

Objektivno i subjektivno poređenje različitih modela za izračunavanje glasnosti

Ivana Čpajak i Sanja Rakić

Sadržaj — U radu je prikazana uporedna analiza više širokopoljnih modela za procenu glasnosti vremenski promenljivih signala. Realizovane su procedure za izračunavanje glasnosti na 4 principa. Analizirano je 50 audio zapisa iz različitih muzičkih žanrova, kao i niz zvučnih zapisa koji nisu ni govorni ni muzički signali. Analiza koja je sprovedena u radu imala je za cilj da pokaže razlike u vrednostima procenjene subjektivne glasnosti koje se dobijaju različitim modelima u zavisnosti od dinamičkih i programskih razlika analiziranih signala. Na izabranom uzorku muzičkih signala i zvučnih efekata, izvršeno je usaglašavanje nivoa glasnosti po svim analiziranim algoritmima. Na tako usaglašanim signalima je zatim izvršeno subjektivno testiranje.

Ključne reči — Nivo glasnosti, objektivno i subjektivno poređenje, različiti programski materijali.

I. UVOD

IZRAČUNAVANJE glasnosti predstavlja postupak koji uz pomoć odgovarajućeg matematičkog modela treba da na adekvatan način poveže zvučni pritisak, kao merljivu fizičku veličinu koja opisuje zvučno polje, i subjektivni doživljaj glasnosti te pobude. Subjektivna glasnost zavisi od objektivnih karakteristika zvučnih signala, njihovih dinamičkih karakteristika, frekvencijskog sadržaja i vremenske obvojnice koji na složen način utiču na subjektivni doživljaj. Međutim i sami slušaoci različito mere najvažnije faktore koji utiču na glasnost, a to su: nivo zvučnog pritiska, frekvencijski sadržaj i trajanje. Stoga, definisanje glasnosti zvuka pokazuje izvesnu promenljivost između slušalaca, BLV (Between Listener Variability), čak i u okviru homogenih grupa [1], dok razlike u godinama, polu, kulturi itd. mogu da dodaju dalje varijacije. Pored toga, individualne procene glasnosti kod iste osobe su dosledne samo donekle, i zavise od doba dana, raspoloženja, pažnje i sl. Ovaj vid varijacija se zove promenljivost samog slušaoca, WLV (Within Listener Variability). Za procenu glasnosti razvijani su algoritmi, od kojih je najpoznatiji Cvikerov metod, a koji su pokazali dobre rezultate u proceni glasnosti kod stacionarnih signala. Za procenu glasnosti vremenski promenljivih zvučnih sadržaja, muzičkih i govornih signala, pokazalo se da složen Cvikerov metod nije efikasan [2]. Potreba za jednostavnom procenom glasnosti pojavila se kao problem u radiodifuziji i vršeni su pokušaji da se jednostavnim

širokopoljnim algoritmima izvrši estimacija subjektivne glasnosti.

Objektivne karakteristike audio signala prate se na osnovu merenja vršnih nivoa signala pomoću tzv. PPM (Peak Programme Meter), koji je pre svega namenjen da spreči da nivoi audia odu suviše visoko, jer onda dolazi do distorzije (npr. u D/A konvertorima). Glasnost je, međutim, više korelisana sa prosečnom energijom nego sa vršnim nivoima, tako da ona može značajno da varira između programa sa istim vršnim nivoima. Stoga su audio inženjeri paralelno sa PPM-om koristili i svoj sluh da bi se perceptivno poravnali nivoi između različitih programskih materijala. Međutim, u današnjim radio i TV stanicama operateri su zamenjeni automatizovanom opremom.

