i=1}^4 SI_{P_{fon}}(i) \quad (3)$$

gde je $SI_{P_{fon}} = \{ SI_{P_{fric}}, SI_{P_{plos}}, SI_{P_{cons}}, SI_{P_{vow}} \}$ predviđena razumljivost govora fonema i predstavlja sumu predviđenih razumljivosti govora frikativ $SI_{P_{fric}}$, ploziva $SI_{P_{plos}}$, suglasnika $SI_{P_{cons}}$ i samoglasnika $SI_{P_{vow}}$.

A. Baza

Baza eksperimenta se sastoji od: a) baze govornih signala i b) baze impulsnih odziva.

a) Baza govornih signala je SMST [9]. SMST je sastavljena od izolovanih reči izgovorenih na srpskom jeziku grupisanih u vlastite imenice, glagole, brojeve, prideve i zajedničke imenice. Programski se, prema slučajnom zakonu, formira jedna od mogućih 100000 kombinacija rečenica u formatu: vlastita imenica, glagol, broj, pridev i zajednička imenica. SMST baza je fonetski balansirana. Frekvencija semplovanja je $f_s = 44.1$ kHz sa 2 BpS.

b) Bazu impulsnih odziva čine audio zapisi impulsnih odziva prostorije „Octagon“, snimljenih od strane „Centre for Digital Music“ sa „Queen Mary“ Univerziteta u Londonu [7]. „Octagon“ prostorija se koristi kao biblioteka, konferencijska sala, izložbeni prostor, za održavanje raznih svečanosti i td. „Octagon“ ima 8 zidova visine 7.5 m koji su prekriveni knjigama, drveni pod i plafon obložen gipsom. Kupola plafona je visine 21 m. Zapremina je 9500 m³.

Merenja impulsnih odziva vršena su pomoću pobudnog log-sweep signala u trajanju od 2 s [7] sa $f_s = 96$ kHz sa 3 BpS. Pri ovom eksperimentu korišćen je zvučnik „Genelec“ 8250A i omnidirekcionni mikrofoni DPA 4006.

B. Testna grupa

Testna grupa sastavljena je od studenata Visoke tehničke škole strukovnih studija iz Niša. Struktura testne grupe je 10 muških i 10 ženskih slušalaca, godina starosti od 19 – 25 (srednja vrednost godina starosti i standardna devijacija su: $\mu = 21.9$ god. i $\sigma = 1.944$ god.).

C. Rezultati

U Tabeli II i Tabeli III prikazane su vrednosti objektivnih akustičkih parametara RT_{30} i STI u 25 mernih tačaka prostorije. U Tabelama IV – VI date su subjektivne razumljivosti govora u mernoj tački MP(3, 4) dobijene MOS testom za sve slušaoce, za slušaoce muškog pola i slušaoce ženskog pola, respektivno. Na Sl. 4 – 6 prikazane su greške procene vrednosti razumljivosti govora u mernoj tački za sve slušaoce, za slušaoce muškog pola i slušaoce ženskog pola, respektivno.

TABELA II. VREME REVERBERACIJE RT_{30} U MERNIM TAČKAMA.

M.P.	RT_{30} (s)				
	1	2	3	4	5
1	1.98	3.39	3.31	3.3	2.12
2	1.83	3.58	2.7	3.37	3.03
3	4.67	2.21	1.76	2.29	2.32
4	2.73	2.69	2.03	3.84	3.35
5	2.01	3.24	2.33	3.07	2.57

TABELA III. INDEKS RAZUMLJIVOSTI GOVORA STI U MERNIM TAČKAMA

M.P.	STI				
	1	2	3	4	5
1	0.511	0.499	0.514	0.488	0.494
2	0.492	0.486	0.491	0.485	0.479
3	0.487	0.527	0.539	0.516	0.485
4	0.490	0.589	0.624	0.564	0.487
5	0.471	0.600	0.831	0.585	0.458

TABELA IV. SUBJEKTIVNA RAZUMLJIVOST GOVORA ZA SVE SLUŠAOCE.

