

**АКУСТИКА  
/  
ACOUSTICS  
(АК/АКІ)**



# Kriterijumi zvučnog komfora u prostorijama za vežbanje i izvođenje muzičkog programa

Dragana Šumarac Pavlović, Tatjana Miljković, Miloš Bjelić, Miomir Mijić

**Apstrakt— Zvučni komfor u prostorima za vežbanje i izvođenje muzičkog programa definisan je kriterijumima iz tri domena: zvučne zaštite, kvaliteta zvučne slike i zaštite privatnosti. U ovim akustički osetljivim prostorima zvučni komfor se ne može regulisati opštim normativima iz više razloga. Radi se o prostorima u kojima se uobičajeno javljaju povišeni nivoi zvuka u prostorijama i istovremeno pooštreni kriterijumi za dozvoljene nivoe ambijentalne buke koja dospeva iz okruženja. U ovom radu dat je pregled različitih parametara koji se koriste u projektovanju ovakvih objekata, koji se oslanja na istraživanja, pre svega subjektivnih preferenci korisnika. Polazeći od spektralnih i dinamičkih karakteristika zvuka pojedinačnih muzičkih instrumenata i kriterijuma za zvučnu izolaciju i zaštitu privatnosti analizirani su mogući dometi zvučne zaštite prostora za vežbanje standardnim pregradnim konstrukcijama. Obzirom da je u planu izgradnja nove zgrade Fakulteta muzičke umetnosti u Beogradu, cilj ovog rada je pregled relevantnih standarda i principa koji se moraju poštovati u projektovanju da bi se u prostorima za muzičko obrazovanje ostvario zvučni komfor u svim njegovim aspektima.**

**Ključne reči—** probne sale, zaštita privatnosti, zvučna zaštita

## I. UVOD

Zvučni komfor, to jest kriterijumi akustičkog dizajna, zaštite privatnosti i zvučne izolacije u prostorijama za vežbu i muzička izvođenja u okviru muzičkih škola i fakulteta predstavljaju složen zadatak obzirom na brojne kriterijume koji treba da budu zadovoljeni. U ovim specifičnim objektima kriterijumi akustičkog dizajna prostorija i zvučne zaštite treba da zadovolje mnogobrojne specifične zahteve kako bi se postigla optimalna funkcionalnost prostora i realizovala podsticajna sredina za razvoj i muzičko izražavanje korisnika,

Dragana Šumarac Pavlović – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: dsumarac@etf.rs).

Tatjana Miljković – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: tm@etf.rs).

Miloš Bjelić – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: bjelic@etf.rs).

Miomir Mijić – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: emijic@etf.rs).

bilo da se radi o prostorima za vežbu ili izvođenje. U ovakvim objektima uobičajeno se nalazi veliki broj prostorija različitih namena, kao što su male sale za probe pojedinačnih instrumenata, grupne sale za probe, velike sale za probe orkestra, studijski prostori i različiti prostori za muzička izvođenja (koncertne sale, operске sale, baletske sale). Osnovni nivo zahteva u projektovanju odnosi se na potreban volumen i dimenzije prostorija, optimalno vreme reverberacije i difuznost zvučnog polja u njima, kao i kontrolu sopstvenih modova prostorije. Sledeći nivo zahteva diktiran je karakteristikama pojedinačnih instrumenata kao što su različite spektralne karakteristike njihovog zvuka, dinamičke opsege pojedinih instrumenata i potrebu da se akustički dizajn optimizuje na osnovu kriterijuma optimalne glasnosti i podrške koju prostorija treba da pruži muzičarima kroz povratnu spregu sopstvenog kontrolnog mehanizma u izvođenju. U slučaju obrazovnih ustanova tema je i podrška koju prostor pruža u obrazovnom procesu. U domenu zvučne zaštite zahtevi se odnose na kriterijume za dozvoljene nivoe buke koja dospeva iz spoljašnje sredine, nivoa buke mašinskih sistema i potrebnu zvučnu izolovanost između prostorija.

Prostorije za vežbanje i muzička izvođenja su prostori u kojima je pažnja korisnika prostora usmerena prvenstveno na zvuk, to jest zvuk je predmet njihovog primarnog interesovanja i rada. Otuda svi kriterijumi za projektovanje treba da se oslanjaju s jedne strane na objektivne kriterijume, a s druge na subjektivna očekivanja korisnika. Korisnici ovakvih prostora, čak i kada imaju jasnu predstavu o tome šta im odgovara, ne mogu uvek svoje zahteve da izraze u korelaciji sa uobičajenim objektivnim parametrima.

Poslednjih 50 godina istraživači tragaju sa psihološkim mernim instrumentima kojima bi se opisao utisak o akustici prostorija i više od 100 godina traju naponi da se predlože objektivne mere koje bi omogućile predviđanje perceptualnih karakteristika. Metodologijama koje su u upotrebi nema sasvim zadovoljavajućeg rešenja. [1,2].

U literaturi postoji generalni stav [3,4,5] da uslovi zvučnog komfora u postojećim prostorima za vežbu i izvođenje muzike ne odgovaraju sasvim potrebama korisnika. Projektanti često nisu upoznati sa kompleksnošću problema koji ne može biti jednostavno obuhvaćen zakonskim normama i projektantskim preporukama.

Zakonskom regulativom u našoj zemlji nisu obuhvaćeni ovakvi prostori u svojoj složenosti [6]. U nacionalnom standardu SRPS U.J6.201 definisana je minimalna potrebna izolovanost između prostorija koje se graniče sa prostorijama za muzičko vaspitanje, i navodi se da je minimalan zahtev  $D_{wmin} = 60$  dB. Za istu poziciju u zgradama za međuspratnu

konstrukciju se traže uslovi  $R'_w \geq 60\text{dB}$  i  $L'_{wn} \leq 55\text{dB}$ .

Zvučni komfor u muzičkim vežbaonicama bio je tema istraživanja u mnogim državama. Tako je u Norveškoj 2014. godine usvojen standard NS 8178 pod nazivom „Acoustic criteria for rooms and spaces for music rehearsal and performanse“ [7]. U njemu su definisane potrebne dimenzije prostorija za muzička vežbanja i vreme reverberacije u njima za različite tipove muzike i različit broj članova muzičke grupe. Razmatrane su tri kategorije muzičkog sadržaja: prirodni instrumenti, muzički instrumenti male snage i muzika koja koristi elektroakustički sistem za ozvučavanje. U Velikoj Britaniji publikovan je vodič za projektovanje škola u kome je jedno poglavlje posvećeno preporukama za dizajn prostora namenjenih muzičkom obrazovanju [4,8].

U ovom radu su kao polazna osnova za utvrđivanje objektivnih kriterijuma poslužili rezultati jednog velikog istraživanja koje je sprovedeno na Škotskoj kraljevskoj akademiji za muziku i dramu (*Royal Scottish Academy of Music and Drama*) [5]. Osnovna ideja ovog istraživanja bila je da se na osnovu subjektivnih stavova korisnika prostora utvrdi koji su najvažniji faktori u dizajnu prostora za muzičko vežbanje, s obzirom da ih studenti aktivno koriste više od 42 sata nedeljno. Cilj istraživanja je bio da se utvrdi optimalno vreme reverberacije, tolerantnost na ambijentalnu buku i fizičke karakteristike prostora u korelaciji sa akustičkim karakteristikama. Istraživanjem je obuhvaćeno 22% ukupnog broja studenata muzike, a izabrani su oni koji sviraju drvene duvačke instrumente, limene duvačke, žičane instrumente i pevači. Istraživanje je sprovedeno u više faza u okviru kojih je namenski napravljeno šest različitih vežbaonica sa velikim razlikama u dimenzijama, akustičkim karakteristikama i različitim nivoima ambijentalne buke.

Rezultati analize su pokazali da je za 86% anketiranih studenata primetna saobraćajna buka, ali da većina smatra da je glavni ometajući faktor buka koja potiče iz susednih vežbaonica i to prevashodno kada je u pitanju isti instrument. Utvrđeno je da su prihvatljivi nivoi buke:

- NC30 za buku ventilacionih sistema
- NC25 za buku saobraćaja
- NC15 za buku od istog instrumenta iz susednih vežbaonica.

Pri tome, preko 90% ispitanika se slaže da je nivo od 45 dB(A) koji potiče od saobraćajne buke manje uznemirujući od nivoa 23 dB(A) koji potiče iz susedne vežbaonice.

Prema mišljenju ispitanika optimalno vreme reverberacije prostorija za muzička vežbanja je oko 0.7 s, pri čemu je poželjno da postoje uslovi za promenljivu akustiku kojom bi se vreme reverberacije moglo menjati u opsegu od 0.5 do 0.9s.

Najzad, utvrđeno je da su optimalne dimenzije prostorija za individualno vežbanje 15-20 m<sup>2</sup>.

U britanskom dokumentu BB93 definisane su kriterijumi za nivo ambijentalne buke koja potiče iz susedne prostorije za nekoliko različitih slučajeva [8]. Oni su prikazani u Tabeli I. Uzimajući u obzir nivo aktivnosti u pobudnoj prostoriji i

stepen tolerancije na nivo ambijentalne buke u tabeli II su definisane minimalne vrednosti izolovanosti između prostorija za različite kombinacije.

TABELA I  
KLASIFIKACIJA PROSTORIJA PREMA TOME KOLIKO SE NIVO ZVUKA U NJIMA  
GENERIŠE, KOLIKO SU OSETLJIVE NA SPOLJAŠNJU BUKU I KRITERIJUM  
DOZVOLJENOG NIVOVA BUKU [8]

Tip prostorije	Nivo na pobudnoj strani	Tolerancija na prijemnoj strani	Kriterijum dB(A)
Učionica	Veoma visok	Niska	35
Mala vežbaonica	Veoma visok	Niska	35
Grupna vežbaonica	Veoma visok	Veoma niska	30
Koncertna slala	Veoma visok	Veoma niska	30
Studio za snimanje	Veoma visok	Veoma niska	30
Kontrolna soba	Visok	Niska	35

U norveškom standardu NS 8178 polazna osnova za utvrđivanje optimalnog vremena reverberacije za zadati volumen prostora su dva podatka: prosečna jačina zvuka  $G$  (*stength*) i tipične vrednosti nivoa zvučne snage koju generišu pojedini instrumenti kada se svira dinamikom *forte*.

Polazeći od ukupnog optimalnog nivoa zvuka i spektralnih karakteristika pojedinih grupa instrumenata moguće je utvrditi domete izolacione moći standardnih pregradnih konstrukcija kako bi se zadovoljili uslovi izolovanosti između susednih prostora za muzičku vežbu.

Za nivo udarnog zvuka u svim prostorijama u kojima se izvodi ili snima muzički program definisan je kriterijum  $L'_{nT,w} \leq 55\text{ dB}$

TABELA II  
MINIMALNE VREDNOSTI MERODAVNE IZOLOVANOSTI ZA RAZLIČITE  
KOMBINACIJE AKTIVNOSTI I TOLERANCIJE NA NIVO AMBIJENTALNE BUKU U  
RAZLIČITIM PROSTORIMA [8]

Min $D_{nT}$ (dB)	Nivo aktivnosti u pobudnoj prostoriji				
	Tolerancija na nivo amb. buke	Visoka	30	35	45
Srednja		35	40	50	55
Niska		40	45	55	55
Veoma niska		45	50	55	60

## II. NIVO ZVUKA I DINAMIČKI OPSEG

Nivo zvuka u prostoriji koji generišu prirodni instrumenti zavisi od više faktora:

- tipa i broja instrumenata,
- načina sviranja, dinamike
- zapremine prostorije
- vremena reverberacije u prostoriji.

Muzički instrumenti se prema načinu generisanja tona dele na žičane instrumente sa gudaalom i okidanjem žica, drvene duvačke, limene duvačke i udaraljke. Način generisanja tona određuje nivo zvuka koji proizvode kao i spektralni sadržaj. Nivo zvuka zavisi takođe i od dinamike sviranja. Kao ilustracija, u tabeli III pobrojane su dinamike sviranja i

prosečni nivoi koji se na taj način postižu.

Vrednosti prikazane u tabeli odgovaraju vrednostima nivoa zvuka simfonijskog orkestra koji se očekuju u koncertnim salama. Iz tabele možemo da zaključimo da je dinamički opseg koji proizvodi simfonijski orkestar u toku kompozicije može dostići nivo od 60 dB. Zbog velikog dinamičkog opsega uobičajeno je da se u proračunima razmatraju nivoi koji odgovaraju dinamici *forte*. Majer je u svojim istraživanjima pokazao da se optimalni nivoi za slušaoce kreću u rasponu 85-92 dB, pa su te vrednosti uzete u standardima kao osnova za proračun svih vrednosti vezanih za zvučni komfor.

TABELA III  
RAZLIČITE DINAMIKE SVIRANJA I PROSEČNI NIVOI ZVUKA KOJI SE POSTIŽU U KONCERTNIM SALAMA [10]

Oznaka dinamike	Značenje	Deskripcija	Približan nivo SPL
<i>ppp</i>	<i>piano pianissimo</i>	ekstremno tiho	45-50 dB
<i>pp</i>	<i>pianissimo</i>	jako tiho	55-60 dB
<i>p</i>	<i>piano</i>	tiho	65-70 dB
<i>mf</i>	<i>mezzo forte</i>	srednje glasno	75-80 dB
<i>f</i>	<i>forte</i>	glasno	85-90 dB
<i>ff</i>	<i>fortissimo</i>	jako glasno	95-100 dB
<i>fff</i>	<i>forte fortissimo</i>	ekstremno glasno	105-110 dB

TABELA IV  
PODACI O ZVUKU RAZLIČITIH INSTRUMENTATA: NIVOI ZVUČNE SNAGE PRI DINAMICI *pp, f, ff*, DINAMIČKI OPSEG  $D$  I  $k$  FAKTOR [4]

Tip instrumenta	$L_w(pp)$ dB	$L_w(ff)$ dB	$D$ dB	$L_w(f)$ dB	$k$
Violina	65	97	32	89	0,8
Viola	68	93	25	87	0,5
Violončelo	67	97	30	90	1,0
Kontrabas	75	97	22	92	1,6
Flauta	77	96	19	91	1,3
Klarinet	74	101	27	93	2,0
Saksofon	87	101	14	98	6,3
truba	87	106	19	101	12,6
trombon	89	109	20	104	25,1

U tabeli IV prikazani su osnovni podaci o nivoima zvučne snage i dinamičkim opsezima koje dostižu pojedinačni instrumenti sa različitim dinamikama sviranja [4].

### III. PRORAČUN NIVOVA ZVUKA U RAZLIČITIM PROSTORIJAMA

Kada je poznat nivo zvučne snage izvora zvuka nivo zvučnog pritiska u prostoriji se može proceniti na osnovu izraza:

$$L_p = L_w + 10 \log \left( \frac{1}{4\pi r^2} \right) + 10 \log \left( \frac{4(1-\bar{\alpha})}{S\bar{\alpha}} \right) \text{ (dB)} \quad (1)$$

Kritično rastojanje u prostorijama u funkciji zapremine i vremena reverberacije prikazano je na slici 1. Vidi se da u prostorijama za vežbanje ono ne prelazi vrednost 0,7 m. To znači da se samo muzičar koji svira nalazi u zoni direktnog zvuka njegovog instrumenta, dok se svi ostali prisutni u

prostoriji nalaze van zone direktnog zvuka. Na osnovu toga proizilazi da se nivo zvuka koji slušaju ostali prisutni može približno odrediti izrazom:

$$L_p = L_w + 10 \log \left( \frac{4(1-\bar{\alpha})}{S\bar{\alpha}} \right) \text{ (dB)} \quad (2)$$

Pojačanje koje prostorija unosi opisuje se veličinom  $G$  koja se naziva jačina (*strength*). Ona je:

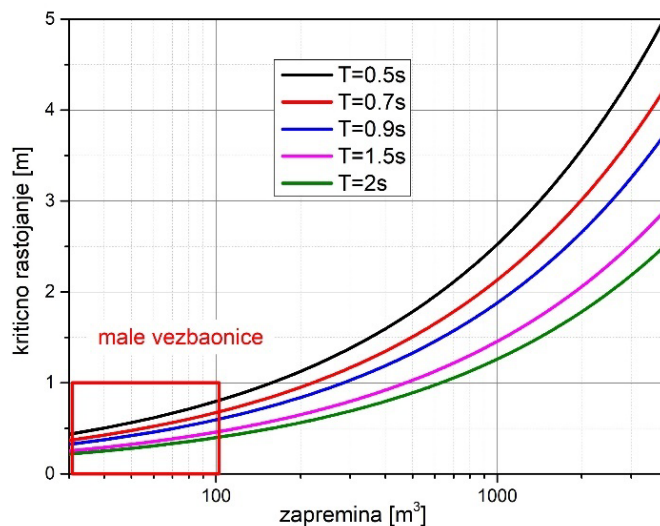
$$G = L_p - L_{\text{dir}}(r_0 = 10\text{m}) \text{ (dB)} \quad (3)$$

$G$  se može proceniti na osnovu srednjeg koeficijenta apsorpcije i ukupne unutrašnje površine u prostoriji na sledeći način:

$$G = 10 \log \left( \frac{4(1-\bar{\alpha})}{\bar{\alpha}S} \right) - 10 \log \left( \frac{1}{4\pi r_0^2} \right) \cong 31 + 10 \log \left( \frac{4(1-\bar{\alpha})}{\bar{\alpha}S} \right) \text{ (dB)} \quad (4)$$

Za svaku geometrijsku formu prostorije može se definisati parametar koji je nazvan "faktor oblika", i koji predstavlja jedan mogući kvantifikator njene geometrijske forme [9]

$$k = \frac{\sqrt[3]{V}}{2S}, \quad S = \frac{V^{2/3}}{k^2} \text{ (m}^2\text{)}, \quad S = \frac{0.16V}{\bar{\alpha}T}, \quad \bar{\alpha} = \frac{0.16V^{1/3}}{k^2T} \quad (5)$$



Slika 1 Kritično rastojanje u prostorijama u funkciji zapremine za različite vrednosti vremena reverberacije

Za jednu formu prostora konstantan je faktor oblika nezavisno od stvarnih dimenzija prostorije. Faktor oblika za paralelopipedne prostorije različitih proporcija kreće se u rasponu 0,35-0,4 [9]. Opseg vrednosti srednjeg koeficijenta apsorpcije u prostorima za vežbu označen je na slici 2.

Nivo zvuka u prostoriji može se izraziti preko vrednosti  $G$  i akustičke snage izvora [2]. Uobičajeno je da se nivo zvuka instrumenata razmatra pri sviranju sa dinamikom *forte*.

$$L_p(f) = L_w(f) + G - 31 \text{ (dB)} \quad (6)$$

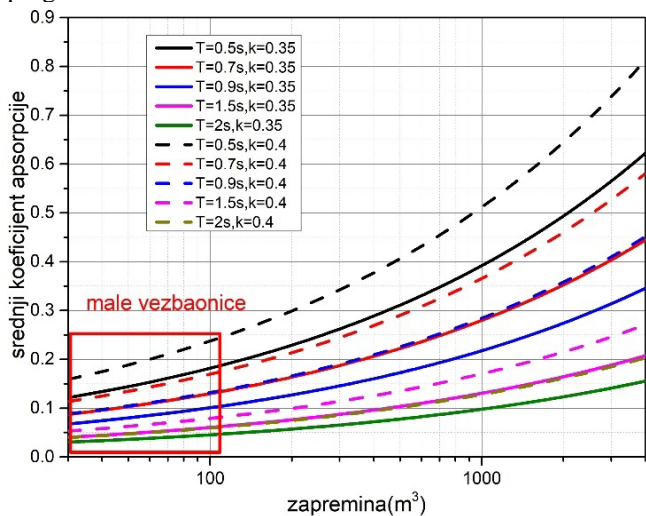
Ukoliko u prostoriji postoji više izvora ukupan nivo zvuka može se proceniti pomoću izraza [11]:

$$L_p(f) = G + 59 + 10 \log \sum_i n_i k_i \text{ (dB)} \quad (7)$$

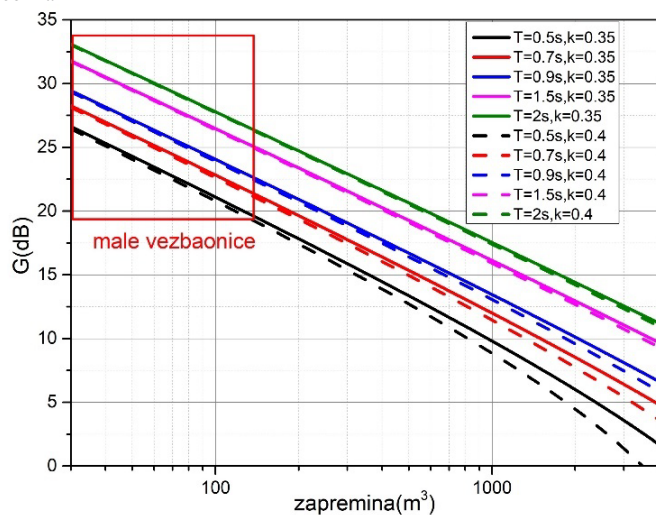
gde je  $n_i$  broj instrumenata iz iste grupe i  $k_i$  faktor definisan u tabeli IV.

Jačina zvuka u prostoriji menja se zavisno od njenog volumena, vremena reverberacije i faktora oblika. Na slici 3 prikazane su promene  $G$  u funkciji zapremine za različite

vrednosti vremena reverberacije i faktora oblika. U vežbaonicama manjih zapremina za raspone vremena reverberacije od 0,5 s do 0,7 s jačina zvuka G ima vrednosti u opsegu 20-30 dB.



Slika 2 Srednji koeficijent apsorpcije u prostorijama u funkciji zapremine za različite vrednosti vremena reverberacije i različite granične vrednosti faktora oblika



Slika 3 Vrednosti parametra G u funkciji zapremine, vremena reverberacije i faktora oblika; označen je opseg vrednosti u kome se nalaze male muzičke vežbaonice.

#### IV. SPEKTRALNE KARAKTERISTIKE INSTRUMENTATA

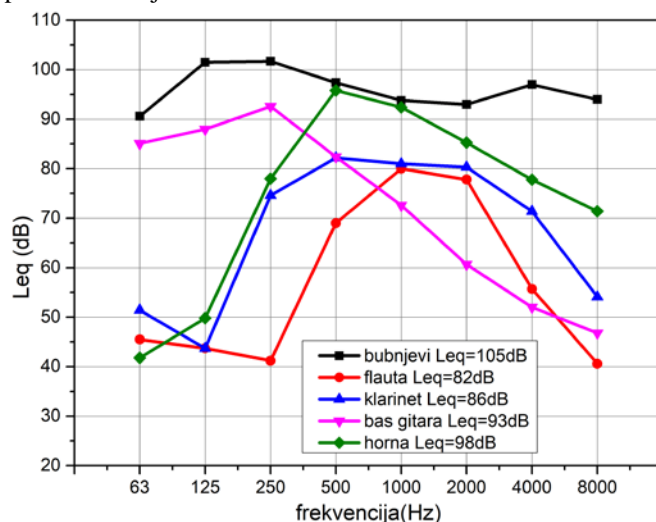
Različiti instrumenti osim što se mogu okarakterisati različitim nivoima zvučne snage i različitim dinamičkim opsegom, značajno se međusobno razlikuju po opsegu frekvencija u kojima dominantno stvaraju zvuk, pa prema tome i po spektralnim karakteristikama. Pored ukupnog nivoa zvuka u prostoriji spektralna karakteristika može značajno da utiče da princip akustičkog dizajna i zvučne zaštite prostora.

Takođe, spektralna karakteristika se menja u zavisnosti od dinamike sviranja. Mayer je utvrdio da je povećanjem dinamike signala menja nagib spektralne karakteristike tako što se dominantno povećava energija na višim harmonicima [11]. To dovodi i do toga da dinamika koji izvođač čuje u zoni direktnog zvuka instrumenta nije ista kao dinamika koju

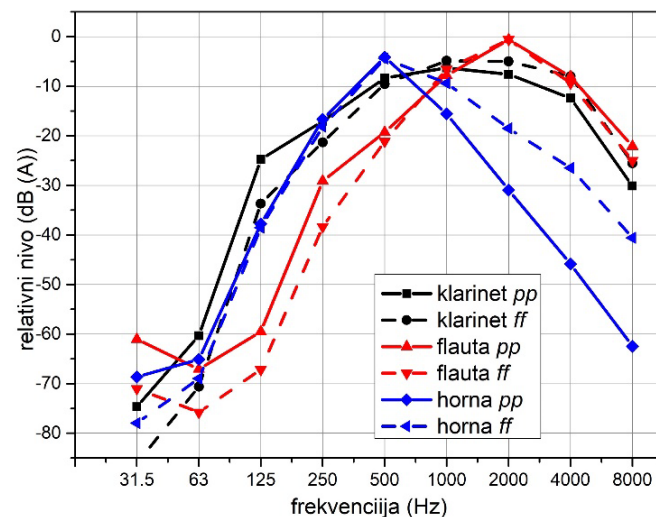
percipiraju slušaoci na udaljenim mestima u sali, pogotovo kada je prisutno značajno smanjenje vremena reverberacije sa frekvencijom.

Na slici 4 prikazane su izmerene vrednosti ekvivalentnog nivoa zvuka u manjim vežbaonicama različitih instrumenata [10]. Ovi podaci mogu se koristiti kao relevantni za procenu nivoa zvučne snage ovih izvora. Razmatrano je nekoliko kategorija instrumenata i to baterija bubnjeva, flauta, klarinet, bas gitara i francuski rog.

Na slici 5 prikazani su normalizovani spektri flaute, klarineta i horne sa dinamikom sviranja *pianissimo* i *fortissimo*. Spektralne karakteristike su ponderisane A filtrom. Kod svih instrumenata uočljivo je manje ili više izdizanje spektra na visokim frekvencijama, što instrumentalisti daje subjektivni doživljaj veće dinamike koju postiže sviranjem.



Slika 4. Spektralne karakteristike i ukupni ekvivalentni nivo za različite instrumente.



Slika 5. Normalizovane spektralne karakteristike sa dinamikom *pianissimo* i *fortissimo* ponderisane A filtrom

#### V. POTREBNA IZOLACIONA MOĆ PREGRADNIH KONSTRUKCIJA

Istraživanja subjektivnog doživljaja muzičara kada vežbaju u malim i srednjim vežbaonicama pokazala su da su u pogledu zvučne izolacije i zaštite privatnosti u objektu kritične pozicije

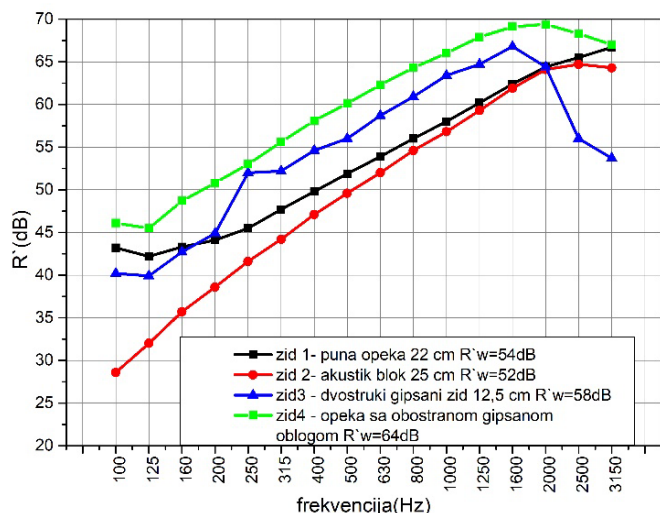
one koje se neposredno graniče sa prostorijama u kojima se pojavljuju isti instrumenti [5]. Kod muzičara koji sviraju određeni instrument dodatno je pojačano analitičko slušanje i osjetljivost na za njih prepoznatljiv zvuk.

U literaturi je ranije konstatovano da su najstrožiji kriterijumi za zvučnu izolovanost između dva susedna prostora u slučaju da u njima vežbaju muzičari koji sviraju isti instrument. U tom slučaju nivo zvuka koji prolazi u susednu prostoriju mora zadovoljavati kriterijum NR15, odnosno nivo zvuka 23 dB(A) [5].

Polazeći od ovih zaključaka analizirane su izolovanosti koje se postižu sa nekoliko standardnih pregradnih konstrukcija:

- ZID 1 - zid od pune opeke (22cm) dvostruko malterisan;
- ZID 2 – zid od blokova debljine 25 cm obostrano malterisan;
- ZID 3 – sendvič zid od dvostrukih gipsanih obloga;
- ZID 4 – zid od pune opeke sa dvostranom gipsanom oblogom od dve ploče gipsa.

Građevinske izolacione moći razmatranih pregradnih konstrukcija prikazane su na slici 6.



Slika 6. Građevinske izolacione moći za četiri tipa pregrada

Za proračune izolovanosti posmatran je slučaj malih vežbaonica veličine 40 m<sup>3</sup> koje deli pregradni zid površine 10 m<sup>2</sup> i u kojima je srednji koeficijent apsorpcije  $\alpha = 0,15$ . Izolovanost D između njih izračunava se na osnovu vrednosti građevinske izolacione moći pregrade koja ih deli, površine te pregrade S i ukupne apsorpcije u prijemnoj prostoriji:

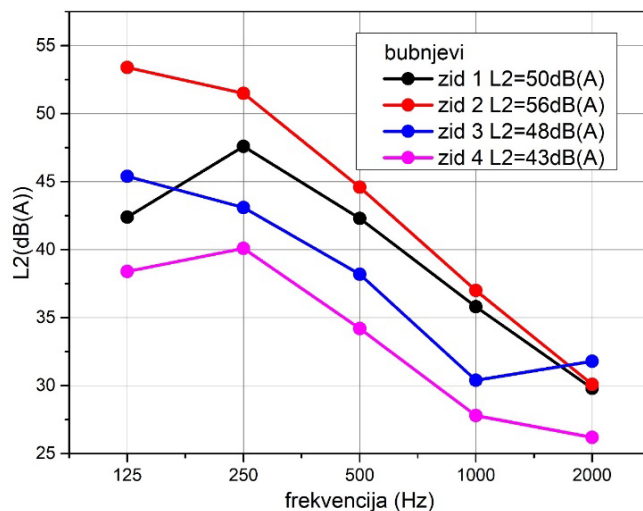
$$D = R' - 10 \log \frac{S}{A_{prijem}}$$

Za usvojene dimenzije i akustičke karakteristike vežbaonice izračunati su nivoi u prijemnoj prostoriji kada su u predajnoj prostoriji usvojeni ekvivalentni nivoi prikazani na slici 5. Za svaki od analiziranih instrumenata prikazane su spektralne karakteristike buke u prijemnoj prostoriji koja nastaje preslušavanjem (slike 7,8,9,10 i 11)

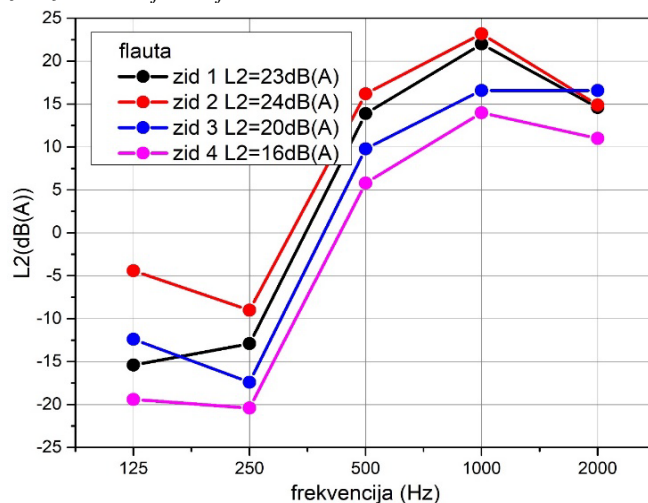
Analizom nivoa zvuka u prijemnoj prostoriji pri pobudama različitim instrumentima pokazano je da standardne pregradne

konstrukcije nisu u mogućnosti da zadovolje stroge kriterijume zvučne izolovanosti. U slučaju baterije bubnjeva ni sa obostranim oblaganjem dvostrukim gipsanim oblogama na potkonstrukciji od mineralne vune nije moguće postići željeni nivo izolovanosti. Šta više, nivoi u prijemnoj prostoriji premašuju i nivoje dozvoljene buke koja potiče od saobraćaja i sistema za ventilaciju. To su situacije u kojima se zadovoljavajuća izolovanost može postići samo uvođenjem tampon zona oko vežbaonica za bučne instrumente.

Prikazani rezultati pokazuju da u slučaju flaute i klarineta, zbog njihove znatno manje zvučne snage i drugačijeg oblika spektra zvuka, moguće je postići traženu izolovanost i sa standardnim zidovima, to jest bez dodatnih gipsanih obloga. U slučaju bas gitare i francuskog roga standardnim pregradama građevinskim nije moguće postići adekvatnu zvučnu izolaciju između prostorija za vežbanje.



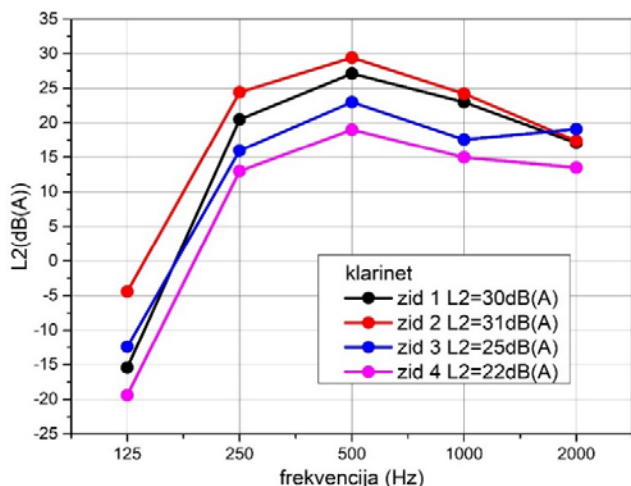
Slika 7. Spektralne karakteristike nivoa zvuka u prijemnoj prostoriji kada je izvor zvuka baterija bubnjeva



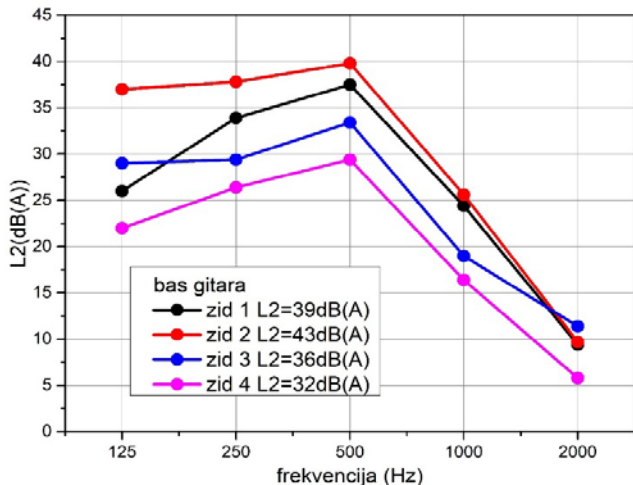
Slika 8. Spektralne karakteristike nivoa zvuka u prijemnoj prostoriji kada je izvor zvuka flauta

## VI. ZAKLJUČAK

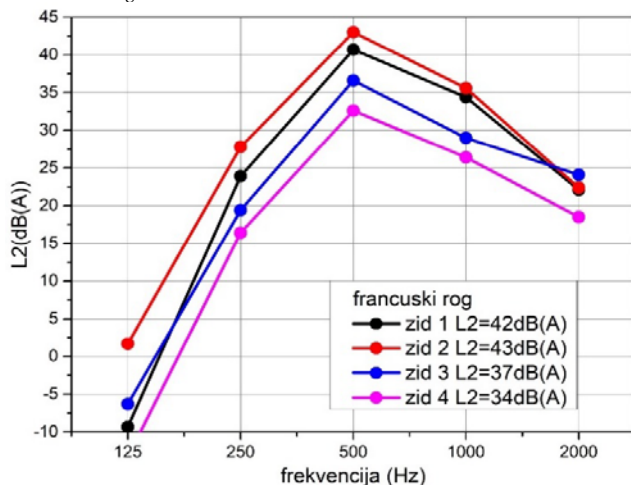
U radu je prikazan pregled poznatih metodologija i standarda koji se odnose na zvučni komfor u zgradama sa prostorijama za vežbanje i izvođenje muzičkog programa.