U emitovanju radio i TV programa pojavljuje se problem usaglašavanja glasnosti različitih programskih materijala. Problemi koji se s tim u vezi sreću u radiodifuziji evoluirali su tokom vremena. Neusaglašenost dinamičkih karakteristika reklamnih i programskih materijala u prvo vreme rešavana je tako što su reklamni materijali stišavani pri emitovanju. Međutim, 1999. godine reklamne agencije su konsultovale inženjersku agenciju da izvede merenje nivoa glasnosti reklamnih blokova. Ispostavilo se da postoje razlike između onoga što agencije dostave stanicama i onoga što one emituju [3]. Reklamne agencije su onda zapretile da će povući svoje reklame ukoliko im stanice opet promene nivo. Radio i TV stanicama je dozvoljeno da jedino provere saglasnost sa ITU/EBU preporukom, koja se odnosi samo na maksimalne PPM nivoe [4]. Kao rezultat nove politike, počele su žalbe gledalaca kako su reklame previše glasne. Jedini izlaz je bio podizanje nivoa običnog programskog materijala pomoću generalne dinamičke kompresije. Problem je ovim bio rešen samo delimično jer su programi kao što su na primer dokumentarni, sa dijalogima snimljenim na ulici, izgubili razumljivost usled primenjene kompresije. Nova podešavanja su dovela do žalbi reditelja dokumentarnih i dramskih filmova, jer su imala negativan uticaj na ravnotežu njihovih umetničkih izbora i izazvala gubitak dramskih efekata ugrožavanjem namernih razlika u nivou. Distribucione kompanije imaju i drugi problem; one moraju da kombinuju nivoe nekoliko stanica tako da gledaoci mogu da menjaju kanale bez strahovitih promena nivoa zvuka. Isti problem koji postoji kod emitovanja reklama se javlja i pri emitovanju savremene muzike. Ova muzika je ozloglašena po (zlo)upotrebi kompresije i klipovanju.

U ovom radu izvršena je analiza glasnosti većeg broja muzičkih signala i različitih zvučnih efekata na bazi 4

Ivana Čpajak, student Elektrotehničkog fakulteta u Beogradu, Srbija (telefon: 381-64-0217200, e-mail: ivanacpajak@yahoo.com)

Sanja Rakić, student Elektrotehničkog fakulteta u Beogradu, Srbija (telefon: 381-64-1998759, e-mail: sanja.rakic@yahoo.com)

algoritma za širokopojasnu procenu glasnosti; pri ovom su praćene i dinamićke karakteristike signala. Posmatrane su razlike pri proceni subjektivne glasnosti signala razlićitog sadržaja i razlićitih dinamićkih karakteristika, koje su dobijene pomoću posmatranih modela. Na bazi tako dobijenih rezultata formirane su dve kolekcije zvućnih signala, jedna sastavljena od 11 segmenata razlićitih muzićkih materijala i druga sastavljena od 15 zvućnih efekata. Pojedinaćni zvućni signali su normalizovani tako da se izjednaće po glasnosti nekom od predloženih metoda, i tako formirani zvućni signali su objektivno i subjektivno testirani.

II. ŠIROKOPOJASNI MODELI ZA PROCENU GLASNOSTI

Za procenu glasnosti realizovana su 4 algoritma za izraćunavanje širokopojasnih parametara koji su u odrećenoj korelaciji sa subjektivnom glasnošću, i to:

- RMS (Root Mean Square) ili srednja kvadratna vrednost signala,
- L_5 [dB] vrednost nivoa signala koja je premašena u 5% vremena,
- Model I prema ITU BS.1770 preporuci [5],
- Model II napravljen po LARM algoritmu [2].

U prva dva modela se ne vrši nikakva frekvencijska korekcija signala. Pri procesiranju signala u Modelu I signal prvo prolazi kroz pre-filtar koji uzima u obzir uticaj glave slušaoca na percepciju signala (glava je modelovana kao rigidna sfera), a zatim se primenjuje filtar koji simulira RLB (Revised Low-frequency B-weighting) težinsku krivu [5]. Nakon ova dva filtriranja koja se zajedno mogu oznaćiti kao K-skalaranje ili R2LB, vrši se izraćunavanje *srednje snage* signala prema formuli:

$$z = \sqrt{\frac{1}{T} \sum_0^T y^2}$$

gde y predstavlja ulazni signal nakon filtriranja, a T vremenski segment nad kojim se vrši raćunanje. Jedinica nivoa glasnosti je LUFs (Loudness Unit relative to Full Scale) ili LKFs (Loudness K-weighted relative to Full Scale) [5].

U Modelu II, signal nakon prolaska kroz RLB filtar i blok za izraćunavanje apsolutne vrednosti signala prolazi kroz asimetrićni filtar propusnik niskih frekvencija koji ima brz linearan uspon i sporo eksponencijalno opadanje (u vremenskom domenu) [2]. Nakon toga *srednja snaga* se raćuna na sledeći naćin:

$$z = \sqrt[p]{\frac{1}{T} \sum_0^T |y|^p}$$

gde y predstavlja signal nakon filtriranja, a p je parametar modela. Za parametar p je uzeta optimalna vrednost, $p=1.5$.

