Izdvajanje signala govornika potiskivanjem pozadinskog šuma pomoću metoda temeljenih na strojnom učenju

Mamić, Marko

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I INFORMACIJSKIH TEHNOLOGIJA OSIJEK

Stručni studij

Izdvajanje signala govornika potiskivanjem pozadinskog šuma pomoću metoda temeljnih na strojnom učenju

Završni rad

Marko Mamić

Osijek, 2022.

SADRŽAJ

1. UVOD

Tijekom pandemije koja je utjecala na način obavljanja posla na radnome mjestu i održavanja nastave u edukacijskim ustanovama, popularnost platformi za udaljenu videokomunikaciju je porasla. Kako zaposlenici, predavači i učenici često nisu bili u zasebnom okruženju pri obavljanju obaveza, korisna su bila programska i sklopovska rješenja za potiskivanje pozadinskih šumova ili smetnji kako bi što kvalitetnije mogli obavljati posao bez nepotrebnih pozadinskih distrakcija. Uz videokomunikaciju, postoje i druga područja koja imaju korist od potiskivanja šuma ili vlastite načine obrade šumova, poput audio i video editora, *noise cancelling* slušalica te samih programskih rješenja za obradu zvučnih zapisa.

Razvoj novih tehnologija i sklopovlja omogućio je razvoj novih metoda za potiskivanje šuma, primjerice metoda temeljenih na strojnom učenju. Bržim sklopovljem omogućeno je potiskivanje šuma u stvarnom vremenu u mnogim aplikacijama.

Za određivanje i potom odvajanje signala govornika zvučnog zapisa od pozadinskih šumova potrebno je odraditi analizu samog zapisa te odrediti frekvencijski i amplitudni raspon uz ostale značajke, ovisno o metodi, kako bi se uspješno odradila transformacija zapisa.

Cilj ovog rada je istražiti koje sve metode za obradu digitalnog signala govora, u svrhu potiskivanja šumova, postoje te koja je razlika između tradicionalnih metoda i novijih metoda temeljenih na strojnom učenju. U radu su testirani tradicionalna metoda (*spectral gating)* i model metode strojnog učenja (PoCoNet arhitektura) nad zvučnim zapisima govornika na engleskom jeziku uz pozadinski šum te vlastitom zvučnom zapisu govornika na hrvatskom jeziku uz pozadinski šum. Odrađena je objektivna i subjektivna evaluacija pročišćenih zvučnih zapisa za obje metode.

1.1 Zadatak završnog rada

Potrebno je implementirati različite metode za potiskivanje pozadinskih šumova, od tradicionalnih do metoda koje se temelje na metodama strojnog učenja. Potrebno je provesti evaluacija pojedinih rješenja te odrediti subjektivnu i objektivnu ocjenu za pojedine metode. Istražiti će se mogućnost implementacije na mobilnom uređaju putem aplikacije.

2. DEFINIRANJE PROBLEMATIKE I PREGLED POSTOJEĆIH RJEŠENJA

Odnos signal/šum je objektivna mjera kvalitete signala. Pozitivna mjera govori da je dominantni dio signala jači nego li šum, dok negativna mjera označuje da je šum jači nego li dominantni dio signala. Osnovna formula dana je relacijom (2-1)

$$SNR_{dB} = 10\,log_{10}\left(\frac{P_{signal}}{P_{šum}}\right)\,[dB] \tag{2-1}$$

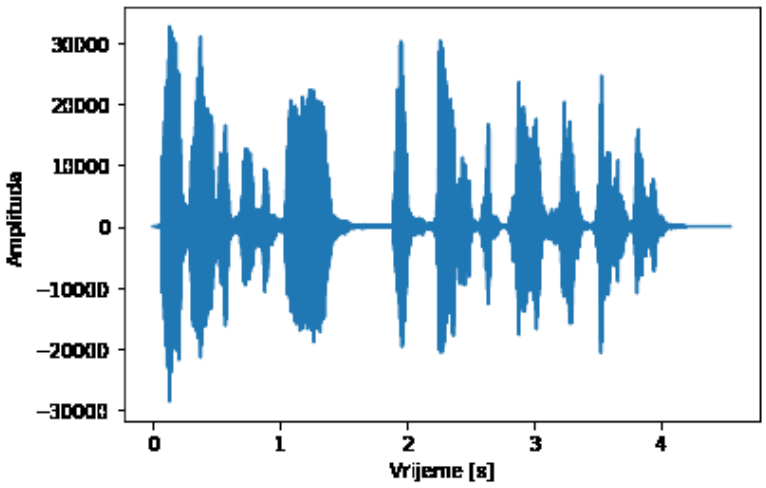
gdje P_{signal} označava srednju snagu signala, a $P_{šum}$ označava srednju snagu šuma te rezultat daje mjeru odnosa signal/šum po logaritamskoj raspodijeli u dB za zvučne zapise.

Dominantni dio signala smatra se dio signala zvučnog zapisa koji se odnosi na zvuk koji treba izolirati te signal kojim se smatra da je u „prvom planu“. U ovome radu dominantni dio signala smatra se signal govornika.