Slika 9. Spektralne karakteristike nivoa zvuka u prijemnoj prostoriji kada je izvor zvuka klarinet



Slika 10. Spektralne karakteristike nivoa zvuka u prijemnoj prostoriji kada je izvor zvuka bas gitara



Slika 11. Spektralne karakteristike nivoa zvuka u prijemnoj prostoriji kada je izvor zvuka bas gitara i sa različitim pregradama između njih

Prikazan je pregled nekih rezultata koji su dobijeni u subjektivnim istraživanjima vezanim za očekivanja muzičara u pogledu zvučnog komfora i definisane su osnovne veličine od kojih se polazi u procesu akustičkog dizajna i projektovanja zvučne zaštite u takvim objektima. Analizirana je zvučna izolovanost i privatnost koja se postiže sa četiri

standardne pregrade za slučaj različitih instrumenata kao izvora zvuka.

Analiza rezultata pokazala je da se pri izboru pregradnih konstrukcija mora voditi računa o realnim nivoima zvuka i spektralnim karakteristikama instrumenata kako bi se pronašla adekvatna rešenja koja obezbeđuju traženu međusobnu izolovanost prostorija. Ta rešenja moraju u određenim slučajevima da podrazumevaju tampon zone između prostorija.

#### LITERATURA

- [1] Weinzierl, S., Vorlander M., „Room Acoustical Parameters as Predictors of Room Acoustical Impression: What Do We Know and What Would We Like to Know?“, Acoustic Australia, DOI 10.1007/s40857-015-0007-6
- [2] SRPS ISO 3382-1:2009, Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces.
- [3] Knofel B., Weisheit L., Troge, J., „Musicians and their practice rooms: What do they think about present room acoustics and would they prefer?“, Conference: Euronoise 2018, 11th European Congress and Exposition on Noise Control Engineering : Reduce Noise to Improve Life
- [4] Rindel, J. H., „Rooms for music – Acoustical needs and requirements“, BNAM2014
- [5] Lamberty, D.C., (1980) „Music Practice Rooms“, *Journal of Sound and Vibration*, Vol 60., No1, p149-155
- [6] SRPS ISO UJ.6.201, „Akustika u zgradarstvu- Tehnički uslovi za projektovanje i gradnje zgrada“, 1989
- [7] NS 8178:2014, „Acoustic criteria for rooms and spaces for music rehearsal and performance“. (In Norwegian). Standard Norge, Oslo, 2014
- [8] BB93: „Acoustic design of School, performance standard“, Education Funding Agency, 2015
- [9] D. Šumarac-Pavlović and M. Mijic, „An approach to numerical quantification of room shape and its function in diffuse sound field model (L)“, *J. Acoust. Soc. Am.* 140 (4), October 2016
- [10] James A., Thompson A., Rees I., „School Music Rooms- designing Beyond BB93“, *Proceedings of the Institute of Acoustics*, 2005
- [11] J. Meyer, „Acoustics and the performance of music“, Springer, 2009

#### ABSTRACT

Sound comfort in the spaces for rehearsing and performing music program is defined by criteria from three domains: sound protection, sound image quality and privacy protection. The problem is especially significant in the buildings of music schools and faculties. In such acoustically sensitive spaces, sound insulation cannot be regulated by general legislation for two reasons. First, these are spaces where high sound levels usually occur compared to standard spaces such as those in other buildings. Second, they have strict criteria for the allowed levels of ambient noise coming from the environment. That is reflected in the special measures of sound insulation, as well as the control of all systems in the building. Starting from the spectral and dynamic characteristics of the musical instruments sound and the criteria for sound insulation and privacy protection, the sound insulation ranges of standard partition structures were analyzed. Special attention is paid to individual and group rehearsal spaces. Given that the construction of a new building for the Music Academy in Belgrade is planned, the aim of this paper is to review all relevant standards and principles that must be respected in order to achieve sound comfort in music education spaces in all its aspects.

#### Criteria for acoustic comfort in the premises for practicing and performing music program

Dragana Šumarac Pavlović, Tatjana Miljković, Miloš Bjelić, Miomir Mijić



# Uticaj estimacije frekvencija harmonika na procenu koeficijenta inharmoničnosti čembala

Tatjana Miljković, Jovana Damnjanović, Jelena Čertić, Dragana Šumarac Pavlović

**Apstrakt**—Inharmoničnost je pojava koja se javlja kod muzičkih instrumenata koji se teorijski smatraju harmoničnim, i predstavlja odstupanje frekvencija parcijala tona od celobrojnih umnožaka osnovne frekvencije tona. Za žičane muzičke instrumente, u literaturi se definiše koeficijent inharmoničnosti, kao mera odstupanja od idealne harmoničnosti. U prethodnom istraživačkom radu pokazano je da automatski algoritam za procenu koeficijenta inharmoničnosti tonova klavira za pojedine tonove iz registra ne vrši dobru estimaciju koeficijenta inharmoničnosti. U ovom radu razmatran je uticaj tačnosti procene frekvencija harmonika na procenu koeficijenta inharmoničnosti. Sprovedena je uporedna analiza dve metode za procenu spektra signala, i to procena spektra na osnovu autoregresivnog AR modela i metodom diskretne Furijeove transformacije DFT. Testiranje predloženih metoda za procenu spektra izvršeno je na realnim tonovima čembala. Ustanovljeno je, na osnovu trenda koeficijenta inharmoničnosti računatog pomoću obe metode za procenu spektra, da je metoda AR modelom superiornija, odnosno da se dobijaju relevantni rezultati za koeficijent inharmoničnosti na celom opsegu tonova od interesa.

**Ključne reči**—Inharmoničnost, žičani instrumenti, čembalo, procena spektra.

## I. UVOD

„Idealna“ žica predstavlja žicu zategnutu na oba kraja, koja se ne odlikuje krutošću. Jedna od glavnih karakteristika takve žice, sa aspekta stvaranja zvuka, je što pri njenom okidanju dolazi do formiranja niza harmonika koji predstavljaju celobrojne umnoške osnovne frekvencije žice. U realnosti nemoguće je posmatrati žicu kao element sa svojstvima idealne fleksibilnosti, već je neophodno uvesti dodatne parametre poput krutosti, debljine, težine i napetosti žice. Krajem XIX veka *John William Strutt*, lord *Rayleigh*, razmatrajući oscilovanje žica klavira, ustanovio je da krutost same žice utiče na parcijale tonova, tako što frekvencije parcijala tonova na klaviru odstupaju od predefinisanih frekvencija parcijala idealne žice [1]. Pojava odstupanja frekvencije parcijala tona od celobrojnih umnožaka osnovne

frekvencije tona naziva se inharmoničnost.

Matematičko-fizički modeli [2], [3], [4] koji u osnovi opisuju inharmoničnost kod klavira, definišu principski jednostavne relacije koje pojavu opisuju preko koeficijenta inharmoničnosti, koji se definiše za svaku žicu instrumenta.

Postoje različite metode i algoritmi za procenu koeficijenta inharmoničnosti na osnovu tonova instrumenta od interesa. *Galebo* i *Askenfelt* [5] predložili su jednu od metoda za procenu koeficijenta inharmoničnosti koja se zasniva na inharmoničnom komb filtru, realizovanom u frekvencijskom domenu. Takođe, tehnike poput određivanja visine tona, kepstralne analize i HPS (*Harmonic Product Spectrum*) su korišćene za estimaciju inharmoničnosti [6], [7]. Jedan od algoritama koji se odlikuje automatskim postupkom za estimaciju koeficijenta inharmoničnosti jeste PFD (*Partial Frequency Deviation*) algoritam [8]. Princip rada algoritma zasniva se na minimiziranju devijacije, koja je nastala kao razlika frekvencija parcijala procenjenih iz spektra tona i frekvencija parcijala računatih na osnovu matematičkog modela. Sam proces minimizacije frekvencijske devijacije se odvija iterativnim postupkom, gde je krajnji produkt algoritma estimirana vrednost koeficijenta inharmoničnosti tona.

U prethodnom istraživačkom periodu analizirani su različiti klaviri na osnovu PFD algoritma za procenu koeficijenta inharmoničnosti. Izvršena je analiza koeficijenta inharmoničnosti procenjena sa različitim brojem parcijala tonova. Takođe, utvrđeno je da se takav algoritam može koristiti za procenu koeficijenta inharmoničnosti tonova klavira za skoro celokupni registar instrumenta [9]. Za pojedine tonove iz opsega klavira, iako je došlo do konvergencije samog algoritma, estimirana vrednost koeficijenta inharmoničnosti nije bila tačna. Razlog za lošu procenu koeficijenta inharmoničnosti pojedinih tonova leži u lošoj proceni frekvencija parcijala tona iz spektra signala.

U ovom radu razmatran je uticaj tačnosti procene frekvencija harmonika na procenu koeficijenta inharmoničnosti. Korišćen je pomenuti iterativni algoritam PFD, pri čemu je akcenat stavljen na prvi korak adaptivnog algoritma, odnosno procenu frekvencija tona iz spektra. Izvršena je uporedna analiza dve metode za procenu spektra signala i to procena spektra na osnovu AR (*Autoregressive*) modela i metodom DFT (*Discrete Fourier transform*). Uticaj tačnosti estimacije frekvencija harmonika na procenu koeficijenta inharmoničnosti testiran je na tonovima čembala. Za potrebe rada izvršeno je snimanje tonova čembala na Fakultetu muzičke umetnosti u Beogradu. Čembalo je izabrano kao instrument za analizu, jer se spektri njegovih tonova odlikuju velikim brojem harmonika. Prisutnost velikog

Tatjana Miljković – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: tm@etf.bg.ac.rs).

Jovana Damnjanović – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: dj193200m@etf.bg.ac.rs).

Jelena Čertić – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: certic@etf.bg.ac.rs).

Dragana Šumarac Pavlović – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: dsumarac@etf.bg.ac.rs).

broja parcijala u spektru tona je značajna, zbog činjenice da su fizičke pojave koje opisuju fenomen inharmoničnosti uočljivije na višim parcijalima tona.

Rad je organizovan kako sledi. U drugom poglavlju date su osnovne razlike između čembala i klavira, poput mehanizma stvaranja zvuka i spektralne odlike tona ovih instrumenata. U narednom poglavlju prikazane su metodologije za procenu spektra korišćene u iterativnom algoritmu za procenu koeficijenta inharmoničnosti. U četvrtom poglavlju prikazani su eksperimentalni rezultati i diskusija dobijenih rezultata. Na kraju izneti su zaključci do kojih se došlo u ovom istraživanju.

## II. OSNOVNE RAZLIKE IZMEĐU ČEMBALA I KLAVIRA

Žičani instrumenti su svi oni instrumenti kod kojih se zvuk stvara vibriranjem žice. S obzirom na različite načine pobuđivanja žice na vibriranje žičani instrumenti se mogu klasifikovati u 3 podgrupe, i to na: gudačke instrumente (violina, violončelo), trzane (gitara, čembalo) i udarne žičane instrumente (klavir) [10].

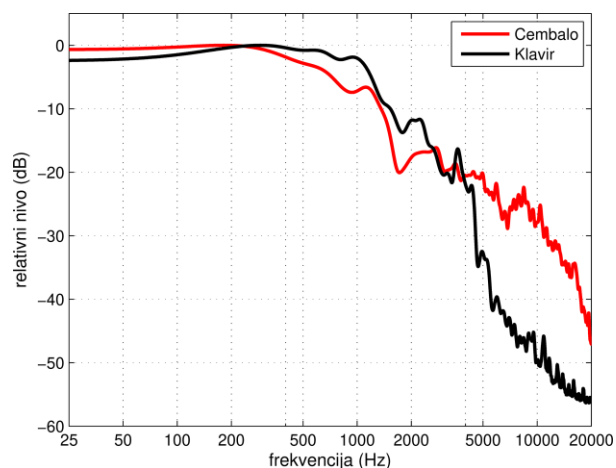
Čembalo je predstavnik žičanih instrumenata, koji spadaju u podgrupu trzanih žičanih instrumenata sa klavijaturom. Iako izgledom podseća na klavir, postoji nekoliko suštinskih razlika između ova dva instrumenta. Jedna od razlika je u materijalu od kojih su sačinjena ova dva instrumenta. Čembalo je izgrađeno od laganog drvenog okvira, tankih žica i nema pedale. Dok je klavir izgrađen od teškog gvozdene okvira, debljih žica gde su neke od njih i obmotane i pedala. Takođe, čembalo može imati jednu, dve ili tri klavijature, koje se nazivaju manuali. Opseg tonova koji se može odsvirati na čembalu je manji u odnosu na klavir i uglavnom iznosi 5 oktava, dok opseg klavira uglavnom iznosi sedam oktava [11].

Značajna razlika između ova dva instrumenta je u mehanizmu nastanka zvuka. Kod klavira zvuk nastaje udarcem čekića o veoma zategnutu žicu, dok kod čembala zvuk nastaje okidanjem blago rastegnute žice perom. Pero koje služi za okidanje žice nalazi se na pokretnom jezičku, smeštenom na vrhu stubića, koji svojim donjim krajem naleže na zadnji kraj dirke. Prilikom pritiska dirke, njen zadnji kraj podiže stubić, tako da pero, u kretanju naviše zakači žicu i proizvede zvuk. Istovremeno, tim dodiranjem sa žicom izbacuje se jezičak iz osnovnog položaja kako pero ne bi pri silaznom kretanju stubića, ponovo zakačilo žicu i time proizvelo novi zvuk. Pri otpuštanju dirke, stubić koji je na donjem kraju opterećen olovom, vraća se u prvobitni položaj, pri čemu posebna opruga vraća i jezičak u vertikalni stav. Pošto se pri vrhu stubića nalazi i mali prigušivač, povratkom u prvobitni položaj on dotiče žicu i prekida njeno treperenje.

Važna karakteristika čembala je u tome da se snagom udara u dirku ne može uticati na jačinu proizvedenog tona, jer trzaj koji čini pero vrlo malo zavisi od te snage. Zato se dinamičke promene mogu ostvariti jedino uključivanjem raznih registara. Pojam registra u ovom smislu ne odnosi se na određeni instrumentalni tonski opseg već predstavlja na određeni način ostvaren zvuk posebnih odlika u dinamici ili tonskoj boji [10]. Ukoliko instrument ima dva manuala, na gornjem manualu svakoj dirci odgovaraju po dve žice, gde je jedna štimovana

na nominalnu tonsku visinu dirke, a druga za oktavu više. Na donjem manualu svaka dirka takođe ima po dve odgovarajuće žice, jednu na nominalnoj tonskoj visini, a drugu za oktavu niže štimovanu. Time, za svaki ton postoje u instrumentu po četiri žice. Osim registara za razne dinamičke promene, postoje i takvi koji zvuku instrumenta daju novu boju, sličnu tonu laute ili harfe.

Još jedna od razlika između čembala i klavira jeste značajno bogatije prisustvo harmonika u tonovima čembala u odnosu na klavir. Na Sl. 1. prikazani su usrednjeni spektri tonova klavira i čembala u okviru celog frekvencijskog opsega. Sa slike se može uočiti da zbog prisustva većeg broja harmonika usrednjeni spektar čembala je izdignutiji u odnosu na spektar klavira za otprilike 30 dB na višim frekvencijama.



Sl. 1. Usrednjeni spektri klavira i čembala u okviru celog frekvencijskog opsega.

## III. METODOLOGIJA PROCENE SPEKTRA I ODREĐIVANJA KOEFICIJENTA INHARMONIČNOSTI

Kako bi se pojam inharmoničnosti tona mogao objektivizirati, stvorena je formulacija na osnovu koje se može odrediti frekvencija  $k$ -tog harmonika odsviranog tona kao:

$$f_k = kf_0 \sqrt{1 + k^2 B}, \quad (1)$$

gde je  $k$  redni broj harmonika,  $f_0$  osnovna frekvencija tona, a  $B$  koeficijent inharmoničnosti nastao pri sviranju tona [3]. Koeficijent inharmoničnosti  $B$  se može izračunati na osnovu (1), gde bi se na osnovu spektra realnog signala odredilo prvih  $k$  harmonika odsviranog tona, a zatim na osnovu njih procenila vrednost  $B$ . Jedan od predloženih algoritama za automatsku procenu koeficijenta inharmoničnosti je PFD algoritam [8]. Iterativni deo PFD algoritma se zasniva na izračunavanju devijacije frekvencije parcijala, gde se proračunava razlika niza frekvencija parcijala dobijenih iz procene spektra i odgovarajućeg niza frekvencija parcijala dobijenih na osnovu jednačine (1) za vrednost  $B$  u tekućoj iteraciji. Na osnovu trenda devijacije frekvencije, donosi se odluka da li se koeficijent  $B$  povećava ili smanjuje [8]. Sam algoritam se pokazao kao dosta robustan u situacijama kada u spektru signala nedostaje neki harmonik ili kada je frekvencija pojedinog parcijala pogrešno procenjena. Međutim, u prethodnom istraživačkom radu ispostavilo se da za pojedine

tonove iz registra klavira dolazi do loše procene koeficijenta  $B$  [9]. Razlog za lošu procenu koeficijenta  $B$  leži u početnom koraku samog algoritma gde se vrši procena spektra i određuju vrednosti prvih  $k$  frekvencija parcijala. Sama procena frekvencija prvih  $k$  harmonika može se realizovati na različite načine, DFT analizom [8], [12], tehnikama zasnovanim na inharmoničnim češljastim filtrima [5], [13], kao i AR modelima [14]. U ovom radu razmatran je uticaj procene frekvencija harmonika na tačnost procene parametra  $B$  na primeru snimljenih tonova jednog čembala. Kao metode procene spektra korišćene su metode: DFT analiza i AR modelovanje.

#### A. Metoda 1 – DFT analiza

DFT analiza predstavlja klasičan pristup u proceni spektra signala. U obradi audio signala, zbog nestacionarnosti signala, često se koristi analiza po segmentima, odnosno „prozorima“. Dodatno, procena tačne pozicije komponenti u spektru se popravljala dodavanjem nula, odnosno računanjem DFT-a u više tačaka nego što je dužina segmenta signala. Dužina prozora, tip prozorske funkcije kao i broj tačaka u kojima se DFT računa su parametri koji utiču na tačnost procene frekvencija harmonika.

Kao ulazni podaci algoritma za procenu koeficijenta  $B$  koriste se odbirci audio signala snimljenog tona čembala.

Na početku analize, signal je prozorovan *Hamming*-ovom prozorskom funkcijom. Naredni korak je računanje spektra prozorovanog signala pomoću DFT. Kako bi frekvencijska rezolucija spektra signala bila dobra, neophodno je niz dopuniti nulama do vrednosti  $2^{16}$ , a zatim izvršiti izračunavanje spektra signala u  $2^{16}$  tačaka. Dodavanjem nula postignuto je lakše procenjivanje spektralnih komponenti signala.

Za posmatrani audio signal na samom početku vrši se procena osnovne frekvencije tona. Osnovna frekvencija tona se određuje tako što se za osnovnu frekvenciju tona proglašuje frekvencija koja odgovara najvećoj amplitudi u spektru za predloženi opseg, definisan na osnovu očekivanih frekvencija za ton koji se analizira. Kod tonova koji pripadaju nižem registru čembala, ova metoda za procenu osnovne frekvencije tona daje loše rezultate. Razlog za lošu procenu leži u tome što u predloženom opsegu za datu osnovnu frekvenciju postoji nekoliko frekvencija čije su pripadajuće amplitude sličnih vrednosti. U tom slučaju, da bi se izbeglo pogrešno procenjivanje osnovne frekvencije za tonove iz nižeg registra procenjuje se frekvencija prvog harmonika tona.

S obzirom na to da je očekivano da se harmonici nalaze na celobrojnim umnošcima osnovne frekvencije tona, moguće je izvršiti manipulaciju nad spektrom u cilju smanjivanja neinformativnih delova spektra signala. U tom kontekstu, spektar se deli na podopsege gde se iz svakog podopsega vrši selekcija 10 spektralnih komponenti sa najvećim amplitudama. Širina podopsega je definisana kao  $3f_1$ , gde  $f_1$  predstavlja procenjenu frekvenciju prvog parcijala tona. Deo procene spektra, u slučaju primene DFT analize se završava na početku samog iterativnog algoritma, gde se još jednom vrši selekcija spektralnih komponenti. Nad već redukovanim

spektrom signala, vrši se selekcija frekvencija parcijala tona  $f_k$  koje odgovaraju maksimalnoj amplitudi spektra unutar zatvorenog intervala. Donja granica intervala predstavlja razliku frekvencije parcijala tona  $f_k$  i pomeraja  $\Delta f$ , dok gornja granica intervala predstavlja zbir frekvencije parcijala tona  $f_k$  i pomeraja  $\Delta f$ . Za vrednost pomeraja  $\Delta f$  u algoritmu korišćena je vrednost  $0.4f_1$ .

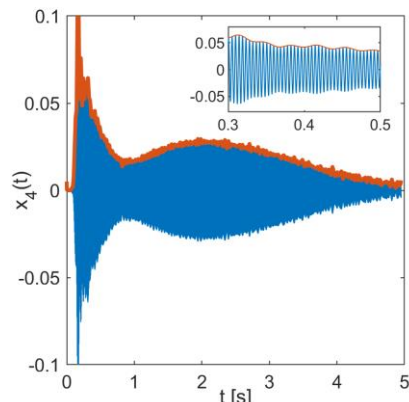
#### B. Metoda 2 – AR model

AR modelovanje podrazumeva da se na osnovu segmenata signala procene koeficijenti modela. Znajući koeficijente modela, za signale kod kojih su spektralne komponente izražene moguće je izvršiti procenu frekvencija spektralnih komponenti. Iako su za modelovanje audio signala, a posebno tonova muzičkih instrumenata, u literaturi preporučeni ARMA modeli [14], za samu procenu spektra AR modeli su se pokazali kao dovoljno dobar metod. Ujedno, AR modeli su jednostavniji za implementaciju jer se koeficijenti modela mogu izračunati *Yule-Walker*-ovim jednačinama [15]. Podešavanjem reda modela moguće je uticati na tačnost procene spektralnih komponenti.

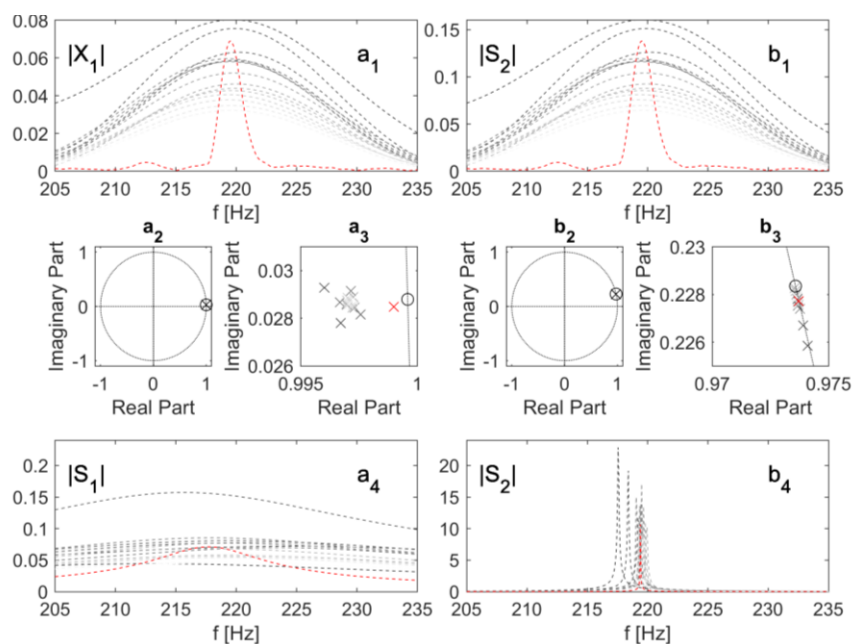
Kao i za slučaj procene spektra metodom DFT analize ulazni podaci algoritma su odbirci audio signala snimljenog tona čembala. Na samom početku potrebno je odrediti red AR modela. Red je empirijski procenjen i variran u zavisnosti od frekvencije pojedinih tonova iz opsega čembala. Zatim je spektar signala podeljen na podopsege čija je širina adekvatna za tri harmonika. Svaki od podopsega je „spušten“ u osnovni opseg i zatim je izvršeno filtriranje i decimacija [14]. Kako faktor decimacije zavisi od osnovne frekvencije tona, korišćene su različite vrednosti faktora decimacije. Krajnji rezultat metode AR modelovanja je procenjivanje frekvencija spektralnih komponenti na osnovu položaja dominantnih polova.

#### C. Analiza uticaja različitih parametara na procenu spektra

Kako bi se ispitao uticaj različitih parametara na procenu koeficijenta inharmoničnosti biće izvršena analiza signala izdvojenog četvrtog harmonika tona nominalne osnovne frekvencije 55 Hz čiji je segment prikazan na Sl. 2. Signal je izdvojen alatom zasnovanom na digitalnoj komplementarnoj filterskoj banci, pogodnoj za analizu audio signala [16].



Sl. 2. Vremenski oblik četvrtog harmonika tona osnovne frekvencije 55 Hz.



Sl. 3. Uticaj različitih parametara na procenu frekvencije četvrtog harmonika tona nominalne osnovne frekvencije  $f_{0N}=55$  Hz. a1) DFT na frekvenciji odabiranja  $f_s$  a2) Polovi AR modela računati na frekvenciji odabiranja  $f_s$  a3) Zumiran detalj slike a2), polovi AR modela, a4) Spektar računat na osnovu modela čiji su polovi prikazani na a2) i a3). b1) DFT na decimiranom signalu. b2) Polovi AR modela računati na decimiranom signalu b3) Zumiran detalj slike b2), polovi AR modela b4) Spektar računat na osnovu modela čiji su polovi prikazani na b2) i b3).

Na Sl. 2 je pored vremenskog oblika četvrtog harmonika crvenom bojom prikazana i njegova anvelopa, dobijena kao moduo Hilbertove transformacije signala. Iako je ovako filtriran signal jednostavniji za analizu, već je na osnovu njegovog vremenskog oblika moguće uočiti određene probleme koji se dodatno usložnjavaju prisustvom većeg broja harmonika. Jedan od problema, koji se može uočiti vizuelnom inspekcijom anvelope harmonika jeste da izdvojeni harmonik nije idealna sinusoida. Činjenica da različiti harmonici istog tona imaju različite anvelope dovodi do toga da u većoj ili manjoj meri njihove anvelope utiču na tačnost procene frekvencije harmonika, a samim tim i na tačnost procene koeficijenta inharmoničnosti. Takođe, signal sa Sl. 2 je u velikoj meri „čist“, jer su filtriranjem potisnuti ostali harmonici, kao i eventualni dodatni šumovi. Kada se analizira ceo ton, to nije slučaj, pa analiza postaje složenija. Ukoliko je cilj da se analiza automatizuje, odnosno bude primenljiva za celokupni registar čembala, postupak se dodatno usložnjava.

Na Sl. 3 prikazani su rezultati dobijeni procenom spektra pomoću dve pomenute metode i za različite parametre za slučaj signala sa Sl. 2. Rezultati sa Sl. 3 koji su označeni slovnim indeksom a) dobijeni su na osnovu obrade signala sa zadatom frekvencijom odabiranja od 48 kHz koja odgovara frekvenciji odabiranja kojom su snimani tonovi čembala. Dok se rezultati koji su označeni slovnim indeksom b) odnose na rezultate dobijene nakon decimacije sa faktorom 8.

U prvom redu Sl. 3, odnosno na graficima a1) i b1) prikazani su amplitudski spektri, rezultati dobijeni DFT analizom. Crnom i sivom bojom je naznačena spektralna komponenta računata na segmentu trajanja 0.1 s sa preklapanjem od 75% i dopunjavanjem nulama do rezolucije od 0.1 Hz. Dok je crvenom bojom rezultat dobijen na osnovu DFT-a sračunatog za prvu sekundu snimljenog signala. Korišćena je *Hamming*-

ova prozorska funkcija. Prikazani su rezultati za nekoliko segmenta signala. Može se uočiti da se pozicije maksimuma, odnosno procene frekvencije harmonika, menjaju od prozora do prozora. Poređenjem Sl. 3 a1) i b1) može se uočiti da decimacija nema značajan uticaj na estimaciju frekvencije.

U drugom redu Sl. 3, na graficima a2), a3), b2) i b3), prikazani su polovi AR modela. Polovi su računati tako da modeluju signal četvrtog harmonika nakon Hilbertove transformacije, pa su dobijeni modeli sa kompleksnim koeficijentima. S obzirom da se analizira samo jedan harmonik, red modela je 1. Na grafiku a2) prikazani su polovi AR modela računati sa frekvencijom odabiranja  $f_s$ , dok su na grafiku b2) prikazani polovi AR modela računati na decimiranom signalu. Kružićem je prikazana pozicija na krugu koja odgovara nominalnoj frekvenciji četvrtog harmonika od 220 Hz. Frekvencija od 220 Hz ne predstavlja „tačnu“ vrednost frekvencije četvrtog harmonika, jer tačna frekvencija odsviranog tona nije unapred poznata, već je poznata njena vrednost do nivoa raspona tona. S obzirom na izraz (1) bilo bi očekivano da je dobijena frekvencija zapravo nešto viša od 220 Hz, ali na osnovu prikazanog rezultata može se zaključiti da je osnovna frekvencija odsviranog tona nešto niža od nominalne vrednosti od 55 Hz. Na a3) i b3) prikazani su zumirani detalji sa grafika a2) i b2), respektivno. Crni i sivi krstići predstavljaju polove računate po prozorima dužine 0.1 s, gde se na osnovu slike može uočiti da se pozicija pola menja od prozora do prozora. Crvenom bojom je prikazan pol modela koji je dobijen na osnovu 1 s signala. Ukoliko se izvrši poređenje grafika a3) i b3) može se uočiti da u ovom slučaju decimacija doprinosi boljoj proceni utoliko što je pol pozicioniran bliže jediničnom krugu, pa je njegov efekat izraženiji. Intuitivno je jasno da decimacija doprinosi boljim rezultatima jer, se jedne strane, odbačen deo spektra koji nije

od interesa, a s druge strane je relativan odnos frekvencije od interesa i frekvencije odabiranja manji, što se vidi i kroz razliku položaja nominalne frekvencije na jediničnom krugu, grafici a2) i b2). U poslednjem redu na graficima a4) i b4), prikazani su spektri dobijeni na osnovu AR modela koji verifikuju zaključak o uticaju decimacije na bolju procenu frekvencije [14].

#### IV. EKSPERIMENTALNI REZULTATI I DISKUSIJA

Na osnovu analize izložene u prethodnom delu, za dalju diskusiju usvojene su dve metode za procenu spektra, dok je za samu procenu koeficijenta inharmoničnosti korišćen iterativni PFD algoritam [10]. Korišćene su metode DFT analiza i AR modelovanje. Obe metode za procenu spektra su definisane tako da imaju odliku automatske obrade, odnosno da daju vrednosti koeficijenta inharmoničnosti za celokupni registar čembala. Algoritam je testiran na bazi formiranoj za potrebe ovog rada. U bazi se nalaze tonovi snimljeni na 5 različitih registara jednog čembala.

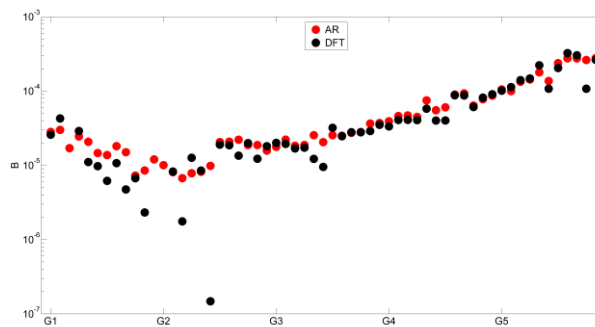
##### A. Baza snimaka

Za potrebe ovog rada napravljena je baza koju čini 310 snimaka tonova odsviranih u 5 različitih registra čembala. Tonovi čembala su snimljeni u prostorijama Fakulteta muzičke umetnosti u Beogradu. U analizi uticaja tačnosti estimacije frekvencija harmonika na procenu koeficijenta inharmoničnosti korišćeni su tonovi iz prva dva registra. Izabrani su tonovi iz ova dva registra, jer njihova upotreba služi za pojačanje izvođačke dinamike, a ne za menjanje boje instrumenta.

##### B. Analiza uticaja različitih parametara na tačnost procene koeficijenta inharmoničnosti

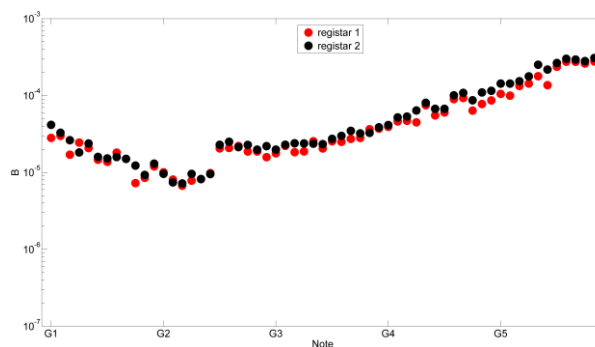
Uticaj tačnosti različitih metoda za estimaciju spektra na procenu koeficijenta inharmoničnosti najbolje se može uočiti poređenjem trenda krive koeficijenta  $B$ . Na Sl. 4 prikazan je grafik koeficijenta inharmoničnosti  $B$  za celokupan registar čembala procenjen na osnovu dve metode za procenu spektra i to DFT-om i AR modelovanjem. Za tonove iz srednjeg i višeg registra čembala koeficijenti  $B$  procenjeni na osnovu dve metode se ne razlikuju značajno, dok za slučaj tonova iz nižeg registra dolazi do razilaženja u vrednostima. Razlog za lošiju procenu koeficijenta  $B$  na osnovu metode DFT jeste u proceni osnovne frekvencije tona. Sam iterativni algoritam iako robustan u slučajevima kada u spektru signala nedostaje neki harmonik ili kada je frekvencija pojedinog parcijala pogrešno procenjena, usled pogrešno procenjene osnovne frekvencije tona kao krajnji rezultat daje loše procenju vrednost koeficijenta  $B$ . Do pogrešno procenjene osnovne frekvencije tona DFT metodom dolazi zbog samog spektra signala, gde je na nižim frekvencijama često osnovni harmonik potisnut u odnosu na ostale harmonike tog tona. Takođe, pri proceni spektra DFT analizom rezolucija po frekvenijskoj osi je ista za ceo opseg koji se analizira, dok se frekvencije osnovnih tonova menjaju logaritamski, te tačnost procene u muzičkom smislu nije ista u celom opsegu. Metoda AR modelovanjem dala je tačnije rezultate za koeficijent  $B$ . Osnovna frekvencija

tona je procenjivana metodom najmanjih kvadrata (*Least square*) nad nekoliko parcijala tona počev od drugog. Time je premošćen problem korišćenje osnovnog parcijala. U najvećem broju snimljenih tonova čembala harmonici nižeg reda su dovoljno jasni, odnosno procena njihovih frekvencija je relativno tačna. Broj harmonika za procenu osnovne frekvencije, kao i broj harmonika na osnovu kojih je izvršena procena parametra  $B$  variran je u zavisnosti od visine tona, jer je za tonove nižih frekvencija broj jasno prisutnih harmonika u snimljenom tonu veći.



Sl. 4. Koeficijent inharmoničnosti čembala dobijen metodama: AR i DFT.

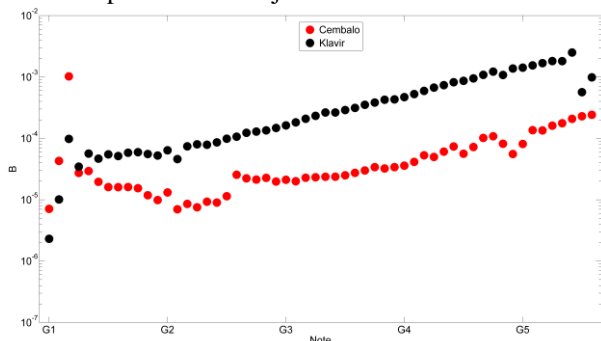
Na Sl. 5 prikazana je kriva koeficijenta inharmoničnosti  $B$  za tonove iz prvog i drugog registra čembala. Koeficijenti  $B$  su procenjeni na osnovu estimiranih frekvencija harmonika pomoću AR modela. Trend krive koji je očuvan i za tonove iz prvog i drugog registra čembala verifikuje tačnost estimacije frekvencija harmonika koja je dobijena na osnovu AR modela.



Sl. 5. Koeficijenti inharmoničnosti čembala iz prvog i drugog registra.