Zavisno od trajanja intervala T nad kojim raćunamo srednju snagu signala, pomoću oba modela možeme dobiti kratkovremenu glasnost (ovde je korišćeno $T = 2$ s) i dugovremenu glasnost (T jednako trajanju analiziranog segmenta).

Kao izlazne vrednosti iz algoritama koji su oznaćeni

kao Model I i Model II dobijaju se deskriptori glasnosti [6] CoG (Center of Gravity) I i CoG II koji odgovaraju dugovremenoj glasnosti, koja se meri jedinicama LU (Loudness Units), kao što je preporučeno za meraće glasnosti u BS.1771 [7]. Parametar CoG ima negativne vrednosti i što je materijal subjektivno glasniji ova vrednost se više približava nuli.

Zajedno sa procenom glasnosti praćena su i dinamićka obeležja signala i to pomoću:

- dinamićkog opsega materijala koji je predstavljen kao razlika između vrednosti koja je premašena u 1% i one koja je premašena u 99% vremena ($L_1 - L_{99}$), i
- parametra Cons (Consistency) koji predstavlja konzistentnost materijala, a raćuna se prema formuli:

$$Cons = 10 \log_{10} \sqrt{\frac{1}{M} \sum_i^M (L_i - L)^2}$$

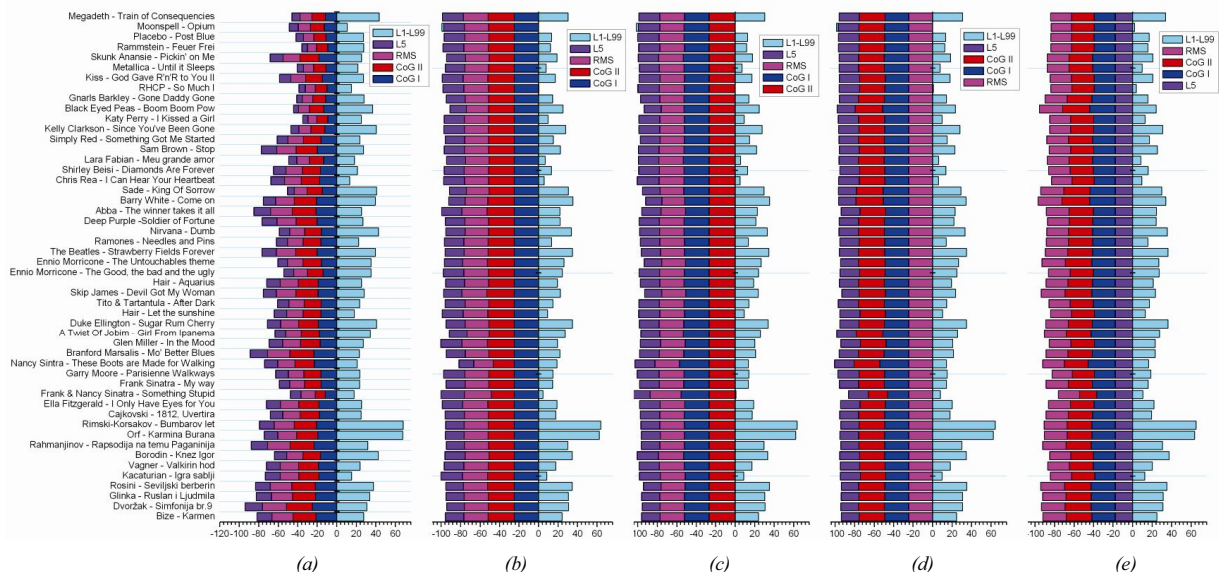
gde L_i predstavlja kratkovremenu glasnost dobijenu za i -ti prozor usrednjavanja (od ukupno M), a L predstavlja dugovremenu glasnost na celom segmentu. Cons se razlikuje za Model I i Model II. Parametar Cons uzima negativne vrednosti i što je materijal konzistentniji (manje varijacije glasnosti od vremenskog prozora do vremenskog prozora), to se mereni Cons više približava 0 LU, što predstavlja maksimalnu konzistentciju.

