# Poredjenje razumljivosti govora na srpskom jeziku pri koriscenju COVID19 maski i filtriranju ekvivalentnim filtrima

Aleksa Nikolic, Uros Stojanovic, Vukan Arsenijevic

## Apstrakt

Epidemija bolesti COVID 19 je pogodila čitav svet početkom 2020. godine. Time je uzdrman svaki aspekt svakodnevnog života, između ostalog, nastava u školama i na fakultetima održavana je online. Sada, u 2022. nastava se vraća u normalan režim rada, a COVID 19 maske su postale standard. U ovom radu bavićemo se uticajem COVID 19 maski na razumljivost govora na srpskom jeziku, kao i uticajem filtara, modelovanih po COVID 19 maskama, koji bi trebalo da pojačaju razumljivost tako što negiraju uticaj maske. U eksperimentu su korišćene 3 vrste COVID 19 maski i plastični vizir. Cilj istraživanja bio je da odredimo da li je razumljivost bolja sa maskama ili sa filtrima koji negiraju njihov uticaj.

Obavljeno je 12 (broj testova) subjektivnih testova u kontrolisanim uslovima. Kao rezultat dobili smo da je razumljivost snimaka sa maskom bolja od razumljivosti filtriranih snimaka osim kod vizira gde je bolja razumljivost filtriranog snimka.

## I Uvod

Zaštitne maske su, za vreme epidemije COVID 19, postale esencijalan deo naše svakidašnjice. Uprkos njihovom svojstvu da štite od prenošenja virusa, maske mogu imati i negativan uticaj na iskustvo slušaoca u svakodnevnom govoru, predavanjima u školi ili na fakultetu, solo pevanju, horskom pevanju i mnogim drugim instancama. Naime, razumljivost govora je važan deo ljudskih interakcija na koji utiče nošenje zaštitnih maski. U ovom radu bavićemo se razumljivošću govora kroz različite vrste maski i pokušaćemo da projektujemo filtre koji bi negirali uticaj datih maski i vratili govor u stanje bez maske. Zatim ćemo uporediti snimke govora kroz masku i iste snimke propuštene kroz filtre koje smo dobili.

## II Metodologija

Prvenstveno je odredjen oblik spektra dugovremenog govora pri korišćenju hirurške, N95 i pamučne maske, kao i zaštitnog vizira. Za odredjivanje oblika spektra koristili smo govorne signale deset različitih govornika. Govornici su varirali po polu, naglasku i dubini glasa. Govor je sniman u kontrolisanim uslovima u gluvoj sobi Elektrotehničkog fakulteta. Svaki od govornika je snimljen u pet različitih slučajeva kako čita isti tekst; sa jednom od tri maske ili vizirom i bez ikakve dodatne opreme.

Nakon određivanja parametara filtara i obrade signala test grupa je na subjektivnim testovima ocenjivala razumljivost govora svakog od navedenih slučajeva. Subjektivni testovi su izvođeni u kontrolisanim uslovima(na slušalicama, bez ambijentalne buke). Snimljene recenice su isečak iz književnog dela “Koreni” Dobrice Ćosića i odabrane su kao simulacija svakodnevnog govora. Ispitanici u test grupi dobili su upitnik u koji su upisivali svoja mišljenja o snimcima koji su im puštani.

Rezultati ovog testa se nalaze u poglavlju pet(V).

## III Postavka eksperimenata

Pri izvođenju eksperimenata korišćene su tri vrste zaštitnih maski: hirurška maska, N95 maska, pamučna maska i zaštitni vizir. Na slici 1. se nalaze maske i vizir korišćeni u eksperimentu. Pri snimanja govora sa vizirom korišćena je i hirurška maska ispod njega.



1. b) c) d)

*Slika 1. Zaštitne maske i vizir*

*a)Hirurška maska b) N95 maska c) Pamučna maska d) Vizir*

Snimanja su odrađena u gluvoj sobi. Govornici su se nalazili 20 centimetara od mikrofona i čitali dati tekst. Svaki od deset govornika (pet muškaraca i pet žena) je čitao tekst sa svakom od maski, vizirom i bez ikakve dodatne opreme ponaosob. Koristeći snimljene signale odredili smo dugovremeni spektar signala za svaku masku i prema njemu modelovali filtre. Nakon obrade signala snimljenih bez maski (normalni govor) dobijenim filtrima dobijeni signali su puštani test grupama u kontrolisanim uslovima.

