

Sveučilište u Splitu
Prirodoslovno-matematički fakultet
Odjel za fiziku

Frana Perišić

**PRIMJENA METODE NEZAVISNIH
KOMPONENATA U SVRHU UTVRĐIVANJA
GOVORNIKA IZ GOVORNIH SNIMKI**

Diplomski rad

Split, 2018.

Ovaj rad, izrađen u Splitu na Prirodoslovno-matematičkom fakultetu, pod mentorstvom doc. dr. sc. Damira Kovačića, predan je na ocjenu Odjelu za fiziku Prirodoslovno-matematičkog fakulteta Sveučilišta u Splitu radi stjecanja zvanja magistre (mag.) fizike, smjera biofizike.

Zahvaljujem se Gorani Glavurčić, Kristini Latinčić, Ivani Ursi, Marku Perkušiću i Dinu Perišiću što su sudjelovali na snimanjima i time pridonjeli izradi ovog rada. Također se zahvaljujem i Tonću Kozini koji mi je pomogao u realizaciji.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Splitu
Prirodoslovno-matematički fakultet
Odjel za fiziku

Diplomski rad

PRIMJENA METODE NEZAVISNIH KOMPONENATA U SVRHU UTVRĐIVANJA GOVORNIKA IZ GOVORNIH SNIMKI

Frana Perišić
Ruđera Boškovića 33, 21 000 Split

SAŽETAK

U ovom radu je na snimke više govornika primijenjena metoda nezavisnih komponenata (FastICA algoritam u Matlabu) u svrhu izdvajanja izvornih signala. Kada bi se izvorni signali mogli uspješno izdvojiti iz bilo koje mješavine signala to bi mogli primijeniti na niz različitih problema, npr. omogućili bi slušnim aparatima da raspoznaju različite govornike i fokusiraju se na određenog govornika, što je poznato pod nazivom '*cocktail party problem*'. Dobiveni rezultati pokazuju da ova metoda funkcionira na snimkama gdje je broj govornika jednak broju mikrofona uz pretpostavku da su izlazne vrijednosti mikrofona linearne mješavine nezavisnih komponenti. Ljudski govor je sam po sebi jako kompleksan, a to posebno dolazi do izražaja kad istovremeno govori više govornika. Otkrivanje mehanizma selektivne pozornosti u mozgu je ključ za rješavanje '*cocktail party problem*'.

(40 stranica, 2 slike, 7 tablica, 36 grafova, 12 literaturnih navoda, jezik izvornika: engleski i hrvatski)

Ključne riječi: *slijepa separacija signala, izvorni signal, nezavisnost, procjena maksimalne vjerojatnosti, negaussivnost, FastICA*

Voditelj: doc. dr. sc. Damir Kovačić

Ocjenitelji: doc. dr. sc. Damir Kovačić, doc. dr. sc. Petar Stipanović, dr. sc. Tonko Kovačević, v.pred.

Rad prihvaćen: 13. 09. 2018.

BASIC DOCUMENTATION CARD

University of Split
Faculty of Science
Department of Physics

Graduation Thesis

APPLICATION OF INDEPENDENT COMPONENT ANALYSIS FOR IDENTIFICATION OF SPEAKERS FROM SPEECH RECORDINGS

Frana Perišić
Ruđera Boškovića 33, 21 000 Split

SUMMARY

In this paper was applied independent component analysis (FastICA algorithm in Matlab) on the records of more speakers for the purpose of extracting the original signals. If the source signals could be successfully extracted from any signal mixture they could apply to a variety of different problems, for example, allow hearing aids to discern different speakers and focus on a particular speaker, known as the '*cocktail party problem*'. The results obtained show that this method works on experimental recordings where the number of speakers is equal to the number of microphones assuming that the output values of the microphone are linear mixtures of independent components. Human speech is very complex by itself, and this is particularly pronounced when multiple speakers speak simultaneously. Detecting a selective attention mechanism in the brain is the key to solving the '*cocktail party problem*'.

(40 pages, 2 figures, 7 tables, 36 graphs, 12 references, original in: english and croatian)

Key words: *blind source separation, source signal, independence, maximum likelihood estimation, nongaussianity, FastICA*

Supervisor: doc. dr. sc. Damir Kovačić

Reviewers: doc. dr. sc. Damir Kovačić, doc. dr. sc. Petar Stipanović, dr. sc. Tonko Kovačević, v.pred.

Thesis accepted: 13. 09. 2018.

SADRŽAJ

| | | |
|---|---|----|
| 1 | UVOD | 1 |
| | 1.1 CILJ ISTRAŽIVANJA | 1 |
| | 1.2 KARAKTERISTIKE GLASA | 1 |
| | 1.2.1 Fundamentalna frekvencija | 1 |
| | 1.2.2 Intenzitet ili jakost glasa | 2 |
| | 1.3 DIGITALIZACIJA ZVUKA | 2 |
| | 1.3.1 Broj kanala | 2 |
| | 1.3.2 Frekvencija uzorkovanje | 2 |
| | 1.3.3 Rezolucija | 2 |
| | 1.4 'COCKTAIL PARTY PROBLEM' | 3 |
| | 1.5 METODA NEZAVISNIH KOMPONENATA | 3 |
| | 1.5.1 Definicija ICA-e | 3 |
| | 1.5.2 Pretpostavke na kojima se temelji ICA-a | 6 |
| | 1.5.3 Primjene ICA-e | 7 |
| 2 | MATERIJALI I METODE | 8 |
| | 2.1 PROCJENA MODELA ICA-E METODOM MAKSIMALNE VJEROJATNOSTI | 8 |
| | 2.2 FASTICA | 10 |
| | 2.2.1 FastICA algoritam | 11 |
| | 2.3 POMOĆNI ALATI | 12 |
| | 2.3.1 Audacity | 12 |
| | 2.3.2 Praat | 12 |
| 3 | REZULTATI | 13 |
| | 3.1 DVA GOVORNIKA I DVA MIKROFONA | 15 |
| | 3.2 PET GOVORNIKA I PET MIKROFONA | 19 |
| | 3.3 PET GOVORNIKA - UMJETNE MJEŠAVINE SIGNALA | 28 |
| 4 | RASPRAVA I ZAKLJUČAK | 38 |
| 5 | LITERATURA | 39 |

1 UVOD

1.1 CILJ ISTRAŽIVANJA

Kada više govornika govori istovremeno u stanju smo 'filtrirati' sve ostale govornike osim onoga kojeg želimo čuti. Ovakva situacija je opće poznata i naziva se '*cocktail party problem*'. U ovom diplomskom radu napravljena su dva odvojena snimanja koja su simulirala '*cocktail party problem*'. Cilj istraživanja je bio primijeniti *metodu nezavisnih komponenata* (eng. independent component analysis – ICA) na mješavine signala koje su izlazna vrijednost mikrofona i kao rezultat dobiti pojedinačne izvore signala (govornike).

1.2 KARAKTERISTIKE GLASA

1.2.1 Fundamentalna frekvencija

Fundamentalna frekvencija F_0 ('pitch', osnovni glas, nulti harmonik) je broj vibracija koje glasnice učine u jednoj sekundi, a izražava se u Hertzima (Hz). Što je veći broj vibracija glasnica, viša je vrijednost fundamentalne frekvencije pa i glas doživljavamo kao viši. Visina glasa je perceptivni fenomen, ovisi o F_0 . Fundamentalna frekvencija je determinirana aktivnošću unutarnjih laringealnih mišića, subglotičkim tlakom te napetošću, masom i duljinom glasnica. Na vrijednost F_0 utječe mnogo faktora: dob, spol, tjelesna konstitucija, socijalno okruženje, oštećenje sluha... [4]. Fundamentalna frekvencija izračunava se po sljedećoj formuli:

$$F_0 = \frac{1}{T} \quad (1)$$

gdje je T period tj. najmanji vremenski interval nakon kojeg vremenska funkcija $x(t)$ kojom se ta promjena opisuje poprima iste vrijednosti:

$$x(t) = x(t + T) \quad (2)$$

Trajanje jednog prosječnog punog valnog titranja kreće se u rasponu 8–12 ms kod muškaraca i 5–8 ms kod žena [5]. Prosječna *fundamentalna frekvencija* muškog glasa iznosi oko 120 Hz, a ženskog 225 Hz [4].

1. 2.2 Intenzitet ili jakost glasa

Intenzitet ili jakost glasa ovisi o amplitudi titranja glasnica te o subglotičkom tlaku zraka pa što su njihove vrijednosti veće intenzitet glasa percipiramo kao glasniji i obratno. Referentni zvučni tlak od $20 \mu Pa$ na frekvenciji od $1000 Hz$ je ujedno i *prag čujnosti* tj. najmanji intenzitet koji zdravo uho može zamijetiti te je ovaj intenzitet u akustici označen kao vrijednost od $0 dB$. Intenzitetski raspon od tek čujnog glasa do najglasnijeg kojeg pojedinac može izvesti iznosi i do $70 dB$. Govorni je intenzitet varijabilan [4].

1.3 DIGITALIZACIJA ZVUKA

Proces pretvaranja analognog audio signala u digitalni sastoji se od tri koraka: *uzorkovanje*, *kvantiziranje* i *kodiranje*. Digitalni audiozapis definiran je *brojem kanala*, *frekvencijom uzorkovanje* i *rezolucijom*. Kvaliteta zvuka ovisi o gore navedenim parametrima.

1.3.1 Broj kanala

Zvuk može biti snimljen u *mono* i *stereo* tehnici. Mono podrazumijeva jednokanalno snimanje odnosno da će prilikom reprodukcije zvuk biti identičan na lijevom i desnom kanalu (zvučnicima). Stereo je dvokanalno snimanje (različiti su lijevi i desni kanal). Broj kanala nam daje informaciju da li je zvuk snimljen mono (1 kanal) ili stereo (2 kanala) tehnikom.

1.3.2 Frekvencija uzorkovanje

Frekvencija uzorkovanje (*eng. sample rate*) definira se kao brzina (učestalost) kojom se zvuk uzorkuje. Npr. Za vrijednost $44100 Hz$ znači da se 44100 puta u sekundi uzme uzorak. Svako vrijeme vremenskoj točki uzorkovanje pridodaje se brojčana vrijednost trenutnog stanja zvuka.

1.3.3 Rezolucija

Rezolucija (*eng. sample depth*) je broj bitova korištenih za mjerenje amplituda (glasnoće) uzorka. Što je veća rezolucija bolji je omjer između najtišeg i najglasnijeg dijela zvuka i bolji je odnos signala i šuma. Npr. Kod 8 bitne rezolucije koristi se 256 različitih varijacija glasnoće, a kod 16 bitne rezolucije koristi se 65536 [4].

1.4 'COCKTAIL PARTY PROBLEM'

'Cocktail party problem' ili 'Cocktail party efekt' je fenomen sposobnosti mozga da se usredotoči na slušnu pozornost (učinak selektivne pozornosti u mozgu) na određeni podražaj dok istodobno filtrira niz drugih podražaja [2]. Definirao ga je i imenovao Colin Cherry u svom radu 1953.g (prepoznavanje poruke jednim ili s dva uha) [3]. Više od pola stoljeća je prošlo od kad je Cherry definirao 'cocktail party problem' međutim i dan danas je ostao neriješen. Posvećeni su brojni naporu u različitim granama istraživanje fiziologija, neurobiologija, psihofiziologija, kognitivna psihologija, biofizika, računalna znanost i inženjering međutim mehanizam još nije razriješen.

1.5 ANALIZA NEZAVISNIH KOMPONENATA

Slijepa separacija izvora (eng. BSS - Blind Source Separation) je tehnika obrade podataka čiji je cilj rekonstrukcija originalnih signala iz promatranih mješavina signala bez poznavanja originalnih signala i načina na koji su signali miješani. *Metoda nezavisnih komponenata* (eng. independent component analysis – ICA) je jedna od metoda koja je najčešće korištena za obavljanje *slijepa separacije izvora* i upravo ona je korištena u ovom radu [6]. Jedan od osnovnih motiva za istraživanje ICA-e bio je 'cocktail party problem'. ICA je izvorno razvijena kako bi se bavila problematikom usko vezanom s 'cocktail party problemom' [6]. Ideja je da postoji n izvora signala (govornika) koji su zabilježeni mikrofonom. Najčešće zbog pojednostavljenja problema broj mikrofona je jednak broju govornika i želimo izdvojiti samo jedan od izvora signala.

1.5.1 Definicija ICA-e (poopćenje za 'cocktail party problem')

Senzori (mikrofoni) kao izlaznu vrijednost daju snimke signala u vremenu mješavine signala koje će se označavati s $x_1(t), x_2(t), \dots, x_m(t)$. Glasovi govornika u 'cocktail party problemu' su generirani različitim fizikalnim procesima tj. različitim govornicima. Svaki snimljeni signal se može prikazati kao linearna kombinacija n izvora signala $s_i(t), i = 1, \dots, n$ koji su međusobno nezavisni:

$$\begin{aligned}
x_1(t) &= a_{11}s_1(t) + \dots + a_{1n}s_n(t) \\
&\vdots \\
x_m(t) &= a_{m1}s_1(t) + \dots + a_{mn}s_n(t)
\end{aligned} \tag{3}$$

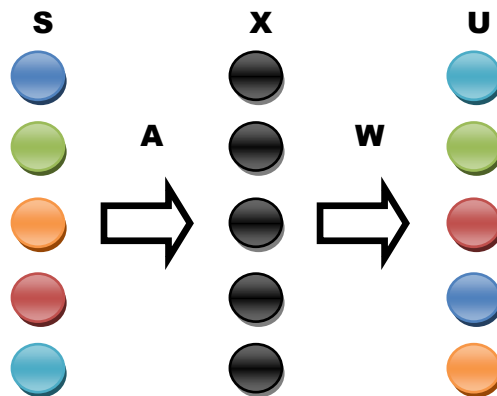
Gdje su a_{ij} težinski faktori tj. relativni doprinosi govornika koji ovise o udaljenosti mikrofona i govornika. U daljnjem razmatranju izostaviti ćemo oznaku vremena jer pretpostavljamo da su svi izvori signala $s_i(t)$ i sve mješavine signala $x_j(t)$ slučajne varijable, a ne vremenski signali [7]. Uobičajeno je koristiti vektorski zapis umjesto sume u sustavu jednadžbi (1). Pridijelimo vektoru \mathbf{x} elemente x_1, \dots, x_m koji označavaju mješavine izvora signala, a vektoru \mathbf{s} izvore signala s_1, \dots, s_n . Slovom \mathbf{A} označit ćemo matricu miješanja čiji su elementi težinski faktori a_{ij} gdje je $i = 1, \dots, m; j = 1, \dots, n$. Koristeći vektorski zapis sustav jednadžbi (3) se može zapisati na sljedeći način [6]:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{4}$$

(Svi vektori se podrazumjevaju kao stupčani vektori što znači da ih je potrebno transponirati da bi dobili redčani vektor.) Statistički model prikazan jednadžbom (4) naziva se *metoda nezavisnih komponenata* ili *model ICA-e*. Model ICA-e je generativni model što znači da opisuje kako se promatrani podaci (mješavine signala) x_i generiraju procesom miješanja komponenti s_i [7]. Početna točka ICA-e temelji se na jednostavnoj pretpostavci da su komponente s_i *statistički nezavisne*. Pretpostavka koja puno pojednostavljuje problem je da je nepoznata matrica miješanja *kvadratna*. Matrica \mathbf{A} je kvadratna ako je broj govornika i broj mikrofona u cocktail party problemu isti ($n = m$). Nakon procjene matrice \mathbf{A} budući je kvadratna možemo izračunati njen inverz koji ćemo označiti slovom \mathbf{W} , $\mathbf{W} = \mathbf{A}^{-1}$. Nakon što smo izračunali inverz matrice \mathbf{A} jednostavno možemo dobiti nezavisne komponente [6]:

$$\mathbf{s} = \mathbf{W}\mathbf{x} \quad (5)$$

Generativni model ICA-e je linearni trenutni model miješanja bez buke i čini jezgru u teoriji ICA-e. Ovim modelom je objašnjeno kako se mješavine signala x_i generiraju procesom miješanja komponenti s_i i postavilo se pitanje kako procijeniti matricu miješanja \mathbf{A} [7].