Šumom signala smatra se dio signala zvučnog zapisa koji se odnosi na zvuk koji treba potisnuti te signal kojim se smatra da je u pozadini. U ovome radu šumom signala smatra se bilo koji zvuk koji ne odgovara zvuku govornika.

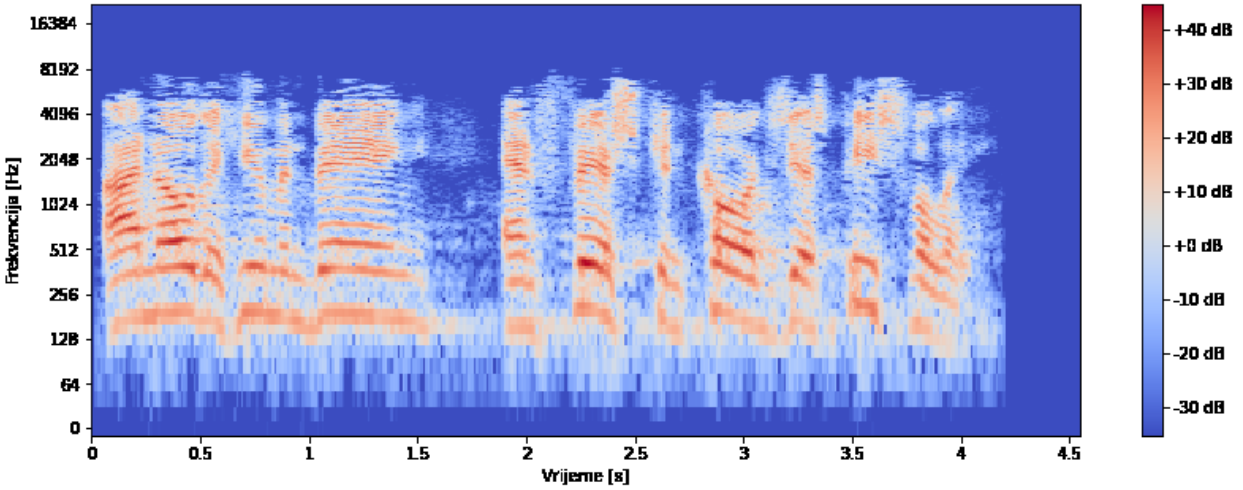
Spektrogram je vizualni prikaz frekvencija zvučnog signala koje se mijenjaju s vremenom te boja skale označuje glasnoću pojedinog dijela signala u dB.

Slika 2.1. prikazuje zvučni zapis u vremenskoj domeni gdje je y-os amplituda a x-os vrijeme u sekundama.



Slika 2.1. Prikaz zvučnog zapisa u vremenskoj domeni

Slika 2.2. prikazuje zvučni zapis u frekvencijskoj domeni gdje je y-os frekvencija u Hz, x-os vrijeme u sekundama te skala boje označuje glasnoću pojedinog dijela signala u dB.



Slika 2.2. prikaz zvučnog zapisa u frekvencijskoj domeni

2.1 Načini potiskivanja šuma iz okoline zvučnih signala

Postoje 3 opća načina na koji se može potisnuti šum iz okoline:

1. Aktivno
2. Pasivno
3. Adaptivno

Sva tri pristupa objašnjavaju kompletna rješenja za implementaciju u slušalice, no fokus ovog rada biti će algoritamski tj. programski dio aktivnog i adaptivnog pristupa.

2.1.1 Aktivno potiskivanje šuma iz okoline

Ovaj način radi na principu da se iz okoline snime uzorci šuma, analiziraju se te transformiraju tako da se dodavanjem transformiranog šuma i originalnog šuma dogodi destruktivna interferencija. Ovaj koncept je nastao 1930-ih, daljnji razvoj se nastavio 1950-ih te su nastale slušalice za pilote koje blokiraju niskofrekventne šumove (zvukovi motora). Moderno aktivno potiskivanje šuma koristi analogne sklopove ili digitalno procesiranje signala. Adaptivni algoritmi analiziraju zvučni zapis pozadinskog šuma, zatim (ovisno o specifičnom algoritmu) generiraju signal koji je suprotne faze ili polarnosti u odnosu na originalni signal. Takav suprotni signal pozadinskog šuma amplitudno je pojačan kako bi odgovarao amplitudi početnog signala te zbrajanjem takva dva signala kreira se destruktivna interferencija u svrhu potiskivanja šuma iz originalnog signala. Ovaj pristup potiskivanju šumova najčešće se koristi pri niskofrekventnim šumovima dok za visokofrekventne šumove korisnost implementacije opada porastom frekvencije [1].

2.1.2 Pasivno potiskivanje šuma iz okoline

Za razliku od aktivnog načina, pasivni ne zahtijeva dodatni sklop ili digitalno procesiranje signala. Zbog toga što nema aktivnog sklopa koji bi transformirao ulazni zvučni signal, pasivni način se oslanja na fizičke barijere, slojeve ili sam oblik uređaja (npr. oblik slušalice). Ovaj način je efikasan kod sprečavanja visokofrekventnih šumova te je lako skalabilan. To ga čini pogodnim za jeftiniju, a kvalitetnu proizvodnju opreme za zaštitu od buke iz okoline, primjerice oprema za rad na gradilištu [1].