S obzirom da se tonovi čembala u spektru odlikuju većim brojem harmonika u odnosu na tonove klavira istih frekvencija, izvršeno je poređenje koeficijenta  $B$  za tonove čembala u odnosu na koeficijent  $B$  za tonove klavira posmatrane na istom frekvenijskom opsegu. Na Sl. 6 prikazani su koeficijenti inharmoničnosti tonova čembala iz prvog registra i polukoncertnog klavira Steinway D7. Na osnovu slike se može zaključiti da postoji sličan trend krive koeficijentata  $B$  za oba instrumenta, iako se ova dva instrumenta razlikuju na osnovu načina nastanka tona. Kod tonova čembala, vrednosti koeficijenta  $B$  su manje za red veličine. Pored velikog broja harmonika tonova čembala u odnosu na klavir, postoji još jedna razlika između ova dva instrumenta koja doprinosi velikim odstupanjima koeficijenta inharmoničnosti. Kod klavira u većinskom frekvenijskom

opsegu jednom tonu odgovaraju tri žice, dok čembalo za svaki ton ima jednu. S obzirom na postojanje tri žice kod klavira, koje se vide u spektru signala, pri proceni spektra uvek se donosi odluka na osnovu jedne spektralne komponente na osnovu čije frekvencije se ulazi u dalju procenu koeficijenta inharmoničnosti. Tim fenomenom se otvaraju nove potrebe u smislu modelovanja uticaja tri žice na procenu spektra, a samim tim i procenu koeficijenta inharmoničnosti.



Sl. 6. Koeficijent inharmoničnosti  $B$  tonova čembala i klavira.

## V. ZAKLJUČAK

U ovom radu razmatran je uticaj procene frekvencija harmonika na tačnost procene koeficijenta inharmoničnosti. Formirana je baza snimaka koja se sastoji od preko 300 tonova odsviranih na čembalu. Čembalo je izabrano kao instrument za analizu, jer se spektri njegovih tonova odlikuju velikim brojem harmonika. Prisutnost velikog broja parcijala u spektru tona je značajna, zbog činjenice da su fizičke pojave koje opisuju fenomen inharmoničnosti uočljivije na višim parcijalima tona. Sem toga sa prisustvom većeg broja harmonika iterativni algoritam bolje procenjuje koeficijent inharmoničnosti. Izvršena je uporedna analiza dve metode za procenu spektra signala i to procena spektra na osnovu AR modela i metodom DFT. Pokazano je da je procenu frekvencije harmonika bolje vršiti na osnovu AR modela, nego metodom DFT, jer trend krive procenjenih koeficijenata inharmoničnosti u tom slučaju ima konzistentnije ponašanje. Poređenjem, koeficijenta inharmoničnosti tonova klavira i čembala uočeno je da na njihove vrednosti između ostalog utiče i broj žica koji je potreban za sviranje jednog tona. Dalja unapređenja postupka za procenu koeficijenta inharmoničnosti biće predlog rešenja za modelovanje uticaja prisustva više žica na koeficijent inharmoničnosti. Izvedeni zaključci su korisni za buduća istraživanja gde će se na ovaj način procenjeni koeficijenti inharmoničnosti koristiti u sintezi i modelovanju tonova žičanih instrumenata koje će predstavljati osnovu subjektivnih testova. Cilj subjektivnih testova bi bio upotpunjavanje istraživanja o uticaju inharmoničnosti žičanih instrumenata na ljudsku percepciju tonova.

## ZAHVALNICA

Ovaj rad je realizovan u okviru projekta TR36026 koji finansira Ministarstvo prosvete, nauke i tehnološkog razvoja Republike Srbije. Zahvaljujemo se kolegama sa Fakulteta muzičke umetnosti na ustupljenim muzičkim instrumentima i formiranju baze snimaka.

## LITERATURA

- [1] J. W. Strutt, *The Theory of Sound*, Macmillan 1894.
- [2] R.W. Young, "Inharmonicity of Plain Wire Piano Strings", *J. Acoust. Soc. Am.*, vol. 24, no. 3, pp. 267–273, 1952.
- [3] H. Fletcher, E. D. Blackham, and R. Stratton, "Quality of piano tones", *J. Acoust. Soc. Am.*, vol. 34, no. 6, pp. 749–761, 1962.
- [4] H. Fletcher, "Normal Vibration Frequencies of a Stiff Piano String", *J. Acoust. Soc. Am.*, vol. 36, no. 1, pp. 203–209, 1964.
- [5] A.S. Galebo, and A. Askenfelt, A. "Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano," *IEEE Trans. Speech Audio Process*, vol. 7, no. 2, pp. 197–203, 1999.
- [6] A. Askenfelt and A.S. Galebo, "Study of the spectral inharmonicity of musical sound," *Acoust. Phys.* vol. 46, no. 2, pp. 121–132, 2000.
- [7] A. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *IEEE Trans. Speech Audio Process*, vol. 11, no. 6, pp. 184–194, 2003.
- [8] J. Rauhala, H.M. Lehtonen, V. Välimäki, "Fast automatic inharmonicity estimation algorithm" *J. Acoust. Soc. Am.*, vol. 121, EL184, doi: 10.1121/1.2719043, 2007.
- [9] T. Miljković, M. Bjelić, D. Šumarac Pavlović, J. Čertić, "Analiza algoritma za procenu koeficijenta inharmoničnosti različitih klavira", Zbornik radova 64. Konferencije ETRAN, pp. AK1.2.1 - AK1.2.6, Beograd, Sep, 2020.
- [10] D. Despić, *Muzički instrumenti*, Univerzitet umetnosti Beograd, 1979.
- [11] A. Bond, *A Guide to the Harpsichord*, Portland, USA, Amadeus Press, 1997.
- [12] S. Dixonb, M. Mauch, and Dan Tidhar "Estimation of harpsichord inharmonicity and temperament from musical recordings", *J. Acoust. Soc. Am.*, vol. 131, no. 1, Pt. 2, pp. 878–887, Jan 2012.
- [13] H.M. Lehtonen "Analysis of Piano Tones using an Inharmonic Inverse Comb Filter", Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08), Espoo, Finland, September 1-4, 2008.
- [14] P.A.A. Esquef, M. Karjalainen, and V. Välimäki, "Frequency-Zooming ARMA Modeling for Analysis of Noisy String Instrument Tones", *EURASIP J. Adv. Signal Process*, pp. 953-967, 2003.
- [15] S. Haykin, *Adaptive Filter Theory*, 4th ed. New Jersey, USA, Prentice Hall, 2002.
- [16] J. Čertić, D. Šumarac Pavlović, I. Salom, "Softverski paket za obradu i analizu audio signala", *TELFOR 2010*, pp. 1269-1272, Beograd, Nov., 2010.

## ABSTRACT

Inharmonicity is a phenomenon that occurs in musical instruments that are theoretically considered harmonic, and it is the deviation of the frequencies of the tone partials from the integer multiple of the fundamental frequency. For stringed musical instruments, the inharmonicity coefficient is defined in the literature as a measure of deviation from the ideal case. In the previous research work, it was shown that an automatic algorithm for estimating the inharmonicity coefficient of piano tones, for individual tones from the register doesn't perform a good estimation of the inharmonicity coefficient. In this paper, the accuracy impact of harmonics frequency estimation on the estimation of inharmonicity coefficient is considered. Comparative analysis of two methods for signal spectrum estimation was performed, namely spectrum estimation based on AR model and DFT method. Testing of the proposed methods for spectrum estimation was performed on real harpsichord tones. It was found that based on the trend of the inharmonicity coefficient calculated using both spectrum estimation methods, the AR model method is superior and performs a more accurate estimation of the inharmonicity coefficient over the whole range of tones of interest.

## Impact of harmonic frequency estimation on the estimation of harpsichord inharmonicity coefficient

Tatjana Miljković, Jovana Damnjanović, Jelena Čertić,  
Dragana Šumarac Pavlović

# Vreme reverberacije energetskog transformatora

Miloš Bjelić, Bogdan Brković, Mileta Žarković, Tatjana Miljković

**Apstrakt**— Energetski transformator je uz generator jedan od najvažnijih elemenata elektroenergetskih sistema. Pouzdanost rada energetskog transformatora direktno je povezana sa njegovim stanjem. Analiza rezultata ispitivanja svodi se na vizuelno poređenje frekvencijskih odziva periodičnih merenja transformatora, dobijenih pomoću specijalizovanih uređaja. Nedostatak ove metode je oslanjanje na inženjersko iskustvo. Objektivniji način detekcije kvarova je izračunavanje matematičkih indikatora zasnovanih na analizi frekvencijskog odziva. Ovi parametri često nisu u stanju da nedvosmisleno ukažu na stepen kvara. Ovaj rad predstavlja novi pristup za otkrivanje unutrašnjih kratkih spojeva transformatora. Predlaže se novi indikator zasnovan na vremenu reverberacije, koje se uobičajno koristi u akustici. Predloženi parametar zasnovan je na proračunu u vremenskom domenu i omogućava pouzdanu detekciju kratkih spojeva transformatora i tačnu procenu njihovog stepena.

**Ključne reči** — Detekcija kvara, energetski transformator, indikator, obrada signala, vreme reverberacije.

## I. UVOD

Energetski transformator (u daljem tekstu ET) je jedan od najvažnijih i najskupljih elemenata u svakom delu i na svakom nivou napona elektroenergetskog sistema. Pored toga, ET je izuzetno složena komponenta na čiji rad utiču termički, mehanički, električni i hemijski procesi. Neki od ključnih delova transformatora koji su od velike važnosti za njegov pouzdan rad su: magnetni krug, ulje, namotaji (stanje geometrije i stanje električnih veza) i čvrsta izolacija namotaja. Deformacije namotaja obično se otkrivaju metodom analize frekvencijskog odziva (FRA) i *sweep* analizom frekvencijskog odziva (SFRA). Standardna SFRA metoda uključuje poređenje frekvencijskih odgovora uzastopnih periodično sprovedenih testova. Ako se dva uzastopna frekvencijska odziva ne razlikuju, tada se uzdužni parametri  $R$ ,  $L$  i  $C$  nisu značajno promenili, što ukazuje da u međuvremenu nije došlo do deformacija namotaja.

U radovima i standardima [1-14] predstavljene su FRA i SFRA metoda i njihov značaj u praćenju i dijagnostici stanja ET. Referenca [1] predstavlja teorijske osnove FRA metode.

Miloš Bjelić – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: bjelic@etf.rs).

Bogdan Brković – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: brkovic@etf.rs).

Mileta Žarković – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: mileta@etf.rs).

Tatjana Miljković – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: tm@etf.rs).

Razvoj akustičkih metoda pomoću vibracija predstavljen je u [2], gde je pokušano da se rezultati FRA metode dopune i povežu sa postojećim IEEE standardima [3], IEC standardom [4] i iskustvom radne grupe CIGRE [5,6]. Dobra interpretacija rezultata FRA metode predstavljena je u [7], gde je pokušano da se odrede pojedinačni indikatori i njihove približne granične vrednosti. Metoda koja grupiše rezultate simulacije FRA metode i povezuje ih sa određenom vrstom kvara predstavljena je u [8]. Eksperimenti na ET u praznom hodu i uslovima kratkog spoja, uz primenu FRA metode predstavljeni su u [9,10]. Detekcija radijalnih deformacija namota pomoću impulsnog napona groma prisutna je u [11]. Uobičajeni statistički indikatori kvarova dati su u [12-14]. U članku [15] je predstavljena nova metodologija za dijagnostiku kvara ET, zasnovana na rezultatima SFRA. Ova metoda upoređena je sa metodom konačnih elemenata pri otkrivanju unutrašnjih i spoljnih kvarova kratkog spoja [16].

Tumačenje rezultata SFRA obično se vrši vizualizacijom slika generisanih od strane specijalizovanih uređaja za takva ispitivanja. Razvoj novih metoda motivisan je potrebom za postizanjem objektivne interpretacije rezultata. Ovaj rad prikazuje neke najčešće korišćene identifikatore koji se primenjuju na dobijene SFRA testove. Takvi identifikatori trebali bi jasnije da kvantifikuju rezultata SFRA metode i omogućiti lakše i nedvosmisleno otkrivanje oštećenja namotaja ET i kvarova magnetnog kruga.

U ovom radu predložena je detekcija kratkih spojeva i određivanje njihove težine vrši se na osnovu vremena reverberacije [17]. Ovaj parametar se najčešće koristi za analizu akustike prostorija. Glavna hipoteza istraživanja u ovom radu je da se broj kratko spojenih namotaja može odrediti na osnovu vrednosti vremena reverberacije. Realizovani eksperimenti obuhvatili su testiranje sa kratkim spojem različitih delova namotaja jednog ET. Merenja su sprovoda pomoću audio kartice. Snimljeni signali se koriste za izračunavanje statističkih parametara koji se obično koriste u literaturi za detekciju kvarova ET. Izvršeno je poređenje između vremena reverberacije i postojećih parametara u pogledu osetljivosti na dužinu kratkih spojeva.

Rad je organizovan na sledeći način. U odeljku 2 dat je pregled najznačajnijih metoda i statističkih parametara koji se koriste za ET dijagnostiku. U ovom poglavlju objašnjen je postupak izračunavanja vremena reverberacije. Eksperimentalna postavka opisana je u poglavlju 3. Rezultati dobijeni korišćenjem zvučne kartice prikazani su u poglavlju 4. Zaključci istraživanja dati su u poglavlju 5.

## II. METODOLOGIJA

### A. FRA i SFRA analiza

U svrhu otkrivanja mehaničkih kvarova ET, mogu se koristiti različiti ulazni signali:

- odskočni signal (FRA analiza),
- *sweep* signal (SFRA analiza).

Kada se ET terminal pobudi sa nekim od ulaznih signala, meri se odziv na izlazu ET. U ovim merenjima, namotaj ET može se posmatrati kao dvopolna mreža koja sadrži otpornike, kondenzatore i kalemove, gde je otpor obično zanemarljiv. Ako se dva upoređena frekventijska odziva (periodična testiranja transformatora) ne razlikuju, to znači da se parametri nisu značajno promenili, tj. nije bilo deformacija namotaja od prethodnog merenja. U slučaju mehaničkih oštećenja namotaja, vrednosti ekvivalentnih kondenzatora i kalemova se menjaju, pa se odzivi dva periodična ispitivanja se razlikuju, što ukazuje na kvar. Frekventijski odziv  $H(j\omega)$  se formira kao odnos Furijeove transformacije izlaznog signala  $V_{output}$  i Furijeove transformacije ulaznog signala  $V_{input}$ :

$$H(j\omega) = \frac{V_{output}(j\omega)}{V_{input}(j\omega)}. \quad (1)$$

Za analizu se najčešće koristi samo amplitudska karakteristika frekventijskog odziva. Ordinata predstavlja amplitudu u decibelima, a apscisa predstavlja frekvenciju na logaritamskoj skali. Ovaj postupak prelaska u frekventijski domen u velikoj meri olakšava analizu i pruža nezavisnost od talasnog oblika pobude. Međutim, to povećava računsku i vremensku kompleksnost samog potapka. Poređenjem izračunatih frekventijskih odziva u uzastopnim ispitivanjima mogu se otkriti različite promene u geometriji namotaja i promene u magnetnom kolu ET. Tumačenje rezultata SFRA oslanja se na ličnu stručnost i može dovesti do različitih zaključaka. Štaviše, konvencionalni SFRA *fingerprint* vrlo je neprecizan u otkrivanju početnih i niskih nivoa mehaničkih kvarova [13]. Da bi se rešili ovi problemi, predloženi su različiti statistički indikatori [6, 12, 13, 14]. Neki od ovih indikatora definisani su u tabeli I. Analiza osetljivosti indikatora može se izvršiti izračunavanjem njihovih vrednosti za različite intenzitete kvara.

TABELA I  
KORIŠĆENI STATISTIČKI INDIKATORI IZ LITERATURE

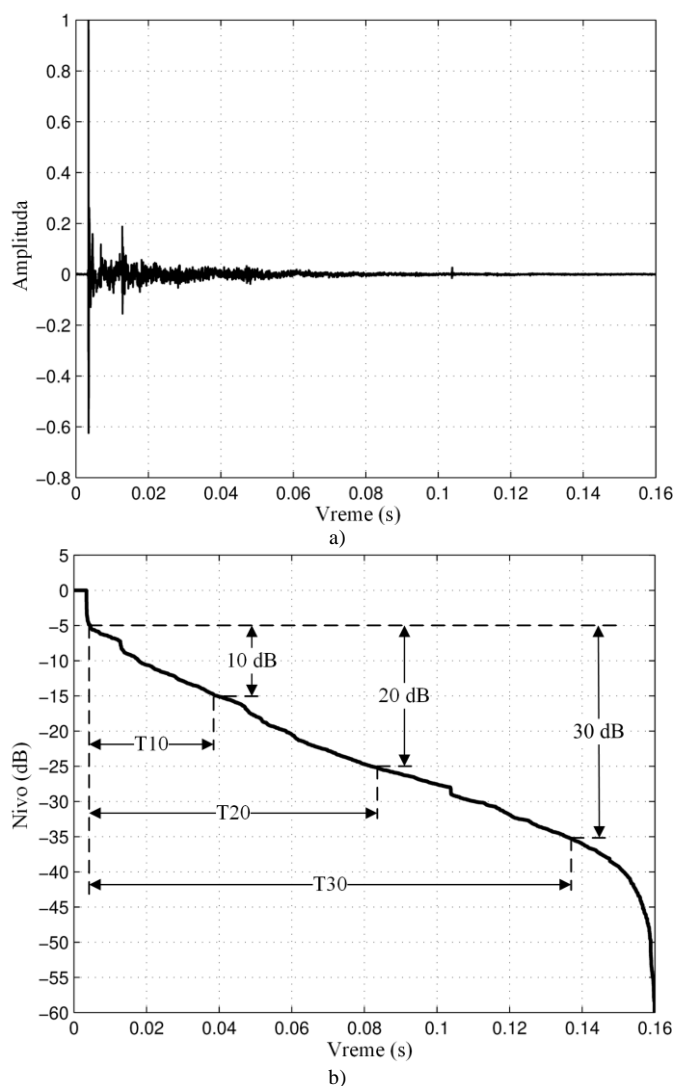
Indikator	Opis
Jaccard Distance (JD)	$\frac{\sum_{i=1}^N (x_i - y_i)^2}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$
Absolute sum of logarithmic error (ASLE)	$\frac{\sum_{i=1}^N  20 \log_{10}(y_i) - 20 \log_{10}(x_i) }{N}$
Absolute average difference (DABS)	$\frac{\sum_{i=1}^N  y_i - x_i }{N}$

( $N$  – ukupan broj podataka;  $x_i$  i  $y_i$  – ulazni, odnosno izlazni podaci;  $\bar{x}$  i  $\bar{y}$  – srednje vrednosti)

### B. Vreme reverberacije

Linearni vremenski invarijantni sistem (LTI) može se jednoznačno opisati frekventijskim odzivom ili impulsnim odzivom u vremenskom domenu [18]. Frekventijski odziv sistema se dobija primenom Furijeove transformacije na impulsnim odzivom u vremenskom domenu. Impulsnim odzivom se može dobiti pobuđivanjem sistema Dirakovim impulsom. Međutim, generisanje Dirakovog impulsa u praksi nije moguće, pa se umesto toga moguće koriste jedinični impuls. Češće se koriste indirektno metode kao što su *sweep* signal (*chirp*) i MLS sekvenca [19]. Impulsnim odzivom se obično koristi u oblastima obrade signala, automatskog upravljanja i akustike.

Impulsnim odzivom prostorije meri se pomoću zvučnika i mikrofona. Soba se može posmatrati kao LTI sistem, gde signal mikrofona predstavlja izlaz sistema, a signal koji se emituje iz zvučnika predstavlja ulaz sistema. Impulsnim odzivom prostorije predstavlja njenu „ličnu kartu“, jer omogućava određivanje gotovo svih akustičkih parametara prostorije. Primer impulsnog odziva prostorije prikazan je na slici 1 a.



Sl. 1. Određivanje vremena reverberacije za prostoriju: a) Impulsnim odzivom, b) Kriva opadanja i princip određivanja T10, T20 i T30



Struktura impulsnog odziva je vrlo složena i njegov talasni oblik ne pruža mnogo informacija o prostoriji. Zvučno polje obično se može analizirati jednostavnim posmatranjem globalnog oblika impulsnog odziva. Zbog toga se impulсни odziv najčešće karakteriše tzv. krivom opadanja zvuka (reverberaciona kriva) [17]. Da bi se smanjio uticaj fluktuacija krive, koristi se tzv. integrisani impulсни odziv, poznat i kao *Schroeder*-ova kriva [17]. Ovaj pristup zasnovan je na uvođenju krive opadanja  $L_R(t)$ , definisane kao:

$$L_R(t) = \int_0^{\infty} h^2(t) dt = \int_0^{\infty} h^2(t) dt - \int_0^t h^2(t) dt, \quad (2)$$

gde je  $h(t)$  impulсни odziv. Na slici 1 b) prikazana je kriva opadanja impulsnog odziva, prikazanog na slici 1 a.

Najčešće korišćeni parametar u akustici prostoriji je vreme reverberacija (u oznaci T60). Vreme reverberacije je definisano kao vreme potrebno da se kriva reverberacije smanji za 60 dB od početne vrednosti, što odgovara smanjenju od milion puta na linearnoj skali. Može se smatrati da nakon vremena reverberacije u sobi nema zvučne energije. Dinamiku od 60 dB je teško postići u praksi, zbog čega se T60 određuje indirektno pomoću parametara T10, T20 ili T30. Tumačenje ovih parametara ilustrovano je na slici 1 b [17]. Vreme reverberacije T60 izračunava se množenjem T30 sa dva, T20 sa tri ili T10 sa šest.

U ovom radu se ispituje mogućnost korišćenja vremena reverberacije kao indikatora za određivanje stanja ET. Za razliku od akustike prostorije, gde su pobuda i impulсни zvučni signali (zvučni pritisak), pobuda i impulсни odziv ET-a su električni signali (napon odnosno struja). Iako su izmerene veličine različite prirode, princip izračunavanja vremena odjeka važi u oba fizička domena. U opštijem smislu, vreme reverberacije može se tumačiti kao vreme smirivanja impulsnog odziva.

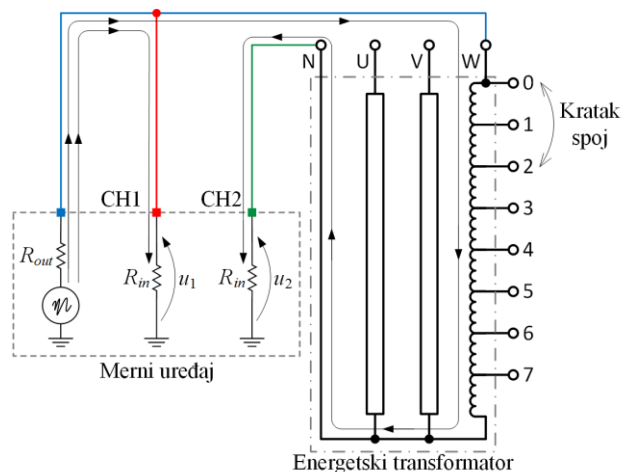
### III. POSTAVKA EKSPERIMENTATA

Ekperimentalna postavka je šematski prikazana na slici 2. ET koji se koristi u postavci je proizveden u Laboratoriji za visoki napon Elektrotehničkog fakulteta, Univerziteta u Beogradu. Parametri transformatora dati su u tabeli II.

TABELA II  
PARAMETRI KORIŠĆENOG ENERGETSKOG TRANSFORMATORA

Parametar	Vrednost
Nazivni napon	10/0.4 kV/kV
Nazivna struja	23/557 A/A
Frekvencija	50 Hz
Tip veze	Yyn0
Broj faza	3

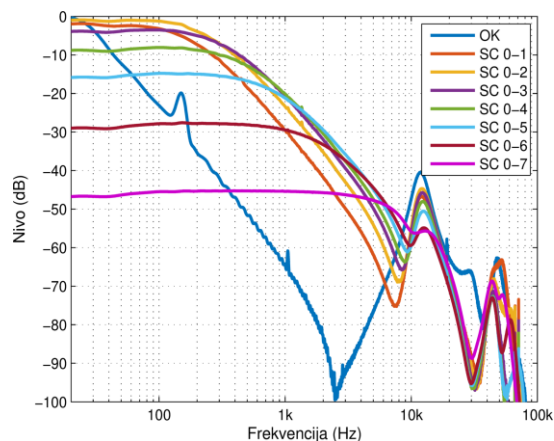
Ovaj ET ima sedam dvostrukih namotaja za sa pristupačnim kontaktima na namotaju faze W, označenim sa 0 do 7 na slici 2. Takva konstrukcija omogućava kratki spoj različitih delova namotaja. U zavisnosti od veza između ovih kontakata, stepen kvara može se menjati, pri čemu kratki spojevi između susednih kontakata odgovaraju najmanjem, a kratki spoj između kontakata 0 i 7 odgovaraju najvećem stepenu kvara.



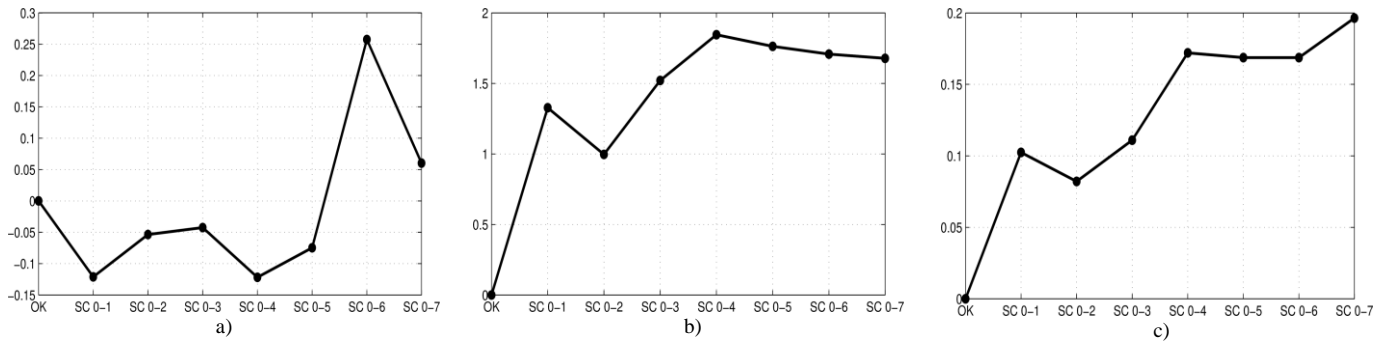
Sl. 2. Šematski prikaz eksperimentalne postavke

Fazni W namotaj se pobuđuje pomoću izvora promenljive frekvencije (plava linija). Ulazni napon ( $u_1$ ) meri se između fazne W i zemlje (Kanal 1 - crvena linija). Napon koji odgovara odzivu ( $u_2$ ) se meri između neutralne tačke transformatora i zemlje (Kanal 2 - zelena linija). Na osnovu ova dva merenja može se dobiti frekvencijski odziv namotaja. Putanje struja u kolu su označene crnim linijama sa strelicama koje pokazuju smer struje.

Ukupno je izvršeno osam merenja pomoću svakog, jedno na zdravom ET (označeno sa OK) i sedam na neispravnom ET (označeno kao SC – *Short Circuit*). Kvarovi su nastajali formiranjem kratkog spoja između kontakata 0 i kontakata 1 do 7 (videti sliku 2). Na taj način formirani su kvarovi različite težine. Pobudni signal je bio logaritamski *sweep* signal. Kao merni uređaj korišćena je audio kartica Steinberg UR22 [21]. Ovaj uređaj predstavlja eksternu zvučnu karticu koja se obično koristi za snimanje audio signala. Cena uređaja je oko 150\$, što je nekoliko redova veličine niže u poređenju sa profesionalnim SFRA analizatorom. Vreme potrebno za jedno merenje je oko 10 s. Maksimalna frekvencija odabiranja zvučne kartice je 192 kHz. Na osnovu toga ovaj uređaj omogućava merenje frekvencijskog odziva u opsegu od 20 Hz do oko 100 kHz. Merenje je realizovano korišćenjem aplikacije, realizovane u programskom paketu Matlab.



Sl. 3. Frekvencijski odziv dobijen pomoću audio kartice



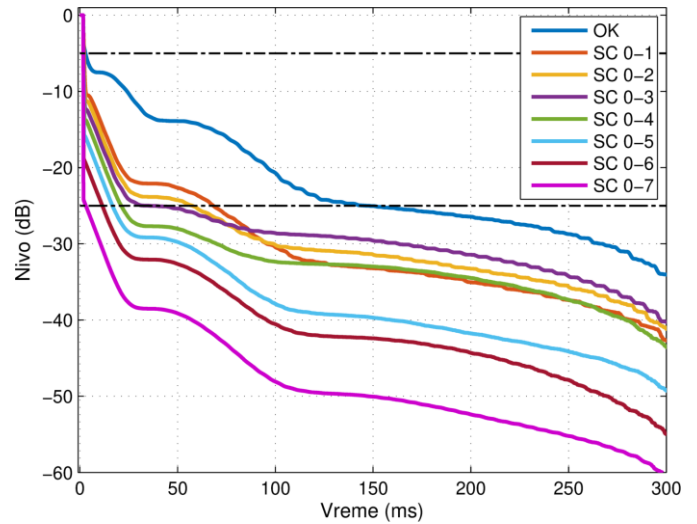
Sl. 4. Statistički indikatori (tabela I): a) Jaccard Distance (JD); b) Absolute sum of logarithmic error (ASLE); c) Absolute average difference (DABS)

#### IV. REZULTATI I DISKUSIJA

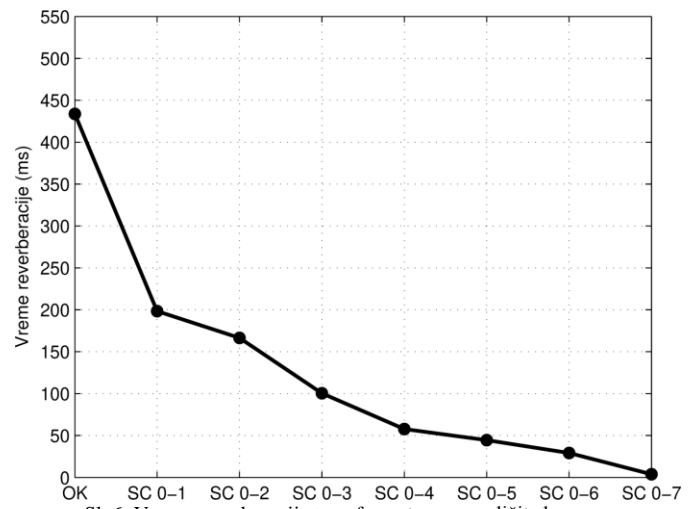
Frekvencijski odziv ET-a sa zdravim namotajem (plava boja) i sa kratkim spojevima (druge boje) prikazani su na slici 3. Povećani nivo u frekvencijskom odzivu primećuje se kada su prisutni kvarovi na transformatoru, što se može pripisati promenjenim vrednostima parametara namotaja (induktivnosti i kapaciteta). ET se može tretirati kao složena *RLC* mreža, sa kratkim spojevima koji utiču na ekvivalentnu impedansu namotaja [22]. Ovaj efekat je najizraženiji na niskim frekvencijama kada su dominantne induktivnosti namotaja. Smanjenje impedanse namotaja dovodi do povećanja napona na CH2 (slika 2), povećavajući tako frekvencijski odziv. Rezonantna frekvencija zadržava vrednost od oko 10 kHz bez obzira na stanje namotaja. Frekvencijski odzivi sa slike 3 dobijeni su korišćenjem izmerenih signala iz kanala 1 (pobuda) i kanala 2 (odziv). Na osnovu ovih signala izračunavaju se statistički pokazatelji definisani u tabeli I. Izračunate vrednosti su prikazane na slici 4 za sve analizirane slučajeve.

Vrednosti parametra JD date su na slici 4 a. Ovaj indikator je jednak 0 za zdrav ET i kreće se od -0.13 do 0.25 kada su prisutni kratki spojevi na namotaju ET. Na osnovu vrednosti ovog indikatora ne može se utvrditi da li na ET postoji kvar. Jasno je da se onda ne može utvrditi ni stepen kvara, odnosno dužina namotaja koja je u kratkom spoju. Indikatori ASLE i DABS jasno ukazuju na kratak spoj namotaja ET. To omogućava vrlo pouzdano otkrivanje kvara. Kao što je vidljivo sa slike 4 b i c, nijedan od dva indikatora se ne menja monotono, pa se stepen kvara ne može pouzdano utvrditi. Ipak, ova dva indikatora imaju očiglednu prednost u poređenju sa JD indikatorom prikazanim na slici 4 a. Oni pružaju najpouzdaniju detekciju kvara i mogu u određenoj meri ukazati na ozbiljnost kvara.

Impulsni odziv ET (tj. namotaja) izračunat je za svaki eksperiment na osnovu snimljenih ulaznih i izlaznih signala. Krive opadanja nivoa signala, koje odgovaraju svim analiziranim scenarijima, dobijaju se na osnovu izraza (3), koristeći izračunate impulsne odzive. Krive opadanja integrisanog impulsnog odziva prikazane su na slici 5. Sve krive pokazuju slično ponašanje tokom prvih 15 ms. Nakon ovog intervala, kriva koja odgovara ispravnom transformatoru sporije opada u odnosu na krive koje odgovaraju ET sa kvarom. Kako se ozbiljnost kvara, tj. broj kratko spojenih navojaka, povećava tako se povećava i nagib krivih opadanja.



Sl. 5. Krive opadanja impulsnih odziva transformatora



Sl. 6. Vreme reverberacije transformatora za različite kvarove

Vreme reverberacije ( $T_{60}$ ) izračunato je prema slici 2 b. Za izračunavanje  $T_{60}$  korišćen je dinamički opseg od 20 dB, tj. parametar  $T_{20}$ . Vrednosti vremena reverberacije dobijene za ispravan ET i analizirane kvarove prikazane su na slici 6. Vrednost  $T_{60}$  za zdrav ET je 434 ms. Kada je prisutan najmanji kvar (slučaj SC 0-1),  $T_{60}$  se smanjuje na 198 ms, što je smanjenje od 55% u poređenju sa ispravnim stanjem. Vrednosti  $T_{60}$  se kreću od 4 ms za najteži kvar (SC 0-7) do 166.5 ms za slučaj kvara SC 0-2. Značajno smanjenje  $T_{60}$  prilikom postojanja kvara omogućava vrlo pouzdano

otkrivanje kvara. Dalje, T60 se monotono smanjuje kako se povećava stepen kratkog spoja. Ovo daje predloženom parametru izrazitu prednost u odnosu na postojeće parametre zasnovane na SFRA u pogledu njegove sposobnosti da jasno ukaže na ozbiljnost kvara.

Upotreba zvučne kartice kao mernog uređaja umesto specijalizovanih SFRA analizatora je povoljnija pre svega u pogledu cene. Međutim, njena prednost nad SFR analizatorom je i u pogledu individualnog trajanja merenja. Trajanje merenja je oko 5 sekundi prilikom korišćenja zvučne kartice u poređenju sa 2 minuta prilikom korišćenja SFR analizatora. Sprovođenjem SFR analize i izračunavanjem predloženog indikatora (vreme reverberacije - T60) pomoću zvučne kartice može se izvršiti pouzdano otkrivanje kvarova i njihova kvantifikacija. Rezultati prikazani u ovom radu jasno ukazuju da primena predložene procedure pomoću jednostavnog audio uređaja obezbeđuje nedvosmisleni indikaciju kvarova energetskog transformatora.

#### V. ZAKLJUČAK

Vreme reverberacije (T60) predloženo je u ovom radu kao novi parametar za detekciju kvara energetskog transformatora. Iako se pretežno koristi u akustici prostorija, vreme reverberacije (T60) u opštijem smislu predstavlja vreme potrebno da se impulsni odziv sistema smiri. Glavna prednost ovog parametra je što njegovo izračunavanje zahteva samo talasne oblike u vremenskom domenu ulaznog i izlaznog signala. Profesionalni SFRA analizatori vrše izračunavanja u frekvencijskom domenu, pa je predloženi postupak proračuna jednostavniji. Realizovani eksperimenti pokazuju sposobnost novog parametra da ukazuje na prisustvo kvara unutrašnjeg namotaja i njegovu težinu, tj. broj kratko spojenih navojaka. Pored toga, predloženi parametar ima prednost u odnosu na postojeće statističke pokazatelje zasnovane na SFRA, posebno u pogledu određivanja težine kvara. Novi indikator T60 jasno ukazuje na različite tipove grešaka bez složenih matematičkih operacija.