Slika 1: Shematski prikaz ICA-e

U drugim modelima ICA-e jedna od prvih pretpostavki je da statistički nezavisne varijable ne smiju imati Gaussovu razdiobu (negaussivnost). Kod osnovnog modela ove pretpostavke nema. Alternativne metode ICA-e ne koriste generativni model, a za tim metodama moramo posegnuti u sljedećim situacijama [7]:

- broj govornika i mikrofona je različit
- postoji pozadinska buka
- miješanje signala je nelinearno

Postoji više metoda za procjenu parametara statističkog modela ICA-e, a to su:

- Maksimiziranje negaussivosti (*eng. Maximization of Nongaussianity*)
- Procjena maksimalne vjerojatnosti (*eng. Maximum likelihood estimation*)
- Minimiziranje zajedničkih informacija (*eng. Minimalization of mutual information*)

Postoji više različitih algoritama za ICA metodu, a u ovom radu je korištena FastICA (Matlab verzija).

1.5.2 Pretpostavke na kojima se temelji ICA-a

1. *Linearno nezavisne varijable su statistički nezavisne*

Za slučajne varijable y_1, y_2, \dots, y_n kažemo da su nezavisne ako poznavanje vrijednosti jedne varijable y_i ne daje nikakve informacije o drugim varijablama y_j za $i \neq j$. Nezavisnost se može definirati *gustoćom vjerojatnosti*. Označimo s $p(y_1, y_2, \dots, y_n)$ *zajedničku funkciju gustoće vjerojatnosti* (eng. probability density function - pdf) od y_i varijabli, a s $p_i(y_i)$ *rubnu funkciju gustoće vjerojatnosti*. Kažemo da su y_i nezavisne varijable ako i samo ako se *zajednička funkcija gustoće vjerojatnosti* može faktorizirati na sljedeći način [7]:

$$p(y_1, y_2, \dots, y_n) = p_1(y_1)p_2(y_2) \dots p_n(y_n) \quad (6)$$

Ako skup izvora signala ima maksimalnu entropiju funkcije gustoće vjerojatnosti onda to podrazumijeva da su signali međusobno nezavisni. Obrnuto ne vrijedi. Skup nezavisnih signala nema nužno funkciju gustoće vjerojatnosti s maksimalnom entropijom [10].

2. *Nezavisne varijable ne smiju imati Gaussovu razdiobu*

U stvarnosti većina setova podataka (u našem slučaju signala) nemaju Gaussovu razdiobu što značajno otežava situaciju. Da naši signali imaju Gaussovu razdiobu lako bi pronašli nezavisne komponente jer kod Gaussove razdiobe svi podaci koji nisu u korelaciji su statistički nezavisni [3, pdf str. 5]. U skladu s tim su napravljene metode za procjenu ICA-e. Kod metode *Maximization of Nongaussianity* za procjenu parametara statističkog modela ICA-e koriste se kumulanti višeg reda (statistika višeg reda) . Kod Gaussove razdiobe kumulanti višeg reda su nula što znači da se u tom slučaju ne može izvršiti procjena. Najviše jedan izvor smije imati Gaussovu razdiobu [3]. Po centralnom graničnom teoremu iz statistike zbroj dviju razdioba koje nisu Gaussove je bliži Gaussovoj razdiobi od originala, što je veći broj uzoraka.

3. *Matrica miješanja je kvadratna*

Broj nezavisnih komponenti jednak je broju promatranih mješavina. Drugim riječima, broj govornika (izvora) jednak je broju mikrofona (senzora). Ova pretpostavka znatno pojednostavljuje problem. Nakon procjene matrice miješanja možemo izračunati njen inverz i jednostavno dobiti nezavisne komponente [7].

1.5.3 Primjene ICA-e

Analiza nezavisnih komponenata je prvi put uvedena (ne pod tim imenom) 1980s od strane J. H'eraulta, C. Juttena, i B. Ansa. ICA-a je u originalu izvedena u svrhu rješavanja '*cocktail party problema*'. Uspješno je primjenjena u mnogim područjima istraživanja koji se bave analizom biomedicinskih podataka kao i u računalnom modeliranju [10]. Neke od primjena ICA-e su [3]:

1. *Separacija artefakta u MEG podacima* - Magnetoencefalografija (MEG) je neinvazivna tehnika pomoću koje se mjeri aktivnost mozga s vrlo dobrom vremenskom rezolucijom i umjerenom prostornom rezolucijom. U radu [12] autori predstavljaju tehniku za odvajanje moždane aktivnosti od artefakata (pomicanje očiju, treptanje) koristeći ICA-u.
2. *Pronalaženje skrivenih faktora u financijskim podacima* - primjerice na tržištu dionica ili trgovini valutama. ICA se koristi za otkrivanje zajedničkih utjecajnih faktora, tj. mehanizama koji utječu na tok cijena [8].
3. *Smanjenje šuma u prirodnim slikama* – šum u slikama je definiran kao neželjeni nasumični signal koji se također uspješno može ukloniti primjenom ICA-e .
4. *Telekomunikacije* – odvajanje korisnikova vlastita signala od signala drugih korisnika u CDMA (Code-Division Multiple Access) mobilnim komunikacijama [9].

2 MATERIJALI I METODE

2.1 PROCJENA MODELA ICA-e METODOM MAKSIMALNE VJEROJATNOSTI

Uobičajena metoda za procjenu modela ICA-e je metoda maksimalne vjerojatnosti (*eng. maximum likelihood*) [11]. Budući da smo pretpostavili da je broj izvora signala jednak broju senzora slijedi da je \mathbf{A} kvadratna matrica. Inverz matrice \mathbf{A} označit ćemo s \mathbf{W} , $\mathbf{W} = \mathbf{A}^{-1}$. Matrica \mathbf{A} definira linearnu transformaciju signala \mathbf{s} . Takva linearna transformacija je obično invertibilna:

$$\mathbf{s} \approx \mathbf{u} = \mathbf{x}\mathbf{W} \quad (7)$$

Potrebno je prilagoditi separacijsku matricu \mathbf{W} tako da dobiveni elementi od \mathbf{u} budu međusobno nezavisne komponente koje odgovaraju govornicima \mathbf{s} . To ćemo postići tako da maksimiziramo entropiju od \mathbf{U} prilagođavanjem separacijske matrice \mathbf{W} :

$$\mathbf{U} = g(\mathbf{u}) = g(\mathbf{x}\mathbf{W}) \quad (8)$$

Pretpostavimo da je funkcija g *kumulativna funkcija distribucije* (*eng. cumulative distribution function - cdf*) od \mathbf{s} . Ako je skup signala \mathbf{x} preslikan na drugi skup \mathbf{U} tada je entropija $H(\mathbf{U})$ od \mathbf{U} dana sljedećim izrazom:

$$\begin{aligned} H(\mathbf{U}) &= H(\mathbf{x}) + \Delta H(\mathbf{x} \rightarrow \mathbf{U}) \\ &= H(\mathbf{x}) + \Delta H(\mathbf{x} \rightarrow g(\mathbf{x}\mathbf{W})) \end{aligned} \quad (9)$$

Gdje je $H(\mathbf{x})$ entropija mješavine signala \mathbf{x} , a ΔH je promjena entropije uzrokovana preslikavanjem $\mathbf{x} \rightarrow \mathbf{U}$. Maksimiziranje entropije $H(\mathbf{U})$ ovisi samo o entropiji ΔH budući je entropija mješavine signala \mathbf{x} fiksna (u daljnjem razmatranju zanemarit ćemo član $H(\mathbf{x})$). Preslikavanje $\mathbf{x} \rightarrow \mathbf{U}$ ovisi o dva člana:

- *kumulativnoj funkciji distribucije* g
- separacijskoj matrici \mathbf{W}

Budući da je oblik *kumulativne funkcije distribucije* g fiksna, ne utječe na maksimiziranje entropije. Slijedi da separacijska matrica \mathbf{W} maksimizira zajedničku entropiju. Promjena

entropije uzrokovana transformacijom $g(\mathbf{x}\mathbf{W})$ može se smatrati omjerom infinitezimalnih volumena povezanih s odgovarajućim točkama u \mathbf{x} i \mathbf{U} . Ovaj omjer je dan kao očekivana vrijednost od $\ln|J|$, gdje J označava apsolutnu vrijednost Jacobijeve determinante. Izostavljajući matematičke detalje dobivamo:

$$H(\mathbf{U}) = \frac{1}{n} \sum_{k=1}^n \sum_{j=1}^N \ln g'_j(u_j^{(k)}) + \ln|W| \quad (10)$$

Sumiranje preko n uzoraka plus $\ln|W|$ je promjena očekivane vrijednosti od $\ln|J|$, a funkcija $g'_j = \frac{\partial g_j}{\partial x}$ je *funkcija gustoće vjerojatnosti* i -tog izvora signala [10]. Funkcija gustoće vjerojatnosti neprekidne slučajne varijable dobije se kao derivacija *kumulativne funkcije distribucije* g [3, str. 16]. Derivacija $H(\mathbf{U})$ daje:

$$\nabla H(\mathbf{U}) = \mathbf{W}^{-1} + \psi(\mathbf{u})\mathbf{x}^t \quad (11)$$

Gdje je:

$$\psi(\mathbf{u}) = \left[\frac{g'(\mathbf{u}_1)}{g''(\mathbf{u}_1)}, \dots, \frac{g'(\mathbf{u}_N)}{g''(\mathbf{u}_N)} \right] \quad (12)$$

Prisjetimo se, cilj nam je bio prilagoditi \mathbf{W} da bi maksimizirali $H(\mathbf{U})$. Ako skup signala ima maksimalnu entropiju pdf-a, onda to znači da su signali međusobno nezavisni. Standardna metoda za to je porast gradijenta (eng. gradient ascent). Metoda se sastoji od iterativnog dodavanja malog udjela gradijenta separacijskoj matrici:

$$\mathbf{W}_{nova} = \mathbf{W}_{stara} + \eta \nabla H(\mathbf{U}) \quad (13)$$

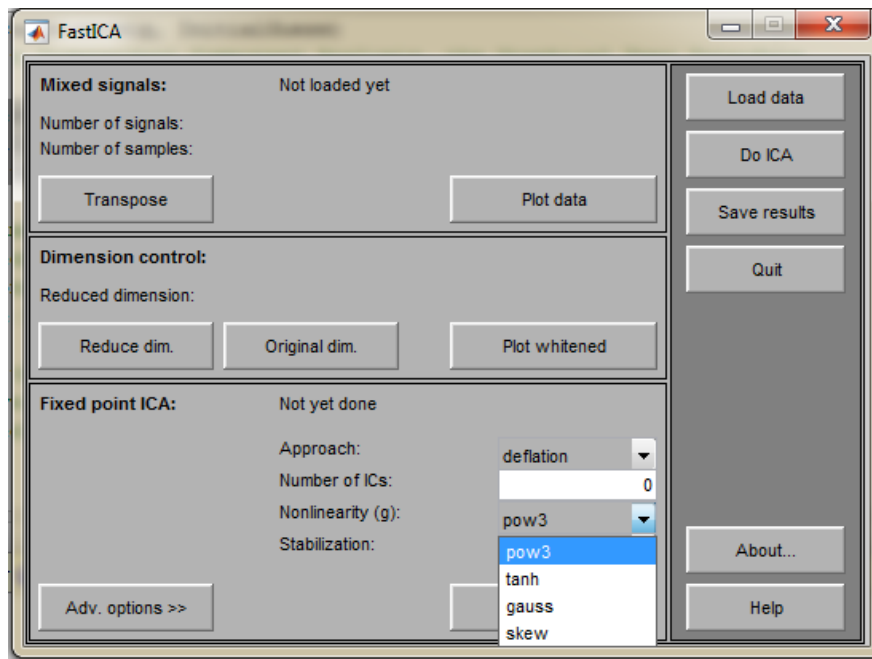
Funkcija $H(\mathbf{U})$ može se maksimizirati porastom gradijenta ili koristeći mnogo efikasniji 'prirodni gradijent' [10].

2.2 FastICA

FastICA (fast fixed-point) algoritam je računalna metoda za obavljanje procjene ICA-e. Njegov autor je Aapo Hyvärinen s Tehnološkog sveučilišta u Helsinkiju [6]. FastICA paket za Matlab je programski paket s grafičkim korisničkim sučeljem koji implementira *fixed-point algoritam* za ICA-u i dostupan je besplatno na internetu [16]. Ima jednostavno grafičko korisničko sučelje i računalno jak algoritam koji je 10 - 100 puta brži od konvencionalnih gradijentnih metoda za obavljanje procjene ICA-e. Obilježja FastICA algoritma koja ga čine boljim od drugih algoritama [6] :

- Konvergira jako brzo zato što je konvergencija kubna (ili barem kvadratna). Kod ostalih ICA algoritama temeljenih na gradijentnim metodama konvergencija je linearna.
- Kontrastno gradijentnim metodama nema odabira veličine koraka parametara što ovaj algoritam čini znatno jednostavnijim.
- Algoritam direktno pronalazi nezavisne komponente bilo koje distribucije koja nije Gaussova. Kod drugih algoritama prvo je potrebno procijeniti funkciju distribucije vjerojatnosti.
- Možemo utjecati na izvođenje algoritma optimizacijom.
- Nezavisne komponente možemo računati jednu po jednu.
- FastICA algoritam je računski jednostavan i zahtjeva malo memorije.

U Matlabu naredbom *fasticag* otvaramo grafičko korisničko sučelje FastICA-e. Slika 2 prikazuje izgled sučelja.



Slika 2 : Grafičko korisničko sučelje FastICA-e. Unutar sučelja možemo birati funkciju pomoću koje ćemo izvršiti procjenu *negentropije* (jedna od metoda za maksimiziranje *negaussivnosti*): *pow3*, *tanh* (najčešće se koristi), *gauss*, *skew*.

2.2.1 FastICA algoritam

Koraci FastICA algoritam za procjenu maksimalne vjerojatnosti [8] su:

1. Centriranje. Izračunavanje korelacijske matrice $C = E\{xx^T\}$.
2. Odabir inicijalne separacijske matrice B .
3. Računanje:

$$y = Bx$$

$$\beta_i = -E\{y_i g(y_i)\}, \text{ za } i = 1, \dots, n$$

$$\alpha_i = -\frac{1}{\beta_i + E\{g'(y_i)\}}, \text{ za } i = 1, \dots, n$$

4. Ažuriranje separacijske matrice:

$$B \leftarrow B + \text{diag}(\alpha_i)[\text{diag}(\beta_i) + E\{g(y)y^T\}]B$$

5. Dekorelacija i normalizacija:

$$B \leftarrow (BCB^T)^{-1/2}B$$

6. Ako ne konvergira vratiti se na korak 3.

2.3 POMOĆNI ALATI

2.3.1 Audacity

Audacity je besplatni digitalni audio uređivač i aplikacijski softver za snimanje [14]. Može se koristiti za naknadnu obradbu svih vrsta zvukova. Jako je jednostavan za korištenje, a pruža široku paletu opcija. Koristili smo ga za sinkroniziranje snimki.