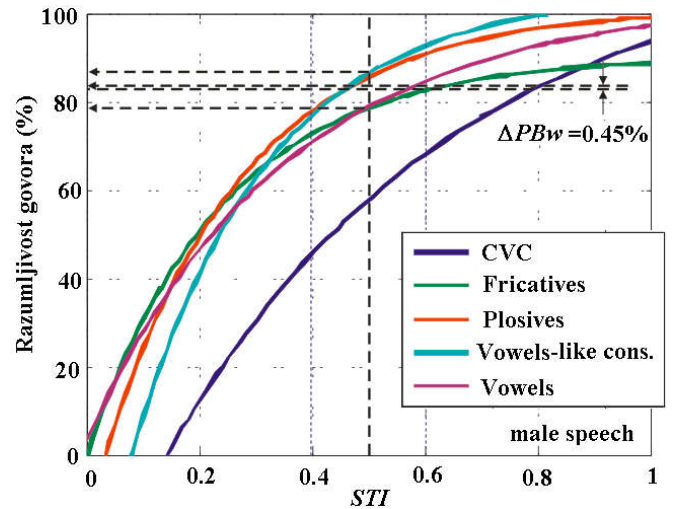
Broj slušalaca	Broj tačnih rečenica	Broj tačnih reči				
		Vlastita imenica	Glagol	Broj	Pridev	Zajednička imenica
20	110	181	180	166	170	147
		844				
		Subjektivna razumljivost govora				
	55 %	84.4 %				

TABELA V. SUBJEKTIVNA RAZUMLJIVOST GOVORA ZA SLUŠAOCE MUŠKOG POLA.

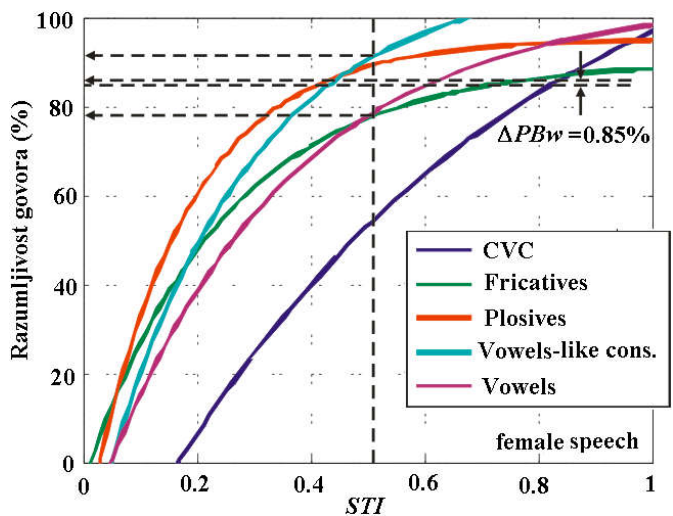
Broj slušalaca	Broj tačnih rečenica	Broj tačnih reči					
		Vlastita imenica	Glagol	Broj	Pridev	Zajednička imenica	
10	58	91	89	80	78	78	
		416					
	Subjektivna razumljivost govora		83.2 %				
		58 %	83.2 %				

TABELA VI. SUBJEKTIVNA RAZUMLJIVOST GOVORA ZA SLUŠAOCE ŽENSKOG POLA.

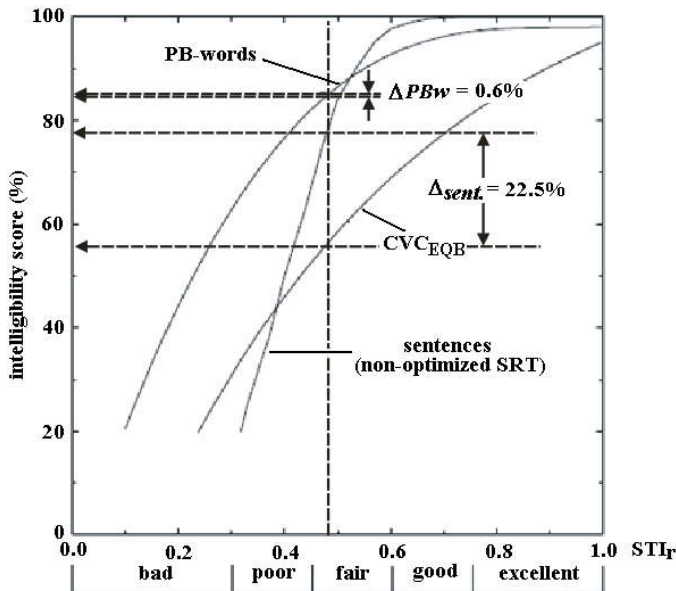
Broj slušalaca	Broj tačnih rečenica	Broj tačnih reči					
		Vlastita imenica	Glagol	Broj	Pridev	Zajednička imenica	
10	52	90	91	86	92	69	
		428					
	Subjektivna razumljivost govora		85.6 %				
		52 %	85.6 %				