Za realizaciju eksperimenata korišćena je audio kartica. S obzirom na svoj jednostavniji dizajn, znatno nižu cenu i kraće vreme potrebno za merenje i obradu signala, zvučna kartica predstavlja vrlo korisnu alternativu profesionalnim SFRA analizatorima za preliminarnu dijagnostiku ET.

Budući rad biće usmeren na korišćenje vremena reverberacije za otkrivanje drugih vrsta unutrašnjih kvarova energetskog transformatora, poput deformacija namotaja. Takođe će se ispitati mogućnost korišćenja predloženog parametra za praćenje starenja energetskog transformatora.

#### ZAHVALNICA

Ovaj rad je realizovan u okviru projekta TR 36026 i projekta III 42009, koje finansira Ministarstvo prosvete, nauke i tehnološkog razvoja Republike Srbije.

#### LITERATURA

- [1] S. Ghani, Y. Thayoob, Y. Ghazali, M. Khiar, I. Chairul, Distribution system performance evaluation accounting for data uncertainty, *Procedia Engineering*, vol.68, pp 469-476, 2013.
- [2] E. Kornatowski, S. Banaszak, Diagnostics of a Transformer's Active Part with Complementary FRA and VM Measurements, *IEEE Transactions on Power Delivery*, vol. 29, no. 3, Berlin, pp 1398-1406, 2014.
- [3] IEEE PC57.149/D8, Guide for the Application and Interpretation of Frequency Response Analysis for Oil Immersed Transformers, 2012.
- [4] IEC 60076-18, Power transformers - Part 18: Measurement of frequency response, 2012.
- [5] CIGRE Working Group A2.26, "Mechanical-condition assessment of transformer windings using frequency response analysis", 2008.
- [6] CIGRE Working Group A2.53 "Advances in the interpretation of transformer Frequency Response Analysis (FRA)", Reference 812, SeETember, 2020.
- [7] J.R. Secue, E. Mombello, Sweep frequency response analysis (SFRA) for the assessment of winding displacements and deformation in power transformers, *Electric Power Systems Research*, vol. 78, pp. 1119-1128, 2008
- [8] A. Abbasi, M. Mahmoudi, Diagnosis and Clustering of Power Transformer Winding Fault Types by Cross-Correlation and Clustering Analysis of FRA Results, *IET Generation, Transmission & Distribution*, vol: 12, pp. 4301-4309, 2018.
- [9] A. Pandya, B. Parekh, Interpretation of Sweep Frequency Response Analysis (SFRA) trace for the open circuit and short circuit winding fault damages of the power transformer, *Electrical Power and Energy Systems*, vol. 62, pp. 890-896, 2014.
- [10] Akshay A. Pandya, B.R. Parekh, Interpretation of Sweep Frequency Response Analysis (SFRA) traces for the open circuit and short circuit winding fault damages of the power transformer, vol. 62, pp. 890-896, 2014.
- [11] Jian Sun, Qing Yang, PeiyuSu, Shilin Wu, Song Chen, Lewei He, Diagnosis of winding fault in three-winding transformer using lightning impulse voltage, *Electric Power Systems Research*, vol. 175, pp. 1-9, 2019.
- [12] Ni Jianqiang, Zhao Zhongyong, Tan Shan, Chen Yu, Yao Chenguo, Tang Chao, The actual measurement and analysis of transformer winding deformation fault degrees by FRA using mathematical indicators, *Electric Power Systems Research*, vol. 184, pp. 1-11, 2020.
- [13] Ali Reza Abbasi, Mohammad Reza Mahmoudi, Application of statistical control charts to discriminate transformer winding defects, *Electric Power Systems Research*, vol. 191, pp. 1-11, 2021.
- [14] Reza Khalili Senobari, JavadSadeh, Hossein Borsi, Frequency response analysis (FRA) of transformers as a tool for fault detection and location: A review, vol. 155, pp. 172-183, 2018.
- [15] Janneth R. Secue, Enrique E. Mombello, New SFRA measurement interpretation methodology for the diagnosis of power transformers, *Electrical Engineering*, vol. 96, pp. 183-198, 2014.
- [16] Hossein Ahmadi, BehroozVahidi, Amin Foroughi Nematollahi, A simple method to detect internal and external short-circuit faults, classify and locate different internal faults in transformers, *Electrical Engineering*, vol. 96, pp. 183-198, 2014.
- [17] M. R. Schroeder, New Method of Measuring Reverberation Time, *The Journal of the Acoustical Society of America* 37(6):409-412, March 1965, DOI: 10.1121/1.1909343.
- [18] A Papoulis, *Circuits and Systems: A Modern Approach*, Oxford University Press. June 8, 1995.
- [19] G.V. Stan, J. J. Embrechts, D. Archambeau, Comparison of different impulse response measurement techniques, *Journal of the Audio Engineering Society* 50(4):249-262, April 2002,
- [20] Tehnička dokumentacija proizvođača Doble M5400 SFRA Analyzer, dostuono na: <https://www.doble.com/product/m5400/>, pristupano 19.6.2021.
- [21] Tehnička dokumentacija proizvođača UR22 Sound Card, dostupno na [http://download.steinberg.net/downloads\\_hardware/UR22/UR22\\_documentation/UR22\\_OperationManual\\_en.pdf](http://download.steinberg.net/downloads_hardware/UR22/UR22_documentation/UR22_OperationManual_en.pdf), pristupano 19.6.2021.
- [22] S. M. Al-Ameri, M. S. Kamarudin, S. Muhammad, et. al., Interpretation of Frequency Response Analysis for Fault Detection in Power Transformers, *Applied Sciences*, vol. 11, no. 7, 2021.

## ABSTRACT

The power transformer and generator are two of the vital elements of electric power systems. The reliability of the power transformer is directly related to its condition. The analysis of the results is based on a visual comparison of the frequency responses of periodic tests of transformers, obtained with the help of specialized devices. The disadvantage of this method is the weakening of the engineering experience. An objective way to detect faults is to calculate mathematical indicators based on frequency response analysis. These parameters are often not able to unambiguously indicate the degree of failure. This paper presents a new approach for detecting internal

short circuits of transformers. A new indicator based on the reverberation time, commonly employed in acoustics, is proposed instead. The proposed parameter is based on time-domain calculation and enables reliable detection of transformer inter-turn short circuits and accurate estimation of their degree.

### **Power Transformer Reverberation Time**

Miloš Bjelić, Bogdan Brković, Mileta Žarković, Tatjana  
Miljković

# Analiza upotrebljivosti ekonomičnog audio hardvera prilikom snimanja impulsnih odziva prostorije

Marko Ličanin, Dejan Ćirić, Darko Mihajlov i Momir Prašćević

**Apstrakt**—Monitoring i analiza buke u životnoj sredini, prema trenutnim standardima i metodologijama zahteva korišćenje opreme visoke preciznosti čija cena može biti izuzetno visoka. Zbog toga je monitoringom teško pokriti šira područja gde je nivo buke potrebno pratiti u većem broju tačaka. Postavlja se pitanje da li je moguće vršiti monitoring koristeći ekonomičnija rešenja čime bi se povećao broj mernih lokacija. Prvi korak u analizi ovakvih rešenja je ispitivanje tehničkih karakteristika dostupnog hardvera koji se može iskoristiti u svrhu snimanja audio signala. Ubrzani razvoj mikro računara omogućio je njihovu integraciju u različitim projektima, gde oni služe kao centralne jedinice za obradu signala. Razvoj i unapređenje novih generacija senzora različitog tipa, imajući u vidu njihove niske cene na tržištu, omogućava praćenje velikog broja fizičkih i hemijskih veličina široj populaciji stručnjaka i entuzijasta. Istraživanje koje je ovde prezentovano odnosi se na analizu rada MEMS mikrofona kao jednog od pomenutih senzorskih uređaja, kada se on u sprezi sa Raspberry Pi mikroracunarom koristi za snimanje impulsnog odziva prostorije. Izvršeno je poređenje rezultata sa onim dobijenim mernom akustičkom opremom u istim uslovima rada.

**Ključne reči**—Impulzni odziv, MEMS mikroskop, Latencija, Ponovljivost merenja, Raspberry Pi.

## I. UVOD

Generacije mikroracunara koje se pojavljuju poslednjih godina poseduju visoke performanse i veliki broj ulazno-izlaznih periferija. Fleksibilnost koju ovu uređaji nose u smislu programabilnosti i skalabilnosti, mala potrošnja energije i niska cena daju veliku prednost u odnosu na komercijalnu mernu opremu [1]. Jedan od tipičnih predstavnika je Raspberry Pi (RPi) [2] mikroracunar (trenutno četvrte generacije) koji ima sledeće karakteristike:

- Broadcom, Quad core Cortex-A72 (ARM v8);
- 8GB LPDDR4-3200 SDRAM;
- 2.4 GHz i 5.0 GHz wireless, Bluetooth 5.0, BLE;
- Gigabit Ethernet;
- 2 USB 3.0 porta; 2 USB 2.0 porta;
- Raspberry Pi standard 40 pin GPIO;

Marko Ličanin – Univerzitet u Nišu, Fakultet zaštite na radu u Nišu, Čarnojevića 10A, 18106 Niš, Srbija (e-mail: marko.licanin@znrfaq.ni.ac.rs).

Dejan Ćirić – Univerzitet u Nišu, Elektronski fakultet u Nišu, Aleksandra Medvedeva 14 18115 Niš, Srbija (e-mail: dejan.ciric@elfak.ni.ac.rs).

Darko Mihajlov – Univerzitet u Nišu, Fakultet zaštite na radu u Nišu, Čarnojevića 10A, 18106 Niš, Srbija (e-mail: darko.mihajlov@znrfaq.ni.ac.rs).

Momir Prašćević – Univerzitet u Nišu Fakultet zaštite na radu u Nišu, Čarnojevića 10A, 18106 Niš, Srbija (e-mail: momir.prascevic@znrfaq.ni.ac.rs).

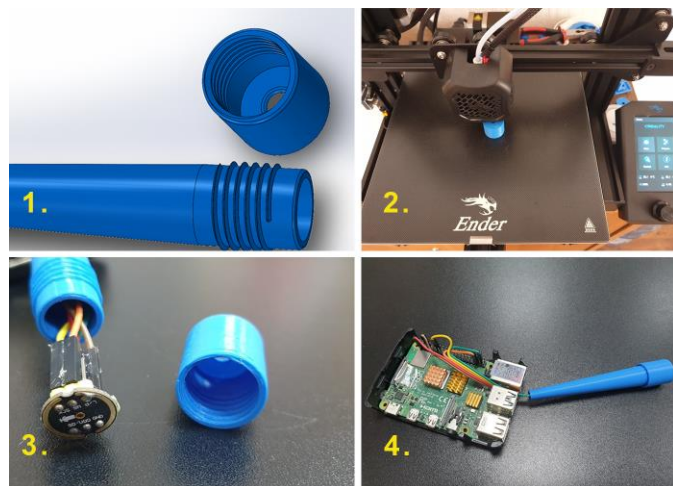
- 2 × micro-HDMI porta;
- 4-polni stereo audio i kompozitni video port;
- 3.5 mm analogni audio konektor.

Procesna moć ovakvog uređaja se približava performansama desktop računara uz značajno nižu potrošnju električne energije, što ga čini izuzetno primenljivim u uređajima koji se mogu napajati baterijski ili solarnom energijom. Upravo iz tog razloga, autori rada vršili su ispitivanje mogućnosti za iskorišćenje RPi pri snimanju audio signala, kao centralne jedinice za monitoring buke. Akustički senzor (mikrofon) koji je kompatibilan sa RPi je MEMS mikroskop INMP441. Vršeno je merenje impulsnog odziva prostorije korišćenjem RPi i INMP441 u dva stanja rada RPi. Snimanja su ponovljena korišćenjem merne akustičke opreme povezane na RPi i izvršeno je poređenje dobijenih rezultata. Za izdvajanje odziva upotrebljena je *swept-sine* tehnika [3][4]. Analiza je urađena tako što su određene vršne vrednosti i latencije impulsnih odziva, i rezultati su predstavljeni grafički.

## II. POVEZIVANJE OPREME I PROCES MERENJA

### A. Povezivanje MEMS mikrofona

INMP441 je minijaturni MEMS mikroskop montiran na okruglu PCB pločicu sa pratećim SMD komponentama. Kako bi se simulirao izgled komercijalnog mikrofona, dizajnirano je kućište u programskom paketu SolidWorks. Ovo je prikazano na Slici 1, označeno žutom bojom i brojem 1. Model kućišta je zatim odštampan na 3D štampaču Creality Ender 3 V2.



Sl. 1. Realizacija MEMS mikrofona sa kućištem i povezivanje sa Raspberry Pi računarom. Brojevi označavaju korake u realizaciji.

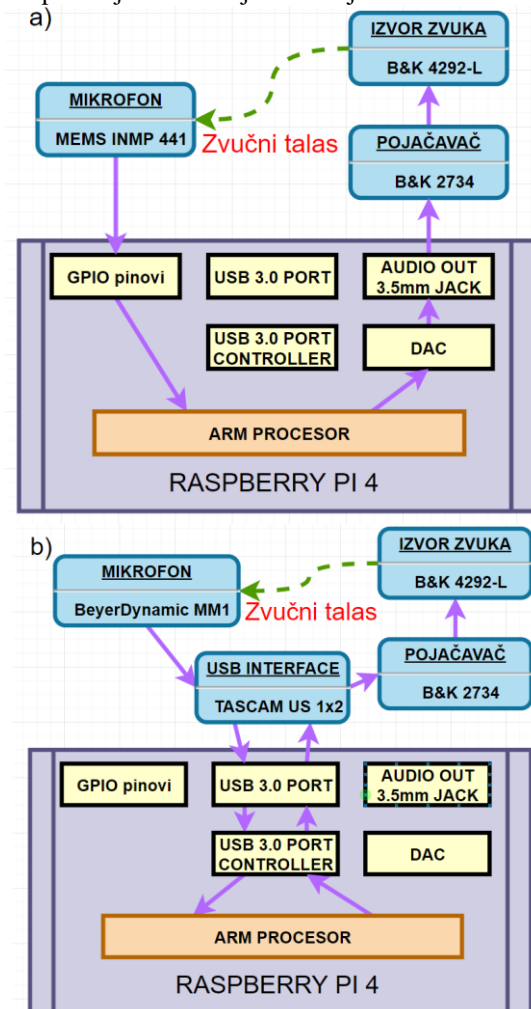
Nakon 3D štampe, montiran je mikروفon u kućište i zaštitna kapsula sa otvorom. MEMS mikروفon je zatim povezan na RPi na odgovarajuće pinove 40-pinskog konektora i u programskom jeziku Python napisana skripta za testiranje procesa snimanja.

### B. Povezivanje ostale opreme

Oprema korišćena u procesu ispitivanja je sledeća :

- MEMS mikروفon InvenSense I2S INMP441;
- Merni kondenzatorski mikروفon BeyerDynamic MM1;
- USB interfejs (zvučna kartica) Tascam US2x1;
- Brüel & Kjær pojačavač tip 2734;
- Brüel & Kjær omni-direkcionni referentni izvor zvuka tip 4292-L;

Navedenom opremom realizovane su dve situacije koje su grafički prikazane na Slici 2. Razlog za to je ispitivanje uticaja različitog hardvera i putanja audio signala na latenciju impulsnog odziva prostoriije, kao i na ponovljivosti latencije prilikom ponovljenih merenja. Situacije su sledeće:



Sl. 2. Grafički prikaz povezivanja mernog lanca u situaciji kada je korišćen MEMS mikروفon (a) i situaciji kada je korišćena merna oprema (b).

- Situacija 1 – MEMS mikروفon INMP 441 povezan je preko pinova sa RPi. I2S drajver uvodi signal sa mikrofona direktno u ARM procesor (Slika 2. a)). Signal reprodukcije

se konvertuje preko DA konvertora i kao analogni vodi na pojačavač. Ova situacija bi trebalo da pokaže minimalnu latenciju s obzirom da je putanja kretanja signala minimalna i da nema mnogo hardvera kroz koji signal prolazi [5].

- Situacija 2 – Na RPi vezan je USB interfejs Tascam US 1x2 (zvučna kartica) preko koga se odvija snimanje i reprodukcija, između interfejsa i procesora postoji i USB 3.0 kontroler koji može uneti određeno kašnjenje, na ulaz USB interfejsa vezan je BeyerDynamic kondenzatorski mikروفon, dok je izlaz odakle se vrši reprodukcija povezan na pojačavač B&K 2734 (Slika 2. b)). Putanja kretanja signala je u ovom slučaju duža u odnosu na situaciju 1, te je očekivano da latencija ima višu vrednost [5].

### C. Proces izdvajanja impulsnih odziva

Kako bi se izdvojili impulsnii odzivi prostoriije korišćena je tehnika snimanja pobudnim signalom promenljive sinusoide razvučene u vremenu (*swept-sine* ili *sweep* tehnika). U programskom paketu Python generisan je *swept-sine* signal i njegov inverzni filter. Signal je reprodukovano preko omni-direkcionalnog referentnog izvora zvuka B&K 4292-L i vršeno je snimanje odziva prostoriije. Snimljeni signal se zatim konvoluira sa inverznim filtrom, čime se dobija impulsnii odziv prostoriije [6].

### D. Tok merenja

Merenja su izvršena podešavajući RPi u dva stanja:

- Stanje 1 - RPi GUI je uključen;
- Stanje 2 - RPi GUI je isključen.

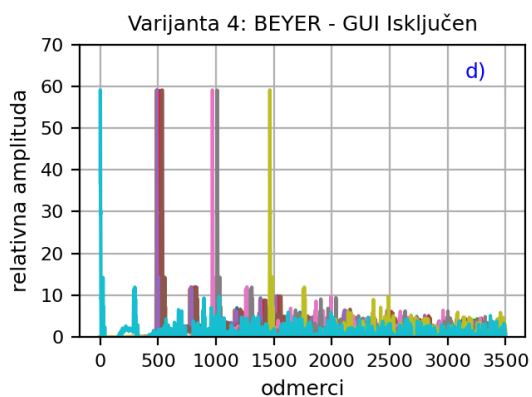
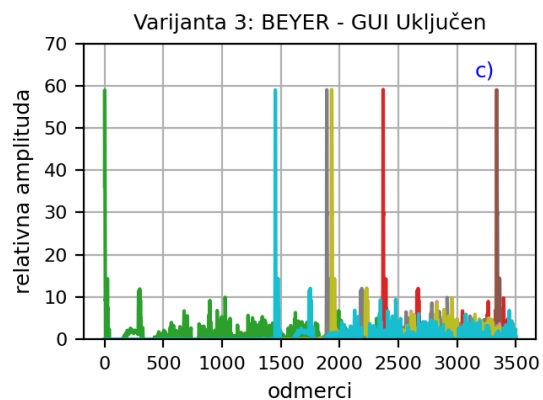
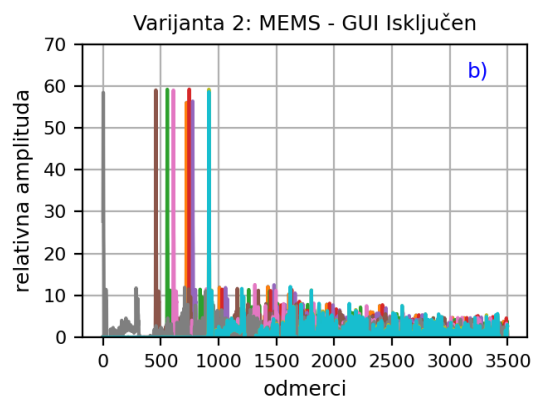
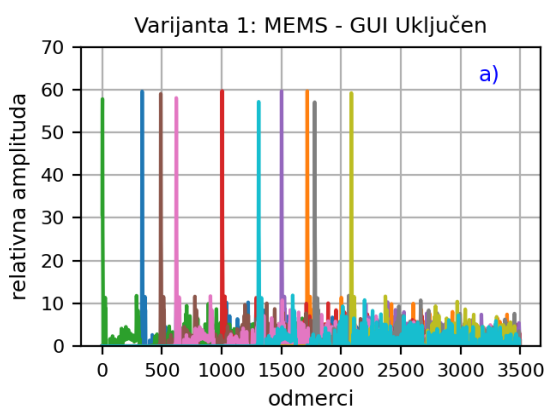
Stanje 1 podrazumeva korišćenje monitora ili udaljeno povezivanje RPi računara preko VNC servisa na pristupni terminal (u ovom slučaju laptop). Linux sistem se kontroliše kroz Desktop aplikaciju i odatle pokreće Python skripta za snimanje i reprodukciju. Stanje 2 podrazumeva ukidanje rada grafičkog interfejsa. RPi računaru se pristupa preko udaljenog terminala (laptop) preko SSH servisa i uređaj kontroliše komandama u Linux terminalu. Ova analiza izvršena je u cilju ispitivanja uticaja rada Linux grafičkog okruženja na performanse rada audio hardvera i RPi računara, prvenstveno kada se koristi MEMS mikروفon u kombinaciji sa RPi. Iz tog razloga izvršeno je snimanje impulsnih odziva u četiri varijante:

- Varijanta 1 (Situacija 1-Stanje 1): Povezan MEMS mikروفon, GUI uključen, 10 snimljenih odziva;
- Varijanta 2 (Situacija 1-Stanje 2): Povezan MEMS mikروفon, GUI isključen, 10 snimljenih odziva;
- Varijanta 3 (Situacija 2-Stanje 1): Povezan BeyerDynamic mikروفon i Tascam USB interfejs, GUI uključen, 10 snimljenih odziva;
- Varijanta 4 (Situacija 2-Stanje 2): Povezan BeyerDynamic mikروفon i Tascam USB interfejs, GUI isključen, 10 snimljenih odziva.

Koristeći interno napisane Python skripte izvršena je obrada podataka i izdvajanje impulsnih odziva, koji su zatim analizirani.

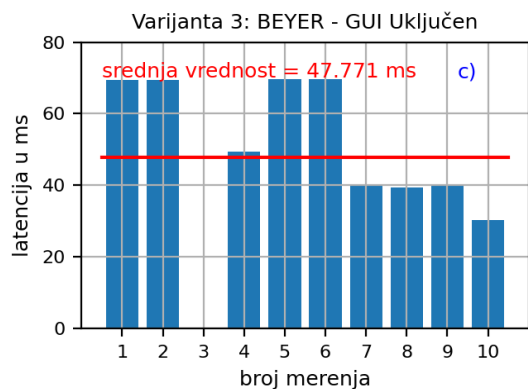
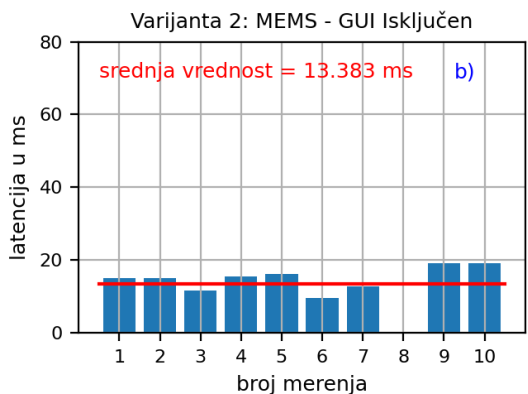
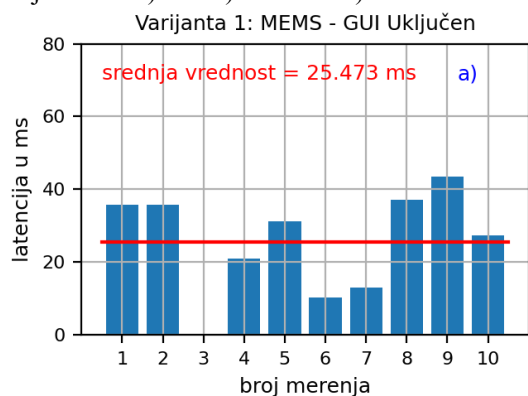
### III. REZULTATI

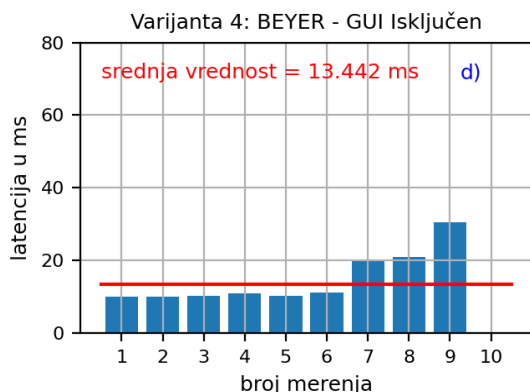
Na Slici 3 prikazani su oblici izdvojenih impulsnih odziva (kao unipolarni signali) za svaku od prethodno pomenutih varijanti i njihov pomeraj u vremenu, odnosno latencija. Latencija je određena kao relativno kašnjenje ostalih impulsnih odziva u odnosu na najmanje zakašnjeni impulsni odziv. Poređenjem varijanti 1 i 2 (Slike 3.a) i 3.b)) može se zaključiti da isključenjem grafičkog interfejsa dolazi do značajno boljeg grupisanja impulsnih odziva i smanjenja latencije. Isti zaključak je moguće doneti ukoliko se uporede rezultati kod varijanti 3 i 4 (Slike 3.c) i 3.d)). Kada se koristi merna oprema umesto MEMS mikrofona, takođe dolazi do smanjenja latencije isključenjem grafičkog interfejsa na RPi. Poređenjem situacija 1 i 2 kada je u pitanju latencija, može se zaključiti da ona ima manju vrednost u korist situacije 1 (poređenje slika 3.a) sa 3.c) i 3.b sa 3.d)



Sl. 3. Raspored impulsnih odziva predstavljen u odnosu na najmanje zakašnjeni odziv kod koga je latencija 0. Rezultati su dati za različite varijante merenja (varijante 1-4 odgovaraju slikama a-d) respektivno).

Analiza ponovljivosti latencije može se bolje uvideti na Slici 4. Primetno je da srednja vrednost latencije impulsnih odziva opada značajno isključenjem grafičkog interfejsa RPi (poređenje slika 4.a) sa 4.b) i 4.c sa 4.d).

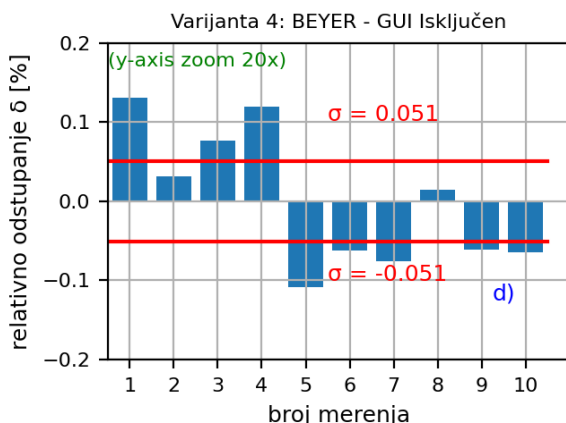
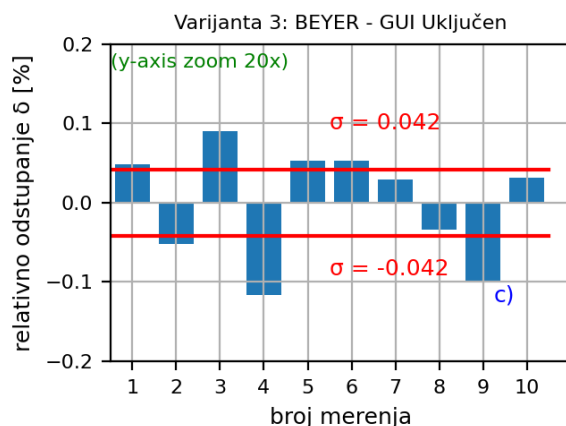
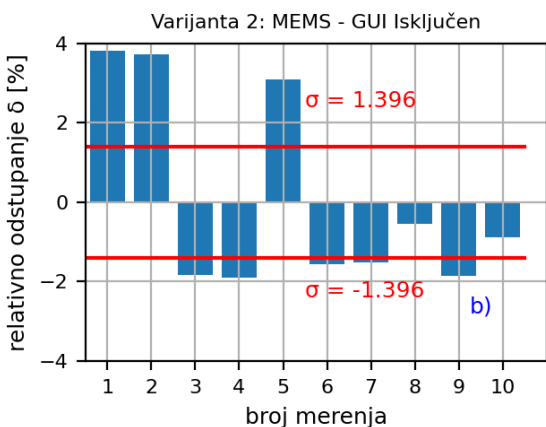
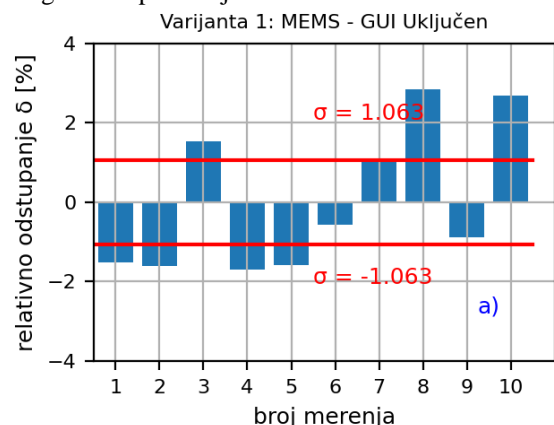




Sl. 4. Latencija impulsnih odziva u milisekundama predstavljena u odnosu na najmanje zakašnjeni odziv. Rezultati su dati za različite varijante merenja (varijante 1-4 odgovaraju slikama a-d) respektivno).

Posmatranjem varijante 2 na Slici 4.b) može se utvrditi da je ponovljivost latencije izuzetno dobra za slučaj kada je grafički interfejs isključen. To dovodi do zaključka da sa stanovišta ponovljivosti latencije, sasvim je validno koristiti MEMS mikrofona u kombinaciji sa RPi, kao i da je neophodno isključiti grafički interfejs mikroračunara RPi.

Poslednja analiza odnosi se na ponovljivost amplituda impulsnog odziva prostorije. Kao i u prethodnim slučajevima, vršeno je merenje u četiri varijante, što je prikazano na Slici 5. Na y osama bar grafika prikazano je relativno odstupanje pojedinačnih amplituda od srednje proračunate vrednosti za 10 merenja. Na x osama obeležen je redni broj merenja impulsnog odziva prostorije.



Sl. 5. Relativno odstupanje amplituda impulsnih odziva od srednje vrednosti proračunate za 10 merenja. Rezultati su dati za različite varijante merenja (varijante 1-4 odgovaraju slikama a-d) respektivno).

Posmatrajući varijante 3 i 4 (Slika 5.c) i 5.d)), i poredeći rezultate sa varijantama 1 i 2, (Slika 5.a) i 5.b)) može se zaključiti da je ponovljivost amplituda značajno bolja kada je u pitanju merna oprema u odnosu na MEMS mikrofona. X – osa je u slučaju prezentovanih rezultata kod merne opreme uveličana 20 puta kako bi se uočila razlika u amplitudama. Takođe se može zaključiti da promena stanja kod RPi (uključenje ili isključenje grafičkog interfejsa) ne utiče na promenu ponovljivosti amplitude impulsnih odziva.

#### IV. ZAKLJUČAK

Izvršena ispitivanja predstavljaju samo jedan od koraka u evaluaciji ograničenja i prednosti koje pruža ekonomični hardver pri korišćenju u audio aplikacijama. Vršena je analiza ponovljivosti latencije i amplitude impulsnih odziva prostorije u četiri varijante. Sagledavan je uticaj grafičkog interfejsa na latenciju i amplitudu impulsnih odziva. Analiza je urađena za slučaj kada je MEMS mikrofona povezan na Raspberry Pi 4, kao i za slučaj kada je kao audio interfejs korišćena USB kartica sa mernim mikrofonom.

Ispitivanje je pokazalo da postoji značajan uticaj grafičkog interfejsa na ponovljivost latencije čime je utvrđeno da je bolje raditi samo sa Linux terminalom kada je audio snimanje u pitanju. Pokazano je MEMS mikrofoni imaju ograničenja po pitanju amplitude kada se rezultati ponovljivosti amplituda porede sa ponovljivošću kod studijske opreme.

U budućnosti treba izvršiti dalja ispitivanja, koja se odnose



na ponovljivost frekvencijskih karakteristika impulsnih odziva, kao i ovde prezentovane rezultate potvrditi novim proširenim setom merenja.

#### ZAHVALNICA

Ovaj rad je podržan od strane Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije. Deo opreme koji je iskorišćen u realizaciji rada nabavljen je u okviru Erasmus + projekta SENVIBE (ref. br. 598241-EPP-1-2018-1-RS-EPPKA2-CBHE-JP) koji je finansiran od strane Evropske unije.

#### LITERATURA

- [1] M. Ličanin, M. Prašćević, D. Mihajlov "Realization of the low-cost noise measurement monitoring station using MEMS microphone technology and micro-PC," In: Proc. of 26th International Conference: Noise and Vibration, pp. 133–136, Niš, Srbija: Fakultet zaštite na radu u Nišu, 2018.
- [2] D. Molloy, *Exploring Raspberry Pi: Interfacing to the Real World with Embedded Linux*, 1st ed. Springdale, Arkansas, US: Willey, June 13, 2016.
- [3] S. Müller S. P. Massarani, "Transfer-Function Measurement with Sweeps" J. Audio Eng. Soc., vol. 49, no. 6, pp. 443–471, 2001
- [4] S. Müller, "Measuring transfer-functions and impulse responses," in *Part I: Acoustic Signals and Systems*, in *Handbook of Signal Processing in Acoustics*, pp. 65–86, New York, USA: Springer Science, 2008
- [5] M. Jelenković, D. Ćirić, M. Ličanin, A. Đorđević "Ponovljivost impulsnog odziva prostorije," In: Zbornik radova konferencije ETRAN, Društvo za ETRAN, Banja Vrućica, Srbija, pp. AK.2.7-1–AK.2.7-4., 2011.
- [6] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-sine Technique," 108th Convention Audio Eng. Soc., abstract in J. Audio Eng. Soc., vol. 48, no. 4, p. 350, 2000.

#### ABSTRACT

Monitoring and analysis of environmental noise, according to current standards and methodology, requires the use of high-precision equipment whose price can be extremely high. Therefore, it is difficult to cover wider areas where noise levels need to be monitored at a large number of points. The question is if it is possible to perform the level supervision with low-cost solutions, which increases the number of measuring locations. The first step in the analysis of such solutions is to examine the technical characteristics of the available hardware that can be used for the purpose of recording audio signals. The accelerated development of microcomputers enables their integration into various projects, where they serve as central signal processing units. The development and improvement of new generations of sensors of different types, having in mind their low prices on the market, enables the monitoring of a large number of physical and chemical quantities by a wide population of experts and enthusiasts. Research presented here refers to the analysis of the operation of the MEMS microphone as one of the mentioned sensor devices, when it is used in conjunction with the Raspberry Pi microcomputer to record the room impulse responses. A comparison of the results with those obtained using the low-cost measurement acoustic equipment in the same working conditions have been done here.

#### **Analysis of the usability of economical audio hardware when recording room impulse responses**

Marko Ličanin, Dejan Ćirić, Darko Mihajlov, Momir Prašćević

# Uticaj COVID 19 zaštitnih maski na razumljivost govora u srpskom jeziku

Miloš Bjelić, Tatjana Miljković, Miomir Mijić, Dragana Šumarac Pavlović

**Apstrakt**—U ovom radu prikazana je analiza uticaja zaštitnih maski za lice na razumljivost govora. Analizirana su tri različita tipa zaštitnih maski koje se koriste u pandemijskim uslovima (pamučna maska, hirurška maska i maska N95). Takođe, analiziran je uticaj zaštitnog transparentnog vizira. Na osnovu govora nekoliko osoba utvrđen je oblik spektra dugovremenog govora u srpskom jeziku u slučaju upotrebe zaštitnih sredstava. Izvršeno je nekoliko subjektivnih testova u kojima je merena logatomska razumljivost govora sa i bez zaštitnih maski. Eksperimenti su organizovani u kontrolisanim uslovima (slušanje preko slušalica) i u prostoriji sa velikim vremenom reverberacije. Pokazano je da se prilikom korišćenja zaštitne maske N95 ostvaruje najbolja razumljivost govora u srpskom jeziku u odnosu na druga dva tipa analiziranih maski.