2.3.2 Praat

Praat je besplatni računalni programski paket za znanstvenu analizu govora u fonetici [15] napravili su ga Paul Boersma i David Weenink sa Sveučilišta u Amsterdamu. Može se koristiti na širokom rasponu operativnih sustava, uključujući različite verzije sustava Unix, Linux, Mac i Microsoft Windows (2000, XP, Vista, 7, 8, 10). Između niza opcija koje nam Praat nudi moguće je iz spektrograma očitati *intenzitet* i *fundamentalnu frekvenciju* za svaku vremensku točku. Također je moguće dobiti listu vrijednosti intenziteta i 'pitch-a' za željeni vremenski period (*intensity listing, pitch listing*). Ovi podaci su potrebni u analizi podataka.

3 REZULTATI

U svrhu istraživanja '*cocktail party problema*' napravljena su dva snimanja. Oba snimanja su se odvijala u *tihoj komori* na Prirodoslovno-matematičkom fakultetu Sveučilišta u Splitu. Broj senzora (mikrofona) je u oba slučaja bio jednak broju govornika koji su sudjelovali u istraživanju. Snimanje se odvijalo tako da su svi govornici istovremeno čitali različite proizvoljne tekstove. WAV je osnovni format zvuka na Windowsima. Ne koristi kompresiju podataka i pogodan je za analizu zvuka i akustičke obrade. Zbog gore navedenog upravo taj format zvuka je korišten. Uobičajeni parametri računalnog WAV formata su broj kanala 2 (stereo), rezolucija 16 bita i frekvencija uzorkovanje 44100 Hz [4]. Vrijednosti parametara snimki dva govornika i dva mikrofona ($m = n = 2$) te pet govornika i pet mikrofona ($m = n = 5$) dane su u Tablici 1. Treći stupac se odnosi na '*umjetnu mješavinu signala pet govornika*' koja će detaljnije biti objašnjena u daljnjem tekstu.

Tablica 1. Vrijednosti parametara snimki: dva govornika i dva mikrofona, pet govornika i pet mikrofona i '*umjetne mješavine signala pet govornika*'.

| $m = n = 2$ | $m = n = 5$ | $m = n = 5$ (<i>umjetna</i>) |
|--------------------------------------|--------------------------------------|--------------------------------------|
| Broj kanala: 1 (mono) | Broj kanala: 1 (mono) | Broj kanala: 1 (mono) |
| Frekvencija uzorkovanje: 96000 Hz | Frekvencija uzorkovanje: 16000 Hz | Frekvencija uzorkovanje: 16000 Hz |
| Rezolucija: 32 bit | Rezolucija: 32 bit | Rezolucija: 32 bit |
| Vrijeme: 30 s | Vrijeme: 30 s | Vrijeme: 87,5 s |

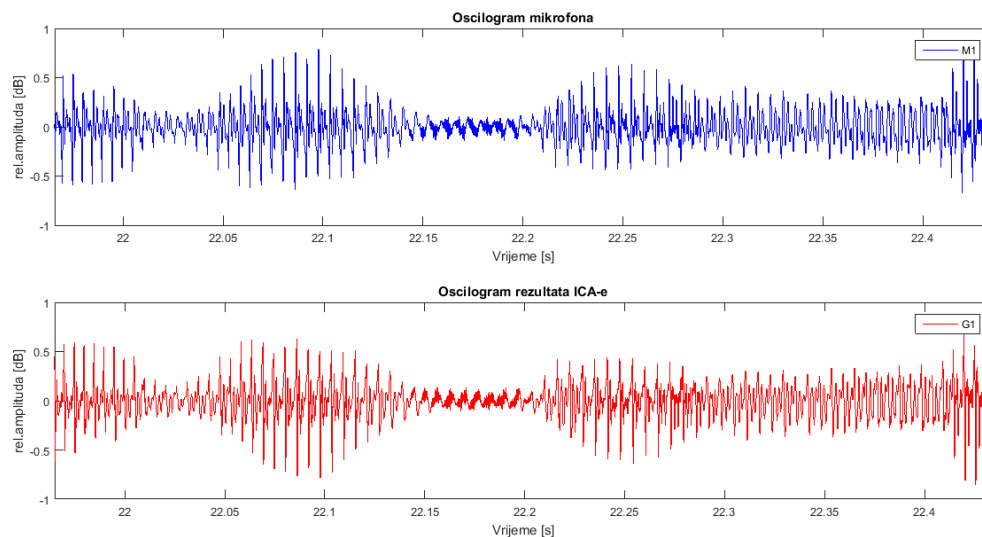
Prvi korak prije provedbe analize je sinkronizacija snimki, a drugi podešavanje veličine ulaznih matrica u sami algoritam ICA. Prvi korak za snimku pet govornika je bio kritičan jer je snimanje provedeno s pet neovisnih mikrofona i pet neovisnih aplikacija, te samim time sve snimke nisu istovremeno počele. Snimke pet govornika s pet mikrofona su sinkronizirane pomoću alata „Audacity“ određivanjem sinkronizacijskog uzorka u jednoj snimci koji je određen kao početak u preostalim snimkama. Drugi korak je učitavanje podataka iz audio zapisa u matrice korištenjem Matlaba i njihovo reduciranje kako bi se dobile matrice istih dimenzija, što je uvjet za ispravan rad algoritma ICA. Dobivena matrica, čiji redci odgovaraju snimkama s mikrofona, predstavlja ulazni podatak za algoritam fastICA na osnovu koje se vrši obrada. Iz dobivenih rezultata ICA-e utvrđuje se koja nezavisna komponenta odgovara

kojem govorniku jer rezultati ICA-e ne moraju biti poredani istim redom kao ulazne mješavine signala, redoslijed je nasumičan [10]. Subjektivnom procjenom (preslušavanjem) utvrđeno je koji rezultat ICA-e (redak izlazne matrice) odgovara kojem govorniku. Preslušavanje je osnovni kriteriji prema kojem se vidi da li su rezultati zadovoljavajući. Za sve tri promatrane situacije uspješno su upareni rezultati ICA-e s ulaznim vrijednostima, te je idućim korakom u analizi uspoređen *intenzitet* i *fundamentalna frekvencija F0* 'svakog para'. Također je prikazan izravan prikaz električnog signala u vremenu – *oscilogram* za svaku mješavinu signala i njoj odgovarajući rezultat ICA-e. Pomoću „Praat-a“ na jednostavan način možemo dobiti listu vrijednosti intenziteta i 'pitch-a' za željeni vremenski period (*intensity listing, pitch listing*). Želimo pronaći korelaciju između 'parova', svaki par čini snimka mikrofona i odgovarajući rezultat ICA-e koji mu je dodijeljen slušanjem. Vrijednosti *intenziteta* za svaki pojedini par dovedene su u korelaciju pomoću naredbe *corrcoef*. Koeficijent korelacije (naredba *corrcoef* u Matlab-u) je numerička mjera neke vrste korelacije, što znači statistički odnos između dvije varijable [1]. Kao izlaznu vrijednost dobijemo *matricu koeficijenata korelacije - R* i *matricu P - vrijednosti* za ispitivanje nulte hipoteze (da nema veze između promatranih pojava). *Matrica koeficijenata korelacije - R* je simetrična matrica dimenzija 2x2. Na glavnoj dijagonali nalaze se jedinice prema konvenciji, a na sporednoj dijagonali nalaze se *koeficijenti korelacije* za uneseni par varijabli. Vrijednosti koeficijenata mogu biti u rasponu od -1 do 1 pri čemu -1 predstavlja *negativnu korelaciju*, 0 predstavlja da uneseni par varijabli *nije u korelaciji*, a 1 predstavlja *pozitivnu korelaciju*. *Matrica P - vrijednosti* je također simetrična i istih dimenzija kao *R*. Na glavnoj dijagonali nalaze se jedinice, a na sporednoj *P- vrijednosti* koje mogu biti u rasponu od 0 do 1 pri čemu 0 predstavlja *značajnu korelaciju* u *R* i nisku vjerojatnost u promatranju nulte hipoteze [13]. Vrijednosti *fundamentalne frekvencije* za svaki pojedini par nisu dovedene u korelaciju koristeći *corrcoef* zato što je krivulja *fundamentalne frekvencije* isprekidana i ima previše nedefiniranih vrijednosti, NaN vrijednosti u Matlabu. Ovakav izgled krivulje *F0* s kakvim se susrećemo u ovom radu nije tipičan. Razlog je što nemamo originalne izvore signala (govornike) već mješavine signala iz kojih rekonstruiramo govornike. Problem je riješen na način da su nedefinirane vrijednosti *F0* aproksimirane s prosječnom vrijednošću te potom uspoređivane aproksimirane krivulje *F0* za svaki par. Izračunata je srednja vrijednost i standardna devijacija svake aproksimirane krivulje *F0* u Excel-u i potom su uspoređeni parovi koje smo povezali subjektivnom procjenom (preslušavanjem snimki). (U tablicama i na grafovima koji se pojavljuju u ostatku rada korištene su sljedeće oznake: M1 – *mješavina*

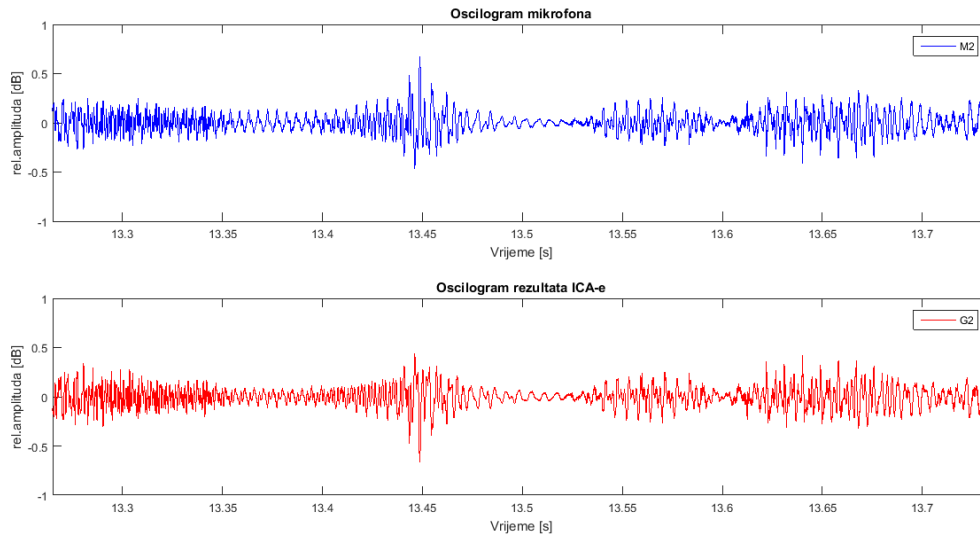
signala na prvom mikrofону, G1 – rezultat ICA-e koji odgovara prvom govorniku, S1 - prvi izvorni signala...analogno su označeni ostali parovi koje smo promatrali.)

3.1 DVA GOVORNIKA I DVA MIKROFONA

Na snimanju su sudjelovala dva govornika (jedna žena i jedan muškarac). Snimanje je obavljeno na profesionalnim mikrofönima koji su bili kalibrirani. Mikrofonu su bili razmješteni tako da je svakom govorniku jedan mikrofön bio znatno bliži nego drugom. Na taj se način htjelo postići da na svakom senzoru kao izlazna vrijednost postoji jedan govornik koji se 'ističe' tj. bolje čuje od drugog.

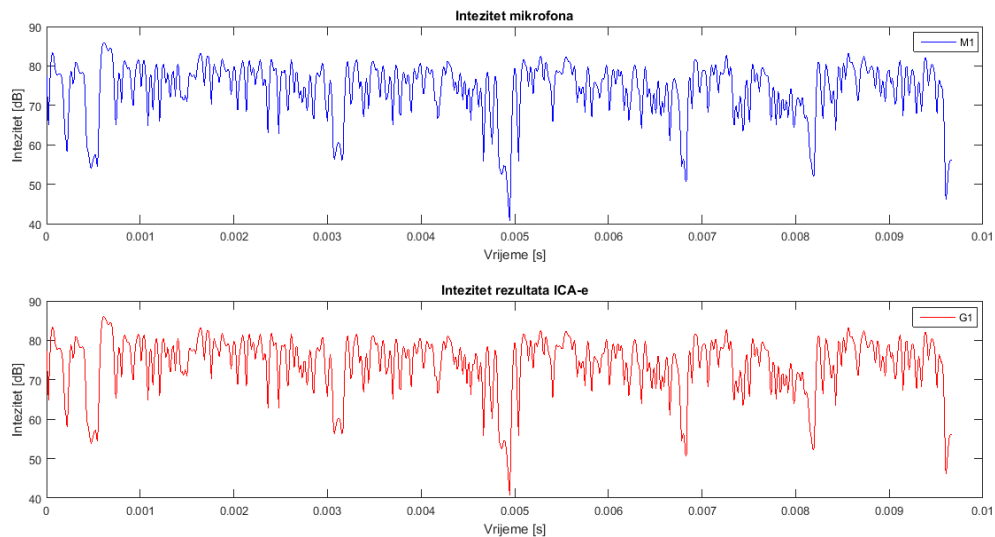


Graf 1: Plavom bojom je označen oscillogram mješavine signala na prvom mikrofönu (M1), a crvenom oscillogram rezultata ICA-e koji odgovara prvom govorniku (G1).

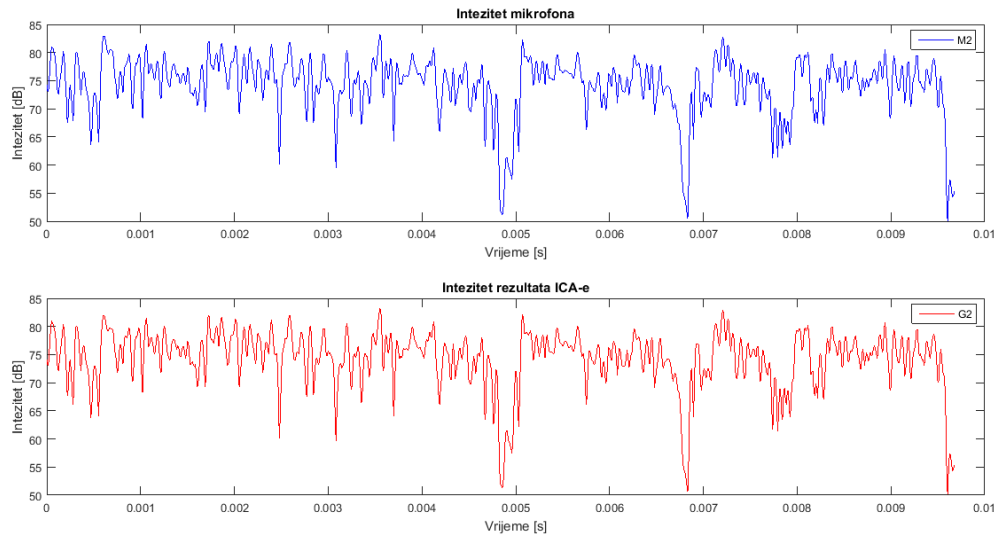


Graf 2: Plavom bojom je označen oscilogram mješavine signala na drugom mikrofону (M2), a crvenom oscilogram rezultata ICA-e koji odgovara drugom govorniku (G2).