Test grupa od 12 ispitanika slušala je po dva snimka, snimak sa maskom i filtriran snimak, za svaku od pomenutih maski. Svoja zapažanja upisivali su u upitnik gde su mogli da štikliraju snimak koji su bolje razumeli i da daju svoj komentar. Ispitivanje se vršilo u zgradi Računskog Centra u odvojenoj prostoriji. Za puštanje snimaka korišćen je audio interfejs Audient iD4 i slušalice HyperX Cloud 2. Slušalice imaju raspon od 15 Hz do 20 kHz čime ispunjavaju zahtev ovog istraživanja.



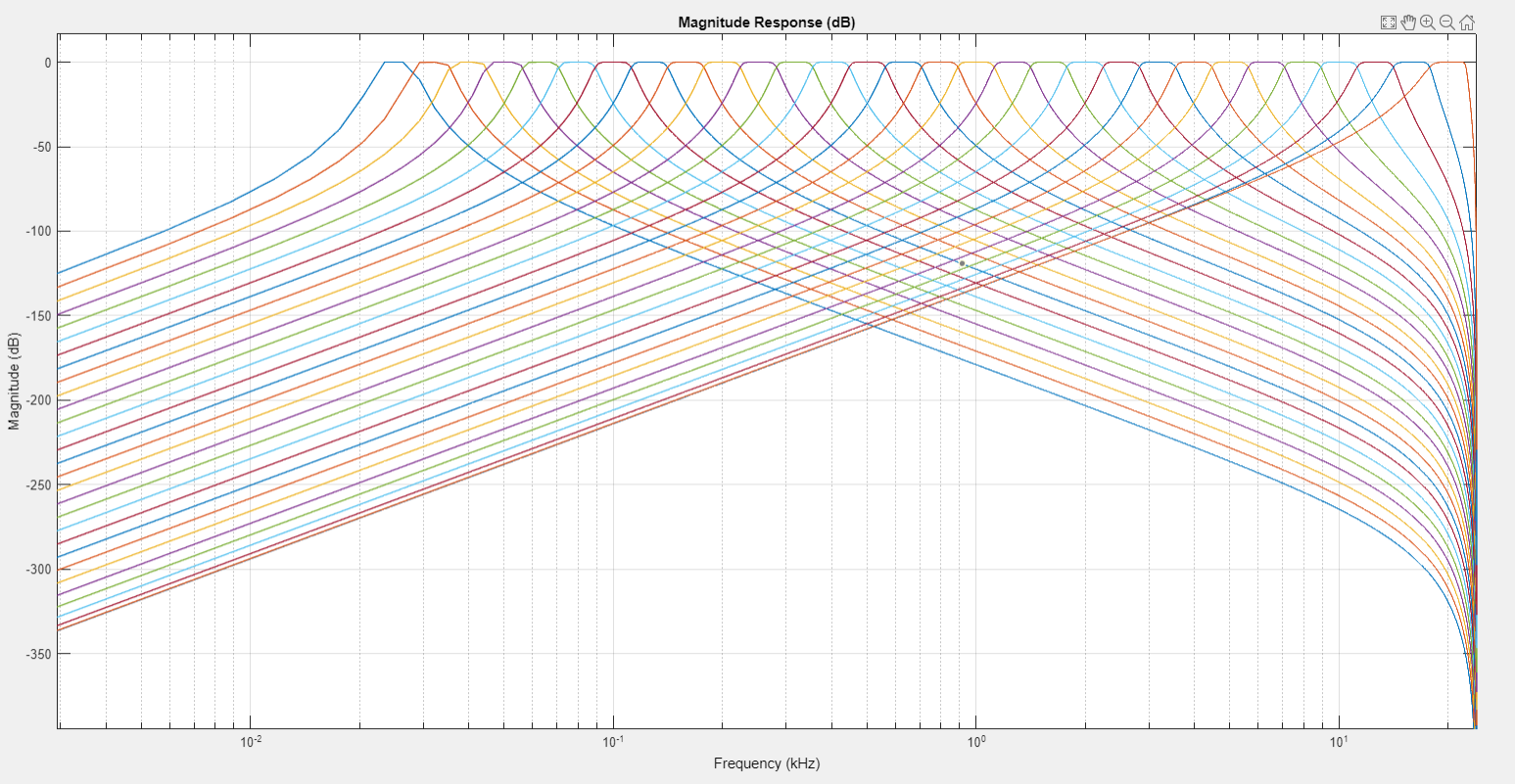
Slika 3. Korišćeni interfejs

Slika 2. Korišćene slušalice

## IV Kodovi i slike iz Matlaba

Za kreiranje potrebnih zvučnih signala koristili smo MATLAB. Ideja našeg koda bila je da projektujemo filtarsku banku sa kojom ćemo množiti bilo koji zvučni signal kako bi dobili njegovu karakteristiku zavisnosti nivoa od frekvencije.

Prvo smo dizajnirali banku filtara koristeći **fdesign.octave** za kreiranje oktavnih filtara. Utvrdili smo da je najefikasnije da postavimo red filtra 8 za broj intervala po oktavi 3. Definisali smo i centralnu frekvenciju od 1000 Hz. Frekvencija odabiranja bila je 48 kHz što odgovara našim snimcima. Funkcijom **validfrequencies** nalazimo sve frekvencije na kojima ćemo crtati. Prolazimo kroz ciklus onoliko puta koliko tih frekvencija imamo i kreiramo filtre funkcijom **design**. Koristeći funkciju **fvtool** možemo prikazati dobijenu banku filtara.