**Gljučne reči** — Covid19, logatomi, N95, razumljivost govora, slabljenje, subjektivni testovi, zaštitne maske.

## I. UVOD

Prijem govornih informacija je proces prepoznavanja pojedinačnih zvučnih simbola, ali i čitavih reči i rečenica kao celina. Razumljivost govora pri prenosu kroz neki fizički kanal smanjivaće se kada neki parazitski signali prekrivaju delove govornog signala manjih amplituda. Vokali su delovi govornog signala koji imaju relativno veliku energiju dok su konsonanti delovi koji su značajno slabiji. Prekrivanje tiših delova govornog signala može nastati na dva načina: kao vremenski uniformno pokrivanje aditivnim šumom i kao vremenski ograničeno pokrivanje refleksijama u prostoriji koje stižu sa kašnjenjem nakon jačih delova govornog signala (vokala) [1]. Reverberacioni proces u značajnoj meri može smanjiti razumljivost govora. Pokazano je da se razumljivost govora manja za 15% u prostorijama sa velikim vremenom reverberacije ( $T_{60} > 2.5$  s) u odnosu na akustički obrađene prostore [2]. U literaturi je pokazano da razumljivost govora pri malim odnosima signal-šum, npr. manjim od 5 dB, može biti ispod 75% [3-4].

U poslednje skoro 2 godine pandemija virusa Covid19 uticala je na život ljudi na celom svetu. Preduzimane su mere predostrožnosti kako bi se smanjilo širenje virusa. Pored

socijalne distance i karantina, upotrebu sredstava za dezinfekciju sprovede se mere koje su podrazumevale obaveznu upotrebu zaštitne opreme, kao što su zaštitne maske, viziri za lice i rukavice. Upotreba zaštitnih maski, u velikoj meri, sprečava širenje virusa. Ove vrste zaštite pored sprečavanja širenja virusa imaju uticaj i na komunikaciju između ljudi. Maske za lice su vidljiva prepreka i utiču na verbalnu komunikaciju, emocionalna ekspresiju i čitanje sa usana, što je vrlo korisno pomagalo u uobičajenoj komunikaciji među ljudima i izuzetno korisno za ljude sa oštećenim sluhom [5].

Zbog navedenih razloga veliki broj istraživača u oblasti akustike bavio se uticajem zaštitnih maski na govornu komunikaciju. Istraživanja su se pre svega bavila uticajem zaštitnih maski na razumljivost govora [2-3], [5-12]. Analizirana je razumljivost u prostorijama, kao što su školske učionice [6]. Pokazano je da zaštitne maske utiču na zamaranje slušalaca prilikom dužeg slušanja, kao posledica smanjene razumljivosti [7]. Analiziran je i uticaj zaštitni maski na govor prilikom reprodukcije govora preko slušalica [3]. Takođe, ispitivan je uticaj zaštitne opreme na usmerenost govornika [8]. Posebno su analizirani uticaji maske na oblike dugovremenog spektra govora [3], [9-10], kao objektivna mera uticaja zaštitne opreme. U nekim istraživanjima posebno je ispitivan uticaj transparentnog zaštitnog vizira na govornu komunikaciju [9, 10]. Često su korišćeni subjektivni testovi kao metod za ispitivanje nekih od hipoteza vezanih za zaštitne maske [11].

Istraživanja koja se mogu pronaći u literaturi se pre svega odnose na ispitivanje uticaja maski na govor u engleskom jeziku. Ideja istraživanja u ovom radu bila je da se sagleda uticaj zaštitnih maski i vizira na komunikaciju u srpskom jeziku. Analizirana je promena oblika dugovremenog spektra govornog signala prilikom korišćenja različitih tipova zaštitnih maski. Ovi eksperimenti imali su za cilj određivanje objektivnog uticaja maski. Sprovedeni su eksperimenti u kojima je određivana razumljivost govora na srpskom jeziku. Razumljivost govora određivana je za slučaj idealnog slušanja govora i slušanje u realnim uslovima u prostoriji. Da bi se odredila razumljivost govora korišćeni su subjektivni testovi.

Rad je organizovan u pet poglavlja. U drugom poglavlju predstavljena je metodologija. U narednom poglavlju prikazana je eksperimentalna postavka. U četvrtom poglavlju prikazani su dobijeni rezultati i njihova diskusija. Na kraju izneti su zaključci na osnovu sprovedenog istraživanja.

Miloš Bjelić – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: bjelic@etf.rs).

Tatjana Miljković – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: tm@etf.rs).

Miomir Mijić – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: emijic@etf.rs).

Dragana Šumarac Pavlović – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: dsumarac@etf.rs).

## II. METODOLOGIJA

Prvi deo istraživanja u ovom radu obuhvata analizu oblika spektra dugovremenog govora. Oblik spektra normalnog govora na srpskom jeziku može se pronaći u literaturi [13], međutim oblik spektra govora sa zaštitnim sredstvima ne postoji. Za određivanje oblika spektra potrebno je snimanje većeg broja govornika koji izgovaraju neutralni govor. Govor je potrebno snimiti kada govornici nose zaštitne maske, odnosno vizir i u slučaju kada govornik ne nosi nikakvu zaštitnu opremu (normalan govor). Ovaj eksperiment predstavlja objektivni način za sagledavanje uticaja zaštitne opreme u pandemijskim uslovima na govornu komunikaciju.

Drugi deo istraživanja predstavlja realizaciju subjektivnih testova u kojima su ljudi ocenjivali razumljivost govora prilikom korišćenja zaštitne opreme. U slučaju izgovora reči bez smisla prepoznavanje izgovorenog teksta je moguće samo u okolnostima kada je omogućeno prepoznavanje svakog pojedinačnog glasa, a za to je potreban dovoljno visok nivo kvaliteta prenosa [14]. Iz toga sledi da samo reči bez smisla mogu dati izvesnu objektivnost u oceni kvaliteta prenosa kroz kanal bez uticaja sofisticiranih mehanizama percepcije koji nadoknađuju nedostatke u govornom signalu. Metoda koja se koristi merenje kvaliteta kanala za prenos govora i kvantifikovanje razumljivosti zasnovana je na emitovanju posebno sastavljenih reči bez smisla koje se nazivaju logatomi [14]. Proces merenje se zasniva na emitovanju logatoma na ulazu u prenosni kanal i beleženju onoga što se čuje na mestu prijema. Utvrđuje se procenat tačno primljenih logatoma i taj podatak predstavlja ocenu kvaliteta prenosa govora koji se naziva „logatomska razumljivost“.

Logatomi se posebno dizajniraju za potrebe merenja razumljivosti, i obično se sastoje od tri ili četiri glasa u redosledu konsonant–vokal–konsonant (CVC) ili konsonant–vokal–konsonant–vokal (CVCV) [1, 14]. Logatomi se prave u grupama od po 50 ili 100. Pri tome se grupe tako formiraju da budu fonetski balansirane (PB word–*Phonetically Balanced word*), što znači da je u jednoj grupi logatoma statistika zastupljenosti pojedinih glasova ista kao u običnom govoru.

Logatomska razumljivost je apsolutni pokazatelj kvaliteta prenosa govora jer se dobija neposrednim slušanjem i zbog toga je u ovom radu izabrana za ispitivanje uticaja zaštitnih maski na razumljivost govora. Sprovedenjem subjektivnih testova u kojima se određuje logatomska razumljivost mogu se odrediti uticaji različitih zaštitnih maski i vizira na govornu komunikaciju.

## III. POSTAVKA EKSPERIMENTATA

U eksperimentima su korišćena 3 tipa zaštitnih maski: pamučna maska, hirurška maska, maska N95 i zaštitni vizir. Na slici 1. prikazani su izgledi korišćene zaštitne opreme. Analiziran je uticaj zaštitnog vizira koji se često koristi u situacijama gde je izloženost virusu velika, kao što su bolnice. Zaštitni vizir se koristi u kombinaciji sa zaštitnim maskama, pa je u ovom radu razmatrana upotreba u kompletu sa zaštitnom hirurškom maskom. Izgled korišćenog vizira dat je na slici 1 d).



Sl. 1. Zaštitna oprema korišćena u eksperimentu: a) hirurška maska, b) N95 maska, c) pamučna maska d) zaštitni vizir

Snimanja govora sa i bez zaštitne opreme obavljena su u anehoičnoj prostoriji („gluva soba“). Govornici su se nalazili 20 centimetara ispred mikrofona i izgovarali potrebne reči, odnosno tekst. Na slici 2 prikazana je postavka snimanja govora u anehoičnim uslovima.



Sl. 2. Snimanje govora sa zaštitnim maskama u gluvoj sobi

U prvom eksperimentu snimani su govornici koji su čitali isti književni tekst u trajanju od 150 sekundi. U eksperimentu je učestvovalo 10 osoba, 5 muških i 5 ženskih. Svi govornici su izgovarali tekst bez maske, a zatim sa svim tipovima maske korišćenih u ovom radu. To znači da je svaki govornik izgovarao je isti tekst 5 puta. Na osnovu snimljenih signala formirane su grupe snimaka prema tipu maske i računati dugovremeni 1/3 oktavni spektri („terčni“ spektri) [15]. Cilj ovog eksperimenta bio je sagledavanje uticaja zaštitne opreme na oblik spektra govora. Za snimanje je korišćen neusmereni mikrofona [16] i AD konvertor sa 24 bita [17].

Da bi se sagledao subjektivni uticaj zaštitne opreme na razumljivost govora bilo je neophodno sprovesti subjektivne testove. U anehoičnim uslovima snimljeno je 5 grupa logatoma od po 50 reči. Svaka od grupa predstavlja jedan od slučajeva koji se analizira (jedan bez maske i 4 sa zaštitnom opremom). Sprovedena su 2 subjektivna testa. Prvi test izveden je u kontrolisanim uslovima, tj. slušanjem snimljenih grupa logatoma preko slušalica. U ovom eksperimentu eliminisan je uticaj spoljašnje sredine, odnosno prostorije i ambijentalne buke. U testu je učestvovalo 12 osoba koje su slušale grupe logatoma preko istih slušalica [18].



Sl. 3. Izgled zvučnika korišćenog za reprodukciju logatoma u učionici

Drugi test realizovan je u prostoriji koja je imala relativno veliku vrednost vremena reverberacije. Dodatno, u prostoriji je postojala ambijentalna buka zbog loše izolacione moći vrata i prozora. Korišćena prostorija je učionica na fakultetu. Grupe logatoma su reprodukovane slušaocima pomoću usmerenog zvučnika koji je prikazan na slici 3. Pozicija zvučnika odgovarala je uobičajenoj poziciji predavača u učionici – na bini ispred table. U eksperimentu je učestvovalo 14 slušalaca, koji su raspoređeni uniformno po prostoriji, kao na slici 4. U oba subjektivna testa slušaoci su pripadali starosnoj grupi od 20 do 30 godina, što predstavlja najbolju starosnu grupu za subjektivna audio ispitivanja.

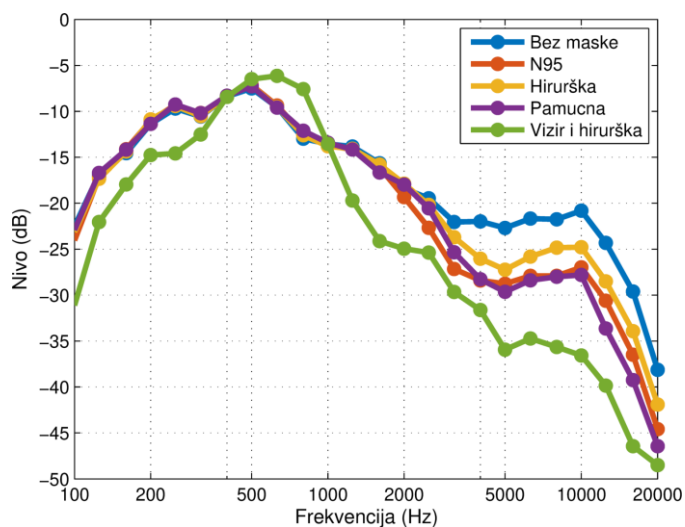


Sl. 4. Realizacija subjektivnog testa u učionici

#### IV. REZULTATI I DISKUSIJA

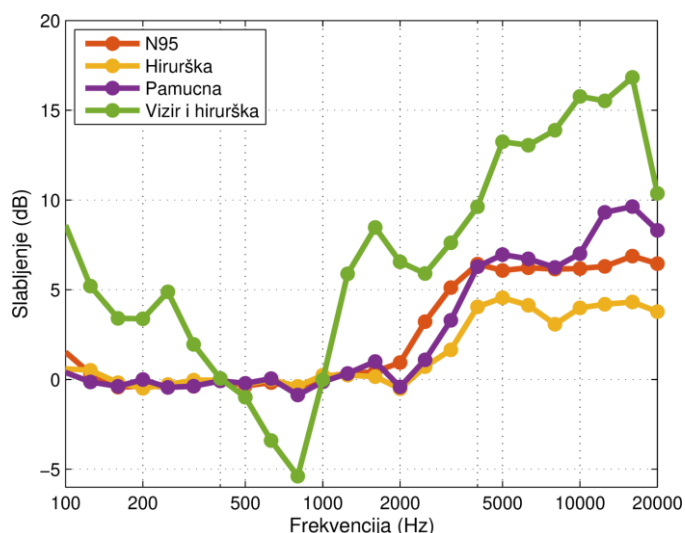
##### A. Dugovremeni spektri govora

Na slici 5 prikazani su usrednjeni dugovremeni spektri za 10 govornika. Spektri su normalizovani tako da ukupan nivo za svaki od spektara iznosi 0 dB. Plavom bojom prikazan je usrednjeni spektar govora bez korišćenja zaštitnih maski, odnosno zaštitne opreme. Maksimalna vrednost nivoa dugovremenog govora odgovara 1/3 opsegu 500 Hz. U oblasti frekvencija ispod 500 Hz spektar opada 10 dB po oktavi, a u oblasti do 3100 Hz spektar govora opada sa 6 dB po oktavi. U opsegu frekvencija između 3100 Hz i 10 kHz spektar je ravan, a nakon 10 kHz se može smatrati da nema značajnih komponenti u spektru dugovremenog govora. Ovakav oblik spektra poklapa se sa podacima o srpskom jeziku koji se mogu pronaći u literaturi [13].



Sl. 5. Izgled 1/3 oktavnih dugovremenih spektara

Dugovremeni spektar govora u slučaju korišćenja zaštitnih maski (hirurška, pamučna i N95) u oblasti frekvencija do 2000 Hz poklapa se sa oblikom spektra za govor bez maske. Prilikom korišćenja hirurške maske nivo spektra govora je za oko 5 dB manji od spektra nivoa govora bez maske, za oblast frekvencija iznad 2 kHz. Maska N95 unosi veće slabljenje u odnosu na hiruršku masku u oblasti visokih frekvencija. Vrednosti slabljenja u odnosu na normalan govor iznosi oko 6 dB. Pamučna maska unosi najveće slabljenje u odnosu na ostala dva tipa maski. Slabljenje u odnosu na govor bez maske iznosi oko 7 dB. Na osnovu vrednosti slabljenja koje unose zaštitne maske može se zaključiti da one za govor deluju kao filter propusnih niskih frekvencija. Očekivano bi bilo da je razumljivost govora obrnuto srazmerna vrednosti slabljenja koje unosi maska. Zbog toga bi razumljivost govora trebala biti najveća za hiruršku masku, zatim za N95 i najmanja za pamučnu masku.



Sl. 6. Slabljenje zaštitnih maski izračunato na osnovu dugovremenih spektara

U slučaju korišćenja zaštitnog vizira (sa hirurškom maskom) oblik spektra je značajno drugačiji od spektra govora bez i sa korišćenjem zasitnih maski. Kako bi se lakše

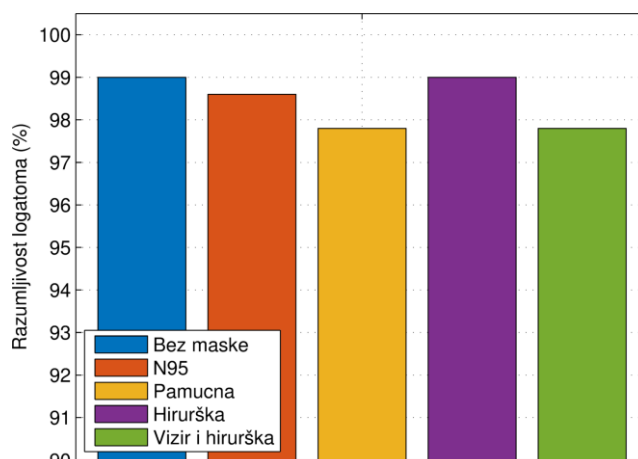
sagledao uticaj zaštitnog vizira na spektar govora izračunato je slabljenje koje unosi zaštitna oprema. Slabljenje je izračunato tako što je od spektra dugovremenog govora bez zaštitnih maski oduzet spektar govora sa maskama (Slika 5). Na slici 6 prikazano je slabljenje koje unose zaštita oprema (maske i vizir). Sa slike se mogu uočiti razlike u vrednostima slabljenja između pojedinačnih maski koje su prikazane i na slici 5.

U obliku spektra dugovremenog govora govornika koji su nosili zaštitni vizir može se uočiti zona između 400 Hz i 1000 Hz u kojoj je koncentrisana najveća energija govornog signala. Opadanje nivoa ka nižim frekvencijama je slično kao kod govora sa zaštitnim maskama, a opadanje ka višim frekvencijama iznosi oko 12 dB po oktavi. Na osnovu oblika spektra zaključuje se da zaštitni vizir deluje kao filter propusnik opsega. Hirurška maska koja je korišćena zajedno sa vizikom doprinosi samo dodatnom slabljenju na visokim frekvencijama. Govor prilikom korišćenja vizira zvuči „telefonski“ zbog efekta filtra propusnika opsega. U literaturi [9, 10] je pokazano da se ovakav oblik spektra javlja prilikom korišćenja zaštitnog vizira i za druge jezike. Međutim, nije dato objašnjenje zašto nastaje ovaj lokalni maksimum u spektru govora. Pokazano je da se takav oblik javlja u eksperimentima u kojima je izvor zvuka čovek [6, 9, 10], ali i veštački glas [10]. To znači da artikulacija kod čoveka prilikom izgovora nije odgovorna za nastanak ovog fenomena. Fenomen koji se javlja je posledica nastanka rezonance u vazдушnom prostoru između lica i zaštitnog vizira, po kraćoj dimenziji vizira. Zbog slabljenja koje unosi vizir na visokim frekvencijama očekivano je da razumljivost govora u ovom slučaju bude najmanja.

#### B. Logatomska razumljivost – slušalice

Za svaku od zaštitnih maski (zaštitne opreme) pripremljena je različita grupa logatoma. Na slici 7 prikazani su rezultati logatomske razumljivosti za subjektivni test u kome su slušaoci slušali logatome preko slušalica. Rezultati su usrednjeni za svaku od grupa logatoma na osnovu rezultata dobijenih za 12 slušalaca za tu grupu. Slušanje preko slušalica predstavlja praktično idealan način slušanja po pitanju ambijentalne buke, tj. odnosa signal šum. Nivo šuma određen je nivoom šuma u gluvoj sobi gde su snimljene grupe logatoma. Odnos signal šum prilikom reprodukcije logatoma preko slušalica u ovo eksperimentu iznosio je 40 dB.

Logatomska razumljivost za slučaj govora bez zaštitne maske iznosi 99%, što predstavlja jako dobru razumljivost. To znači da slušaoci u proseku nisu razumeli samo jednu reč od 50. Razumljivost za grupu logatoma koja je snimljena kada je govornik nosio hiruršku zaštitnu masku iznosi takođe 99%. Na osnovu toga može se zaključiti da ovaj tip zaštitne maske ne kvari razumljivost govora. U slučaju korišćenja maske N95 logatomska razumljivost iznosi 98.6%, što takođe predstavlja jako dobru razumljivost govora. Nešto manja logatomska razumljivost dobijena je za slučaj korišćenja dvoslojne pamučne maske i u slučaju korišćenja zaštitnog vizira sa hirurškom maskom. Razumljivost u ovim slučajevima je ista i iznosi 97.8%.



Sl. 7. Logatomska razumljivost za slušanje preko slušalica

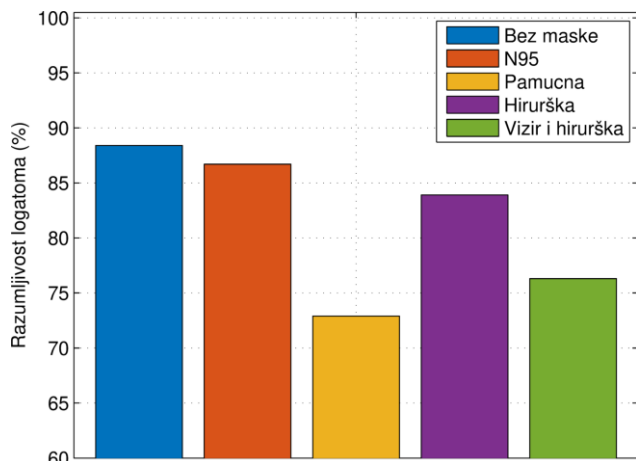
Razlike u razumljivosti dobijene za različite tipove zaštitne opreme su relativno male pre svega zbog velikog odnosa signal šum, odnosno idealnih uslova slušanja. Vrednosti su nešto manje za pamučnu masku i zaštitni vizir, ali u ovim slučajevima slušaoci, u proseku, nisu razumeli samo dve reči. Rezultati dobijeni u ovom eksperimentu mogu pokazati koje grupe maski mogu značajnije uticati na razumljivost govora, ali su procenti razumljivosti jako slični pa se neki značajniji zaključci ne mogu napraviti. Dodatno, na osnovu rezultata u ovom subjektivnom testu ne može se sagledati uticaj zaštitnih maski na razumljivost govora u svakodnevnim uslovima. U svakodnevnom životu nivo ambijentalne buke je veći, a često i značajno veći, od nivoa ambijentalne buke u ovom eksperimentu. Ambijentalna buka može maskirati delove govornog signala, što je značajno pre svega na višim frekvencijama gde zaštitne maske dodatno slabe nivo govornog signala.

#### C. Logatomska razumljivost – učionica

U ovom subjektivnom testu 14 slušalaca slušalo je u učionici površine 100 m<sup>2</sup> grupe logatoma, koji su snimljeni sa različitim zaštitnim maskama. Učionica nije akustički obrađena, što se može videti na slici 4. Vreme reverberacije iznosi oko 2 sekunde za 1/1 oktavnih frekvencijskih opsega sa centralnim frekvencijama 250 Hz, 500 Hz, 1000 Hz i 2000 Hz. Nivo ambijentalne buke u prostoriji iznosio je 36.5 dBA. Merenje ambijentalne buke izvršeno je u toku trajanja eksperimenta, u sredini prostorije. Odnos signal šum prilikom reprodukcije logatoma iznosio je maksimalno 11 dB. Na slici 8 prikazani su rezultati logatomske razumljivosti u učionici za različite tipove zaštitnih maski. Rezultati pojedinačnih slušalaca su usrednjeni za odgovarajuće grupe logatoma.

Vrednost logatomske razumljivosti u prostoriji za slučaj bez korišćenja zaštitne opreme iznosi 88.4%. Vrednost razumljivosti u ovom eksperimentu je za oko 10% manja nego u slučaju eksperimenta sa slušalicama. Najbolja razumljivost ostvarena je kod slušalaca koji su sedeli u prednjoj zoni učionice i maksimalno je iznosila 94%. Za slušaoca koji su sedeli u zadnjem delu sale razumljivost je značajno manja i minimalno iznosi 82%. U slučaju korišćenja zaštitne opreme

najbolja razumljivost ostvarena je za masku N95. Logatomska razumljivost za masku N95 iznosi 86.7%, što je približno vrednosti koja je ostvarena bez korišćenja maski. Kao i u prethodnom slučaju slušaoci koji su sedeli bliže zvučniku imali su veću uspešnost prilikom prepoznavanja logatoma. Vrednosti razumljivosti za ovaj tip maske u prostoriji kreću se u opsegu od 74% do 94%.



Sl. 8. Logatomska razumljivost za slušanje u Sali

Za slučaj u kom je korišćena hirurška zaštitna maska logatomska razumljivost je nešto manja u odnosu na prethodna dva slučaja i iznosi 83.9%. Raspodela razumljivosti po prostoriji za hiruršku masku ne poklapa se sa raspodelom sa slučajem maske N95 i slučajem bez korišćenja maske. U ovom slučaju vrednosti razumljivosti su potpuno slučajne u odnosu na rastojanje slušalaca od zvučnika preko koga su emitovani logatomi. Ostvarena razumljivost govora u slučaju u kom je korišćen zaštitni vizir zajedno sa hirurškom maskom iznosi 76.3%. Ova vrednost je za 22% manja u odnosu na slučaj razumljivosti govora bez korišćenja zaštitne maske. Vrednost razumljivost je očekivano manja u odnosu na slučaj kada je korišćena samo hirurška zaštitna maska. Vrednost je manja za 7.6%, što se može pripisati uticaju vizira. Raspodela razumljivosti je slučajna po prostoriji, i ne zavisi od rastojanja slušaoca od zvučnika. Najmanja razumljivost od svih analiziranih slučajeva postignuta je u slučaju korišćenja dvoslojne pamučne maske i iznosi samo 72.9%. Ova vrednost razumljivosti je manja za čak 25.5% u odnosu na razumljivost govora ostvarenu za slučaj bez zaštitne opreme.

Neke od grešaka koje su pravili slušaoci su nastajale zamenom parova /f/-/h/, /f/-/p/ i /d/-/g/. Fonemi /g/ i /d/ pripadaju grupi ploziva i u artikulacionom smislu su slični jer se sastoje iz dva dela okluzije i eksplozije. Kod fonema /d/ eksplozija stvara šumnu energiju koja se karakteriše sa više koncentrata gotovo ravnomerno do 8 kHz [1]. Prilikom izgovora fonema /g/ pojavljuju se 2 koncentrata energije, veći na 2000 Hz i manji na oko 5 kHz. Zbog slabljenja zaštitnih maski iznad 2000 Hz slušaoci slovo /d/ mešaju sa slovom /g/. Fonem /f/ pripada grupi frikativa. Energija kod šumnog fonema /f/ se prostire od 500 Hz do gotovo 10 kHz [1]. Zbog uticaja maski u ovom frekvencijskom području razumljivost ovog fonema može biti narušena.

Posmatrajući oblik usrednjenih spektara sa slike 5 zaključuje se da maska N95 unosi veće slabljenje u odnosu na hiruršku masku. Međutim, na osnovu rezultata subjektivnog testa realizovanog u realnim uslovima (npr. slušanje predavanja u učionici) zaključuje se da je najmanji uticaj na razumljivost govora ima maska N95. Maska N95 skoro da ima isti uticaj na spektar dugovremenog govorna kao i pamučna maska, ali je razumljivost govora u učionici prilikom korišćenja pamučne maske za 14% manja u odnosu na masku N95. Takođe, slabljenje koje unosi zaštitni vizir sa hirurškom maskom je značajno veće u odnosu na sve tipove maski, ali je razumljivost govora u tom slučaju veća od razumljivosti za pamučnu masku. Na osnovu toga zaključuje se da posmatranje maske (odnosno zaštitne opreme) kao filtra za govorni signal nije dobar način za sagledavanje uticaja maski na razumljivost govora. Ponašanje maske kao filtra propusnika niskih frekvencija nije zanemarljiv uticaj na razumljivost, ali nije jedini. U literaturi se mogu pronaći podaci i o drugim uticajima maske na govor [13], pored filtriranja. Pored unošenja slabljenja na visokim frekvencijama maske utiču i na način izgovora slova, odnosno na artikulaciju. Pamučna maska je pripijena uz lice i postoji konstantno trenje usana o masku. Prilikom dužeg govora ovo stvara ometanje i otežava otvaranje usta prilikom izgovora određenih slova. Dodatno, ova maska se pomera prilikom govora. Sličan uticaj ostvaruje na artikulaciju ima i hirurška maska. Ova dva tipa maske prilikom dužeg govora stvaraju osećaj otežanog disanja, pa i to utiče na način izgovora određenih fonema. Za razliku od toga maska N95 ima čvrstu formu i ne pomera se zajedno sa ustima. Ona je odvojena fizički od usana i ne opterećuje prilikom dužeg govora. Zbog toga je i logatomska razumljivost za ovu masku najveća.

Kako bi se sagledalo da li su neki govornici usled nedovoljne koncentrisanosti ili neobučenosti za slušanje logatoma uticali na globalne rezultate izvršena je detaljnija analiza. Za svakog slušaoca izvršena je normalizacija vrednosti razumljivosti za svih 5 grupa logatoma. Normalizacija je urađena tako da su vrednosti logatomske razumljivosti za svaku od grupa podeljene sa maksimalnom vrednošću razumljivosti. Nakon toga izvršeno je usrednjavanje razumljivosti za svaku od grupa za sve slušaoce. Dobijene su sledeće vrednosti: bez maske 98.1%, maska N95 96.2%, hirurška maska 93.1%, vizir sa hirurškom maskom 84.6% i pamučna maska 80.76%. Međusobni odnos normalizovanih vrednosti je isti kao u slučaju bez normalizacije vrednosti. To znači da slušaoci nisu uticali na dobijene rezultate.

## V. ZAKLJUČAK

U ovom radu analiziran je uticaj zaštitne opreme koja se koristi u vreme pandemije virusa COVID19 na razumljivost govora u srpskom jeziku. Analizirana su 3 tipa maski i jedan zaštitni vizir. Pokazano je da se dugovremeni spektar govora na srpskom jeziku menja upotrebom zaštitnih maski. Na frekvencijama iznad 2 kHz zaštitne maske unose slabljenje od 3 dB do čak 10 dB. Zaštitni vizir pored slabljenja na visokim

frekvencijama ističe oblast frekvencija između 400 Hz i 1000 Hz. Rezultati subjektivnih testova realizovanih slušanjem preko slušalica pokazali su da uticaj zaštitnih maski postoji. Zbog idealnih uslova slušanja dobijeni rezultati ne pokazuju uticaj zaštitne opreme u svakodnevnim uslovima. Zbog toga je sproveden subjektivni test gde su slušaoci u prostoriji slušali pripremljene grupe logatoma, koji su prethodno snimljeni sa zaštitnim sredstvima u gluvoj sobi. Rezultati su pokazali da je razumljivost najveća za masku N95, zatim slede hirurška maska, zaštitni vizir i pamučna maska. Dobijeni rezultati poklapaju se sa rezultatima iz literature dobijenim za engleski jezik. Rezultati se ne poklapaju sa vrednostima slabljenja dugovremenih spektara govora, što ukazuje da maske ne utiču samo kao filter za govorni signal već utiču i na artikulaciju. Dobijeni rezultati pokazuju značajno smanjenje razumljivosti govornika sa određenim tipovima zaštitnih maski u reverberantnim prostorima sa ambijentalnom bukom.

#### ZAHVALNICA

Ovaj rad je realizovan u okviru projekta TR 36026 koga finansira Ministarstvo prosvete, nauke i tehnološkog razvoja Republike Srbije. Autori žele da se zahvale studentkinjama Maji Đaković i Melaniji Milenković na pomoći prilikom realizacije eksperimenata, kao i svima koji su učestvovali u snimanju govora i u subjektivnim testovima.

#### LITERATURA

- [1] S. T. Jovičić, "Govorna komunikacija – fiziologija, psihoakustika i percepcija," pp. 29–72, NAUKA, Beograd, Srbija, 1999.
- [2] P. Bottalico, S. Murgia, G. E. Puglisi, A. Astolfi, K. I. Kirk, "Effect of masks on speech intelligibility in auralized classrooms," *The Journal of the Acoustical Society of America*, vol. 148, pp. 2878- 2884, November, 2020.
- [3] Y. Choi, "Acoustical measurements of masks and the effects on the speech intelligibility in university classrooms," *Applied Acoustics*, vol. 180, pp. 1-8, May, 2021.
- [4] N. Prodi, C. Visentin, A. Feletti, "On the perception of speech in primary school classrooms: Ranking of noise interference and of age influence," *The Journal of the Acoustical Society of America*, vol. 133, pp. 255- 268, November, 2012.
- [5] V. V. Ribeiro, A.P. Dassisti-Leite, et al, "Effect of Wearing a Face Mask on Vocal Self-Perception during a Pandemic," *Journal of Voice*, pp. 1-7, October, 2020.
- [6] M. Caniato, A. Marzi, A. Gasparella, "How much COVID-19 face protections influence speech intelligibility in classrooms?" *Applied Acoustics*, vol. 178, pp. 1-14, March, 2021.
- [7] M. Nobrega, R. O. Mariana, M. L. Christiane, A. Nobrega, "How face masks can affect school performance," *International Journal of Pediatric Otorhinolaryngology*, vol. 138, pp. 1-2, May, 2021.

- [8] C. Pörschmann, T. Lübeck, J. M. Arend, "Impact of face masks on voice radiation," *The Journal of the Acoustical Society of America*, vol. 148, pp. 3663- 3670, December, 2020.
- [9] R. M. Corey, U. Jones, A. C. Singer, "Acoustic effects of medical, cloth, and transparent face masks on speech signals," *The Journal of the Acoustical Society of America*, vol. 148, pp. 2371- 2375, October, 2020.
- [10] S. R. Atcherson, B. R. McDowell, M. P. Howard, "Acoustic effects of non-transparent and transparent face coverings," *The Journal of the Acoustical Society of America*, vol. 149, pp. 2249- 2254, April, 2021.
- [11] S. R. Atcherson, L. L. Mendel, W. J. Baltimore, C. Patro, S. Lee, M. Pousson, M. J. Spann, "The Effect of Conventional and Transparent Surgical Masks on Speech Understanding in Individuals with and without Hearing Loss," *Journal of the American Academy of Audiology*, vol. 28, no. 1, pp. 58- 67, 2017.
- [12] M. Magee, C. Lewis, G. Noffs et al., "Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols," *The Journal of the Acoustical Society of America*, vol. 148, pp. 3562- 3568, December, 2020.
- [13] M. Vojnović, M. Mijic, "The influence of the oxygen mask on longtime spectra of continuous speech," *The Journal of the Acoustical Society of America*, vol. 102, no. 4, pp. 2456-2458, October, 1997.
- [14] E. Zwicker, H. Fastl, "Psychoacoustics: Facts and Models," 2nd edition, Springer-Verlag, Beograd, New York, 1999.
- [15] *Specification for Octave, Half-Octave, and Third Octave Band Filter Sets*, ANSI S1.11, 49 CFR 227, June 2009.
- [16] Tehnička dokumentacija proizvođača, dostupno na mreži: <http://www.nti-audio.com/Portals/0/data/en/MiniSPL-Measurement-Microphone-Product-Data.pdf>, pristupano 19.6.2021.
- [17] Tehnička dokumentacija proizvođača, dostupno na mreži: [http://download.steinberg.net/downloads\\_hardware/UR22/UR22\\_documentation/UR22\\_OperationManual\\_en.pdf](http://download.steinberg.net/downloads_hardware/UR22/UR22_documentation/UR22_OperationManual_en.pdf), pristupano 19.6.2021.
- [18] Tehnička dokumentacija proizvođača, dostupno na mreži: <https://www.akeg.com/Headphones/Professional%20Headphones/K92.html>, pristupano 19.6.2021.

#### ABSTRACT

This paper presents an analysis of the impact of face masks on speech intelligibility. Three different types of protective masks used in pandemic conditions (cotton mask, surgical mask and N95 mask) were analyzed, as well as a protective transparent visor. Based on the speech of several people, the shape of the spectrum of long-term speech in the Serbian language in the case of the use of protective equipment was determined. Several subjective tests were performed in which the logatom intelligibility of speech with and without protective masks was measured. The experiments were organized in controlled conditions (listening through headphones) and in a room with a large reverberation time. It has been shown that the best intelligibility of speech in the Serbian language is achieved when using the N95 protective mask in relation to the other two types of analyzed masks.