III. ANALIZA PO MUZIĆKIM ŽANROVIMA

Napravljena je kolekcija od 50 mono audio zapisa koji su svrstani u 6 razlićitih žanrova. Zapisi su trajanja od 1.5 do 3 minuta. Pre obrade izvršena je normalizacija svakog muzićkog zapisa u odnosu na njegovu vršnu vrednost. Za svaku numeru izraćunate su vrednosti 4 parametara (RMS, L_5 , CoG I i CoG II) za procenu glasnosti i 3 parametara ($L_1 - L_{99}$, Cons I i Cons II) dinamićkih karakteristika. Vrednosti parametara unutra svakog posmatranog žanra su usrednjene i prosećne vrednosti su prikazane u tabeli 1. U tabeli je data i zastupljenost numera u okviru pojedinih žanrova.

TABELA 1: PROSEĆNE VREDNOSTI PARAMETARA.

Vrsta muzike	Parametar					
	Klasićna	Džez	Filmska	Pop/rok	Savremena pop/rok	Metal
Broj segmenata	11	9	6	12	5	7
RMS [dBFS]	-20,4	-17,5	-16,8	-17,3	-10,7	-12,6
L_5 [dB]	-15,0	-13,2	-11,9	-12,6	-6,4	-8,9
CoG I [LUFs]	-20,6	-18,1	-16,8	-17,9	-11,4	-12,9
CoG II [LUFs]	-22,3	-18,7	-18,9	-19,4	-12,9	-14,3
$L_1 - L_{99}$ [dB]	36,6	26,2	27,4	28,3	28,9	26,1
Cons I [LU]	-9,3	-8,3	-6,6	-8,3	-3,5	-5,0
Cons II [LU]	-10,5	-9,9	-8,3	-9,6	-4,7	-5,8



Sl. 1. Rezultati merjenja nad kolekcijom od 50 segmenata različitih muzičkih žanrova: (a) realno stanje; (b) normalizovani po CoG I; (c) normalizovani po CoG II; (d) normalizovani po RMS vrednosti; (e) normalizovani po vrednosti L_5 .

Iz table 1 se može videti da najveću prosečnu glasnost ima, kao što je i očekivano, savremena pop/rok muzika. Ovaj žanr pokazuje i najveću konzistentnost, što za posledicu ima prilično konstantnu glasnost tokom trajanja numere. Klasična muzika ima najmanju prosečnu glasnost, ali ispoljava i najmanju konzistentciju, što potvrđuje njen veliki dinamički opseg.

Razlike u vrednostima pokazivanja modela I i II pri izračunavanju CoG za razmatranu kolekciju su u proseku 1.5 LU, a najveća razlika od 2 LU se javlja u slučaju filmske muzike. Pri računanju parametra Cons prosečna razlika između pokazivanja ova dva modela je 1.3 LU, a u najgorem slučaju je 1.7 LU, opet kod numera filmske muzike.

Na sl. 1 su prikazani posmatrani parametri za procenu glasnosti, RMS, L_5 , CoG I i CoG II (sa leve strane ordinatne ose) i dinamičke karakteristike L_1 - L_{99} (sa desne strane) za svih 50 numera, pri čemu su vrednosti parametara kumulativno nalepljene jedna na drugu. Pri tom, prateći tabelu 1, prvih 11 segmenata su numere klasične muzike, narednih 9 su primeri džez muzike itd.

Slučaj pod (a) na sl. 1 predstavlja realno stanje, tj. situaciju kada su segmenti normalizovani samo u odnosu na svoju vršnu vrednost. Može se lako konstatovati velika raznolikost među numerama, što znači da kada bi ove numere bez dodatnog procesiranja bile puštene u program razlike u glasnosti bi bile lako uočljive. Ispitano je kako bi poravnanje po pojedinim parametrima uticalo na poravnanje nivoa glasnosti ispitivanih segmenata pri čemu su vrednosti parametara prikazane kao što je napomenuto.