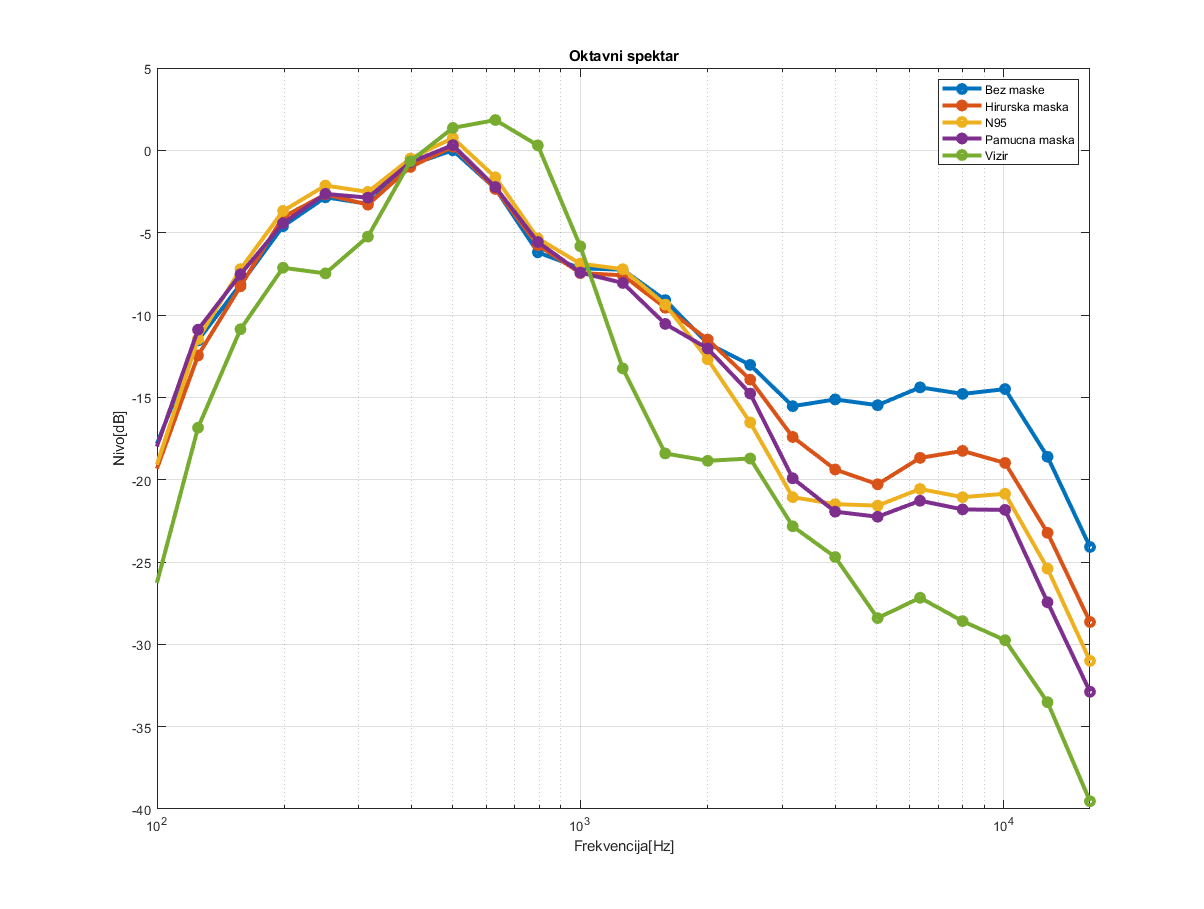


Slika 4. Prikaz oktavnih filtara

Na datoj slici (*Slika 4)* možemo videti banku oktavnih filtara gde su segmenti jednaki po dužini na logaritamskoj skali. Na linearnoj skali svaki sledeći segment bio bi duplo duži od prethodnog.

Kada govorne signale pomnožimo dobijenom bankom filtara, a zatim i izvršimo **RMS** za svaki segment dobijamo oktavni spektar za svaki signal (*Slika 5).*

Na oktavnom spektru (*Slika 5)* možemo primetiti da su karakteristike svih maski, osim vizira, gotovo jednake za frekvencije manje od 1000 Hz. Tek negde nakon 2000 Hz počinju da se razlikuju i možemo primetiti drastičnu razliku nivoa posle 3000 Hz. Na segmentu od 3000 Hz do 10 kHz vidimo da vrednosti prestaju tako brzo da opadaju ali jasno je da je signal bez maske najvišeg nivoa dok su ispod njega redom hirurška, N95 i pamučna maska. Nakon 10 kHz ponovo vidimo drastičan pad nivoa.

Oktavni spektar vizira razlikuje se od ostalih maski. Nivo je niži od ostalih za frekvencije manje od 400 Hz ali u intervalu od 400 Hz do 1 kHz uspeva da dostigne znatno viši nivo od ostalih maski. Ovo možemo opravdati pojačanom rezonancom pri korišćenju vizira između govornika i plastike. Ovim se stvara efekat pojačanja sličan onom koji koristie akustične gitare, na primer.Nakon ovog intervala, spektar vizira naglo opada skroz do 16 kHz. U ovom segmentu vizir ima najniži nivo od merenih maski.