#### Impact of COVID19 face masks on speech intelligibility in the Serbian language

Miloš Bjelić, Tatjana Miljković, Miomir Mijić, Dragana Šumarac Pavlović

# Izdvajanje režima praznog hoda motora sa unutrašnjim sagorevanjem na osnovu audio zapisa

Marko Milivojčević, Emilija Kisić, Dejan Ćirić

**Apstrakt**—U ovom radu je predstavljen postupak za prepoznavanje i izdvajanje režima praznog hoda motora sa unutrašnjim sagorevanjem putničkih vozila na osnovu audio zapisa prikupljenih na ulazu u podzemnu garažu. Analiza audio zapisa je realizovana u vremenskom domenu, kako bi se omogućila obrada signala sa što manjim zahtevima u pogledu potrebne procesorske snage akvizicionog sistema sa baterijskim napajanjem. Celokupan postupak prepoznavanja i izdvajanja audio zapisa koji odgovara režimu praznog hoda motora sa unutrašnjim sagorevanjem putničkih vozila realizovan je upotrebom programskog jezika *Python*. Cilj predstavljenog postupka je priprema velikog broja prikupljenih audio signala za dalju analizu i obradu u pogledu izdvajanja karakterističnih audio obeležja i upotrebu u obuci neuronskih mreža.

**Ključne reči**—akustičke karakteristike, audio zapis, vremenski domen, motori sa unutrašnjim sagorevanjem, akvizicija, *python*.

## I. UVOD

Primene algoritama veštačke inteligencije na audio signale vremenom postaju sve brojnije. Klasifikacija zvukova, detekcija audio događaja i prepoznavanje audio scena su primeri zadataka koji se uspešno realizuju u praksi primenom mašinskog učenja. U tom kontekstu, mašinsko i duboko učenje se mogu koristiti za prepoznavanje vrste motora sa unutrašnjim sagorevanjem na osnovu zvuka koji motori generišu. Naime, zvuk ovih motora se razlikuje u zavisnosti od pogonskog goriva koje koriste motori (benzin ili dizel). Činjenica da ljudsko uvo na osnovu zvuka koji motor generiše može da prepozna o kojoj vrsti motora se radi, poslužila je kao osnovna ideja da se napravi sistem za automatsko prepoznavanje vrste motora na osnovu zvuka koji motori generišu.

U cilju realizacije jednog ovakvog sistema najpre je bilo potrebno napraviti akvizicioni sistem za prikupljanje audio signala koji će činiti bazu na osnovu koje će se sistem obući nekim od algoritama mašinskog/dubokog učenja. Kako bi se formirala kvalitetna baza audio signala, bilo je neophodno prikupiti veliki broj audio signala generisanih radom motora

Marko Milivojčević – Akademija tehničko-umetničkih strukovnih studija Beograd, odsek: Visoka škola elektrotehnike i računarstva strukovnih studija, Vojvode Stepe 283, 11000 Beograd, Srbija (e-mail: [markom@viser.edu.rs](mailto:markom@viser.edu.rs)).

Emilija Kisić – Akademija tehničko-umetničkih strukovnih studija Beograd, odsek: Visoka škola elektrotehnike i računarstva strukovnih studija, Vojvode Stepe 283, 11000 Beograd, Srbija (e-mail: [emilijai@viser.edu.rs](mailto:emilijai@viser.edu.rs)).

Dejan Ćirić – Univerzitet u Nišu, Elektronski fakultet u Nišu, Aleksandra Medvedeva 14, 18000 Niš, Srbija (e-mail: [dejan.ciric@elfak.ni.ac.rs](mailto:dejan.ciric@elfak.ni.ac.rs)).

sa unutrašnjim sagorevanjem putničkih vozila. Pomenuta banka uzoraka bi se koristila za dalju analizu i izdvajanje karakterističnih audio obeležja sa ciljem da se isključivo na osnovu akustičkih karakteristika odredi pogonsko gorivo. Pomenuti akvizicioni sistem koristi mikrofonski pozicioniran u oblasti ispod motornog prostora putničkog vozila [1, 2] nakon što se detektuje prisustvo vozila.

U prethodnim fazama ovog istraživanja su analizirani spektri signala audio zapisa koji potiču od motora pokretanih različitim pogonskim gorivima pri čemu je uzet u obzir položaj mikrofona iznad poklopca motora i u oblasti ispod motornog prostora [1]. Analiziran je i uticaj položaja mikrofona u oblasti ispod motornog prostora na karakteristiku audio zapisa, gde je utvrđeno da vremenski oblik signala minimalno varira bez obzira na tačku u kojoj je postavljen merni mikrofonski [2]. Ovo je omogućilo prikupljanje uzoraka bez obzira na položaj vozila u trenutku kada je ono detektovano i kada se zaustavi iznad mikrofona.

Osnovni zadatak akvizicionog sistema je prikupljanje audio zapisa motora sa unutrašnjim sagorevanjem koji se nalazi u režimu praznog hoda tj. u neopterećenom režimu, što je detaljnije opisano u [3]. I pored velikog broja uspešno prikupljenih audio snimaka, u formiranoj bazi su se pojavili snimci sa događajima koji ne sadrže samo željeni režim rada motora već i druge modove rada motora. Zbog toga je bilo potrebno da se u okviru postojećih snimaka detektuje, a zatim i izdvoji režim praznog hoda motora sa unutrašnjim sagorevanjem, što predstavlja ključni motiv ovog rada. Kako je akvizicioni sistem baziran na baterijskom napajanju, radi svoje mobilnosti, cilj je bio napraviti što jednostavniji postupak koji bi zahtevao minimalnu procesorsku snagu. Na osnovu toga je doneta odluka da se postupak izdvajanja režima praznog hoda motora bazira na obradi audio signala u vremenskom domenu, odnosno na anvelopi signala.

Rad je podeljen u više poglavlja, gde su karakteristični slučajevi prikupljenih audio signala i metod njihove obrade prikazani u poglavlju gde je opisana metodologija. U poglavlju gde su opisani rezultati data je ilustracija uspešne detekcije promene režima rada motora samo na osnovu vremenskog oblika signala, kao i nove, izdvojene signale koji predstavljaju željeni režim rada motora. Zaključci istraživanja su sumirani u poglavlju koje sledi, iza kojeg je dat spisak korišćene literature.

## II. PREGLED LITERATURE

Analiza dostupnih radova iz oblasti akustike koji za izvor zvuka imaju motore sa unutrašnjim sagorevanjem dala je kao rezultat podelu radova na nekoliko kategorija. Tako su



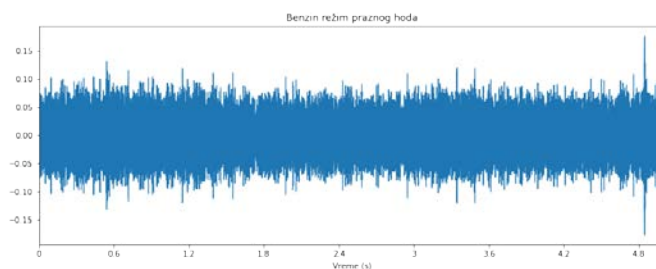
izdvojeni radovi koji za cilj imaju detekciju-monitoring buke i brojanje vozila u saobraćaju [4, 5, 6], zatim detekciju ili predikciju nepravilnosti rada motora vozila [7, 8, 9] i klasifikaciju vozila po kategorijama (putnička, teretna, motocikli) na osnovu zvuka [10, 11]. U većini ovih radova su razvijeni namenski akvizicioni sistemi sa određenim stepenom obrade prikupljenih audio uzoraka. Osnovna razlika takvih sistema i akvizicionog sistema koji je korišćen za prikupljanje uzoraka za ovaj rad je prvenstveno pozicija i kvalitet korišćenog mikrofona, kao i nivo obrade pojedinačnih prikupljenih uzoraka pre skladištenja na memorijski medij. Posmatrajući poziciju mikrofona moguće je uočiti da se u radovima čiji je predmet detekcija-monitoring buke kao i klasifikacija vozila u saobraćaju mikrofoni pozicioniraju pored saobraćajnice, dok se kod radova fokusiranih na detekciju nepravilnosti rada motora mikrofoni pozicioniraju u oblasti iznad motornog prostora, pri čemu je poklopac motora otvoren.

### III. METODOLOGIJA

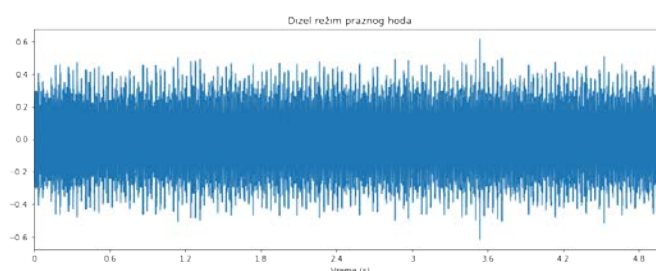
Za prikupljanje velikog broja audio zapisa motora sa unutrašnjim sagorevanjem razvijen je akvizicioni sistem koji funkcioniše autonomno. Sistem je u potpunosti mobilan jer je napajanje baterijsko, a prikupljeni zapisi se skladište na memorijsku karticu. Detekcija prisustva vozila iznad mikrofona je bazirana na ultrazvučnim sensorima koji mere rastojanje po vertikalnoj i horizontalnoj osi u različitim ravnima što onemogućava detekciju malih objekata tj. objekata koji nisu vozila. Dodatna provera, nakon detekcije ultrazvučnim sensorima, se realizuje uključanjem mikrofona i merenjem nivoa zvučnog polja. U sistemu je kreirana logika koja započinje snimanje audio zapisa tek ukoliko je detektovani zvuk iznad zadatog praga. Kako je mikروفon pozicioniran na podlozi iznad koje se vozilo krakotrajno zaustavlja dok vozač ne preuzme karticu za ulazak u podzemnu garažu i dok se ne otvori rampa, ekperimentalno je određen minimalni nivo zvučnog polja tj. prag. U sistemu je takođe definisano trajanje snimanja od 5 sekundi koje je određeno na osnovu empirijske procene zadržavanja vozila u režimu praznog hoda.

Inicijalno postavljanje sistema na ulaznu rampu podzemne garaže je pokazalo da sistem detektuje isključivo vozila i da se na audio zapisima nalaze samo signali koji potiču od motora sa unutrašnjim sagorevanjem, ali da vremena zadržavanja vozila iznad mikrofona znatno variraju. Za najveći broj vozila su dobijena dva audio snimka po vozilu pri čemu prvi snimak, bez izuzetka, predstavlja režim praznog hoda motora-stacionaran režim (Sl. 1 i 2), a drugi, odnosno u pojedinim slučajevima poslednji, snimak delimično sadrži režim praznog hoda motora nakon čega sledi povećavanje broja obrtaja radilice i režim delimičnog opterećenja motora u cilju ubrzanja vozila (Sl. 3, 5 i 6). Na uzorku od 50 vozila ostali slučajevi nisu zabeleženi bez obzira na uticaj autopercionice koja se nalazila u neprednoj blizini u nedeljenom zatvorenom prostoru. Dodatno je, paralelno sa autonomnim radom sistema, manuelno vođena evidencija tipa pogonskog goriva kako bi se uočila eventualna mogućnost greške tj. beleženja audio snimaka koji bi bili neupotrebljivi u odnosu na buku okoline koja je prisutna u zatvorenom prostoru i koja najvećim delom potiče od ventilacije garaže. Ovakav slučaj se nije dogodio u praksi

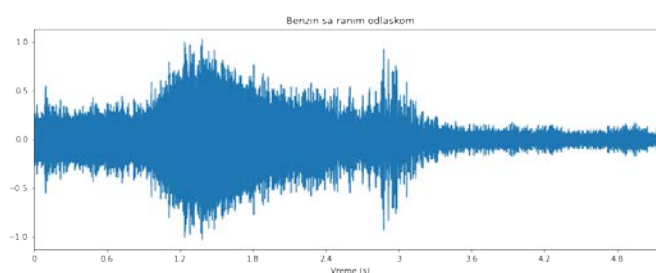
zbog pravilno podešnog praga koji uslovljava početak beleženja snimka. Posmatrajući vremenski oblik signala, na Sl. 3, 4, 5 i 6 može se uočiti snimljena buka tek nakon odlaska vozila.



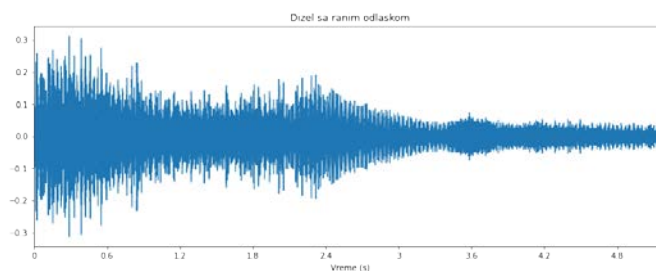
Sl. 1. Audio signal motora koji koristi benzin u praznom hodu, bez promene režima.



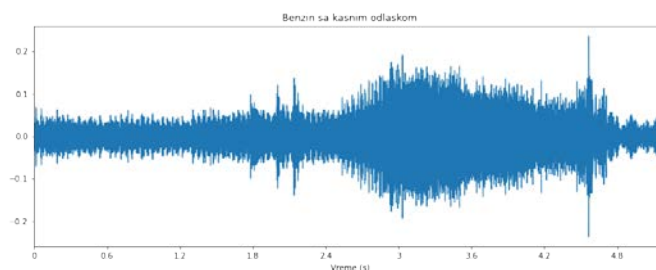
Sl. 2. Audio signal motora koji koristi dizel u praznom hodu, bez promene režima.



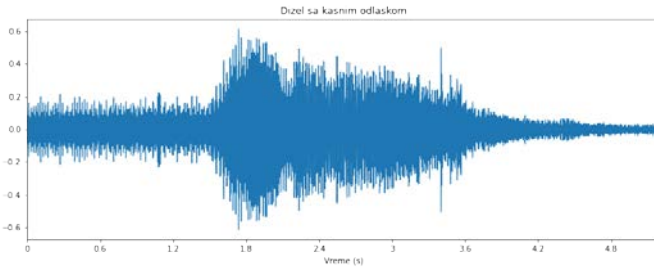
Sl. 3. Audio signal motora koji koristi benzin sa ranom promenom režima rada.



Sl. 4. Audio signal motora koji koristi dizel sa ranom promenom režima rada.



Sl. 5. Audio signal motora koji koristi benzin sa kasnom promenom režima rada.



Sl. 6. Audio signal motora koji koristi dizel sa kasnom promenom režima rada.

U cilju formiranja banke uzoraka audio snimaka stacionarnog režima rada motora i na osnovu analize vremenskog oblika signala kada vozilo odlazi, uočeno je da je moguće izdvojiti željeni režim rada „isecanjem“ dela snimka pre nego što vozilo krene da napušta poziciju tj. pređe u režim delimičnog opterećenja.

Uzimajući u obzir da se sistem napaja baterijski, da obradu celokupnog audio zapisa od 5 s radi u realnom vremenu i da se iz vremenskog oblika signala jasno mogu uočiti događaji, izdvajanje željenog režima rada je bazirano na vremenskom domenu radi smanjenja potrebne procesorske snage. Obrada u vremenskom domenu je podeljena u tri faze: formiranje anvelope amplitude signala, određivanje trenutka sečenja i isecanje signala do prethodno određenog trenutka.

Zvučni zapis motora je sniman u audio kvalitetu 48000 odbiraka u sekundi, 16 bita po odbirku, mono. Najbolje rezultate prilikom formiranja anvelope su dali parametri sa veličinom prozora gde se beleži maksimalna vrednost signala dužine 4000 odbiraka, a pomeraj 3000 odbiraka, tj sa preklapanjem susednih prozora od 1000 odbiraka. Dobijene anvelope su prikazane na Sl. 7 do 12.

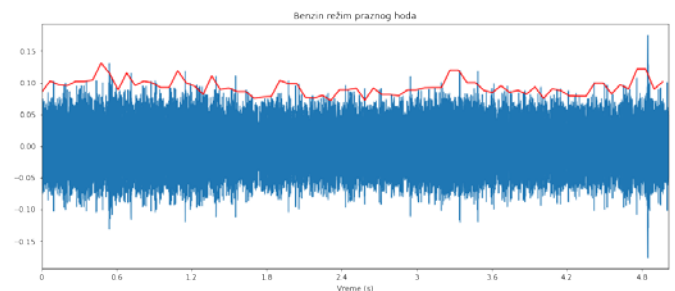
Kako bi se izdvojio samo stacionarni deo koji odgovara praznom hodu motora iz svakog snimljenog audio signala iz baze, bilo je potrebno odrediti prag (vremenski trenutak) nakon kojeg se nestacionaran deo signala odbacuje. S obzirom na prirodu problema, stacionaran deo signala se uvek javlja na početku signala (Sl. 9 do 12), tako da je bilo jasno da je prag potrebno naći u nekom trenutku nakon početka signala, odnosno u prvom trenutku kada signal prestaje da bude stacionaran. Kao što je gore već objašnjeno, slučajevi gde se prazan hod javlja kasnije (u sredini ili na kraju signala) ne postoje, tako da je bilo jasno da je potrebno odrediti samo jedan prag.

Na osnovu analize vremenskih oblika signala koji su snimljeni, došlo se do zaključka da u trenutku kada signal prestaje da bude stacionaran dolazi do naglog porasta njegove amplitude, a samim tim u tom trenutku dolazi i do veoma primetnog porasta njegove anvelope. Osnovna ideja za pronalaženje praga pomoću kojeg bi se izdvojio samo stacionaran deo signala jeste da se najpre odredi anvelopa signala kako je gore opisano, a zatim da se računa razlika između trenutne i prethodne vrednosti anvelope za sve vrednosti anvelope. Dok je signal stacionaran, očekivano je da razlika između trenutne i prethodne vrednosti anvelope bude mala. U trenutku kada signal prestaje da bude stacionaran, dolazi do skoka u anvelopi, što znači da razlika između trenutne vrednosti anvelope u trenutku kada je signal prestao da bude stacionaran i prethodne vrednosti anvelope dok je signal još uvek stacionaran mora biti znatno veća nego razlika između trenutne i prethodne vrednosti anvelope u trenucima dok je signal stacionaran. Prvi trenutak gledajući

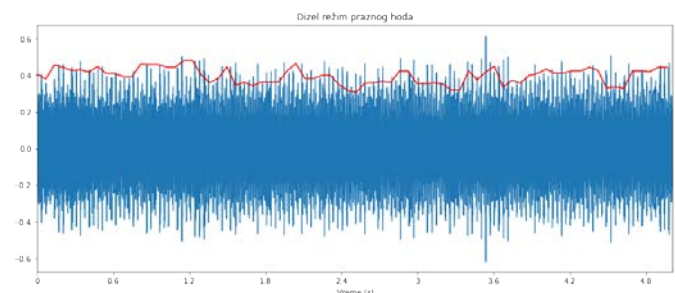
sa leva na desno (od početka do kraja signala) kada dođe do porasta razlike između trenutne i prethodne vrednosti anvelope bi trebalo da bude trenutak za postavljanje praga. Ukoliko signal anvelope označimo kao  $Env(t)$ , a vrednost praga kao  $t_L$ , prag možemo izračunati pomoću sledeće formule:

$$t_L = \min\{Env(t) - Env(t - 1) > 0.1\} \cdot \left(\frac{t_s}{N_f}\right) \quad (1)$$

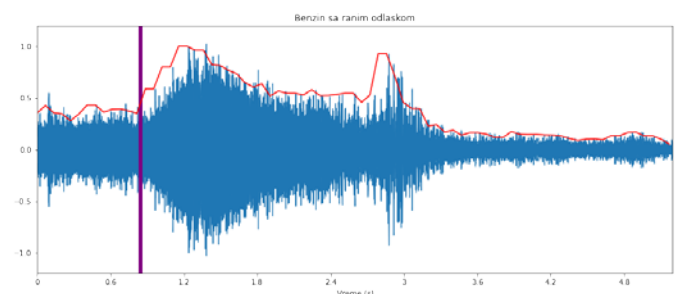
gde je sa  $t_s$  označeno trajanje celog signala, a sa  $N_f$  broj frejmova u kojima su računati maksimumi signala pomoću kojih je određena anvelopa signala. Vrednost 0.1 je određena empirijski. S obzirom na to da je potrebno postaviti prag u prvom trenutku nakon kojeg dolazi do skoka u amplitudi, gledajući od početka do kraja signala (sa leva na desno), potrebno je uzeti najmanju vrednost koja zadovoljava uslov  $Env(t) - Env(t - 1) > 0.1$ . Takođe, kako bi se dobio tačan vremenski trenutak za postavljanje praga, a uslov  $\min\{Env(t) - Env(t - 1) > 0.1\}$  vraća frejm anvelope u kojem dolazi do skoka u anvelopi koji označava prelazak iz stacionarnog u nestacionaran deo signala, bilo je neophodno pomnožiti dobijenu vrednost frejma anvelope sa trajanjem celog signala i podeliti sa ukupnim brojem frejmova u anvelopi. Na ovaj način dobija se željeni vremenski trenutak za postavljanje praga  $t_L$ , odnosno trenutak do kojeg se nalazi stacionaran deo signala koji želimo da izdvojimo.



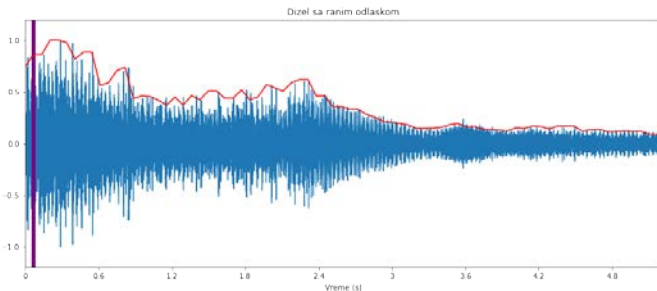
Sl. 7. Anvelopa audio signala motora koji koristi benzin u praznom hodu, bez promene režima rada.



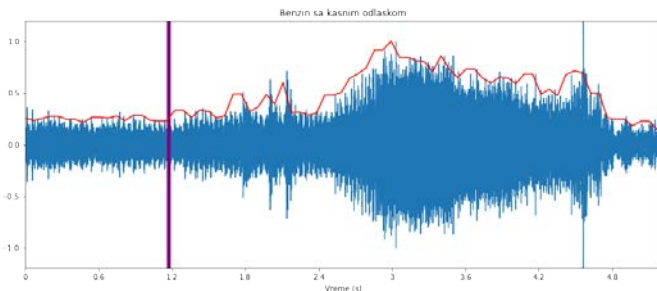
Sl. 8. Anvelopa audio signala motora koji koristi dizel u praznom hodu, bez promene režima rada.



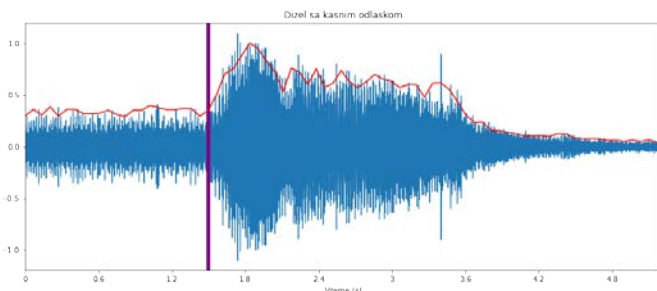
Sl. 9. Anvelopa audio signala motora koji koristi benzin sa ranom promenom režima rada (prag je označen vertikalnom linijom).



Sl. 10. Anvelopa audio signala motora koji koristi dizel sa ranom promenom režima rada (prag je označen vertikalnom linijom).



Sl. 11. Anvelopa audio signala motora koji koristi benzin sa kasnom promenom režima rada (prag je označen vertikalnom linijom).



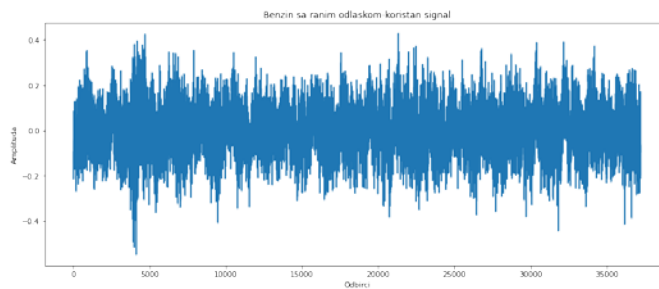
Sl. 12. Anvelopa audio signala motora koji koristi dizel sa kasnom promenom režima rada (prag je označen vertikalnom linijom).

Dobijeni vremenski trenuci koji označavaju prag (trenutak do koga se izdvaja deo signala praznog hoda motora koji se koristi za banku uzoraka) je na Sl. 9 do 12 označen ljubičastom vertikalnom linijom. Na snimljenim uzorcima gde nema promene režima rada motora, vremenski trenutak odsecanja nije mogao biti određen na opisani način. U tom slučaju, čitav audio zapis se koristi kao prazan hod motora i koristi se za dalje analize i obradu.

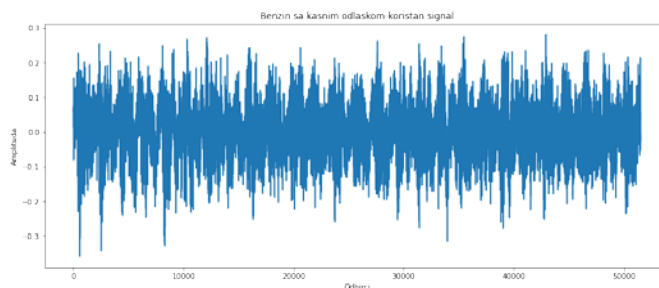
#### IV. REZULTATI

Obrada signala u vremenskom domenu, u vidu formiranja anvelope signala, kao i određivanje vremenskog trenutka kada je potrebno iseći signal na osnovu oblika anvelope, rezultirala je podatkom za svaki audio signal iz snimljene baze koji sadrži trenutak do kojeg je motor u režimu praznog hoda tj. do kojeg je signal stacionaran. Na osnovu tog podatka su iz postojećih audio zapisa upotrebom jednostavne funkcije izdvojeni željeni signali u celoj bazi. Izdvojeni signali su predstavljeni u wav formatu kao i originalni audio zapisi akvizicionog sistema. Radi ilustracije, vremenski oblici karakterističnih izdvojenih audio signala predstavljeni su na Sl. 13, 14 i 15. pri čemu se na horizontalnoj osi nalazi broj odbiraka. Na Sl. 13 je prikazan izdvojeni signal koji odgovara originalnom snimku motora koji koristi benzin sa ranim odlaskom vozila, dok je na Sl. 14. takođe prikazan snimak

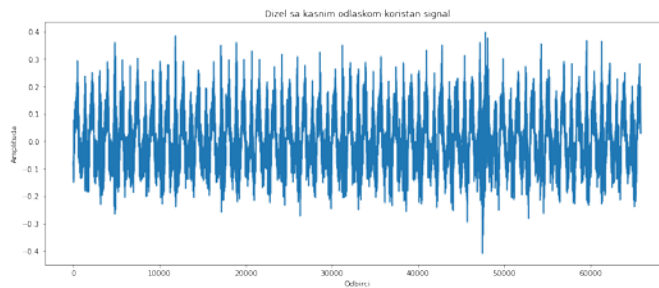
motora koji koristi benzin samo sa kasnijim odlaskom vozila. Posmatrajući vremenski oblik signala može se uočiti da su signali veoma slični pri čemu je jedina razlika u broju odbiraka tj. trajanju izdvojenog signala, što odgovara događajima vidljivim na Sl. 9 i 11.



Sl. 13. Izdvojeni audio signal motora koji koristi benzin u praznom hodu sa ranom promenom režima rada.



Sl. 14. Izdvojeni audio signal motora koji koristi benzin u praznom hodu sa kasnom promenom režima rada.



Sl. 15. Izdvojeni audio signal motora koji koristi dizel u praznom hodu sa kasnom promenom režima rada.

Na Sl. 15. je prikazan izdvojeni signal motora koji koristi dizel pri čemu je signal izdvojen za slučaj sa kasnijim odlaskom vozila ( Sl. 12.). Za slučaj kada vozilo koristi dizel i kada se dogodio rani odlazak vozila gotovo na samom početku snimka, definisana tačka sečenja je takođe bila veoma blizu početka signala (Sl. 10.) što znači da nije detektovan dovoljno dugačak stacionarni režim rada motora, pa je odgovarajućom funkcijom provere trajanja stacionarnog režima ovaj snimak odbačen bez izdvajanja signala. Opisana metoda, odnosno obrada signala u vremenskom domenu dala je jako dobre rezultate, jer je omogućila da se iz svih audio signala iz baze izdvoje željeni, korisni delovi signala, koji predstavljaju prazan hod motora, na potpuno automatizovan način, što je i bio cilj ovog rada.

#### V. ZAKLJUČAK

Posmatrajući broj snimaka sa signalima koji predstavljaju isključivo režim praznog hoda vozila u odnosu na broj uzorkovanih vozila moguće je uvideti da je akvizicioni sistem prikupio barem jedan takav snimak za svako vozilo. Razvoj dodatne obrade zabeleženih signala, a koji ne predstavljaju isključivo režim praznog hoda motora u toku čitavog trajanja

snimka, doprinelo je povećanju broja željenih audio snimaka, odnosno uvećanju banke uzoraka. Akvizicioni sistem sa dodatnom obradom signala za izdvajanje stacionarnog režima rada motora je dao identične rezultate bez obzira na tip pogonskog goriva. Upoređujući izdvojene signale na Sl. 13, 14 i 15 sa signalima koji su izvorno snimljeni kao režim praznog hoda motora (Sl. 1 i 2) mogu se videti veoma male razlike koje su rezultat vremenskog skaliranja prilikom grafičkog prikaza, odnosno jedina suštinska razlika je u dužini trajanja originalnog signala i korisnog dela signala koji je izdvojen pomoću opisane metode.

Ovakav način prikupljanja audio uzoraka, uz dodatnu obradu signala predstavljenu u ovom radu, veoma efikasno formira veliku količinu "čistih" audio snimaka. Na ovaj način napravljena je kvalitetna baza audio signala što će u budućem radu omogućiti razvoj sistema koji će nekim algoritmom mašinskog učenja, odnosno nekom od metoda klasifikacije prepoznavati o kojoj vrsti motora se radi.

#### ZAHVALNICA

Prezentovano istraživanje je realizovano zahvaljujući gospodinu Neđi Petijeviću ispred firme Novi Dom doo u Beogradu koji je omogućio pristup ulaznoj rampi podzemne garaže uz poštovanje svih bezbednosnih procedura. Ovaj rad je podržan od strane Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije.

#### LITERATURA

- [1] M. Milivojčević, F. Pantelić, D. Ćirić, "Comparison of frequency characteristic of sound generated by internal combustion engines depending on fuel," Proc. 26th Noise and Vibration, Niš, Serbia, pp. 115-120, 6-7 December 2018.
- [2] M. Milivojčević, F. Pantelić, D. Ćirić, "Pozicioniranje mikrofona prilikom snimanja audio karakteristika motora putničkih vozila" Proc. 63rd National Conference on Electrical, Electronic and Computing Engineering ETRAN, Srebrno Jezero, Serbia, pp. 58-62, 3-6 June 2019.
- [3] M. Milivojčević, D. Ćirić, „Izdvajanje značajnih akustičkih karakteristika motora sa unutrašnjim sagorevanjem“ Proc. 64rd National Conference on Electrical, Electronic and Computing Engineering ETRAN, Belgrade, Serbia, 2020.
- [4] J. F. Forren and D. Jaarsma, "Traffic monitoring by tire noise," Proc. Conference on Intelligent Transportation Systems, pp. 177-182, 1997.
- [5] R. Lopez-Valcarce, C. Mosquera, R. Perez-Gonzalez, "Estimation of road vehicle speed using two omnidirectional microphones: A

maximum likelihood approach", *EURASIP J. Appl. Signal Process.*, p.1059, 2004.

- [6] B. Barbagli, G. Manes, R. Facchini, "Acoustic sensor network for vehicle traffic monitoring," Proc. 1st International Conference on Advances in Vehicular Systems, Technologies and Applications, Venice, Italy, pp. 1-6, 24-29 June 2012.
- [7] V. Singh, N. Meena, "Engine fault diagnosis using DTW, MFCC and FFT," Proc. First International Conference on Intelligent Human Computer Interaction (IHCI 2009), Allahabad, India, pp. 83-94, 20-23 January 2009.
- [8] S. K. Lee, S. J. Kim, "Internal combustion engine sound-based fault detection and diagnosis using adaptive line enhancers." *Proc. Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 222, no. 4, pp. 593-605, Apr. 2008.
- [9] J.L. Firmino, J.M. Neto, A.G. Oliveira, J.C. Silva, K.V. Michina, M.C. Rodrigues, "Misfire detection of an internal combustion engine based on vibration and acoustic analysis," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, no. 43, art. no. 336, 2021.
- [10] J. George, L. Mary, Riyas K S, "Vehicle detection and classification from acoustic signal using ANN and KNN," Proc. Int. Conf. on Control Communication and Computing (ICCC), pp. 436-439, 2013.
- [11] Y. Na, Y. Guo, Q. Fu, Y. Yan, "An acoustic traffic monitoring system: design and implementation," 2015 IEEE 12th Int. Conf. on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Int. Conf. on Autonomic and Trusted Computing and 2015 IEEE 15th Int. Conf. on Scalable Computing and Communications and its Associated Workshops, pp. 119-126, 2015.

#### ABSTRACT

In this paper, procedure for recognition and extraction of idling mode of the internal combustion engines of passenger vehicles based on audio recordings acquired at the entrance of underground parking is presented. Audio recording analysis is carried out in the time domain in order to have signal processing with as little as possible requirements regarding necessary processing power of the acquisition system with battery power supply. Overall procedure for recognition and extraction of useful parts of audio recordings related only to idling mode of the internal combustion engines of passenger vehicles is realized using *Python* programming language. The main goal of the presented procedure is the preparation of large number of acquired audio signals for further analysis and processing in the context of extraction of characteristic audio features and application of these features in neural network training.

#### Extraction of the idling mode of the internal combustion engine based on the audio recording

Marko Milivojčević, Emilija Kisić, Dejan Ćirić

# Whispered Speech Recognition Based on DTW algorithm and $\mu$ FCC feature

Branko R. Marković and Jovan Galić

**Abstract**—This paper presents the results of normal and whispered speech recognition using the  $\mu$ FCC ( $\mu$ -law Frequency Cepstral Coefficients) feature. This feature uses a warping frequency function and it is applied at the front-end of ASR. The Dynamic Time Warping algorithm is used at the back-end of the ASR system. All experiments were performed using the part of the Whi-Spe database. Four scenarios are analyzed: normal/normal, whisper/whisper, normal/whisper and whisper/normal in the speaker dependent mode. The results confirmed an expected improvement in recognition of whispered speech compared to the standard LFCC and MFCC features.

**Keywords**—  $\mu$ FCC ( $\mu$ -law Frequency Cepstral Coefficients); whispered speech; DTW (Dynamic Time Warping); speech recognition.

## I. INTRODUCTION

THE speech has different modes and one of standard classification is: whisper, soft speech, normal (neutral), loud and shout [1]. Very interesting is whispered speech because is quietly different compared to normal, and at same time is intelligible and in most cases easy to understand. A lot of researches who were involved in normal speech recognition also are trying to apply different tools for whisper [2-4]. They made different results with more or less success.

This paper analyzes different scenarios related to whispered and normal speech. For this purpose the DTW algorithm is used with specific warping scale ( $\mu$  warping).

The DTW (Dynamic Time Warping) algorithm [5] is known as “old” pattern matching method for back-end ASR systems. There are many different new method like HMM (Hidden Markov Models), DNN (Deep Neural Networks), SVM (Support Vector Machines) etc. but for quick and valuable compression DTW is still very successful. Many researchers use the DTW as a method for initial classification of patterns, and then use other methods for more precise results.

For this research speech patterns from the Whi-Spe database [6] are used. The database contains 10,000 patterns which are representation of 50 different words spoken in normal and whispered mode. Five male and five female

Branko R. Marković is with the Faculty of Technical Science Čačak, University of Kragujevac, Department of Computer Science and Software Engineering, Čačak, Svetog Save 65, Serbia (e-mail: brankomarko@yahoo.com)

Jovan Galić is with the Faculty of Electrical Engineering, University of Banja Luka, Department of Electronics and Telecommunications, Banja Luka, Bosnia and Herzegovina (e-mail: jovan.galic@etf.unibl.org)

volunteers were included in this recording. For recording the special acoustical room is used where noise is suppressed.

All experiments are based on three types of features: Mel-frequency cepstral coefficients (MFCC) plus delta, Linear frequency cepstral coefficients plus delta and  $\mu$ FCC cepstral coefficients (LFCC) plus delta. For all experiments the following training/test scenarios are examined: comparison between normal and normal patterns (N/N scenario), comparison between whisper and whisper patterns (W/W), comparison between normal and whisper patterns (N/W) and comparison between whisper and normal patterns (W/N).