Na temelju preslušavanja rezultata utvrđeno je koji mikrofón upariti s kojim rezultatom ICA-e. Iz priloženih oscilograma Graf 1 i Graf 2 vidi se da se osim 'slušanjem' njihovi signali preklapaju i u vremenu.



Graf 3: Plavom bojom je označen intenzitet mješavine signala na prvom mikrofónu (M1), a crvenom intenzitet rezultata ICA-e koji odgovara prvom govorniku (G1).

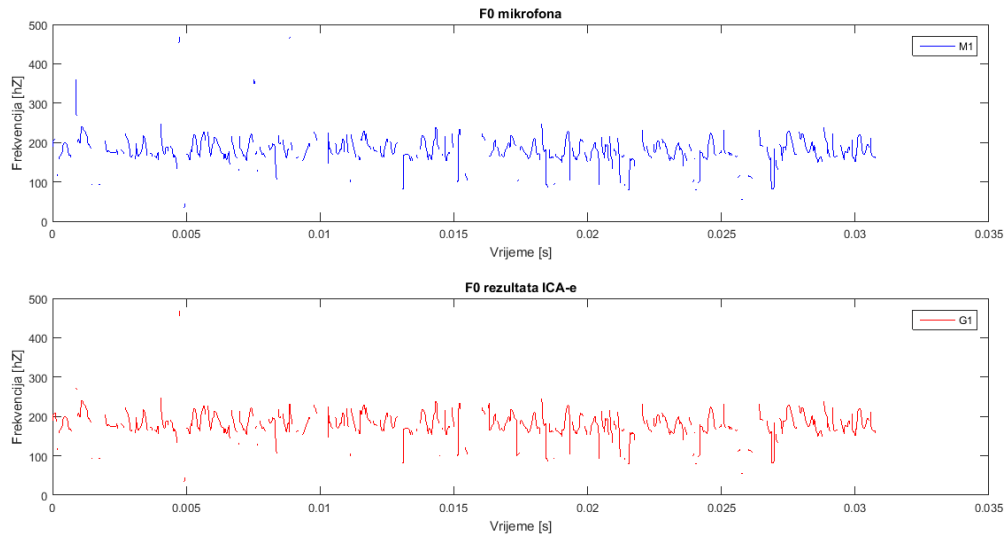


Graf 4: Plavom bojom je označen intenzitet mješavine signala na drugom mikrofona (M2), a crvenom intenzitet rezultata ICA-e koji odgovara drugom govorniku (G2).

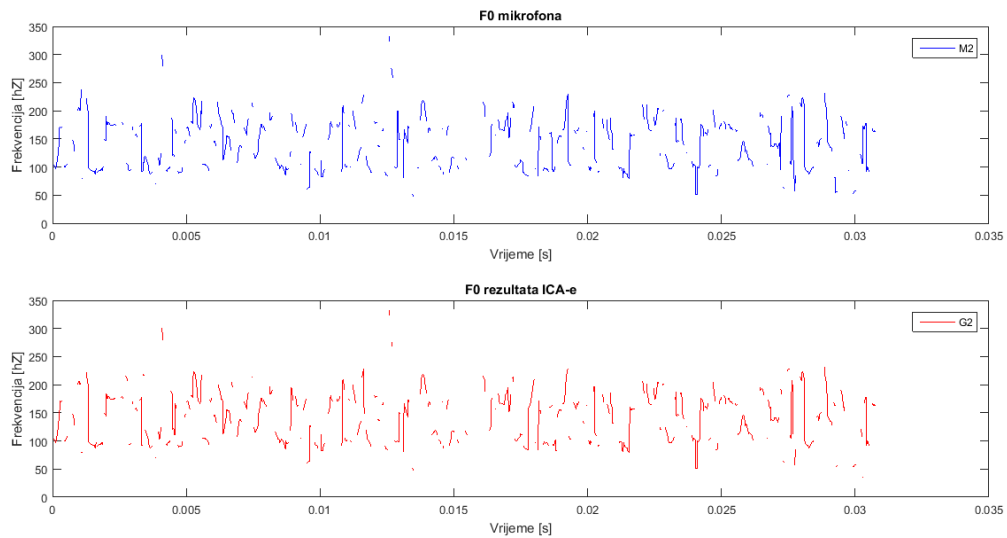
Tablica 2: Vrijednosti koeficijenta korelacije i P-vrijednosti za parove M1 i G1 te M2 i G2.

| | | |
|-------|--|--|
| M1_G1 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 0,9999 \\ 0,9999 & 1,0000 \end{bmatrix}$ |
| M2_G2 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 0,9981 \\ 0,9981 & 1,0000 \end{bmatrix}$ |

Uspoređujući grafički *intenzitete* parova mikrofona – govornik na Grafovima 3 i 4 vidi se da su u korelaciji. Rezultati *coeffcor-a* dani su u Tablici 2. R je približno jedan, a P je nula za oba slučaja što ukazuje na *značajnu pozitivnu korelaciju*.



Graf 5: Plavom bojom označena je fundamentalna frekvencija mješavine signala na prvom mikrofону (M1), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara prvom govorniku (G1).



Graf 6: Plavom bojom označena je fundamentalna frekvencija mješavine signala na drugom mikrofону (M2), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara drugom govorniku (G2).

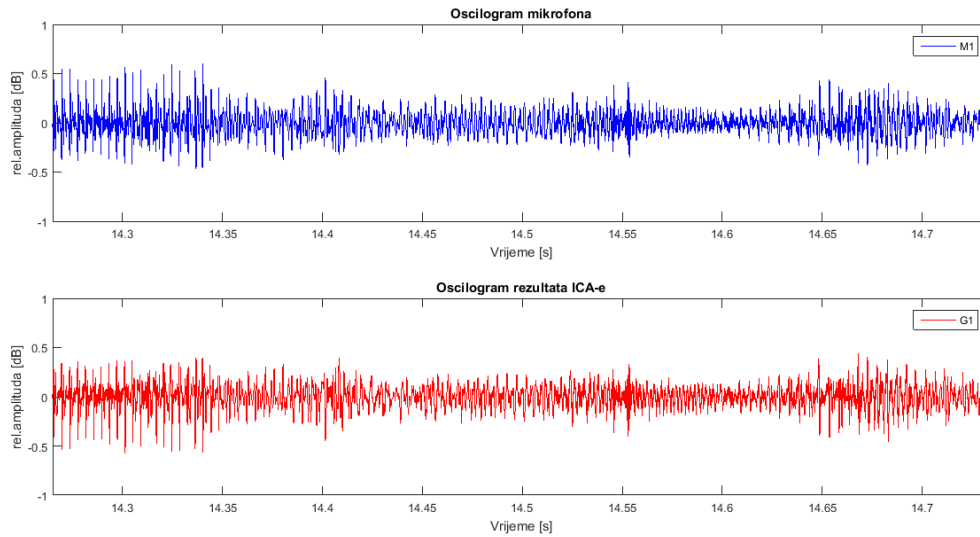
Tablica 3: Rezultati *aproksimirane fundamentalne frekvencije* za parove M1 i G1 te M2 i G2.

| | |
|----------------------|----------------------|
| M1 aproks. F0 | G1 aproks. F0 |
| 178,396 ± 30,412 Hz | 177,668 ± 28,068 Hz |
| M2 aproks. F0 | G2 aproks. F0 |
| 139,721 ± 36,298 Hz | 139,728 ± 36,181 Hz |

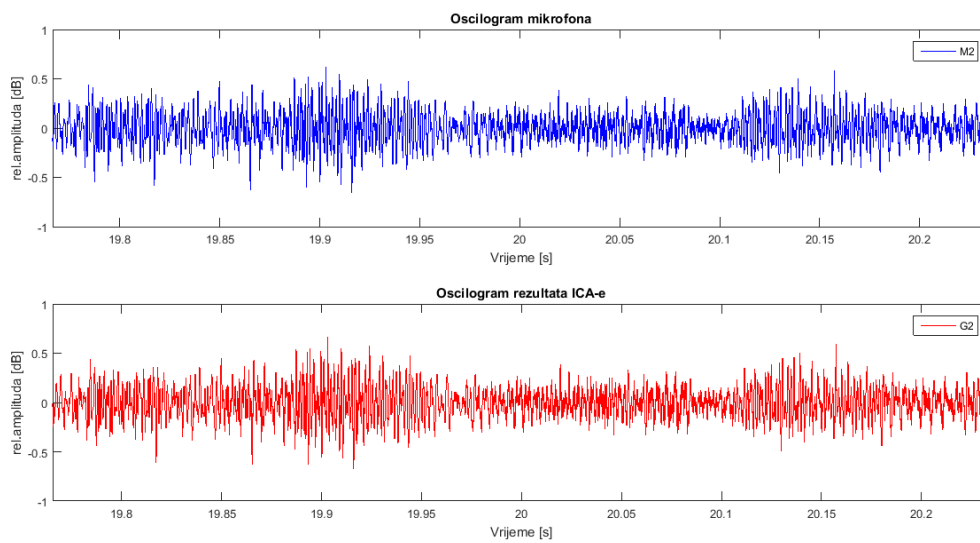
Uspoređujući grafički *fundamentalnu frekvenciju* parova mikrofona – govornik na Grafu 5 i Grafu 6 vidi se sličnost. Zbog rascjepkanosti krivulje bilo je nemoguće primijeniti *coeffcor*. Izračunata je *srednja vrijednost* i *standardna devijacija aproksimiranih krivulja fundamentalnih frekvencija (aproks. F0)* parova govornik - mikrofona i iz priloženih rezultata u Tablici 3 vidi se podudarnost. Osim preslušavanjem iz podataka u Tablici 3 vidi se da govornik jedan odgovara ženskom glasu, a govornik dva odgovara muškom glasu što je dodatna potvrda prethodno navedenog.

3.2 PET GOVORNIKA I PET MIKROFONA

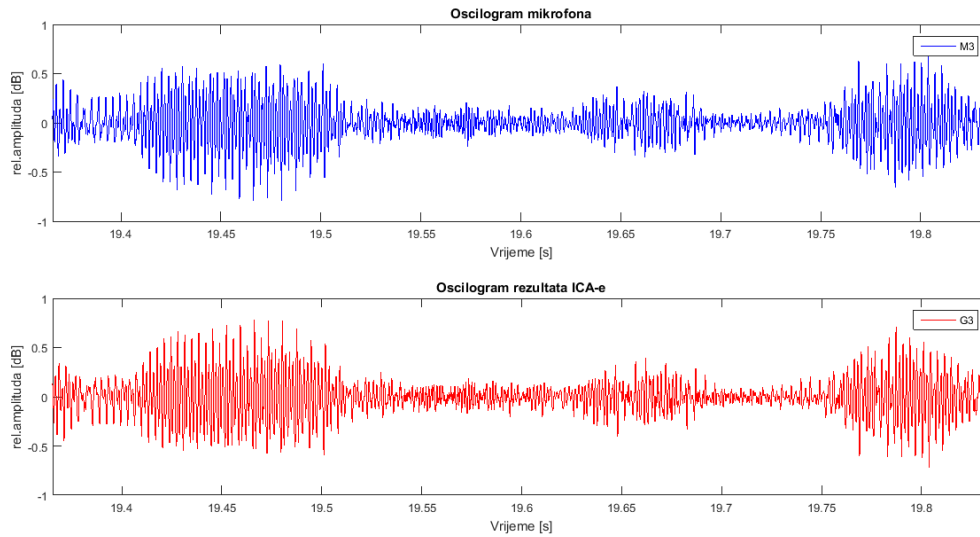
Na snimanju je sudjelovalo pet ljudi (četiri žene i jedan muškarac). Za razliku od prvog snimanja na ovom snimanju su kao senzori poslužili pametni telefoni, različitih karakteristika i proizvođača. Na svakom pametnom telefonu bila je instalirana aplikacija *Smart Voice Recorder* pomoću koje je snimano. Unutar aplikacije postoji opcija kalibracije i postavke koje su na svakom mobitelu bile podešene jednako. Pametni telefoni su bili razmješteni po *tihoj sobi* tako da je svakom govorniku jedan mobitel bio znatno bliži nego svim drugim govornicima. Razlog je isti kao i kod prvog snimanja.



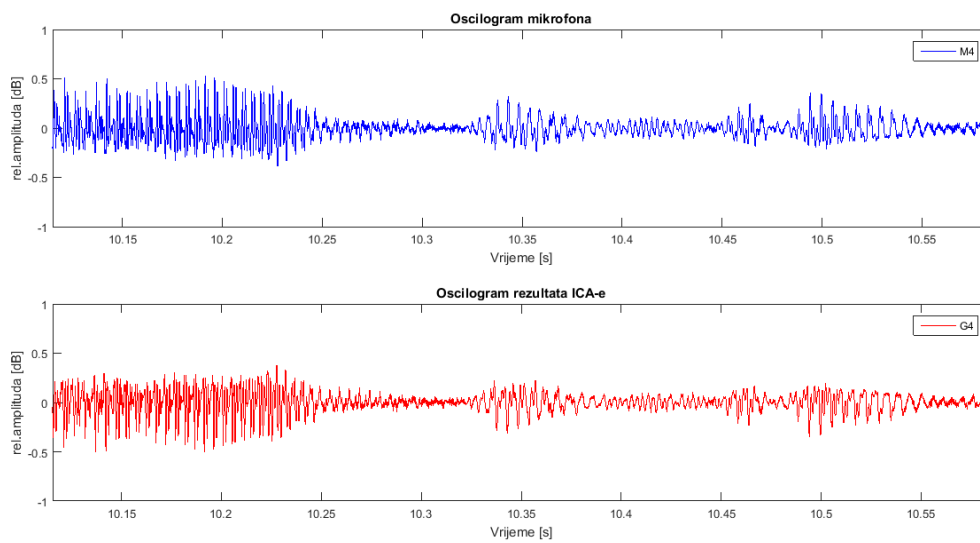
Graf 7: Plavom bojom je označen oscilogram mješavine signala na prvom mikrofona (M1), a crvenom oscilogram rezultata ICA-e koji odgovara prvom govorniku (G1).



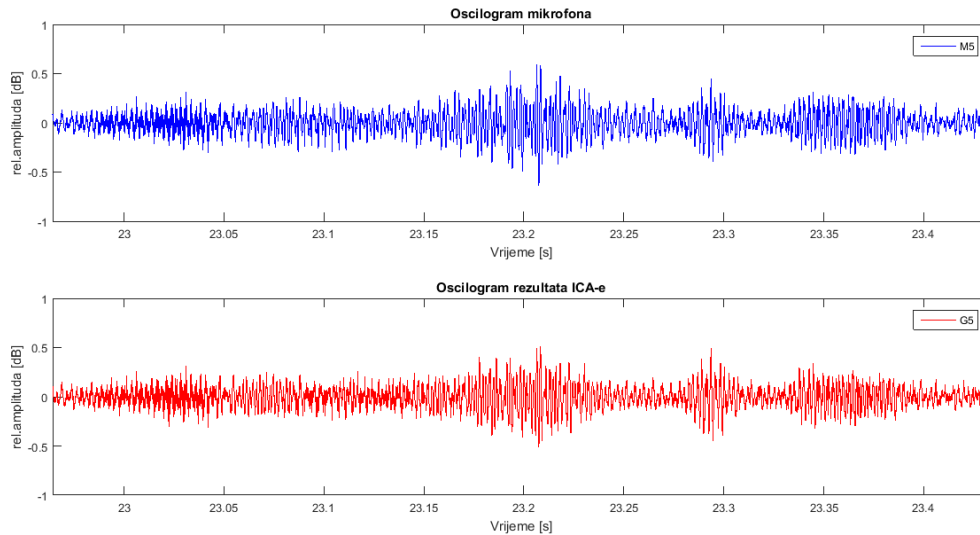
Graf 8: Plavom bojom je označen oscilogram mješavine signala na drugom mikrofona (M2), a crvenom oscilogram rezultata ICA-e koji odgovara drugom govorniku (G2).