Slika 5. Oktavni spekatar

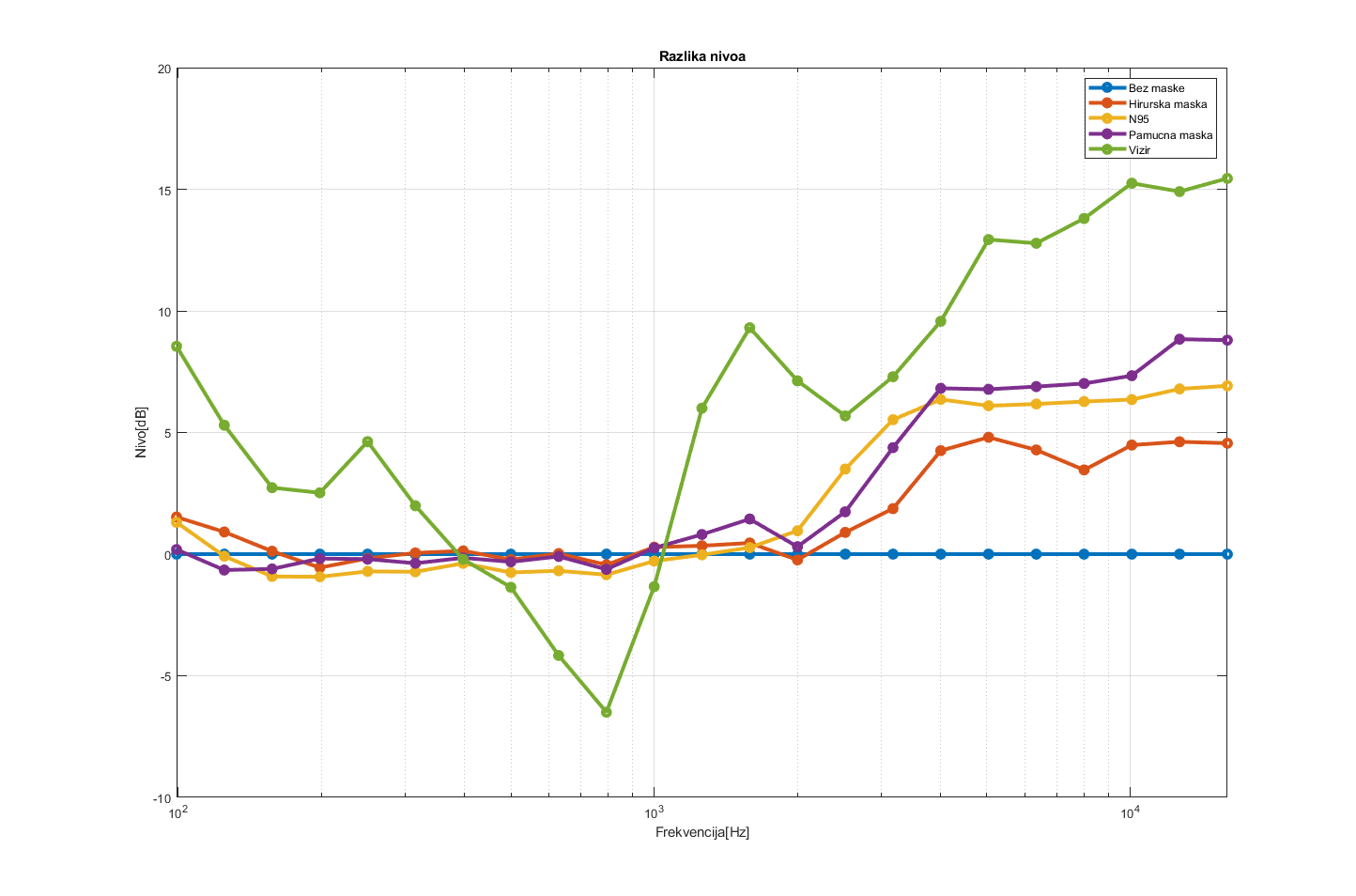
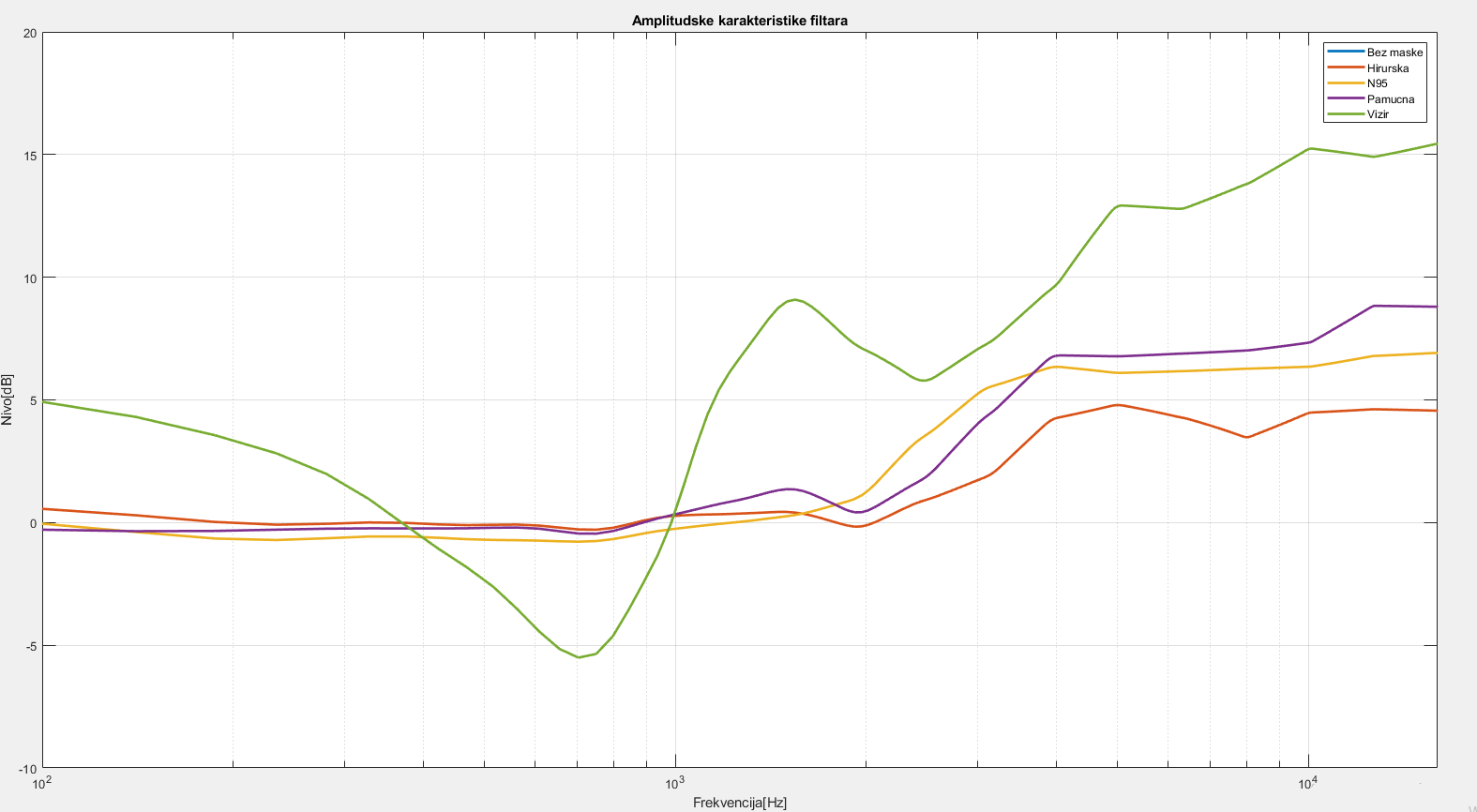
**Razliku nivoa** (*Slika 6)* ćemo prikazati tako što ćemo od svakog odbirka u oktavnom spektru oduzeti odgovarajuće vrednosti odbiraka signala bez maske prikazanog plavom bojom.