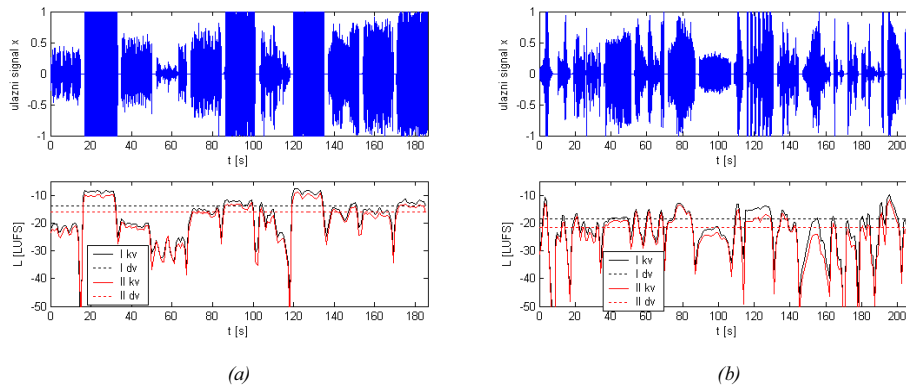
Ukoliko izvršimo virtuelnu normalizaciju svih segmenata po parametru CoG I, što je predstavljeno na sl. 1 (b), ili prema parametru CoG II, što je predstavljeno na sl. 1 (c), primećujemo da ova dva modela pokazuju približno slične varijacije. Na osnovu rezultata zaključujemo da ukoliko poravnamo glasnost segmenata

po Modelu I, tako normalizovani segmenti će dati približno poravnate nivoe glasnosti kada se analiziraju algoritmom primenjenim u Modelu II. Važi i obrnuto. Ovako poravnati segmenti imaju približno ujednačene RMS vrednosti sa par izuzetaka, dok vrednosti L_5 i dalje variraju. Ako sada „normalizujemo“ segmente prema RMS vrednosti (sl. 1(d)) situacija je nešto lošija nego u prethodnim slučajevima, naročito u delu sa džez muzikom. Vrednosti L_5 u ovoj situaciji deluju „upegljanije“. Na sl.1 (e) prikazano je šta se dešava ako se izvrši normalizacija prema vrednosti L_5 . Lako je konstatovati da je ovo najlošija strategija ako se za cilj ima ujednačen nivo glasnosti.

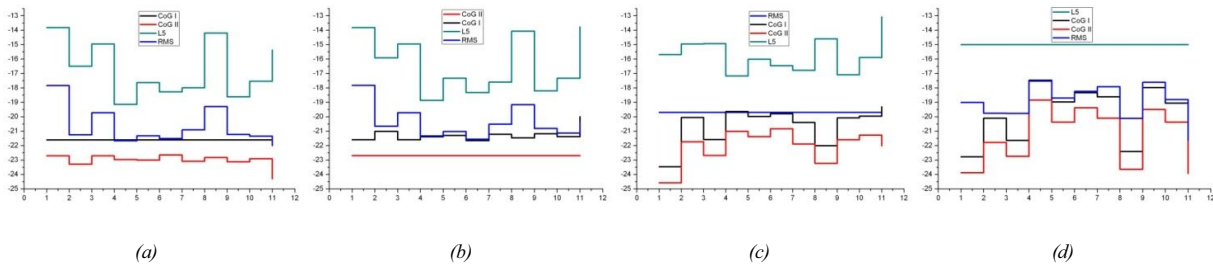
IV. NORMALIZACIJA I SUBJEKTIVNA ANALIZA

Posmatranjem rezultata prethodne analize iz opisane kolekcije je izabrano 11 karakterističnih muzičkih segmenata trajanja 15 s. Dodatno je uzeto 15 segmenata raznih zvučnih efekata (npr. aplauz, smeh, zvuk aviona, koraka...) trajanja od 5 do 17 s. Svaki od segmenata je normalizovan u odnosu na njegovu vršnu vrednost i zatim su napravljena 2 zvučna zapisa, jedan od muzičkih segmenata, a jedan od zvučnih efekata. Na sl. 2 prikazani su vremenski oblici ova 2 seta zvučnih zapisa kao i vrednosti procenjene glasnosti (kratkovremene i dugovremene) po Modelu I i Modelu II.