Graf 9: Plavom bojom je označen oscilogram mješavine signala na trećem mikrofону (M3), a crvenom oscilogram rezultata ICA-e koji odgovara trećem govorniku (G3).

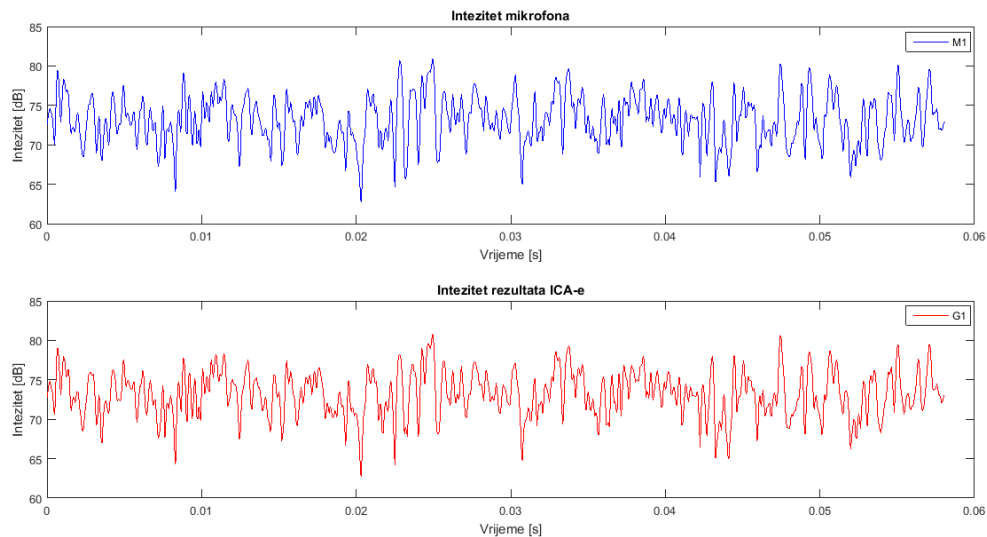


Graf 10: Plavom bojom je označen oscilogram mješavine signala na četvrtom mikrofону (M4), a crvenom oscilogram rezultata ICA-e koji odgovara četvrtom govorniku (G4).

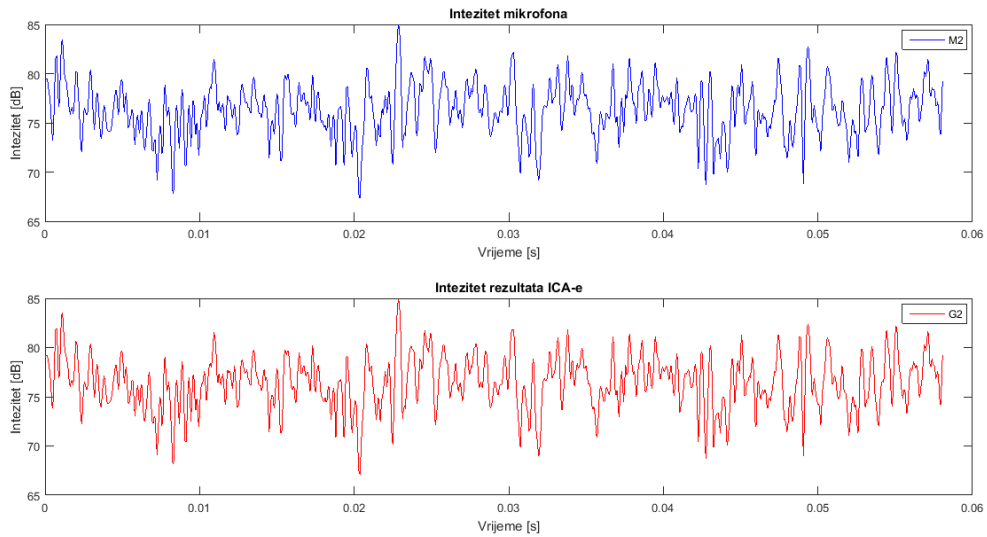


Graf 11: Plavom bojom je označen oscilogram mješavine signala na petom mikrofону (M5), a crvenom oscilogram rezultata ICA-e koji odgovara petom govorniku (G5).

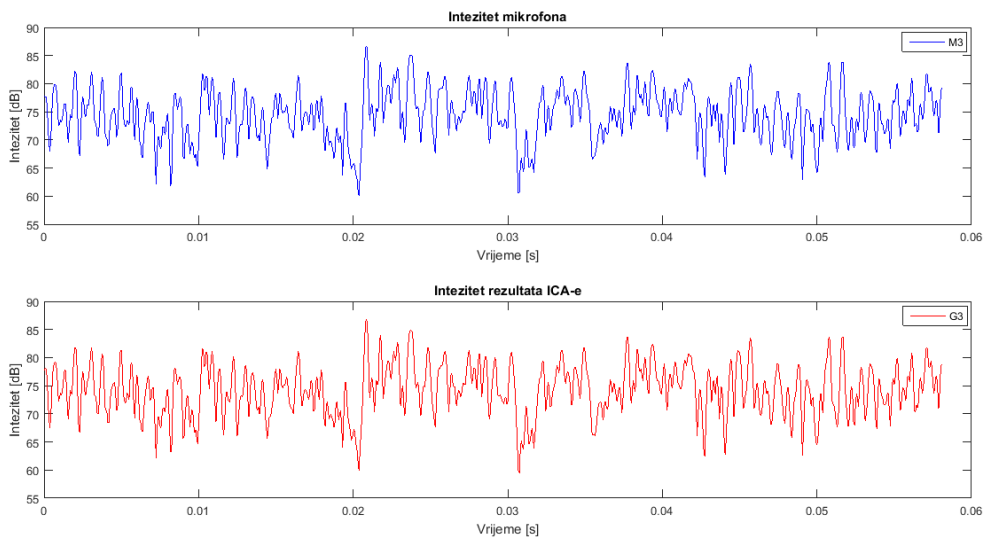
Kao i u prethodnom slučaju na temelju preslušavanja rezultata utvrđeno je koji mikrofón upariti s kojim rezultatom ICA-e. Iz priloženih oscilograma Graf 7 – Graf 11 vidi se je se da se osim 'slušanjem' njihovi signali preklapaju i u vremenu.



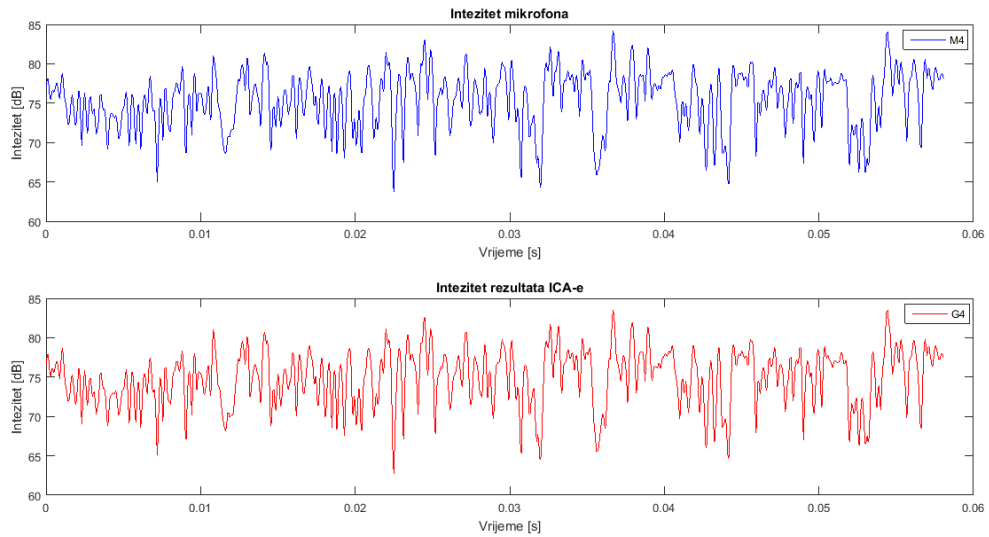
Graf 12: Plavom bojom je označen intenzitet mješavine signala na prvom mikrofónu (M1), a crvenom intenzitet rezultata ICA-e koji odgovara prvom govorniku (G1).



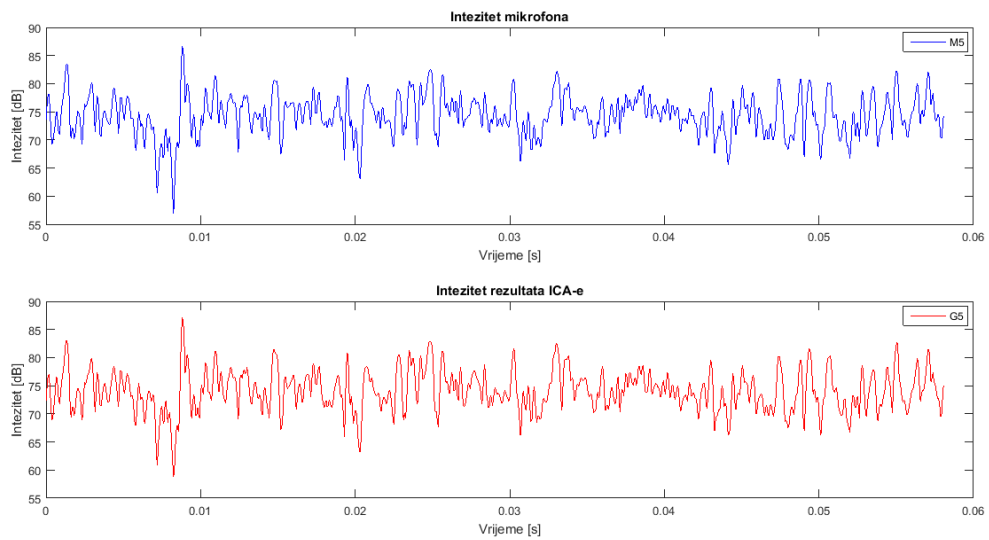
Graf 13: Plavom bojom je označen intenzitet mješavine signala na drugom mikrofону (M2), a crvenom intenzitet rezultata ICA-e koji odgovara drugom govorniku (G2).



Graf 14: Plavom bojom je označen intenzitet mješavine signala na trećem mikrofону (M3), a crvenom intenzitet rezultata ICA-e koji odgovara trećem govorniku (G3).



Graf 15: Plavom bojom je označen intenzitet mješavine signala na četvrtom mikrofону (M4), a crvenom intenzitet rezultata ICA-e koji odgovara četvrtom govorniku (G4).

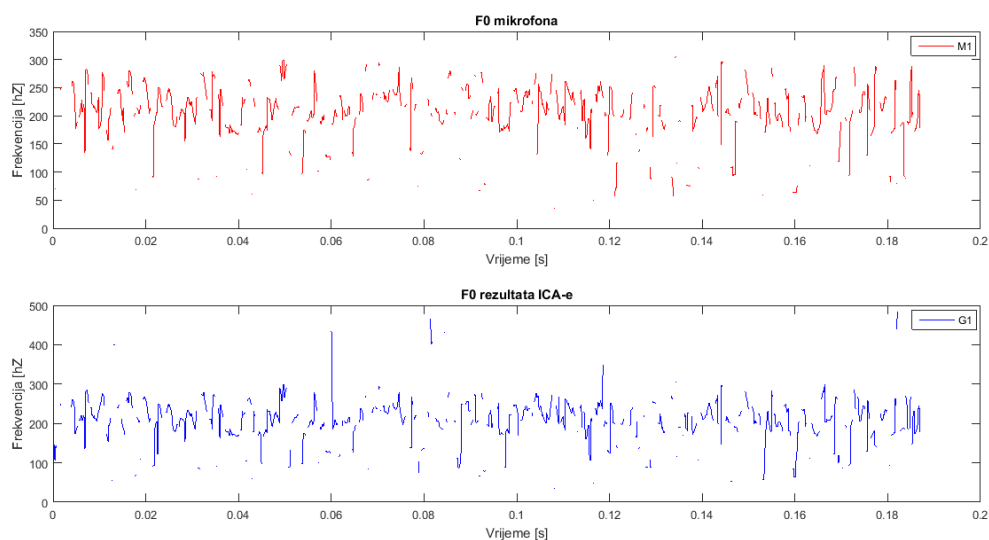


Graf 16: Plavom bojom je označen intenzitet mješavine signala na petom mikrofону (M5), a crvenom intenzitet rezultata ICA-e koji odgovara petom govorniku (G5).

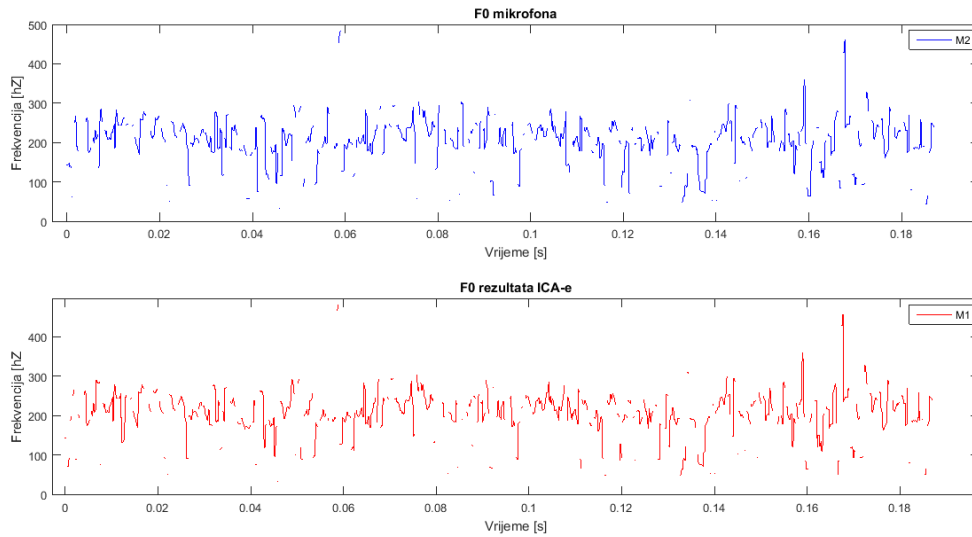
Tablica 4: Vrijednosti *koeficijenta korelacije* i *P-vrijednosti* za parove M1 i G1, M2 i G2, M3 i G3, M4 i G4 te M5 i G5.

| | | |
|--------------|--|--|
| M1_G1 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 0,9779 \\ 0,9779 & 1,0000 \end{bmatrix}$ |
| M2_G2 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 0,9945 \\ 0,9945 & 1,0000 \end{bmatrix}$ |
| M3_G3 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 0,9945 \\ 0,9945 & 1,0000 \end{bmatrix}$ |
| M4_G4 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 0,9968 \\ 0,9968 & 1,0000 \end{bmatrix}$ |
| M5_G5 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 0,9758 \\ 0,9758 & 1,0000 \end{bmatrix}$ |

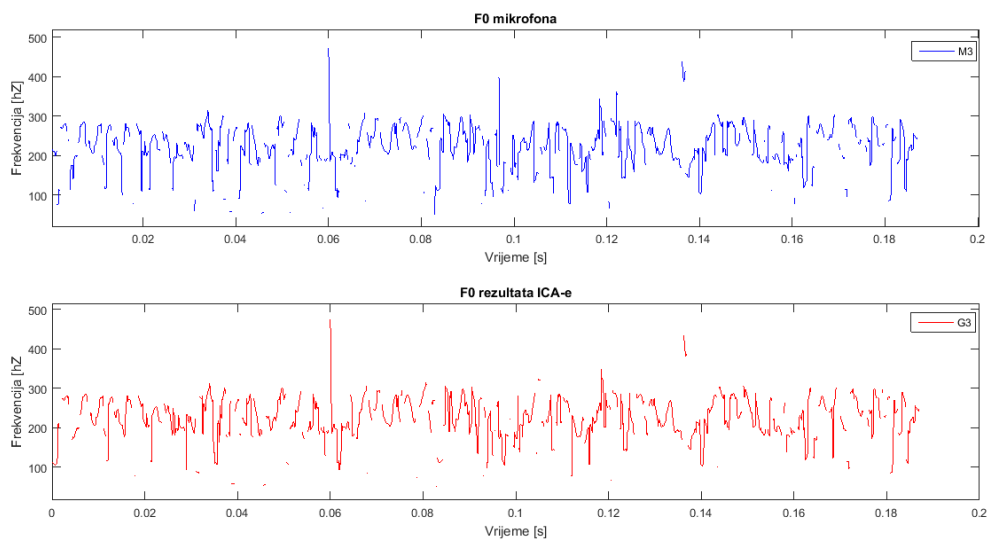
Uspoređujući grafički *intenzitete* parova mikrofona – govornik od Grafa 12 – Grafa 16 vidi se da su u korelaciji. Rezultati *coeffcor-a* dani su u Tablici 4. R je približno jedan, a P je nula za svih pet slučajeva što ukazuje na *značajnu pozitivnu korelaciju*.