The paper has the following structure: the second part explains how to obtain MFCC and LFCC feature vectors from the initial wave files. The third part explains how to obtain  $\mu$ FCC feature vectors. The fourth part shows the results of experiments for all mentioned features. The final remarks and hints for further research are presented at the conclusion.

## II. MFCC AND LFCC FEATURES EXTRACTION

Mel-frequency cepstral coefficients are traditionally very popular feature for speech characterization. The mel-frequency scale (Fig. 1) emulates human’s ear perception. The frequency in mel is calculated using the following equation:

$$f[mel] = 2595 * \log_{10}(1 + f[Hz]/700) \quad (1)$$

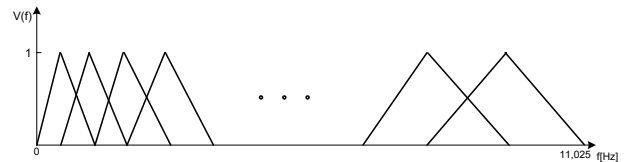


Fig. 1. Filters based on mel scale

Linear-frequency scale has the same shape for all filters (Fig. 2). The Linear frequency cepstral coefficients shows some advantage compare to MFCC in case of speaker identification in whisper [7].

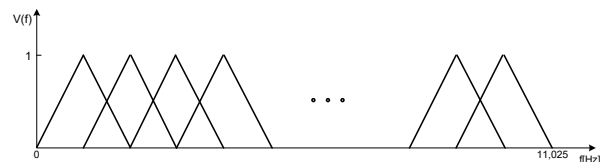


Fig. 2. Filters based on linear scale

The way to obtain MFCC and LFCC features is depicted in Fig. 3. It is a process of getting usually three types of vectors as an output: vectors of cepstral coefficients, vector of cepstral and delta cepstral coefficients and vectors of cepstral, delta

cepstral and delta-delta cepstral coefficients. For this research the first two types are used.

The inputs are wave files from Whi-Spe database [6]. All patterns are recorded with sampling rate of 22050 Hz, 16 bits per sample. Only difference between MFCC and LFCC features is in the scale which is used: MFCC counts the log energy over mel scale while LFCC counts the log energy over linear scale.

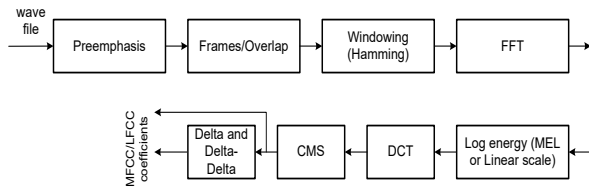


Fig. 3. Block diagram for MFCC/LFCC based features

The preprocessing assumes a several steps to get the feature vectors from an initial wave file (Fig. 3). The following steps are performed: preemphasis, framing with overlap, Hamming window, Fast Fourier Transformation (FFT), Log energy over a specific scale, DCT (Discrete Cosine Transform) and CMS (Cepstral Mean Subtraction) normalization.

The first is preemphasis block and it produces a spectrally flattened signal. Then, in the framing/overlap block, the signal as an output of preemphasis, is divided into frames. Each frame contains 512 samples, and then it is overlapped 50%. In next block the frames are weighted with the Hamming window.

The next step is the FFT, which calculates short time spectra of the signal. Then, the Log energy is calculated over the specific scale (mel or linear).

Finally, the Discrete Cosine Transformation with CMS are applied to produce the cepstral coefficients.

The CMS is a normalization method and is very important approach for whispered speech recognition [10,11].

For calculation of the first derivative (Delta), three neighboring frames are included.

Based on the mel scale and the preprocessing, two types of vectors are produced:

- vector containing 12 MFCCs and
- vector containing 24 coefficients (12 MFCCs and 12 Delta MFCCs).

Similarly, based on the linear scale the following vectors are obtained:

- vector containing 12 LFCCs and
- vector containing 24 coefficients (12 LFCCs and 12 Delta LFCCs).

These types are used in all experiments.

### III. $\mu$ FCC FEATURE EXTRACTION

The reason to involve  $\mu$ FCC feature is in the following: due to unvoiced nature of whispered speech the spectrum is relatively flat. Some significant part of whispered information is in higher part of speech spectrum. The mel scale, due to its

nature, is not able to “catch” these information. The linear scale shows better performances for part of higher frequencies but has worse resolution for lower frequencies. So, as a compromise, the new, warping function of frequency is involved [8,9].

This function is called  $\mu$ -law and originally is involved for speech compression and expanding in Japan and North America. The  $\mu$ -law is defined by the following equation:

$$f_{\mu} = f_N \frac{\ln(1 + \mu \frac{f}{f_N})}{\ln(1 + \mu)} \quad (2)$$

where  $f_N = f_S / 2$  ( $f_S$  is the sampling frequency), and for these experiments  $f_N = 11025$  Hz. The  $\mu$  is a positive number and can have different values. For this research  $\mu$  takes values  $\{0,1,2\}$ . Fig. 4 shows the warping functions of  $\mu$ -law for these three values  $\{0,1,2\}$  [9]. For  $\mu = 0$ , the scale is linear, and practically the feature are LFCC, as mentioned before.

The  $\mu$ FCC feature should make some compromise between linear and mel scale with focus to improve whispered speech recognition.

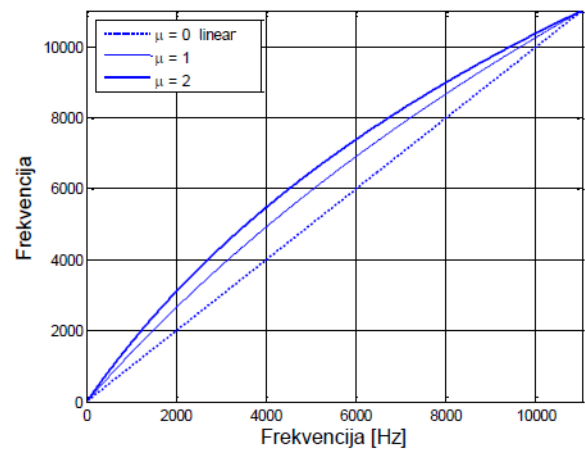


Fig. 4. Warping functions

In order to obtain  $\mu$ FCC features the block diagram from Fig. 5. is used. It is clone to earlier mention diagram for LFCC and MFCC feature (Fig. 3). The main difference is usage of warping function over frequencies when Log energy is calculated.

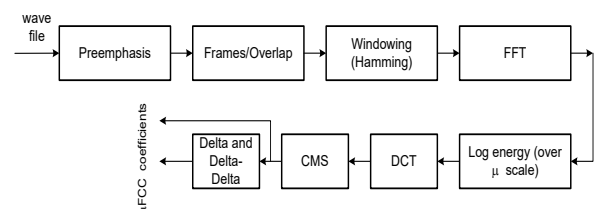


Fig. 5. Block diagram for  $\mu$ FCC based features

For these experiments two values for  $\mu$  are used: 1 and 2. So, for  $\mu=1$  two types of vectors are considered:

- vector containing 12  $\mu$ FCCs ( $\mu$ -law Frequency Cepstral Coefficients) and
- vector containing 24 coefficients (12  $\mu$ FCCs and 12 Delta  $\mu$ FCCs).

Similarly, for  $\mu=2$  two types of vectors are produced. All these vectors are used in all experiments.

#### IV. RESULTS

For the purpose of testing these different features a software package is developed using the MATLAB. There are two parts of this software: the first one converts the targeted wave files from the Whi-Spe database into the set of MFCC, LFCC, and  $\mu$ FCC feature vectors (two type of vectors for all of them, and for  $\mu$ FCC two different values of  $\mu$  are used). The second part compares feature vectors using the DTW algorithm.

The DTW algorithm uses the dynamic programming and it allows finding an optimal path between the starting and ending points. The speech patterns are represented by a set of feature vectors. Two set of patterns are used: the first set of 50 patterns is used as a reference, and all other patterns (nine sets, each of 50 patterns) are used as test data. For a local constraint the type I is implemented [12]. No global constraints are used.

This research used two (of ten) speakers from Whi-Spe database: one female (Speaker1) and one male (Speaker6).

For all types of feature vectors mentioned before, the results are expressed as the Word Recognition Rate (WRR). Four scenarios in speaker dependent mode are analyzed: normal/normal (denoted as N/N), whisper/whisper (W/W), normal/whisper (N/W) and whisper/normal (W/N). Tables I and II shows results for “Speaker1” and “Speaker6” using four different vector features with 12 cepstral coefficients.

TABLE I  
WORD RECOGNITION RATE FOR “SPEAKER1” USING 12 CEPSTRAL COEFFICIENTS

Scenario	LFCC	$\mu$ FCC ( $\mu=1$ )	$\mu$ FCC ( $\mu=2$ )	MFCC
N/N	98.89	99.11	99.56	99.78
W/W	95.56	96.89	97.11	97.78
N/W	81.56	84.75	85.11	78.22
W/N	67.78	66.44	64.00	46.44

TABLE II  
WORD RECOGNITION RATE FOR “SPEAKER6” USING 12 CEPSTRAL COEFFICIENTS

Scenario	LFCC	$\mu$ FCC ( $\mu=1$ )	$\mu$ FCC ( $\mu=2$ )	MFCC
N/N	96.44	98.22	98.67	99.33
W/W	89.33	92.89	95.11	95.11
N/W	62.00	65.56	68.00	68.00
W/N	51.33	50.89	49.78	36.44

Tables III and IV give results for “Speaker 1” and “Speaker 6” using all mentioned feature vectors with 12 cepstral and 12 delta cepstral coefficients.

TABLE III  
WORD RECOGNITION RATE FOR “SPEAKER1” USING 12 CEPSTRAL AND 12 DELTA CEPSTRAL COEFFICIENTS

Scenario	LFCC	$\mu$ FCC ( $\mu=1$ )	$\mu$ FCC ( $\mu=2$ )	MFCC
N/N	98.67	98.89	99.56	99.56
W/W	96.00	97.11	97.11	97.78
N/W	82.00	84.44	84.89	78.22
W/N	66.44	68.00	64.44	45.78

TABLE IV  
WORD RECOGNITION RATE FOR “SPEAKER6” USING 12 CEPSTRAL AND 12 DELTA CEPSTRAL COEFFICIENTS

Scenario	LFCC	$\mu$ FCC ( $\mu=1$ )	$\mu$ FCC ( $\mu=2$ )	MFCC
N/N	96.22	97.78	98.44	99.11
W/W	88.67	93.11	94.22	94.89
N/W	63.11	66.44	67.33	67.56
W/N	51.33	50.89	49.33	37.56

Based on the results from Tables I and II can be concluded that the MFCC feature gives very good results for match scenarios (Normal/Normal and Whisper/Whisper). But for mismatch scenarios  $\mu$ FCC feature is giving better results than MFCC (about 8% for N/W scenario, 30% for W/N scenario – for Speaker1 and about 28% for W/N scenario - for Speaker6). Also, for match scenarios  $\mu$ FCC feature gives better result than LFCC feature for both speakers.

The results in Tables III and IV are based on vectors with 12 cepstral coefficients plus 12 delta cepstral coefficients. In some cases there are improvements related to the word recognition rate with these vectors, but they are not significant (i.e. for Speaker1 and W/N scenario,  $\mu$ FCC feature ( $\mu=1$ ) gives 1,5% better result than for cepstral).

In general, based on results from Tables I-IV it is easy to conclude that the Speaker1 has better results in all scenarios compared to Speaker6. Speaker1 is better “whisperer”. Hence, on Fig. 6. the results of Speaker1 are depicted for the feature vectors which contains 12 cepstral coefficients.

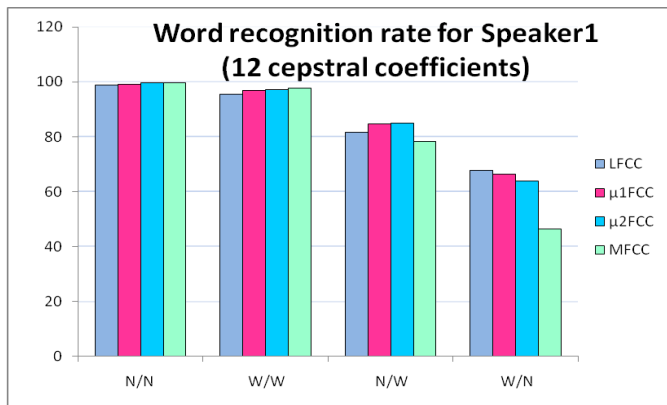


Fig. 6. WRR for Speaker1 using vectors of 12 cepstral coefficients

As it expected, the best results are for N/N, W/W, N/W and W/N scenarios, respectively. With  $\mu 1FCC$  the  $\mu FCC$  feature where  $\mu=1$ , is denoted. Similarly,  $\mu 2FCC$  means  $\mu=2$ .

It is interesting that  $\mu FCC$  feature for N/W scenario gives better results than LFCC and MFCC.

Fig. 7 shows results of Speaker1 for all scenarios when vectors are containing 12 cepstral plus 12 delta cepstral coefficients.

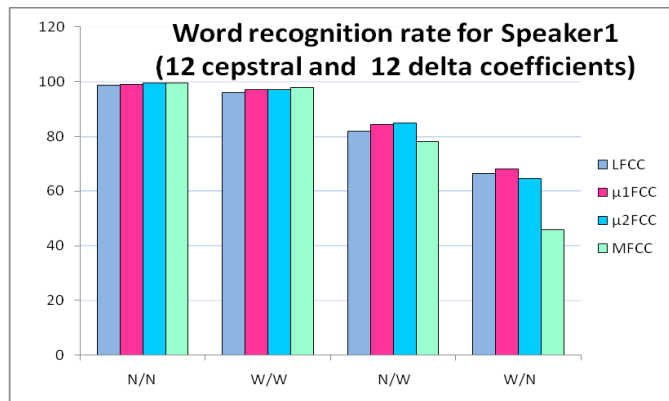


Fig. 7. WRR for Speaker1 using 12 cepstral plus 12 delta coefficients

If Fig. 6 and Fig. 7 are compared the similar trend for all scenarios and all features is evident. Only, for W/N scenario LFCC and  $\mu 1FCC$  changed their places.

## V. CONCLUSION

As it expected, the best recognition results are obtained for Normal/Normal scenario and they are above 99% when MFCC feature is used. Also for Whisper/Whisper scenario the WRR is the best with MFCC. So, for match scenarios MFCC

gives good results.

When mismatch scenarios are analyzed  $\mu FCC$  allows better results than MFCC. This is especially visible for W/N scenario where the improvement is from 28% to 30%. Obviously, these results are optimistic and give a hint to make more detailed research how the values of  $\mu$  cause different WRR.

Comparing the length of feature vectors (12 cepstral coefficients vs. 24 coefficients -12 cepstral and 12 delta) the results are similar. The reason behind it can be "clean" speech in Whi-Spe database, while the delta parameters are usually efficient for noisy speech.

Further analysis may include all ten speakers from Whi-Spe database, and also more different values for  $\mu$ . Instead of {0,1,2} values it can be numbers with decimal point [9]. That should provide new interesting results.

## REFERENCES

- [1] C. Zhang, J.H.L. Hansen, "Analysis and classification of Speech Mode: Whisper through Shouted," *Interspeech 2007*, 2007, pp. 2289-2292
- [2] S.T. Jovičić, Z.M. Šarić, "Acoustic analysis of consonants in whispered speech," *Journal of Voice*, 22(3), 2008, pp. 263-274.
- [3] Matsuda M., Kasuya H. (1999). "Acoustic nature of the whisper", *Proc. Eurospeech 99*, 1, 1999, pp. 137-140.
- [4] S.T. Jovičić, "Formant feature differences between whispered and voiced sustained vowels," *ACUSTICA - Acta Acustica*, 84(4), 1998, pp. 739-743.
- [5] L. Rabiner, B-H. Juang, "Fundamentals of speech recognition", Prentice Hall, New Jersey, 1993
- [6] B. Marković, S.T. Jovičić, J. Galić, Đ. Grozdić, "Whispered Speech Database: Design, Processing and Application, 16<sup>th</sup> International Conference, TSD 2013, I. Habernal and V. Matousek (Eds.): TSD 2013, LNAI 8082, Springer-Verlag Berlin Heidelberg, pp. 591-598
- [7] Fan X., Hansen J.H.L. Speaker identification with whispered speech based on modified LFCC parameters and feature mapping. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*, 2009. pp. 4553-4556.
- [8] Sklar B. *Digital Communications: Fundamentals and Applications*: 2nd edition. Prentice-Hall. 1988. 776 p.
- [9] J. N. Galić, S. T. Jovičić, V. D. Delić, B. R. Marković, D. S. Šumarac Pavlović, Đ. T. Grozdić, "HMM-based Whisper Recognition using  $\mu$ -Law Frequency Warping", *SPIIRAS Proceedings*, Issue No 3(58), 2018, pp. 27-52, ISSN 2078-9181, DOI 10.15622/sp.58
- [10] J. De Veth, L. Boves, "Channel Normalization Techniques for Automatic Speech Recognition over the Telephone", - *Speech Communication*, 25, pp. 149-164, 1998.
- [11] Grozdić Đ., Jovičić S., Šumarac-Pavlović D., Galić J., Marković B. (2017). "Comparison of Cepstral Normalization Techniques in Whispered Speech Recognition", *Advances in Electrical and Computer Engineering*, Vol. 17. Number 1, 2017, pp 21-26.
- [12] Sakoe H. and Chiba S. (1978). "Dynamic programming optimization for spoken word recognition", *IEEE Trans. Acoustics, Speech, Signal Proc.*, pp 43-49, 1978



# The Experiments in SVM-based Whispering Speaker Identification

Jovan Galić, Branko Marković and Đorđe T. Grozdić

**Abstract**—This paper presents results of automatic speaker recognition in normally phonated (neutral) and whispered speech, based on Support Vector Machines (SVM) and Whi-Spe speech database. The performance of the recognizer is examined in matched N/N (Neutral/Neutral) and W/W (Whispered/Whispered) train/test scenarios for different types of kernels (Radial basis function, Polynomial, Linear, and Sigmoid). The best accuracy is obtained with a polynomial kernel (96,12% for neutral speech and 92,16% in case of whispering). The influence of the size of training data on the performance of the recognizer is examined, as well.

**Keywords**—Speaker recognition; Whispered speech; Whi-Spe database; MFCC; SVM algorithm.

## I. INTRODUCTION

The technology of automatic speech and speaker recognition has made significant progress in the last two decades. Still, some disadvantages remain. The key imperfection is the considerable degradation of the performance in adverse conditions [1]. As well, speech technologies are designed for recognition of the most commonly used mode of phonation, i.e. neutral speech. Speech mode can be classified into the five main categories: whispered speech, soft speech, normally phonated speech (neutral speech), loud speech and shouted speech [2].

Nowadays, whisper is often used in a daily life, especially over the mobile phones. There are multiple reasons to use whisper: when someone doesn't like to disturb others, when the loud speech is prohibited or unpleasant, when the information to speak is secret, when someone wishes to hide identity etc. Also, whisper can be produced due to health problems: it may happen after laryngitis or rhinitis [3]. Whisper as a speech mode is characterized by a lack of glottal vibration, noisy excitation of the vocal tract and in general, the changes of the vocal tract structure.

Jovan Galić is with the Faculty of Electrical Engineering, University of Banja Luka, Department of Electronics and Telecommunications, Banja Luka, Patre 5, Bosnia and Herzegovina (e-mail: jovan.galic@etf.unibl.org)

Branko Marković is with the Faculty of Technical Science Čačak, University of Kragujevac, Department of Computer Science and Software Engineering, Čačak, Svetog Save 65, Serbia (e-mail: brankomarko@yahoo.com).

Đorđe T. Grozdić is with Grid Dynamics Holdings, Inc, Blvd. Mihajla Pupina 115, Belgrade, Serbia; and School of Electrical Engineering, University of Belgrade, Blvd. Kralja Aleksandra 73 (e-mail: djordjegrozdic@gmail.com).

There are main differences between neutral and whispered speech. It was determined that formant frequencies for whispered vowels are substantially higher than for the neutral voice [4]. Also, compared to neutral speech, whisper has less energy, longer durations of speech and silence intervals, flatter spectrum and lower sound pressure level (SPL) [2]. Despite of these “weaknesses”, the intelligibility of whisper is pretty high [5]. But, non-linguistic information (like age, sex, emotions or identity), is still a big challenge for research in whispered speech.

The oscillogram and spectrogram of the short sentence "Govor šapata." ("Whispered speech" in English), uttered in neutral and whispered speech are depicted in Figures 1 and 2, respectively. Because of the lack of sonority, the difference in amplitude intensities can be observed, especially for vowels [6]. The analysis of spectrograms shows that some parts of the spectrum are well preserved in whisper, especially in the case of unvoiced consonants and plosives. Moreover, the spectrogram shows that the harmonic structure of vowels is lost in the case of a whisper [7].

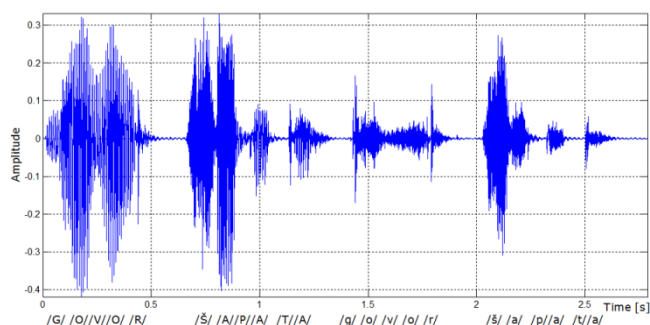


Fig. 1. The waveform of sentence "Govor šapata" in neutral phonation (capital letters) and whispered phonation (small letters).

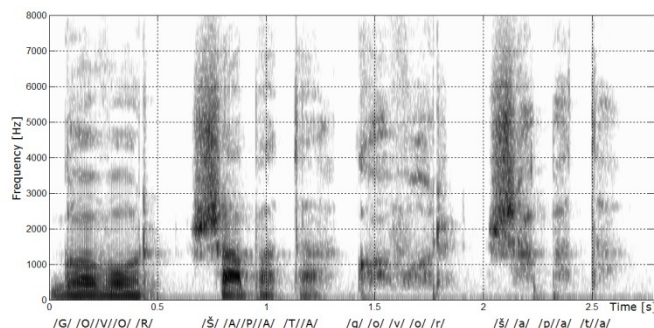


Fig. 2. The spectrogram of sentence "Govor šapata" in neutral phonation (capital letters) and whispered phonation (small letters).

Whispering speaker identification is a great challenge for state-of-the-art speaker recognition systems. In a range of speech modes from whisper to shouted, whispered speech has the most negative influence on the performance of Automatic Speech/ Speaker Recognition (ASR) systems [2]. Recently, the use of Gaussian Mixture Models and K-means algorithms has been analyzed in whispering speaker ID for mel and exponential frequency scales [8]. Also, formant gap features showed higher accuracy in speaker verification compared to baseline features [9].

Automatic speaker recognition can be classified into identification and verification. Methods for speaker ID can be divided into text-independent and text-dependent [10]. For a text-independent ASR system, models for a particular speaker are irrespective of uttered speech, whereas in a text-dependent system the performance of speaker recognition depends on uttered phrases. In this research, text-dependent closed set speaker identification based on Support Vector Machines (SVM) and Mel-Frequency Cepstral Coefficients (MFCC) was analyzed.

The goal of the study presented in this paper is to analyze speaker identification accuracy for neutral and whispered speech, and classification based on SVM. This paper is organized in the following manner. In Section II the classification based on SVM is shortly discussed. The basic characteristics of the ASR system and speech database used for speaker recognition are described in Section III. The results of conducted experiments are given in Section IV whereas concluding remarks and directions for future research are stated in Section V.

## II. SUPPORT VECTOR MACHINES

The SVM classifier is a relatively simple machine-learning algorithm that minimizes the structural risk [11]. Initially, the SVM classifier was introduced for linearly separable classes of objects. The separation of classes is obtained with an  $n$ -dimensional hyperplane that maximizes the margin between classes (circles and squares), as depicted in Figure 3. The margin is labeled as  $M$  and support vectors are hatched.

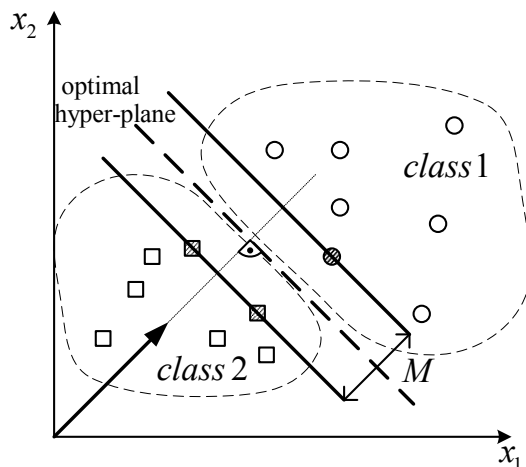


Fig. 3. Determination of hyperplane in SVM for linearly separable classes.

However, the classes are not linearly separable in most practical applications. To overcome that limitation, a non-linear transformation is performed on a feature vector. The mapping into high-dimensional feature space (in which linear separation is expected) is performed by using kernel function. Each function that satisfies necessary properties (Mercer's theorem) can be used as a kernel [12]. The most used types of kernels are:

- Radial basis function kernel (adjustable parameter  $\gamma$ )

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right); \gamma = \frac{1}{2\sigma^2}. \quad (1)$$

- Polynomial kernel (adjustable parameters are the slope  $\alpha$ , the constant term  $c$  and the polynomial degree  $d$ )

$$K(x_1, x_2) = (\alpha x_1^T x_2 + c)^d. \quad (2)$$

- Linear kernel (adjustable parameter  $c$ )

$$K(x_1, x_2) = x_1^T x_2 + c. \quad (3)$$

- Hyperbolic Tangent (Sigmoid) kernel (adjustable parameter are  $\alpha$  and the constant term  $c$ )

$$K(x_1, x_2) = \tanh(\alpha x_1^T x_2 + c). \quad (4)$$

Because SVM is a static classifier, meaning that it works with fixed-size input data, the application in speech/speaker recognition has some restrictions. Some hybrid solutions were developed to overcome that limitation [13]. Another issue is multiclass classification, which is commonly solved by using one of the two following techniques. The first technique includes comparison of each class against all the others (one-vs-all) and the second technique confronts each class against all the others separately (one-vs-one). In this study, a one-vs-all comparison strategy is used.

## III. SYSTEM FOR RECOGNITION

### A. Speech database

One of the real problems related to the whispered speech research is a shortage of an extensive speech databases. There are some of them developed so far [14-17] for Japanese, Mandarin and English. In order to do this research, the Serbian speech database called Whi-Spe (abbreviation of *Whispered Speech*) is used [18]. The database contains two parts: the first one has recordings of whispered words, and the second one has recordings of the same words uttered with neutral speech. The vocabulary of 50 different words is divided in three groups: basic colors (6 words), numbers (14 words) and phonetically balanced (30 words). For recordings of the Whi-Spe ten volunteers (5 female and 5 male) uttered the vocabulary ten times in both speech modes, neutral and whisper. Hence, the speech database contains 10.000 represents of words in form of wave files and the total duration is 2 hours.

More details about the Whi-Spe database regarding segmentation procedure and quality control can be found in [18].

### B. ASR system

Because SVM-based classifiers need feature vectors of fixed dimension, the variation in the duration of input speech utterances must be uniform. The two most common approaches for making a fixed number of frame windows for SVM classifier are using variable window size (with constant overlapping factor) and fixed window size (with variable overlapping factor). This causes some loss of information, especially in long speech utterances. In this paper, segmentation based on variable window size is chosen, using 13 overlapping windows, same as in the SVM-based speech recognition [19].

### C. Feature vector extraction

The most common features used in ASR systems are Mel Frequency Cepstral Coefficients (MFCC). The diagram for obtaining the MFCC feature vector is depicted in Fig. 4.

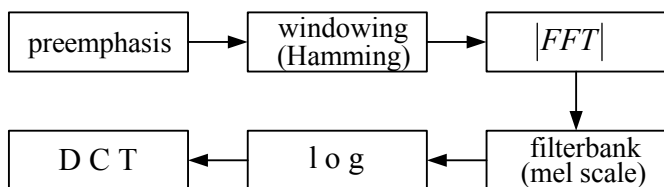


Fig. 4. The diagram for MFCC feature vector generation.

The generation of MFCC feature vectors includes the following steps: preemphasis, framing with overlap and Hamming windowing, application of the Fast Fourier Transformation (FFT), using the mel scale, calculating log energy and finally obtaining the cepstral, delta cepstral and delta-delta cepstral coefficients (based on Discrete Cosine Transformation - DCT).

The MFCC feature vector is obtained by using static features (13) along with time derivatives (delta and delta-delta) and cepstral mean normalization (39 in total). Lastly, each speech utterance from the database is represented with a vector of 507 coefficients (13 x 39) and later used as an input to the SVM classifier. The speech recognizer is developed with Python (version 3.6) using the Scikit learn package.

## IV. RESULTS

In the initial experiment baseline speaker ID performance are evaluated in 2 matched train/test scenarios: N/N and W/W. In order to have a more reliable evaluation of the performance, 10-fold cross-validation was done. The average accuracy is used as a metric for performance evaluation.

Firstly, the influence of kernel selection on recognizer performance was analyzed. The experiments were done for 4 kernels: radial basis function (RBF), polynomial (with degree  $d=3$ ), linear and sigmoid.

The results are graphically presented in Figure 5.

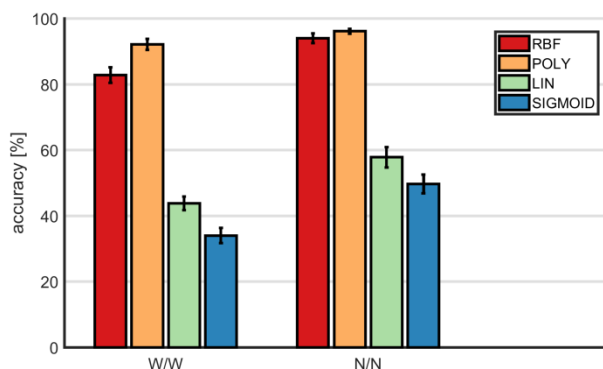


Fig. 5. Results of speaker identification (average accuracy with standard deviation) in neutral (N/N) and whispered speech (W/W) depending on type of kernel.

As it can be seen from Figure 5, the best results in whispering speaker ID are obtained for the polynomial kernel (92,16%). As well, the best result in recognition of speakers in neutral phonation is obtained using the polynomial kernel (96,12%). The experiments showed that speaker identification performance for RBF kernel is poor compared to poly kernel. Linear and sigmoid kernels are practically useless and show poor performance, especially for speaker identification in whisper mode.

As well, the influence of the percentage of the database used for training (i.e. number of training instances) on classification performance was examined. In this experiment, the polynomial kernel was used because it showed the best recognition in the previous experiment. The results are presented in Figure 6. An important part of bar graph is emphasized (higher than 75%).

For the speaker recognition in neutral mode the recognition performance starts from 94,71% (for training 75% of full capacity database, i.e. 3750 utterances) and reaches final 96,12% (for 95%, i.e. 4750 utterances). On the other hand, whispering speaker recognition performance is in the range of 89,49% (for 75%) up to 92,16% (for 95%). As observed, the saturation effect can be seen for both neutral and whisper speech modes (saturation effect is less for whispered speech).

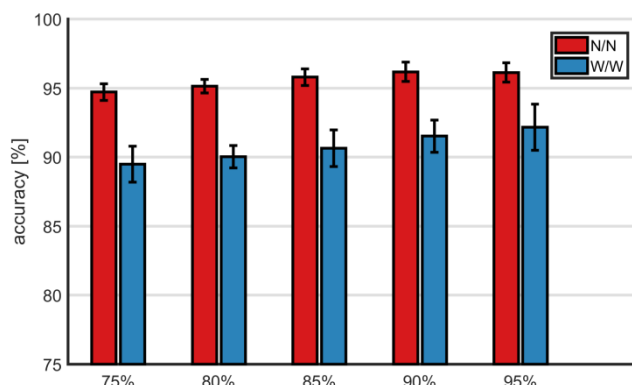


Fig. 6. Results of speaker identification (average accuracy with standard deviation) in dependence of percentage of database used for training for neutral (N/N) and whispered speech mode (W/W).

The results in N/W (neutral/whisper) train/test scenario for polynomial kernel and each speaker are presented in Table I. As can be seen, compared to N/N scenario the performance are degraded and very high difference between different speakers can be observed. Similar observation is found in whispering speaker identification with neutral trained HMM models [20].

TABLE I  
ACCURACY FOR N/W SCENARIO (%)

Speaker	Accuracy
Speaker 1	65,2
Speaker 2	9,2
Speaker 3	21,4
Speaker 4	16,4
Speaker 5	62,0
Speaker 6	27,2
Speaker 7	25,4
Speaker 8	29,4
Speaker 9	33,4
Speaker 10	4,0
Average	<b>29,36</b>

## V. CONCLUSION

Whispered speaker recognition is, by all means, a serious challenge for modern ASR systems.

As expected, whispering speaker ID was shown to be more difficult than recognition of a speaker that utters neutral speech. In this paper experiments on speaker identification in neutral and whisper mode for Whi-Spe speech database and SVM algorithm were conducted.

Future studies will be focused on examining more robust feature vectors in whispering speaker identification. Also, the application of the Teager energy operator [21-23] has shown improvements in robustness for whispered speech recognition, so there are reasonable expectations that it could help in speaker recognition as well.

As well, different machine learning algorithms are going to be analyzed (Gaussian mixture models and Neural networks).

## ACKNOWLEDGEMENT

This research is supported by the project "Razvoj Internet of Things (IoT) aplikacija primjenom optičko-bežičnih tehnologija" of Ministry for Scientific and Technological Development, Higher Education and Information Society of Republic of Srpska.

## REFERENCES

- [1] J. Holms, W. Holms, *Speech Synthesis and Recognition*, Taylor & Francis, London, United Kingdom, 2001.
- [2] C. Zhang, J. H. L. Hansen, "Analysis and classification of Speech Mode: Whisper through Shouted," *Proceedings of Interspeech 2007*, pp. 2289-2292, 2007.
- [3] T. Ito, K. Takeda, F. Itakura, "Analysis and recognition of whispered speech," *Speech Communication*, vol. 45, no. 2, pp. 139-152, 2005.
- [4] Y. Swerdlin, J. Smith, J. Wolfe, "The effect of whisper and creak vocal mechanisms on vocal tract resonances," *Journal of Acoustical Society of America* 127, pp. 2590-2598, 2010.
- [5] V. C. Tarrter, "Identifiability of vowels and speakers from whispered syllables", *Perception & Psychophysics*, vol. 49, no. 4, pp. 365-372, 1991.
- [6] J. Galić, B. Popović, D. Šumarac Pavlović, "Whispered Speech Recognition using Hidden Markov Models and Support Vector Machines," *Acta Politechnica Hungarica*, vol. 15, no. 5, pp. 11-29, 2018.
- [7] J. Galić, S.T. Jovičić, Đ. Grozdić, B. Marković, "Constrained Lexicon Speaker Dependent Recognition of Whispered Speech," *Proceedings of International Symposium on Industrial Electronics and Applications (INDEL)*, pp. 180-184, 2014.
- [8] A. Singh, A. M. Joshi, "Speaker Identification Through Natural and Whisper Speech Signal," *Optical and Wireless Technologies, Lecture Notes in Electrical Engineering*, vol. 546, pp. 223-231, 2020.
- [9] A. R. Naini, A. Rao, P. K. Ghosh, "Formant-gaps features for speaker verification using whispered speech," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 6231-6235, 2019.
- [10] S. Furui, "50 Years of progress in speech and speaker recognition," *Proceeding of the International Conference Speech and Computer SPECOM, Patras, Greece*, pp. 1-9, 2005.
- [11] P. Clarkson, P. J. Moreno, "On the use of support vector machines for phonetic classification," *IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP 99*, vol. 2, pp. 585-588, 1999.
- [12] Z. Čirković, Z. Banjac, "Jedna primena SVM klasifikatora u verifikaciji govornika nezavisno od teksta," *Proceedings of the International Symposium Infoteh Jahorina*, vol. 11, pp. 833-837, 2013.
- [13] Z. Qu, Y. Liu, L. Zhang, M. Shao, "A Speech Recognition System Based on a Hybrid HMM/SVM Architecture," *First International Conference on Innovative Computing, Information and Control (ICICIC) Beijing, China*, pp. 100-104, 2006.
- [14] N. Kawaguchi, K. Takeda, S. Matsubara, I. Yokoo, T. Ito, K. Tatara, T. Shinde, F. Itakura, "Ciair speech corpus for real world applications," *International Joint Conference of the 5th Symposium on Natural Language Processing*, pp. 288-295, 2002.
- [15] P. X. Lee, D. Wee, H. S. Yin Toh, B. P. Lim, N. Chen, B. Ma, "Whispered Mandarin Corpus for Speech Technology Applications," *Proceedings of Interspeech 2014*, pp. 1598-1602, 2014.
- [16] T. Tran, S. Mariooryad, C. Busso, "Audiovisual corpus to analyze whisper speech," presented at the *International Conference on Acoustics, Speech and Signal Processing*, pp. 8101-8105, 2013.
- [17] F. Cummins, M. Grimaldi, T. Leonard, J. Simko, "The chains corpus: Characterizing individual speakers," *Proceedings of International Conference on Speech and Computer SPECOM, St. Petersburg, Russia*, pp. 421-435, 2006.
- [18] B. Marković, S. T. Jovičić, J. Galić, Đ. Grozdić, "Whispered speech database: design, processing and application," In: Habernal, I., Matousek, V. (eds.), *TSD 2013, LNAI 8082*, Springer-Verlag Berlin Heidelberg, pp. 591-598, 2013.
- [19] J. M. Garcia-Cabellós, C. Peleaz-Moreno, A. Gallardo-Antolin, F. Perez-Cruz, F. Diaz-de-Maria, "SVM classifiers for ASR: A discussion about parameterization," *Proceedings of 12th European Signal Processing Conference*, pp. 2067-2070, 2004.
- [20] X. Fan, J. H. L. Hansen, "Speaker identification within whispered speech audio stream," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 5, 1408-1421, 2011.
- [21] Đ. Grozdić, "Application of Neural Networks in Whispered Speech Recognition", PhD dissertation, University of Belgrade, Belgrade, Serbia, 2017.
- [22] Đ. Grozdić, S. Jovičić, M. Subotić, "Whispered speech recognition using deep denoising autoencoder", *Engineering Applications of Artificial Intelligence*, vol. 59, pp. 15-22, 2017.
- [23] Đ. Grozdić, S. Jovičić, "Whispered speech recognition using deep denoising autoencoder and inverse filtering", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 12, pp. 2313-2322, 2017.