Zatim ćemo dobijene vrednosti pomnožiti koeficijentom -1. Time ćemo dobiti niz vrednosti za svaki signal koji će nam pomoći u projektovanju kasnijih filtara u ovom radu.

Dobijene vrednosti nivoa za signal bez maske biće 0.

Pomoću razlike nivoa možemo od signala sa maskom dobiti čist govorni signal bez maske. Ovo postižemo korišćenjem **FIR2** (Finite Impulse Response) filtara.

**FIR2** funkcija fir2(n,f,m) vraća FIR filtar n-tog reda sa frekvencijskom i magnitudnom karakteristikom datom parametrima f i m. Ova funkcija vrši linearnu interpolaciju traženih frekvencija, a zatim nalazi koeficijente filtra koristeći Furijeovu transformaciju i Hammingov prozor.

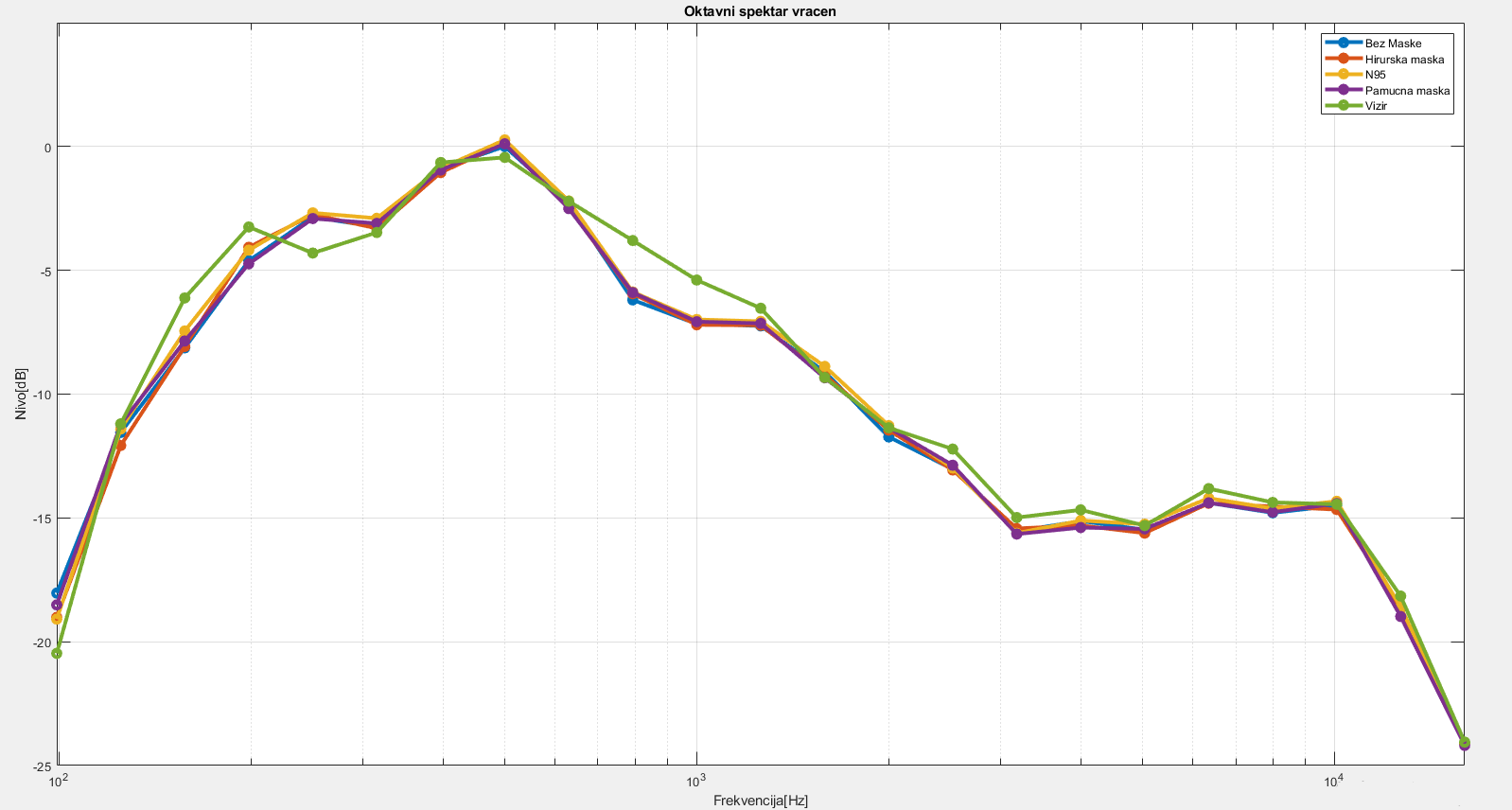


Slika 7. Amplitudska karakteristika filtra

Slika 6. Razlika nivoa

Nakon ovoga, možemo pogledati oktavni spektar dobijenih, filtriranih signala. Ovaj spektar (*Slika 8)* dobijamo na isti način kao i prvi oktavni spektar (*Slika 5)*. Ovde možemo primetiti da se vrednosti svakog odbirka skoro idealno poklapaju sa vrednostima odbiraka signala bez maske, osim odbiraka vizira koji su malo lošiji.

Pre prvog ispitivanja ljudi, imali smo grešku u kodu. To je doprinelo lošem rezultatu vraćenog oktavnog spektra za sve maske osim hirurške. Kasnije smo tu grešku ispravili i dobili rezultat koji smo tražili (*Slika 8).*

**Da bi dobili približnije odbirke signala vizira moramo povećati red filtra, što bi bilo previše obzirom da je trenutni red N=500.