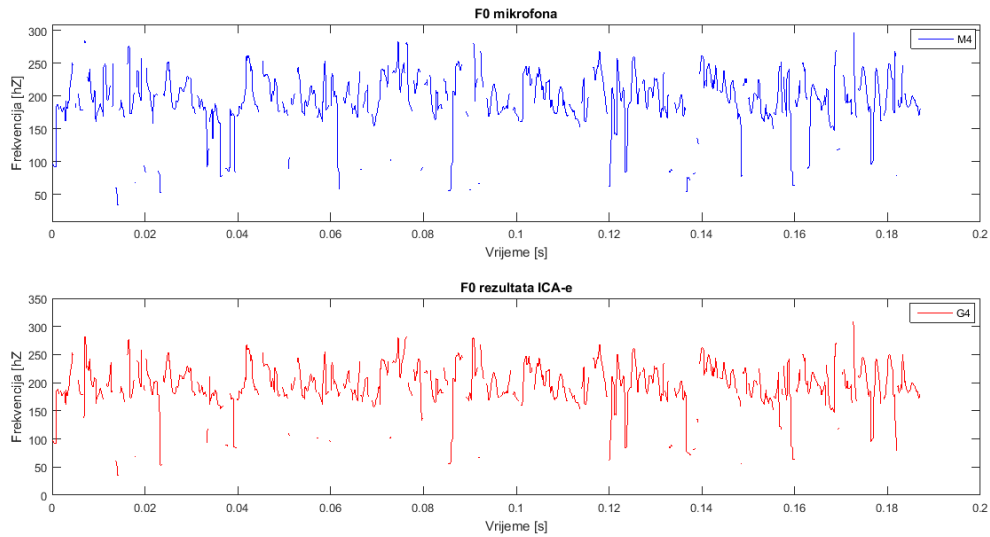
Graf 17: Plavom bojom je označena fundamentalna frekvencija mješavine signala na prvom mikrofona (M1), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara prvom govorniku (G1).



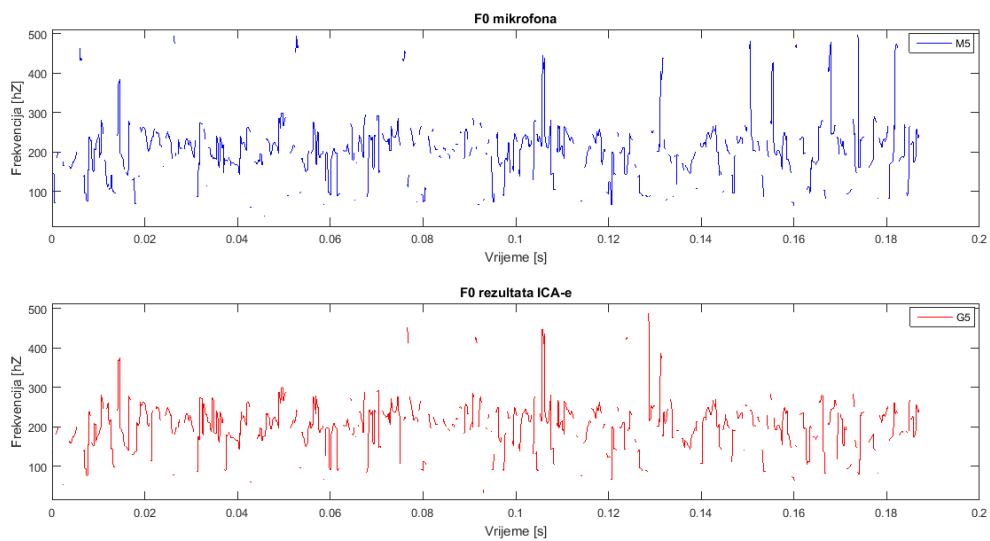
Graf 18: Plavom bojom je označena fundamentalna frekvencija mješavine signala na drugom mikrofону (M2), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara drugom govorniku (G2).



Graf 19: Plavom bojom je označena fundamentalna frekvencija mješavine signala na trećem mikrofону (M3), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara trećem govorniku (G3).



Graf 20: Plavom bojom je označena fundamentalna frekvencija mješavine signala na četvrtom mikrofону (M4), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara četvrtom govorniku (G4).



Graf 21: Plavom bojom je označena fundamentalna frekvencija mješavine signala na petom mikrofону (M5), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara petom govorniku (G5).

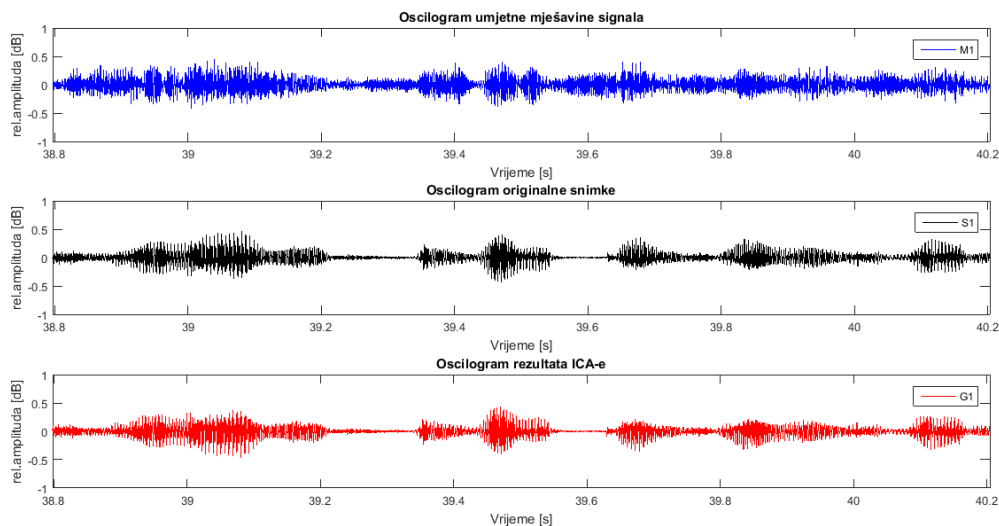
Tablica 5: Rezultati *aproksimirane fundamentalne frekvencije* za parove M1 i G1, M2 i G2, M3 i G3, M4 i G4 te M5 i G5.

| | |
|----------------------|----------------------|
| M1 aproks. F0 | G1 aproks. F0 |
| 202,309 ± 44,297 Hz | 203,761 ± 48,335 Hz |
| M2 aproks. F0 | G2 aproks. F0 |
| 201,679 ± 52,107 Hz | 200,472 ± 51,407 Hz |
| M3 aproks. F0 | G3 aproks. F0 |
| 219,215 ± 55,283 Hz | 220,388 ± 52,719 Hz |
| M4 aproks. F0 | G4 aproks. F0 |
| 190,708 ± 42,069 Hz | 191,8851 ± 39,333 Hz |
| M5 aproks. F0 | G5 aproks. F0 |
| 203,274 ± 66,892 Hz | 200,112 ± 51,950 Hz |

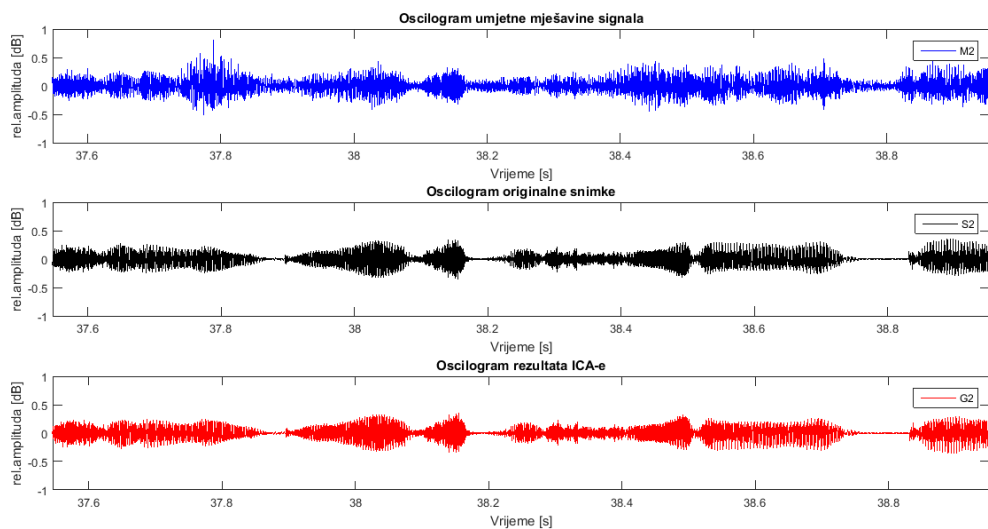
Uspoređujući grafički *fundamentalnu frekvenciju* parova mikrofona – govornik Graf 17 - Graf 21 vidljiva je sličnost. Zbog rascjepkanosti krivulje bilo je nemoguće primijeniti *coeffcor*. Izračunata je *srednja vrijednost* i *standardna devijacija aproksimiranih krivulja fundamentalnih frekvencija (aproks. F0)* parova govornik-mikrofon i iz priloženih rezultata u Tablici 5 vidimo podudarnost.

3.3 PET GOVORNIKA - UMJETNE MJEŠAVINE SIGNALA

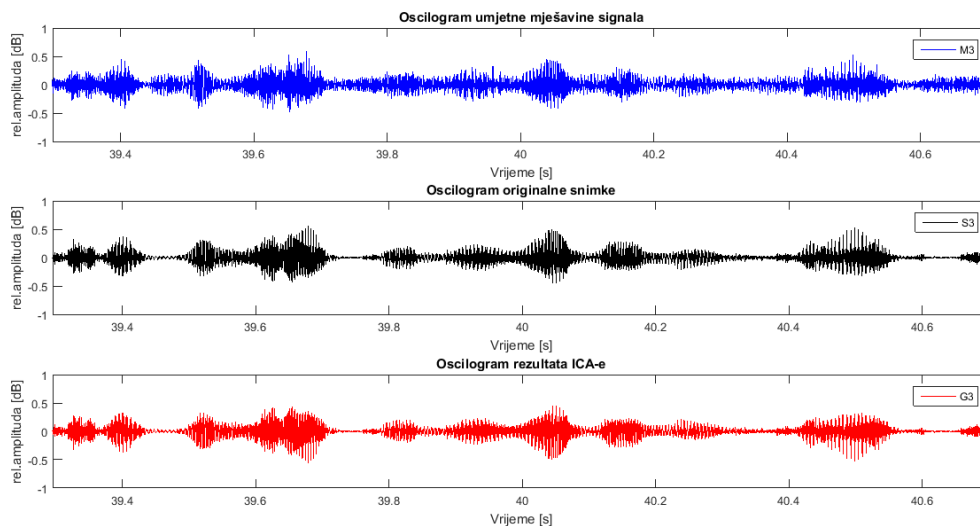
Svakog od pet govornika se snimalo pojedinačno na pet mikrofona. Svrha ovog snimanja je bila usporediti rezultate ICA-e stvarne situacije s rezultatima ICA-e *umjetno načinjenim mješavinama signala*. Između pet snimki svakog govornika odabrane su one najbolje. *Umjetne mješavine signala* napravljene su na sljedeći način. Napravljen je sustav od pet linearnih jednadžbi gdje je svaka jednadžba linearna kombinacija pet pojedinačnih govornika. Cilj je bio da *umjetne mješavine signala* budu što sličnije stvarnim pa je svakoj jednadžbi jedan govornik malo 'pojačan' (množenje s većim skalarom), a ostali 'utišani' (množenje s manjim skalarom). Skalari su na prvu odabrani intuitivno pa su nakon preslušavanja po potrebi modificirani. Očekuju se puno bolji rezultati budući da je poznata matrica miješanja **A** čiji elementi su gore navedeni skalari. U prethodnom slučaju matricu miješanja **A** bilo je potrebno procijeniti. Budući da su u ovom slučaju poznati izvori signala, dobivene rezultate uspoređivalo se sa njima, a ne sa mješavinama signala kao u predhodna dva slučaja.



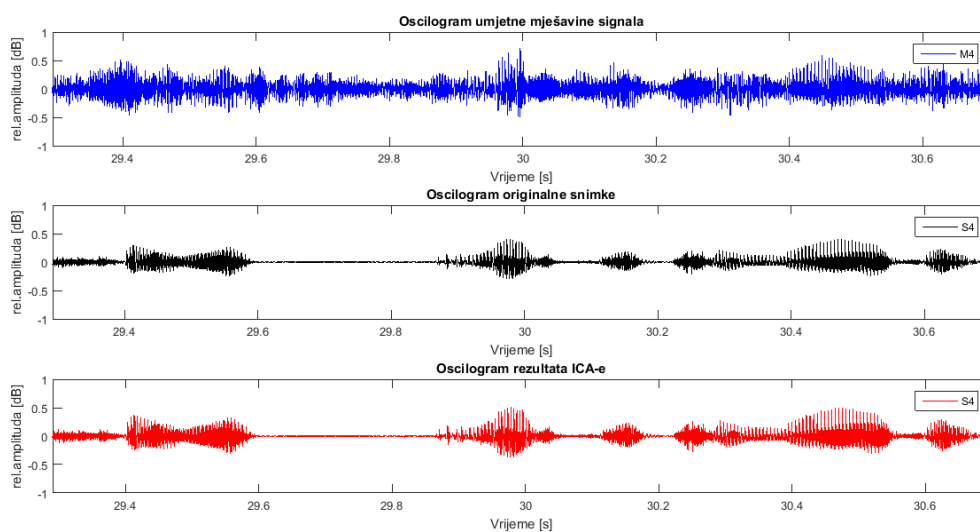
Graf 22: Plavom bojom je označen oscilogram mješavine signala na prvom mikrofону (M1), crnom bojom je označen oscilogram prvog izvornog signala (S1), a crvenom oscilogram rezultata ICA-e koji odgovara prvom izvoru (G1).