Slika 5. Greška procene razumljivosti govora na osnovu razumljivosti PB reči za slušaoce muškog pola.



Slika 6. Greška procene razumljivosti govora na osnovu razumljivosti PB reči za slušaoce ženskog pola.



Slika 4. Greška procene razumljivosti govora na osnovu razumljivosti rečenica i PB reči za sve slušaoce.

III. ANALIZA REZULTATA

Na osnovu rezultata prikazanih u Tabelama II – V i na Sl. 4 – 6 zaključuje se sledeće:

a) statistička analiza akustičkog parametra RT_{30} je pokazala da vrednost ovog parametra za analiziranu oktagonu prostoriju pripada opsegu intervala $2.788 \text{ s} \pm 0.707 \text{ s}$, odnosno opsegu $2.081 \text{ s} \div 3.495 \text{ s}$. Optimalna vrednost RT_{30} za prostorije zapremine 9500 m^3 je 2 s [10]. U izabranoj mernoj tački MP (3, 4), određena je vrednost ovog parametra 2.19 s .

b) statistička analiza akustičkog parametra STI je pokazala da vrednost ovog parametra za analiziranu prostoriju pripada opsegu intervala 0.528 ± 0.075 , odnosno opsegu $0.453 \div 0.603$. U izabranoj mernoj tački vrednost ovog parametra je $STI = 0.516$ i predstavlja medijanu svih vrednosti u skupu mernih tačaka. Ove vrednosti parametara STI pripadaju opsegu prihvatljive razumljivosti govora.

c) dobijena vrednost objektivnog parametra za sve slušaoce u posmatranoj mernoj tački je $STI = 0.48$. Procenjene vrednosti subjektivne razumljivosti govora na osnovu razumljivosti rečenica i PB reči (Sl. 4.) su 77.5% i 85% , respektivno. Rezultati dobijeni MOS testom su pokazali da je subjektivna razumljivost govora, dobijena na osnovu: a) razumljivosti rečenica, 55% i b) razumljivosti reči, 84.4% . Greške procene razumljivosti govora su $\Delta_{sent} = 22.5 \%$ i $\Delta PBw = 0.6 \%$.

d) vrednost objektivnog parametra za muške slušaocce u posmatranoj mernoj tački je $STI_{male} = 0.502$. Procenjena vrednost subjektivne razumljivosti govora na osnovu razumljivosti fonema: frikativa, ploziva, suglasnika i samoglasnika, (Sl. 5.), se nalazi u opsegu 78.5 % ÷ 87 %. Njena srednja vrednost je 82.75 %. MOS test je pokazao da je subjektivna razumljivost govora, za muške slušaocce, dobijena na osnovu: a) razumljivosti rečenica, 58 % i b) razumljivosti reči, 83.2 %. Greška procene razumljivosti govora izračunata na osnovu razumljivosti fonema iznosi $\Delta PBW_{male} = 0.45$ %.

e) objektivni parametar za ženske slušaocce u posmatranoj tački prostorije ima vrednost $STI_{female} = 0.513$. Procenjena vrednost subjektivne razumljivosti govora na osnovu razumljivosti fonema (Sl. 6) pripada opsegu 78 % ÷ 91.5 %. Njena srednja vrednost je 84.75 %. Odgovarajuća vrednost subjektivne razumljivosti, dobijena MOS testom, je 85.6 %. Greška procene razumljivosti govora, za ženske slušaocce, dobijena na osnovu razumljivosti fonema iznosi $\Delta PBW_{female} = 0.85$ %. Subjektivna razumljivost govora, izračunata na osnovu MOS testa sa rečenicama je 52 %.