Slika 8. Vraćen oktavni spektar

## V Rezultati i Diskusija

Rezultati **prvog merenja** (pre ispravke greške) ispitivanja za četiri tipa maski dati su u tabeli i dijagramu. Međutim, pokazalo se da je mogućnost komentarisanja dosta pomogla u razumevanju rezultata. Naime, za prve tri maske (**hirurška, N95, pamučna**), 10 od 12 ispitanika dalo je komentar da je snimak sa maskom prijatniji za slušanje, nevezano za njihov izbor snimka po razumljivosti. Objašnjenje za ovo jeste da filtrirani snimci imaju više šuma, kao i dosta izraženije visoke frekvencije (više od 1Khz). To doprinosi neprijatnom zvuku šumnih bezvučnih suglasnika S,Š,C,Č i P. (Ovo je produkt greške u kodu).

Ovo je donekle i očekivano obzirom da je ispitivanje subjektivno i ne koristi logatome već lični utisak, a ispitanici verovatno imaju različito razumevanje pojma “razumljivost”.

Rezultati za **vizir** dosta se razlikuju od rezultata za prve 3 zaštitne maske. Naime, kod vizira možemo videti da je čak 10 od 12 ispitanika izabralo filtriran snimak. Uprkos komentarima da na filtriranom

snimku postoji više šuma, ispitanici su većinom taj snimak birali kao razumljiviji. Ovo se može objasniti karakteristikom oktavnog spektra vizira (*Slika 5*) koja pokazuje da vizir ima niži nivo od ostalih maski za frekvencije između 100 Hz i 500 Hz. U ovim frekvencijama ljudski glas je najprisutniji i ljudsko uho ga najbolje registruje.

Grafik 1. Razumljivost govora različitih maski

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Tip snimka | |
|  | | Sa maskom | Filtriran |
| Tip Maske | Hirurška | 7 | 5 |
| N95 | 8 | 4 |
| Pamučna | 8 | 4 |
| Vizir | 2 | 10 |

Tabela 1. Broj ispitanika koji su izabrali određeni snimak pre ispravki

Nakon ispravke greške u kodu, izvršili smo testiranje još 5 ispitanika. Rezultati **drugog merenja** su sledeći:

Grafik 2. Razumljivost govora različitih maski nakon ispravke greške

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Tip snimka | |
|  | | Sa maskom | Filtriran |
| Tip Maske | Hirurška | 3 | 2 |
| N95 | 1 | 4 |
| Pamučna | 2 | 3 |
| Vizir | 0 | 5 |

Tabela 2. Broj ispitanika koji su izabrali određeni snimak nakon ispravki

Ovi rezultati su izvršeni u kratkom vremenskom roku sa dostupnom opremom u kući i članovima porodica.

Komentari za **pamučnu** masku su da su signali jako slični, ali filtrirani signal zvuči ravnije i prirodnije. Što se tiče **N95**, filtrirani i nefiltrirani snimci su vrlo razgovetni i sličnog kvaliteta. Ali na našem malom uzorku dominira filtrirani snimak. Hirurška maska praktično nije imala razlike između signala osim što se u filtriranom uzorku na momente čuje šum. Ispitanici su naglasili da kod ove maske ima najmanje razlike između dva snimka. Rezultati za **vizir** su praktično ostali isti kao i u prvom ispitivanju iako je filtar ovaj put realizovan mnogo bolje. Ovaj put, doduše, ispitanici su naglasili veliku razliku u korist filtriranog signala.

## VI Zakljucak

Rezultati prvog i drugog ispitivanja imaju svoje razlike. Primetićemo da je rezultat za hiruršku masku ostao praktično isti obzirom da je inicijalno samo ona pravilno projektovana.

Rezultati za N95 i pamučnu masku su malo bolji obzirom da su snimci sa tim maskama svakako prilično razgovetni sami po sebi.

Najveća razlika u razumljivosti ostaje kod vizira. Pre ispravke greške signal sa vizirom imao je mali broj glasova, da bi nakon ispravke taj broj glasova spao na 0. Komentari sugerišu da je filtrirani snimak ubedljivo razumljiviji nego snimak sa maskom.

Možemo izvesti zaključak da filtrirani snimci zvuče prirodnije i blago razumljivije po našim ispitanicima iako na određenim mestima unose minimalan šum ili imaju izraženije visoke tonove.