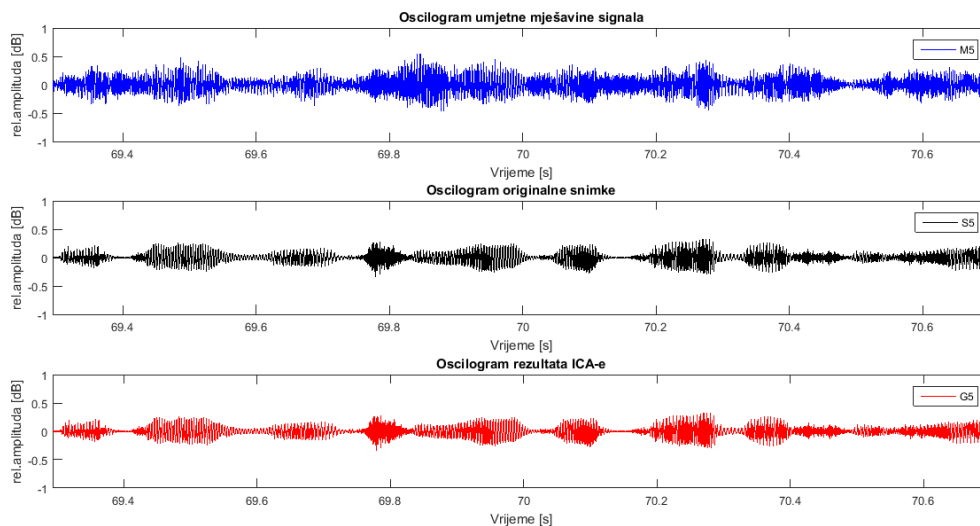
Graf 23: Plavom bojom je označen oscilogram mješavine signala na drugom mikrofону (M2), crnom bojom je označen oscilogram drugog izvornog signala (S2), a crvenom oscilogram rezultata ICA-e koji odgovara drugom izvoru (G2).



Graf 24: Plavom bojom je označen oscilogram mješavine signala na trećem mikrofону (M3), crnom bojom je označen oscilogram trećeg izvornog signala (S3), a crvenom oscilogram rezultata ICA-e koji odgovara trećem izvoru (G3).



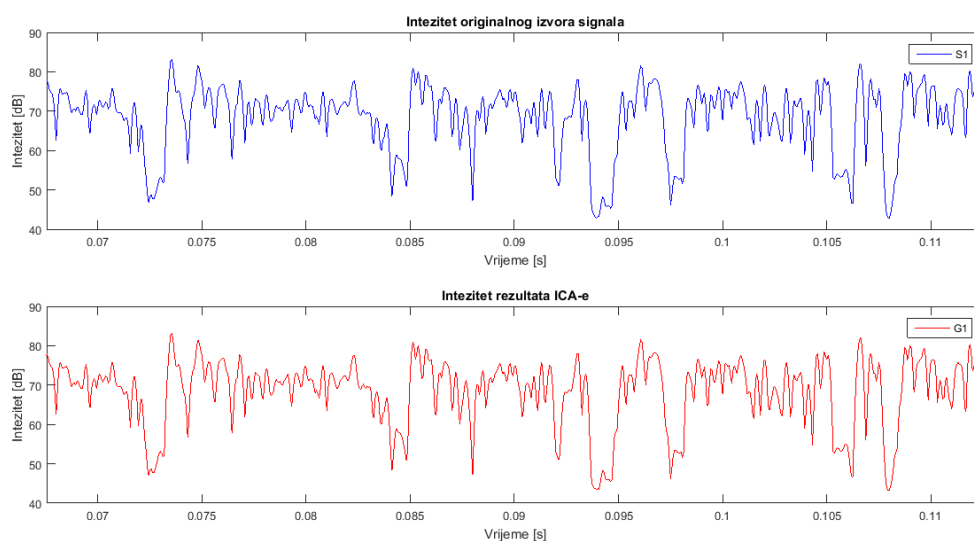
Graf 25: Plavom bojom je označen oscilogram mješavine signala na četvrtom mikrofону (M4), crnom bojom je označen oscilogram četvrtog izvornog signala (S4), a crvenom oscilogram rezultata ICA-e koji odgovara četvrtom izvoru (G4).



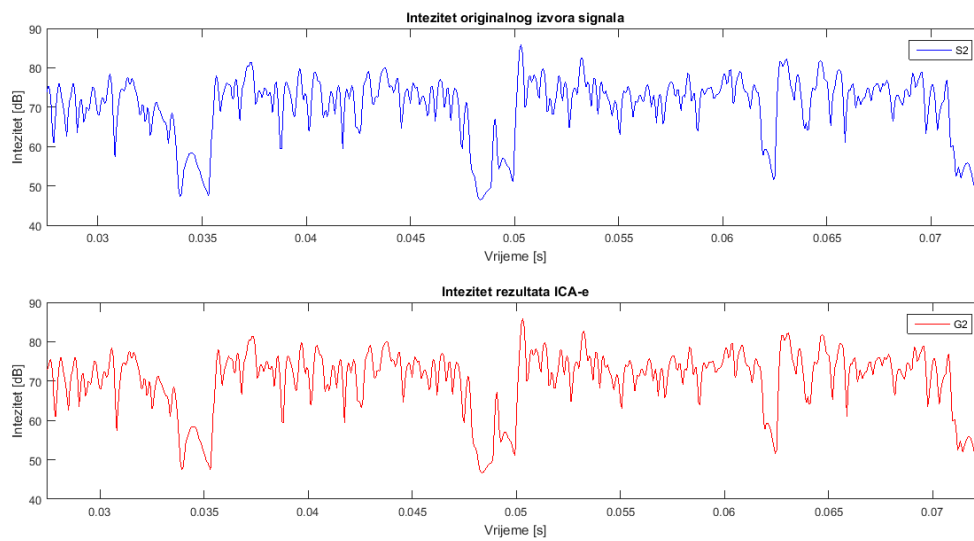
Graf 26: Plavom bojom je označen oscilogram mješavine signala na petom mikrofону (M5), crnom bojom je označen oscilogram petog izvornog signala (S5), a crvenom oscilogram rezultata ICA-e koji odgovara petom izvoru (G5).

Kao i u prethodna dva slučaja na temelju preslušavanja rezultata utvrđeno je koji mikrofón upariti s kojim rezultatom ICA-e. Iz priloženih oscilograma Graf 22 – Graf 26 vidi se je da se osim 'slušanjem' njihovi signali preklapaju i u vremenu.

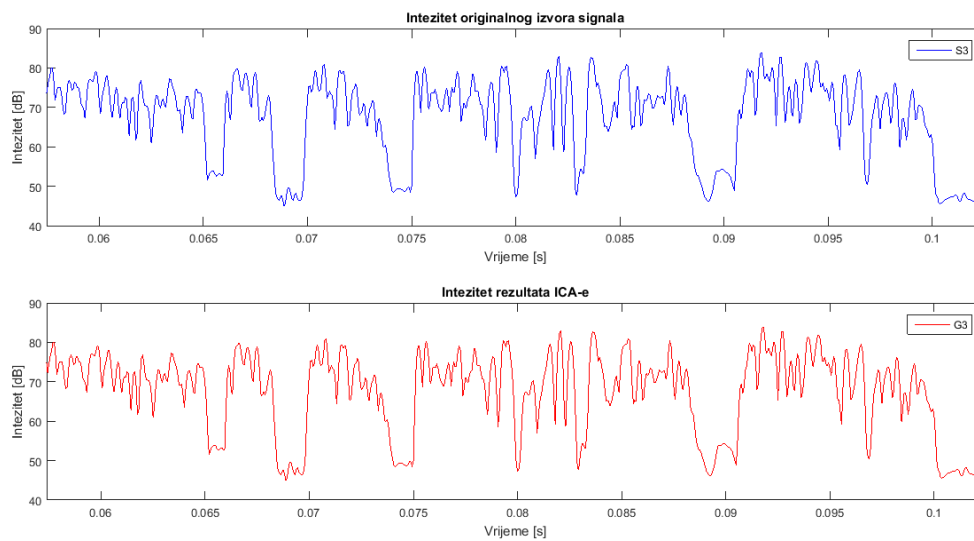
U daljnjem razmatranju uspoređivani su izvorni signali sa rezultatima ICA-e budući da su izvorni signali u ovom primjeru bili poznati.



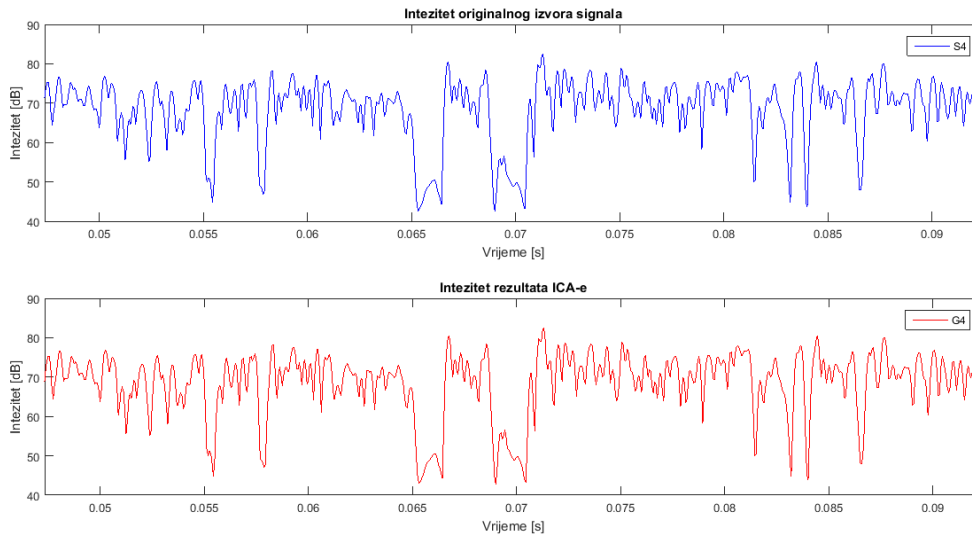
Graf 27: Plavom bojom je označen intenzitet prvog izvornog signala (S1), a crvenom intenzitet rezultata ICA-e koji odgovara prvom izvoru (G1).



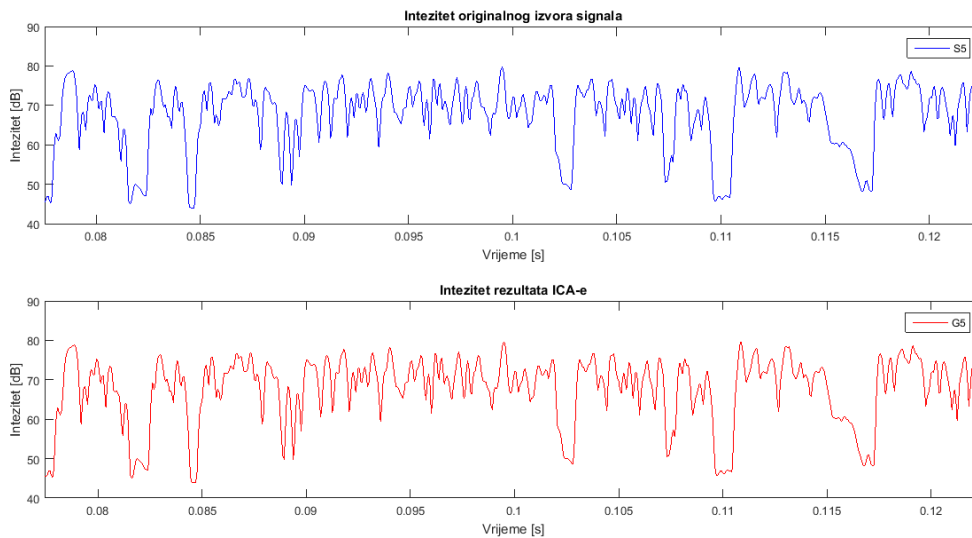
Graf 28: Plavom bojom je označen intenzitet drugog izvornog signala (M2), a crvenom intenzitet rezultata ICA-e koji odgovara drugom izvoru (G2).



Graf 29: Plavom bojom je označen intenzitet trećeg izvornog signala (S3), a crvenom intenzitet rezultata ICA-e koji odgovara trećem izvoru (G3).



Graf 30: Plavom bojom je označen intenzitet četvrto izvornog signala (S4), a crvenom intenzitet rezultata ICA-e koji odgovara četvrto izvoru (G4).

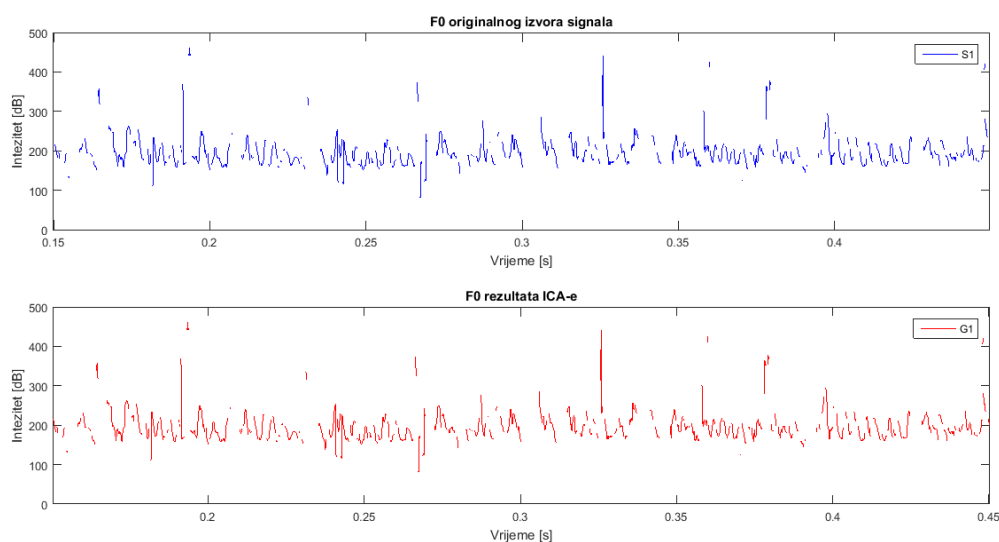


Graf 31: Plavom bojom je označen intenzitet petog izvornog signala (S5), a crvenom intenzitet rezultata ICA-e koji odgovara petom izvoru (G5).