Skup vrednosti objektivnog parametra $STI = \{0.48, 0.502, 0.513\}$ pripada opsegu koji procenjuje prihvatljivu razumljivost govora, razumljivost rečenica: 92 % ÷ 95 % i razumljivost reči: 78 % ÷ 87 % (Tabela I) [6]. Rezultati MOS testa su pokazali da se procenat subjektivne razumljivosti fonema, kod svih slušalaca zajedno (84.4 %), kao i kod muških i ženskih slušalaca odvojeno (83.2 % i 85.6 %), nalazi u procenjenom opsegu. Korišćene rečenice u MOS testu su za sve slušaocce zajedno testirane, kao i za slušaocce muškog i ženskog pola odvojeno testirane, po navedenom kriterijumu, nerazumljive. Njihova subjektivna razumljivost je 55 %, 58 % i 52 %.

IV. ZAKLJUČAK

Na osnovu objektivnog akustičkog parametra, indeksa razumljivosti govora STI , određenog iz impusnog odziva oktogonalne prostorije, procenjene su vrednosti subjektivne razumljivosti govora u njoj. Komparacijom ovih vrednosti i vrednosti dobijenih MOS testom razumljivosti rečenica sastavljenih od izolovanih srpskih reči iz baze SMST, uočena je greška procene subjektivne razumljivosti govora. Greška procene iznosi: a) $\Delta_{sent} = 22.5$ % i $\Delta PBW = 0.6$ % kod svih slušalaca prilikom testiranja razumljivosti rečenica i reči, b)

$\Delta PBW_{male} = 0.45$ % kod slušalaca muškog pola pri testiranju razumljivosti reči i c) $\Delta PBW_{female} = 0.85$ % kod slušalaca ženskog pola pri testiranju razumljivosti reči.

LITERATURA

- [1] H. Kuttruff, "Room Acoustics", E&FN Spon, London, 2000.
- [2] Lj. Subotić, D. Sredojević, I. Bjelaković, "Fonetika i fonologija: ortoepska i ortografska norma standardnog srpskog jezika", Filozofski fakultet, Novi Sad, 2012.
- [3] H. J. M. Steeneken, "On measuring and predicting speech intelligibility", Academisch Proefschrift, Soesterberg, 1992.
- [4] Peutz, V. M. A. "Articulation Loss of Consonants as a criterion for Speech Transmission in a Room", *J. AUDIO – ENG. SOC.* Vol 19, p. 915 - 919, 1971.
- [5] Steeneken, H. J. M., "Apsychical Method for measuring Speech Transmission Quality", *J. ACOUST. SOC. AM.*, Vol 19, 1980.
- [6] International Electrotechnical Commission IEC 60268-16 – International Standard: "Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index", Switzerland:IEC, 2011.
- [7] R. Stewart, M. Sandler, Database of omnidirectional and B _ format room impulse response, ICASSP – 88, 2010, pp. 165 – 168..
- [8] H. J. M. Steeneken, T. Houtgast, Basics of the STI - measuring Method, In Past Present and Future of the Speech Transmission Index, edited by Sander j. Wijngaarden, 13 – 34. The Netherlands: TNO Human Factors, 2002.
- [9] Z. Milivojević, D. Kostić, Z. Veličković, D. Brodić, "Serbian sentence matrix test for speech intelligibility measurement in different reverberant conditions, UNITECH 2016, Gabrovo.
- [10] ISO 3382: 2000: Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters.

ABSTRACT

In the first part of the paper is described an experiment in which, on the basis of objective parameter STI acoustic impulse response estimated speech intelligibility in the analyzed acoustic environment. MOS test was conducted and determined by the subjective assessment of speech intelligibility. The results are presented in tables and graphs. In the second part of the paper is a comparative analysis of the results and determining the error estimates intelligibility of speech on the basis of objective parameter STI .

PRECISION OF ESTIMATE OF SPEECH INTELLIGIBILITY BASED ON THE ACOUSTIC IMPULSE RESPONSE

Violeta Stojanović, Zoran Milivojević, Zoran Veličković