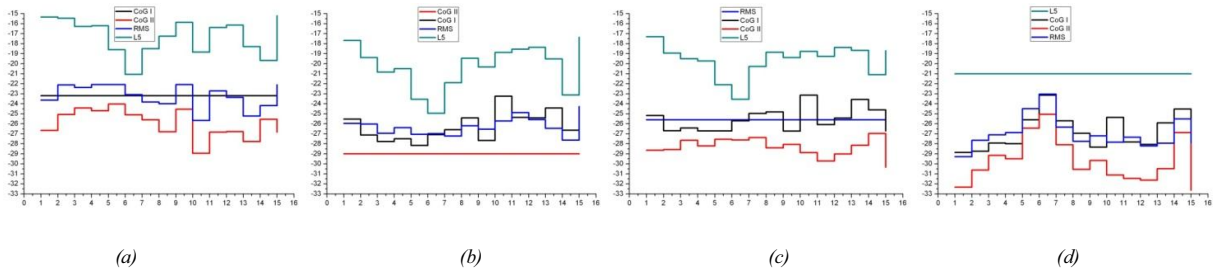
Za svaki od segmenata muzičke prirode su izračunata analizirana 4 parametra (RMS, L_5 , CoG I i CoG II). Potom je izvršeno usaglašavanje glasnosti svih segmenata po jednom od ovih parametara i od tako dobijenih segmenata napravljen je novi zvučni zapis. Ovo je ponovljeno za sva 4 parametra tako da su dobijena 4 nova signala. Identičan postupak je sproveden i nad segmentima koji predstavljaju zvučne efekte. Za svaki novoformirani signal čija je glasnost poravnata po jednom od parametara izračunate su varijacije ostala 3 parametra i one su prikazane na slikama 3 i 4.



Sl. 2. Vremenski oblik signala i pokazivanje Modela I i Modela II za: (a) muzički zapis, (b) zapis sa zvučnim efektima



Sl. 3. Normalizovanje muzičkih segmenata po: (a) CoG I; (b) CoG II; (c) RMS; (d) L5



Sl. 4. Normalizovanje segmenata sa zvučnim efektima po: (a) CoG I; (b) CoG II; (c) RMS; (d) L5

Ovako formirani zvučni zapisi su preslušavani i preliminarni subjektivni testovi su pokazali da se između signala koji su normalizovani po parametrima CoG I i CoG II ne uočavaju značajne perceptivne razlike, što pokazuju male varijacije samih parametara koje se kreću oko 0.3 LU za muzičke i oko od 2.5 LU za zvučne efekte. Odstupanja pri normalizaciji na L_5 za preostala 3 parametra kod muzičkih signala su u proseku 2.4 LU, a kod zvučnih efekata 3.3 LU. Iako objektivne varijacije parametara pri normalizaciji na L_5 pokazuju značajno veća odstupanja, subjektivno se ne percipira primetna razlika u glasnosti zbog relativno velike dinamike signala. Za dalju analizu ostaje da se osmisle precizniji subjektivni testovi.

LITERATURA

- [1] E. Skovenborg, R. Quesnel, and S.H. Nielsen, "Loudness Assessment of Music and Speech", in *Proceedings of the AES 116th Convention*, Berlin, Preprint 6143, 2000.
- [2] E. Skovenborg and S.H. Nielsen "Evaluation of Different Loudness Models with Music and Speech", in *Proceedings of the AES 116th Convention*, Berlin, Preprint 6143, 2004.
- [3] E.M. Grimm, R. van Everdingen, and M. J. L. C. Schöpping "Toward a Recommendation for a European Standard of Peak and LKFS Loudness Levels", *SMPTE Motion Imaging Journal*, April 2010.

- [4] ITU-R Rec. BS.645-2 "Test Signals and Metering to be Used on International Sound-Programme Connections", 1992.
- [5] ITU-R Rec. BS.1770 "Algorithms to Measure Audio Programme Loudness and True-Peak Level", 2006.
- [6] E. Skovenborg and T. Lund "Loudness Descriptors to Characterize Programs and Music Tracks", in *Proceedings of the AES 125th Convention*, San Francisco, 2008.
- [7] ITU-R Rec. BS.1771 "Requirements for loudness and true-peak indicating meters", 2006.

ABSTRACT

Comparative analysis of different single-band models for loudness estimation of non-stationary signals is shown in this paper. Procedures for calculating loudness based on 4 principles are realized. 50 audio segments from different musical genres and some sound effects are analyzed. The purpose of this analysis is to show differences between estimated subjective loudness values obtained with used models, which are caused by dynamic and program differences between analyzed signals.

OBJECTIVE AND SUBJECTIVE COMPARISON OF DIFFERENT MODELS FOR CALCULATION OF LOUDNESS

Ivana Čpajak and Sanja Rakić