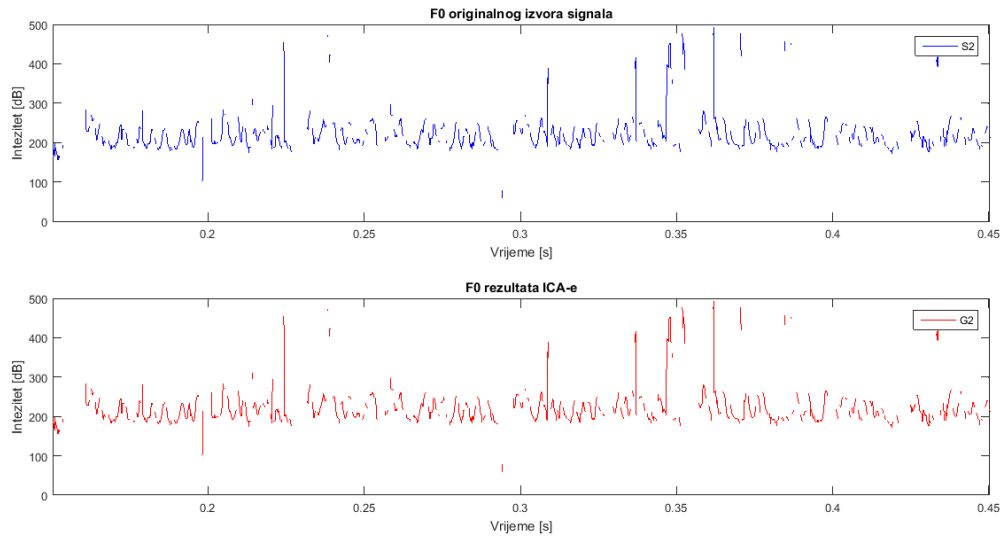
Tablica 6: Vrijednosti *koeficijenta korelacije* i *P-vrijednosti* za parove S1 i G1, S2 i G2, S3 i G3, S4 i G4 te S5 i G5.

| | | |
|--------------|--|--|
| S1_G1 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 1,0000 \\ 1,0000 & 1,0000 \end{bmatrix}$ |
| S2_G2 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 1,0000 \\ 1,0000 & 1,0000 \end{bmatrix}$ |
| S3_G3 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 1,0000 \\ 1,0000 & 1,0000 \end{bmatrix}$ |
| S4_G4 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 1,0000 \\ 1,0000 & 1,0000 \end{bmatrix}$ |
| S5_G5 | $P = \begin{bmatrix} 1,0000 & 0,0000 \\ 0,0000 & 1,0000 \end{bmatrix}$ | $R = \begin{bmatrix} 1,0000 & 1,0000 \\ 1,0000 & 1,0000 \end{bmatrix}$ |

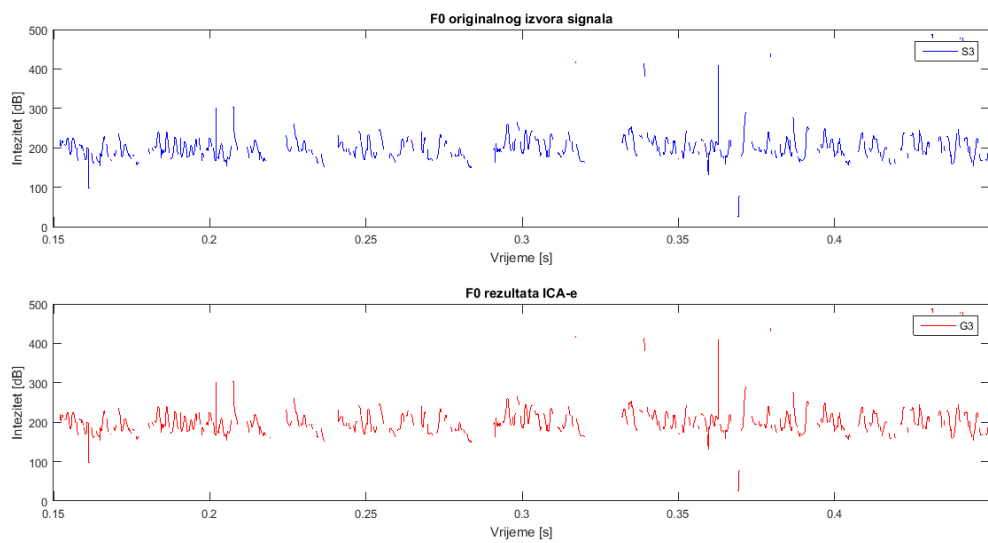
Uspoređujući grafički *intenzitete* parova izvorni signal – rezultat ICA-e koji odgovara tom izvoru od Grafa 27 – Grafa 31 vidi se je da su u korelaciji. Rezultati *coeffcor-a* dani su u Tablici 6. R je jedan, a P je nula za svih pet slučajeva što ukazuje na *savršenu pozitivnu korelaciju*.



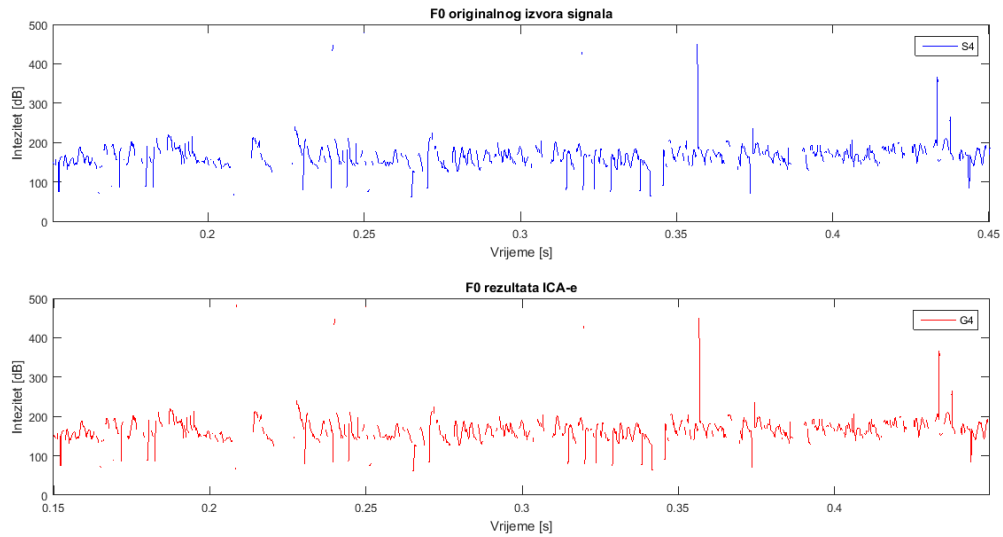
Graf 32: Plavom bojom je označena fundamentalna frekvencija prvog izvornog signala (S1), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara prvom govorniku (G1).



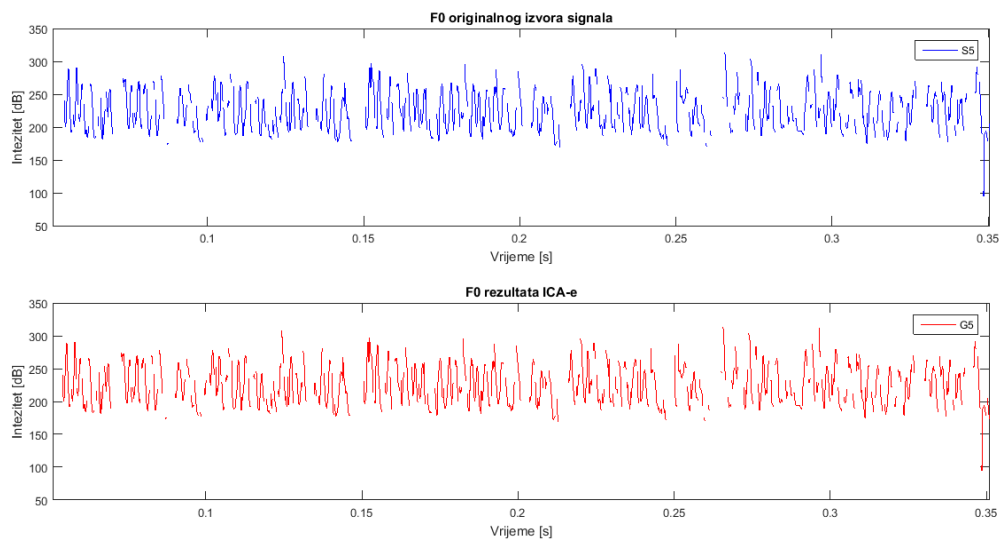
Graf 33: Plavom bojom je označena fundamentalna frekvencija drugog izvornog signala (S2), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara drugom govorniku (G2).



Graf 34: Plavom bojom je označena fundamentalna frekvencija trećeg izvornog signala (S3), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara trećem izvoru (G3).



Graf 35: Plavom bojom je označena fundamentalna frekvencija četvrtog izvornog signala (S4), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara četvrtom izvoru (G4).



Graf 36: Plavom bojom je označena fundamentalna frekvencija petog izvornog signala (S5), a crvenom fundamentalna frekvencija rezultata ICA-e koji odgovara petom izvoru (G5).

Tablica 7: Rezultati *aproksimirane fundamentalne frekvencije* za parove S1 i G1, S2 i G2, S3 i G3, S4 i G4 te S5 i G5.

| | |
|----------------------|----------------------|
| G1 aproks. F0 | S1 aproks. F0 |
| 196,305 ± 29,501 Hz | 196,298 ± 29,507 Hz |
| G2 aproks. F0 | S2 aproks. F0 |
| 218,803 ± 36,706 Hz | 218,785 ± 36,658 Hz |
| G3 aproks. F0 | S3 aproks. F0 |
| 200,151 ± 28,915 Hz | 200,147 ± 28,920 Hz |
| G4 aproks. F0 | S4 aproks. F0 |
| 166,132 ± 27,028 Hz | 165,326 ± 26,745 Hz |
| G5 aproks. F0 | S5 aproks. F0 |
| 224,431 ± 23,814 Hz | 224,436 ± 23,809 Hz |

Uspoređujući grafički *fundamentalnu frekvenciju* parova izvorni signal – rezultat ICA-e koji odgovara tom izvoru iz Grafova 32 - 36 vidi se sličnost. Zbog rascjepkanosti krivulje bilo je nemoguće primijeniti *coeffcor*. Izračunata je *srednja vrijednost* i *standardna devijacija aproksimiranih krivulja fundamentalnih frekvencija (aproks. F0)* parova izvorni signal – rezultat ICA-e koji odgovara tom izvoru i iz priloženih rezultata u Tablici 7 vidi se podudarnost.

4 RASPRAVA I ZAKLJUČAK

Metodom ICA uspješno su obrađena tri slučaja: 2 govornika, 5 govornika stvarna situacija i *'umjetna mješavina signala pet govornika'*. Za sva tri slučaja dobiveni su identificirani izvori signala tj. ekstrahiran govor pojedinog govornika s tom razlikom da je za treći slučaj kao što je i bilo očekivano dobiven najbolji rezultat. Prva dva promatrana slučaja su primjer stvarnih životnih situacija. Pomicanje govornika, okretanje glave, udaljavanje od svog mikrofona, približavanje drugom mikrofona, promjena intenziteta govora govornika itd. razlozi su zbog kojih su unutar identificiranih izvora signala i dalje prisutni utišani ostali izvori. Dobiveni rezultati u prva dva slučaja su pored ovoga zadovoljavajući tj. dobiveni su audio signali na kojima se jasno čuje pojačan svaki posebni govornik. Kod trećeg slučaja situacija je očekivano savršena, tj. kao rezultat su dobiveni audio signali na kojima se čuje samo jedan govornik. Na dobivenim podacima je vršena daljnja analiza. Rezultati *coeffcor-a* za intenzitete *parove mikrofona – govornik* ukazuje na *značajnu pozitivnu korelaciju* kod prva dva slučaja, a kod trećeg slučaja ukazuju na *savršenu pozitivnu korelaciju između parova izvorni signal – odgovarajući rezultat ICA-e*. Vrijednosti fundamentalne frekvencije za svaki pojedini par nisu dovedene u korelaciju koristeći *coeffcor* zbog isprekidanosti fundamentalne frekvencije. Problem je riješen tako da su nedefinirane vrijednosti F_0 aproksimirane s prosječnom vrijednošću te potom uspoređivane *aproksimirane krivulje fundamentalne frekvencije (aprosk. F_0)* za svaki par. Rezultati srednje vrijednosti i standardne devijacije *aproksimiranih krivulja fundamentalnih frekvencija* svih parova u sva tri slučaja ukazuju na sličnost. Dobiveni rezultati pokazuju da metoda nezavisnih komponenata funkcionira na pokusnim snimkama gdje je broj govornika jednak broju mikrofona uz pretpostavku da su izlazne vrijednosti mikrofona linearne mješavine nezavisnih komponenti. U slučajevima da je broj govornika i mikrofona različit, da postoji pozadinska buka ili da je miješanje signala nelinearno potrebno je posegnuti za alternativnim metode ICA-e, a ne koristi se generativni model kao u ovom radu. U potrazi za rješenjem *'cocktail party problema'* znanstvenici su u potrazi preko pola stoljeća. Otkrivanje mehanizma selektivne pozornosti u mozgu je ključ za rješavanje ovog problema. Što se primjene metode tiče ovdje je razmatrana samo jedna situacija, ekstrakcija govornika, ali ova metoda je univerzalna i primjenjiva na čitav spektar ljudskih djelatnosti, od medicine, ekonomije, vojska itd.

5 LITERATURA

- [1] Boddy, R., & Smith, G. (2009). *Statistical methods in practice: for scientists and technologists*. John Wiley & Sons.
- [2] Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica*, 86(1), 117-128.
- [3] Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5), 975-979.
- [4] Heđever Mladen, *Osnove fiziološke i govorne akustike (interna skripta)*(2010), Sveučilište u Zagrebu edukacijsko-rehabilitacijski fakultet
- [5] Heđever Mladen, *Govorna akustika (interna skripta)* (2010), Sveučilište u Zagrebu edukacijsko-rehabilitacijski fakultet
- [6] Hyvärinen, Aapo, and Erkki Oja. "Independent component analysis: algorithms and applications." *Neural networks* 13.4-5 (2000): 411-430.
- [7] Hyvärinen, Aapo, Juha Karhunen, and Erkki Oja. What is Independent Component Analysis: Restrictions in ICA,"Independent Component Analysis. John Willey & Sons." Inc. New York(2001), pages 152-154
- [8] Hyvärinen, Aapo, Juha Karhunen, and Erkki Oja. *Other Applications. Financial application*, "Independent Component Analysis. John Willey & Sons." Inc. New York(2001), pages 441-443
- [9] Ristaniemi, T. and Joutsensalo, J. (1999). On the performance of blind source DMA downlink. In Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99), pages 437–441, Aussois, France.
- [10] Stone, James V. "Independent component analysis: an introduction." *Trends in cognitive sciences* 6.2 (2002): 59-64.

- [11] Stone, James V. Independent component analysis. This article was originally published online in 2005 in Encyclopedia of Statistics in Behavioral Science, John Wiley & Sons, Ltd and republished in Wiley StatsRef: Statistics Reference Online, 2014.
- [12] Vigário, R., Jousmäki, V., Hämäläinen, M., Hari, R., and Oja, E. (1998): Independent component analysis for identification of artifacts in magnetoencephalographic recordings, Advances in Neural Information Processing Systems, volume 10

WEB:

- [13] Anonymus. *Coeffcor*. Pristupljeno 2.9.2018.g. Dostupno na: <https://www.mathworks.com/help/matlab/ref/corrcoef.html>.
- [14] Anonymus. (Ažurirano 2018) *Audacity*. Pristupljeno 10.8.2018.g. Dostupno na: <https://www.audacityteam.org/>
- [15] Boersma, P. i Weenink, D. *Praat*. Pristupljeno 12.8.2018.g. Dostupno na: <http://www.fon.hum.uva.nl/praat/>
- [16] Aapo Hyvärinen. (Ažurirano 9.2.2011) *FastICA*. Pristupljeno 10.8.2018.g. Dostupno na: <https://research.ics.aalto.fi/ica/software.shtml/>