# Cepstrum-Based Pitch Detection of Industrial Product Sound

Dejan Ćirić, Marko Janković, Marko Milenković, Miljan Miletić

**Abstract**—Various audio features can be extracted from audio signals. One of very important ones is pitch. Different algorithms and methods have been proposed in literature to detect pitch. Among them, cepstrum-based pitch detection as a frequency domain method has often been used in practice. Cepstrum is calculated as the inverse Fourier transform of the logarithm of signal spectrum. The fundamental frequency and pitch in this way is estimated as the maximum value of cepstrum in the defined segment. Here, pitch of some industrial products (compressors and DC motors) are estimated by applying the modified cepstrum-based algorithm. The detected pitch values can be used to make a distinction between different working conditions of these products such as different rotation-per-minute (rpm).

**Index Terms**—Cepstrum analysis; Pitch detection; Peak finding; Audio Feature; Audio signals; Industrial product sound.

## I. INTRODUCTION

PITCH detection is a common task present in a number of researches mostly related to speech, since the pitch (or fundamental frequency) is one of the most important parameters of speech. Thus, detection of pitch can be found in different speaker recognition and identification systems, speech synthesis systems, telecommunication systems, etc. [1-3]. In addition, pitch is one of the audio features (attributes) used for audio classification, detection and recognition by applying machine or deep learning [4].

Pitch can be detected in the time or frequency domain. One of simple time domain algorithms (methods) is the zero crossing rate method. The most important methods in the time domain are typically based on auto-correlation using a hypothesis that the auto-correlation function of a periodic signal is also periodic and that these two periods are coincident [5].

Regarding the frequency domain methods for pitch detection, one of the most popular is cepstrum-based method. Power cepstrum has some similar properties with the complex

cepstrum obtained by homomorphic deconvolution [5], with the main difference that phase information is lost in the power cepstrum, which is called in the rest of the paper cepstrum.

Cepstrum, complex cepstrum, and homomorphic deconvolution have been applied in various areas such as audio processing, speech processing, geophysics, radar, medical imaging, etc. [5]. Some of the applications of both cepstrum and complex cepstrum include restoration of old phonograph recordings [6], cepstral pitch detector, speech recognition and speaker identification.

The periodicity present in a signal that is related to the pitch can be estimated from the cepstrum. Comparing with some other methods for pitch detection, the cepstrum method is able to provide accurate and robust results, but it is computationally complex [5].

This paper presents potentials for using pitch as an audio feature of sound of certain industrial products such as compressors and DC motors. The goal is to investigate if such a feature is able to provide clear distinction between different compressors or DC motors. The pitch is estimated by using cepstrum-based algorithm modified in a sense that peak finder is applied to the obtained cepstrum. Different compressors are related to compressors working with different rotations-per-minute (rpm), while different DC motors are related to different types of DC motors used in the automotive industry.

## II. PITCH DETECTION ALGORITHMS

Pitch is an important attribute of some audio signals such as speech signals. In speech, as a consequence of the vocal fold vibrations, the signal waveform contains certain periodicity translated into “pitch step” in the time domain and “fundamental frequency” or pitch in the frequency domain. However, pitch as an audio feature can also be of significance in machine and deep learning applied to a variety of audio signals including those containing sounds of industrial products, e.g. DC motors, home appliances or internal combustion engines of passenger vehicles [7].

There are various algorithms for pitch detection divided according to different criteria. Thus, there are block based and event based algorithms [9]. In the block based algorithms, the signal is sliced into small segments assuming that the pitch remains constant during the segments. On the other hand, event based algorithms use pitch marking or epoch detection. Here, pitch is not assumed to be constant over several pitch cycles. This is why these algorithms are able to track fast pitch changes even during the segments [10].

According to domain in which the algorithms are applied,

Dejan Ćirić is with the University of Niš, Faculty of Electronic Engineering in Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia (e-mail: dejan.ciric@elfak.ni.ac.rs).

Marko Janković is with the University of Niš, Faculty of Electronic Engineering in Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia (e-mail: marko.p.jankovic@elfak.rs).

Marko Milenković is with the Faculty of Arts, University of Niš, Kneginje Ljubice 10, 18105 Niš, Serbia (e-mail: lionheartmarko@gmail.com).

Miljan Miletić is with the Technical College of Applied Studies in Kragujevac, section in Kruševac, Kosančićeva 36, 37000 Kruševac, Serbia (e-mail: miljan.miletic@vhts.edu.rs).

they can be divided in three groups: time domain, frequency domain and hybrid group (combining time and frequency domain) [11]. Algorithms in the time domain use characteristics of a signal in the time domain, e.g. amplitude fluctuations, zero-crossing and auto-correlation attributes. This group contains the algorithms such as parallel processing time-domain method [12], data reduction method [13], modified auto-correlation (AUTO-C), average magnitude difference function (AMDF) and YIN [14].

Frequency domain algorithms are based on the property that periodicity in the time domain is translated into series of peaks (impulses) in the frequency domain. This group of algorithms contains methods such as harmonic product spectrum (HPS), cepstrum-based pitch detection, linear predicting coding (LPC) and sawtooth waveform inspired pitch estimator (SWIPE) [15].

Hybrid detectors combine both the time and frequency domain algorithms. In that regard, frequency domain algorithms can yield temporary spectral aligned sound waves, and after that, auto-correlation methods are applied to determine the pitch period [11]. Hybrid pitch detection methods include pitch estimation filter with amplitude compression (PEFAC) [16], YAAPT [17], multi-band summary correlogram (MBSC) [18] and BaNa [19].

#### A. AUTO-C as Time Domain Method

Among the time domain pitch detection methods, the most used one is the autocorrelation approach. It is based on finding the highest value of the auto-correlation function. Here, the auto-correlation function (AF) of a signal  $s(n)$  ( $n=0, 1, \dots, N-1$ ) is defined as

$$AF(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} s(n)s(n+k), \quad k = 0, 1, \dots, N-1, \quad (1)$$

where  $N$  is the signal (or frame) length, while  $k$  is the lag index. The pitch is detected at the location of the peak of auto-correlation function.

#### B. Cepstrum-based Algorithm as Frequency Domain Method

Cepstrum  $C(m)$  can be calculated as the inverse Fourier transform of the logarithm of Fourier transform of the target signal,  $s(n)$ :

$$\begin{aligned} S(k) &= \log \left\{ \sum_{n=0}^{N-1} s(n) \cdot \exp^{-j\frac{2\pi}{N}nk} \right\} \\ C(m) &= \frac{1}{N} \left\{ \sum_{k=0}^{N-1} S(k) \cdot \exp^{j\frac{2\pi}{N}mk} \right\}. \end{aligned} \quad (2)$$

The pitch is detected at the location of cepstrum maximum calculated as given in (2). It is presented in the literature that the cepstrum-based method is sensitive to noise in the target signal [11].

#### C. PEFAC as Hybrid Method

The pitch in the PEFAC algorithm is detected by convolving the power spectral density of the signal in the log-frequency domain with the filter summing the energy of the pitch harmonics [11]. The model at the time moment  $t$  of a

perfectly periodic signal (having fundamental frequency  $f_0$ ) in the power spectral density domain can be expressed as

$$Y_t(f) = \sum_{k=1}^K a_{k,t} \delta(f - kf_0) + N_t(f), \quad (3)$$

where  $N_k(f)$  is the power spectral density of the undesired noise and  $a_{k,t}$  is the power of the  $k$ -th harmonic. The signal model gets the following form in the logarithmic domain

$$Y_t(q) = \sum_{k=1}^K a_{k,t} \delta(q - \log(k) - \log(f_0)) + N_t(q), \quad (4)$$

where  $q = \log(f)$ . The energies of the signal components in this domain can be combined by convolving  $Y_t(q)$  with the impulse response filter.

The filter  $h(q)$  can suppress the noise with smoothly varying spectra, but this is not the case for high amplitude narrowband noise. This is why the spectrum compression is applied before the convolution with the filter  $h(q)$

$$Y_t'(q) = Y_t(q)^{a_t(q)} \quad (5)$$

where  $t$  represents time index and  $a_t(q)$  represents the compression exponent [11].

### III. METHODS APPLIED

For the purpose of carrying out this research, audio signals with different spectral contents are selected. Some of them have known pitch and harmonics distribution, such as the trumpet sound given in Fig. 1, while in the other signals (containing the sounds of certain industrial products) these parameters are unknown. The sound of trumpet is tonal sound with pronounced pitch, and the main characteristic of such a sound is periodicity, which can be seen in the time domain as presented in Fig. 2.

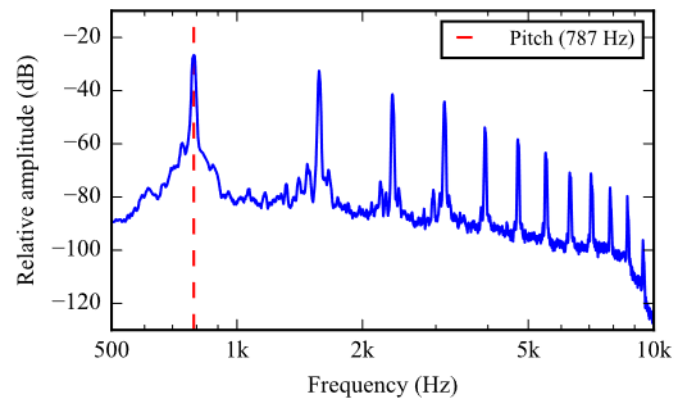


Fig. 1. Spectrum of trumpet sound consisting of fundamental ( $f_0$ ) at frequency of 787 Hz and partials that have a harmonic distribution relative to the fundamental.

Cepstrum-based pitch detection uses signal periodicity. In this case, the periodicity refers to a particular waveform of the specific length that is repeated throughout the signal, and it is reflected in a discrete spectrum containing prominent peaks equally distributed throughout the frequency range.

The cepstrum-based pitch detection algorithm consists of five main steps described below. In the first step, the signal is

transformed from the time to the frequency (spectral) domain using the Fourier transform. By applying the relevant function in Python, two variables are obtained - spectrum (given in complex numbers) and frequency vector. In the second step, the logarithm of the spectrum magnitude is calculated. The basic idea of cepstrum is to transfer the periodicity from the logarithmic representation of the spectrum to the time domain. For that reason, in the third step, the inverse Fourier transform is applied over the logarithm of the spectrum magnitude data. The cepstrum of the trumpet sound from Figs. 1 and 2 obtained in the described way is shown in Fig. 3.

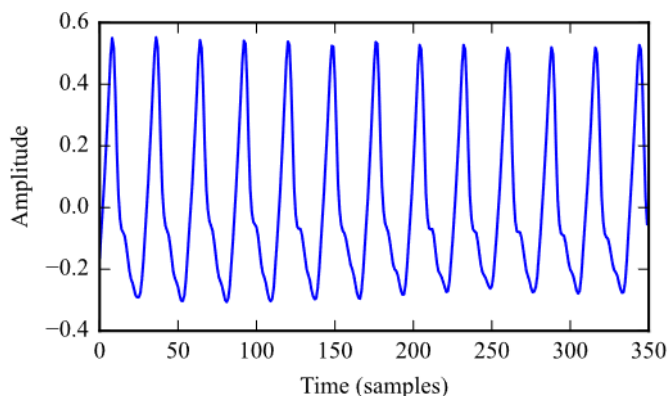


Fig. 2. Trumpet sound shown in time domain.

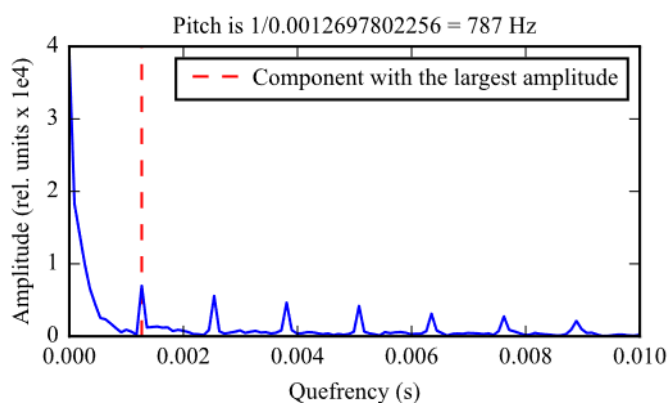


Fig. 3. Cepsrum of trumpet sound.

As can be seen in Fig. 3, the  $x$ -axis is titled quefrency, which is an anagram of frequency and it is related to time scale. The reason is in performing the inverse Fourier transform, that is, in inverting the frequency by this transformation. As a consequence, the highest frequencies are located at the beginning of the  $x$ -axis, while the lowest frequencies are located at the end of the  $x$ -axis. Frequencies are converted into quefrencies by taking  $1/\text{frequency}$ .

The fourth step is related to finding the maximum value of the cepstrum, that is, finding the quefrency value at which the cepstrum maximum occurs. In majority of cases, the maximum amplitude of the cepstrum is located at the beginning of the  $x$ -axis – at the zero quefrency or its vicinity, which greatly complicates the pitch detection in an automated manner. There are several approaches to overcome this

problem. One of them is to limit the quefrency band in which the search of cepstrum maximum is carried out. In this way, an error can be introduced in estimating the pitch of some sounds, such as sound of certain industrial products, since it is not known in advance in which frequency band to expect the pitch. The second approach is not to detect the pitch at the absolute maximum of the cepstrum, but instead to skip the cepstrum maximum at the zero quefrency, and to use the second largest value of the cepstrum for the pitch estimation. Such an approach leads to an error of pitch detection in the case where the cepstrum do not have the maximum value at the zero quefrency.

In this paper, an alternative approach to solve the mentioned problem is applied. Thus, the function for finding peaks (*find\_peaks*) from the library *scipy* in *Python* is used. This function finds all local maxima by simple comparison of neighboring values, with the ability to define large number conditions for the peak properties. Due to the fact that there are no neighboring values on the left side of the cepstrum maximum at the zero quefrency (or in close vicinity), this maximum is automatically not considered as a peak.

When the relevant maximum of the cepstrum is selected, the quefrency value of that maximum is converted into frequency representing the estimated pitch ( $\text{pitch} = 1/(\text{quefrency of cepstrum maximum})$ ).

To check the described algorithm, the fundamental frequency from the trumpet sound is removed by filtering, see the spectrum shown in Fig. 4.a). The determined maximum of the cepstrum is located at the same quefrency position, see Fig. 4.b), as before removing the fundamental component.

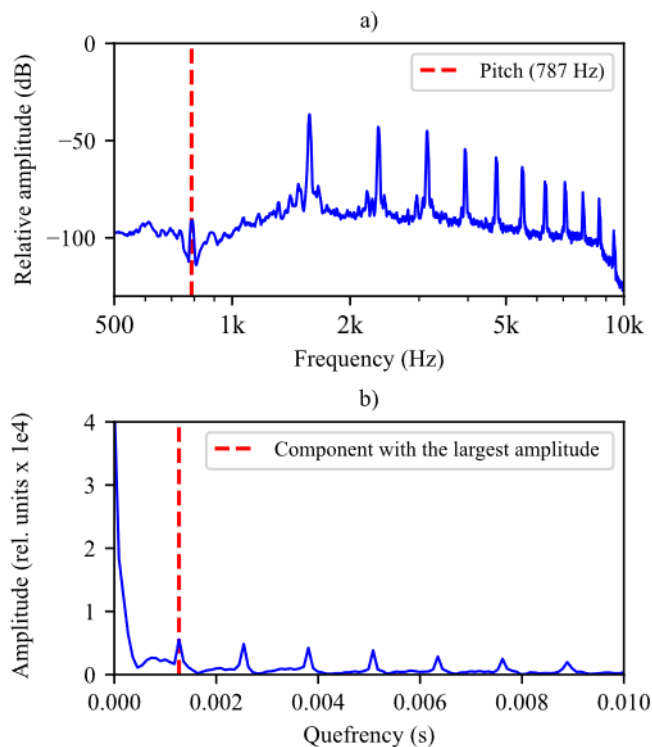


Fig. 4. Spectrum a) and cepstrum b) of trumpet sound after removing the fundamental component located at 787 Hz.

For the purpose of this research, the estimation of pitch by the modified cepstrum-based algorithm described above is done on different sounds of industrial products including fridge compressors and DC motors. The sound of fridge compressors were recorded in the semi-anechoic chamber in three working cycles (modes of operation) having different rpms (4000 rpm, 2400 rpm and 1300 rpm). Sounds of DC motors were also recorded in the semi-anechoic chamber (not the same one used for the compressor recording) within the production line. The recording was done on two different types of DC motors (here denoted type A and type B), in two different directions of rotation (here denoted direction 1 and direction 2) and for two different conditions regarding the failure (without failure and with certain failure). The analysis of recorded audio signals is carried out using the scripts developed in Python 3.8.

#### IV. ANALYSIS OF DETECTED PITCH

Th

e potentials for applying the cepstrum-based pitch detection in making difference between different samples or working conditions of certain industrial products is investigated here focusing on spectrum and cepstrum of the sounds of these products, that is, on one-figure value of the detected pitch.

The first product whose sound is analyzed is fridge compressors. The target for this product is to consider if it is possible to make a distinction between three different working conditions of compressors – three different rpms (4000, 2400 and 1300). The spectrum and cepstrum of the compressors having 4000 rpm are shown in Fig. 5. The pitch estimated by the described procedure is 6400.05 Hz.

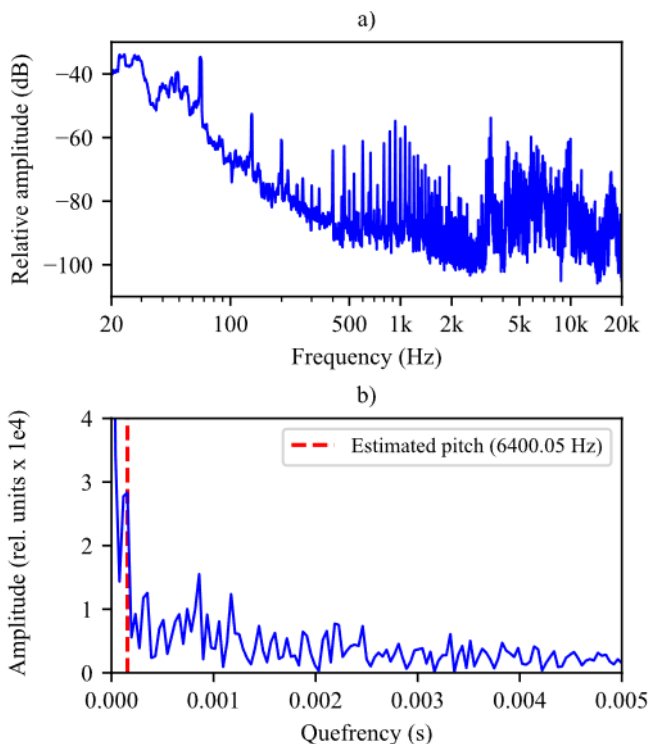


Fig. 5. Spectrum a) and cepstrum b) of fridge compressor sound at 4000 rpm.

When the rpm is changed from 4000 to 2400, it causes certain changes in the spectrum, but also in the cepstrum and consequently in the detected pitch. Fig. 6 shows the spectrum and cepstrum of the fridge compressor sound at 2400 rpm, where the estimated pitch is 2756.25 Hz.

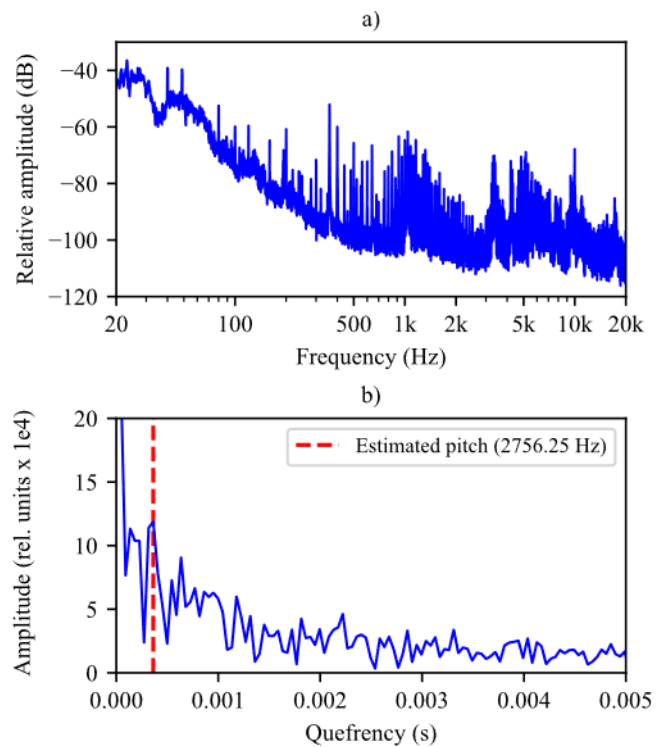


Fig. 6. Spectrum a) and cepstrum b) of fridge compressor sound at 2400 rpm.

By changing the rpm to 1300, the detected pitch is changed to 5120.04 Hz. The spectrum and cepstrum for that rpm are presented in Fig. 7. Regarding the cepstrum, not only the relevant maximum of the cepstrum, but also its pattern is changed by changing rpm. This is why it seems reasonable to introduce at least one more attribute that will reflect dissimilarity of the cepstrum pattern. This attribute can be related either to decay of the cepstrum envelope, cepstrum energy within certain queffency limits or even a vector containing the cepstrum values within pre-defined queffency limits.

The next step in the analysis includes DC motors. Spectrum and cepstrum of the same DC motor of type A, but in two opposite directions of rotation are shown in Fig. 8. Comparing the spectra, it can be concluded that overall pattern is similar for both directions of rotation, with certain differences in particular frequency bands and at particular frequencies. The patterns of cepstrum are also similar, but still having some differences for different directions of rotation. The estimated pitch for both directions is the same, 2666.33 Hz, since the maximum value of cepstrum is located at the same queffency.

Comparison of spectra and “cepstra” of sounds of two DC motors of different types (A and B) is presented in Fig. 9. Shapes of the spectra given in Fig. 9.a) and Fig. 8.a) for the motor A are similar, with certain differences, since different



motors of the same type A are used for the analysis. The same situation exists in the “cepstra” from Fig. 8.b) and Fig. 9.b). In spite of these differences, the estimated pitch for two motors of the same type A is the same (2666.73 Hz), while the estimated pitch for the motor of type B is 888.91 Hz.

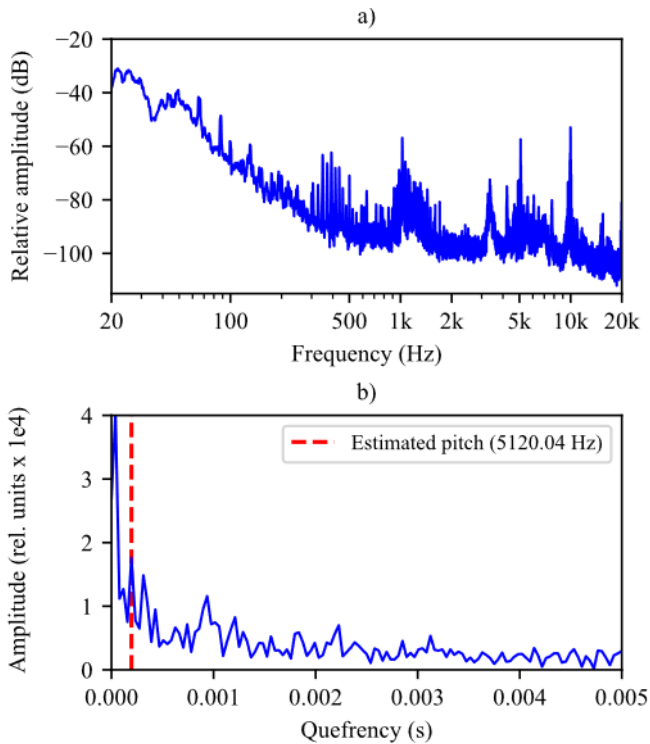


Fig. 7. Spectrum a) and cepstrum b) of fridge compressor sound at 1300 rpm.

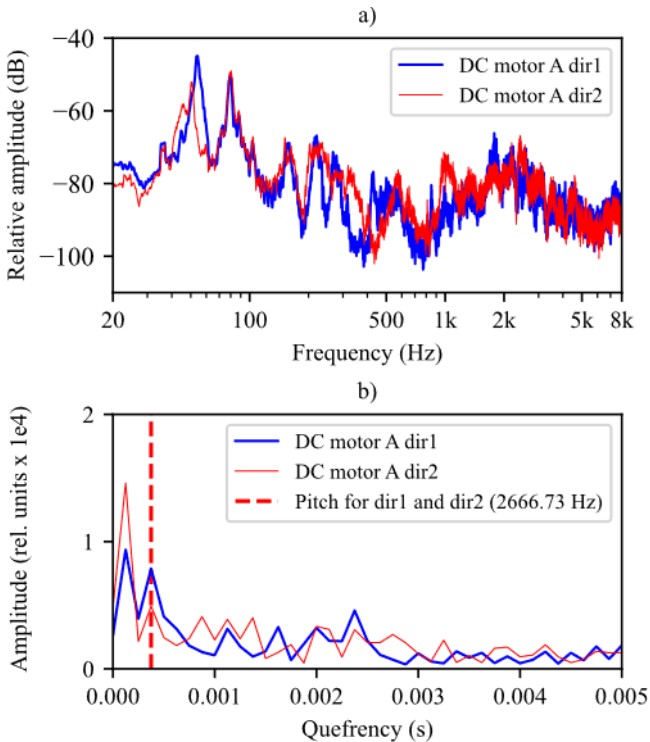


Fig. 8. Spectrum a) and cepstrum b) of DC motor A in two opposite directions of rotation.

When detected pitch values for DC motors with and without failures are analyzed, the differences between motors depend on the failure itself. In some cases, the failure causes change of periodicity or pseudo-periodicity of sound waveform leading to a certain change of the estimated pitch. Fig. 10 illustrates one of such cases presenting spectra and “cepstra” of DC motors without failure (OK motor) and with failure (NOT OK motor). The estimated pitch for OK motor is 2666.73 Hz and for NOT OK motor is 1333.36 Hz. Here, the differences between the spectra and “cepstra” are larger in comparison to the previously analyzed two cases of DC motors.

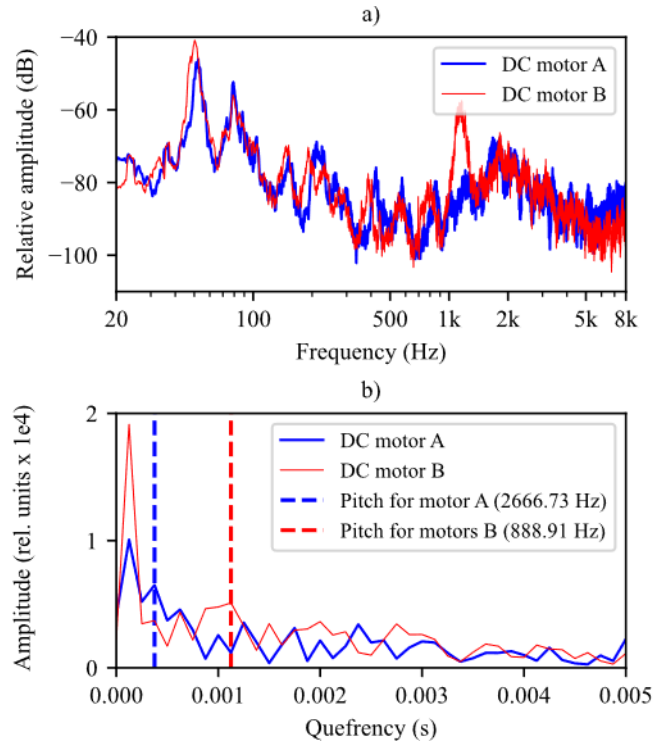


Fig. 9. Spectrum a) and cepstrum b) of different DC motors (motors of type A and B) and direction of rotation 1.

## V. CONCLUSION

This paper analyses potentials for making a difference between working conditions or states of two industrial products, fridge compressors and DC motors, based on pitch of their sounds. The pitch is detected using the cepstrum-based algorithm.

The results show that there are conditions and states where the estimated pitches are different for different conditions (states). However, there are also cases where the pitch only could not be used for making a difference between conditions (states) of these products. Even in such cases, the patterns or shapes of the “cepstra” show certain differences for different conditions (states). This is why a measure calculated from the cepstrum can be introduced in addition to pitch that can be used as a new audio feature.

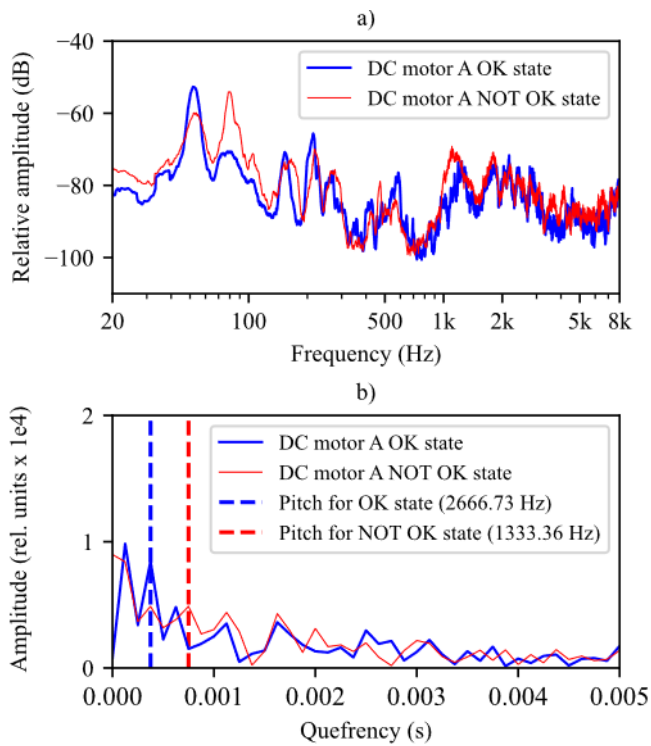


Fig. 10. Spectrum a) and cepstrum b) of OK DC motor (without failure) and NOT OK DC motor (with failure).

#### ACKNOWLEDGMENT

This research was supported by the Science Fund of the Republic of Serbia, 6527104, AI-Com-in-AI.

#### REFERENCES

[1] S. Makov, A. Minaev, I. Grinev, D. Cherhyshov, B. Kudruavcev, V. Mladenovic, "A spectral-based pitch detection method," Proc. AIP Conference (XV International Scientific-Technical Conference "Dynamic of Technical Systems"), Rostov-on-Don, Russia, 2188, 050005, 11-13 September, 2019.

[2] B. Gold, N. Morgan, D. Ellis, *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*, 2nd ed. USA: Wiley, 2011.

[3] X. Zhao, Y. Shao, D. Wang, "CASA-Based robust speaker identification," *IEEE Transactions on Audio Speech and Language Processing*, vol. 20, no. 5, pp. 1608-1616, July, 2012.

[4] A. B. Nielsen, L. K. Hansen, U. Kjems, "Pitch based sound classification," Proc. IEEE International Conference on Acoustics,

Speech and Signal Processing (ICASSP), Toulouse, France, vol. 3, 14-19 May, 2006.

[5] K. Pratibha, H.M. Chandrashekar, "Estimation and tracking of pitch for noisy speech signals using EMD based autocorrelation function algorithm," Proc. 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, pp. 2071-2075, 19/20 May, 2017.

[6] T.G. Stockham, Jr., T.M. Cannon, R.B. Ingebretsen, "Blind deconvolution through digital signal processing," *Proc. IEEE*, vol. 63, pp. 678-692, April, 1975.

[7] M. Milivojčević, D. Ćirić, "Extracting relevant acoustic features of the internal combustion engines," Proc. 7th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN), Belgrade, Čačak, Niš, Novi Sad, Serbia, pp. 49-53, 28-29 September, 2020.

[8] A. V. Oppenheim, R. W. Schaffer, "From frequency to quefrency: A history of the cepstrum," *IEEE Signal Processing Magazine*, pp. 95-106, September, 2004.

[9] H. Chander, B. Singh, R. Khanna, "Short time cepstrum analysis method for pitch estimation of an arbitrary speech signal using MATLAB," *International Journal of Research in Electronics and Computer Engineering*, vol. 5, no. 3, pp. 1-4, July-September, 2017.

[10] T. Ananthapadmanabha, B. Yegnanarayana, "Epoch extraction of voiced speech," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-23, no. 6, pp. 562-575, December, 1975.

[11] L. Sukhostat, Y. Imamverdiyev, "A comparative analysis of pitch detection methods under the influence of different noise conditions," *Journal of Voice*, vol. 29, no. 4, pp. 410-417, July, 2015.

[12] B. Gold, L. Rabiner, "Parallel processing techniques for estimating pitch periods of speech in the time domain," *Journal of the Acoustical Society of America*, vol. 46, no. 2, pp. 442-448, August, 1969.

[13] N. J. Miller, "Pitch detection by data reduction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 23, no. 1, pp. 72-79, February, 1975.

[14] A. de Cheveigne, H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917-1930, April, 2002.

[15] A. Camacho, "SWIPE: A sawtooth waveform inspired pitch estimator for speech and music," Ph.D. dissertation, University of Florida, Gainesville, Florida, USA, 2007.

[16] S. Gonzalez, M. Brookes, "A pitch estimation filter robust to high levels of noise (PEFAC)," Proc. 19th European Signal Processing Conference - EUSIPCO, Barcelona, Spain, pp. 451-455, 29 August - 02 September, 2011.

[17] K. Kasi, S. A. Zahorian, "Yet another algorithm for pitch tracking," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing - ICASSP, Orlando, Florida, USA, 13-17 May, 2002.

[18] L. N. Tan, A. Alwan, "Multi-band summary correlogram-based pitch detection for noisy speech," *Speech Communication*, vol. 55, pp. 841-856, 2013.

[19] H. Ba, N. Yang, I. Demirkol, W. Heintzelman, "BaNa: a hybrid approach for noise resilient pitch detection," Proc. IEEE Statistical Signal Processing Workshop (SSP), Ann Arbor, Michigan, USA, 04-07 August, 2012.