2.1.3 Adaptivno potiskivanje šuma iz okoline

Ovaj način je vrlo sličan načinu rada aktivnog potiskivanja, no razlika je u tome što adaptivni način koristi napredne algoritme kako bi točno prepoznao šum te može procijeniti pogodnost uhu uz adaptaciju za „curenje“ (eng. *leakage)* zvuka. Uz to, adaptivni način također prepoznaje što korisnik sluša kako bi mogao odabrati pogodan algoritam za tu specifičnu svrhu, primjerice razlika između govora i glazbe [1].

2.2 Pregled postojećih rješenja za programski aktivni pristup otklanjanja šuma zvučnih zapisa

U sljedećim točkama, predstaviti će se neka od poznatijih softverskih rješenja, od kojih neki nude i aplikacijsko programsko sučelje (eng. *Application Programming Interface* - API) za razvojne programere kako bi ta rješenja lakše implementirali u vlastite aplikacije.

2.2.1 KRISP

Uz standardno potiskivanje šuma, KRISP programsko rješenje nudi i druge značajke koje uključuju izolaciju glavnog govornika (ako ima više glasova u zvučnom zapisu) i uklanjanje jeke (eng. *echo cancellation*) što pomaže pri uklanjanju efekta „odzvanjanja“. S obzirom na to da je rješenje namijenjeno implementaciji u videokonferencijske platforme, nudi rješenje i za odabir virtualne pozadine. Uz te značajke, nude i statističke podatke o razgovorima, od vremena koje je provedeno pričajući do drugih metrika koji prikazuju koliko je koji govornik aktivno sudjelovao u razgovoru. Kao jednu od statistika nude i opciju prikaza statističkih podatka o trajanju buke koje je KRISP prepoznao te čak i trajanje pojedine razine glasnoće buke [2].

2.2.2 Audo.ai

Tvrtka Audo nudi cijelo studijsko rješenje Audo Studio koje obuhvaća cjelokupni program za digitalnu obradu zvuka uz pomoć metoda temeljenih na strojnom učenju. Također, nude i API koji omogućuje lakšu implementaciju u druga programska rješenja s naglaskom na jednostavnost implementacije. Također, nude Magic Mic – *open source* aplikaciju za Linux sustave koja uklanja pozadinski šum na razini cijelog računalnog sustava neovisno o korištenoj komunikacijskoj aplikaciji [3].

2.2.3 Audacity

Audacity je besplatna, *open source, cross-platform* aplikacija za snimanje i obradu zvučnih zapisa. Pogodna je za početnike zbog jednostavnosti opcija. Uz snimanje i obradu, nudi izvoz (eng. *export*) u različite formate, dodavanje efekata, analizu spektrograma i podršku za dodatne *plugin-ove*. Podržava odabir frekvencije uzorkovanja te odabir razina kodiranja zvučnih zapisa [4].

2.3 Pregled metoda za potiskivanje šumova zvučnih zapisa

Tradicionalne metode većinom koriste algoritme koji identificiraju frekvencije na kojima se nalazi većina snage pozadinskog šuma te ih potom oduzmu od originalnog signala. Većina takvih

pristupa koriste statične filtre kao što su niskopropusni i visokopropusni pojasni filtri, koji imaju specifične parametre kako bi izolirali onaj dio signala koji se smatra dominantnim. Ovi algoritmi najbolje rade s deterministčkim signalima tj. u situacijama gdje se točno zna koju vrstu šuma treba potisnuti te koji se dio signala smatra dominantnim, s obzirom na to da je lakše prepoznatljiv šum koji želimo izolirati. U stvarnosti, ovi filtri nisu efikasni u situacijama kada se šum poklapa sa signalom kojeg treba izolirati [5].

Za razliku od tradicionalnih metoda, metode temeljene na strojnom učenju [6] temelje se na velikim skupovima podataka s primjerima signala sa šumom i čistog signala. Kako bi efikasnost samog modela treniranog nad tim podacima bila što veća, sam set podataka treba imati uzorke signala specifičnih za pojedinu svrhu, primjerice ako je cilj napraviti model koji razdvaja pozadinsku buku prolazećih auta, set podataka treba sadržavati takve signale. Ako je cilj napraviti model koji bi trebao biti šire namjene, set podatka treba imati što veći izbor različitih pozadinskih šumova i primjera signala. Prikladan model za procesiranje zvučnih zapisa u svrhu potiskivanja pozadinskih šumova jesu povratne neuronske mreže (eng. *Recurrent Neural Network* - RNN) [6]. Takve mreže mogu razlikovati i „razumjeti“ sekvencijalni tip podataka, kao što su zvučni zapisi i tekst. Razlog zbog kojeg su takve mreže efikasne u procesiranju sekvencijskih podataka je mogućnost prepoznavanja uzorka kroz vrijeme.

Kod odabira metoda bitno je uzeti u obzir potrebne računalne resurse i vrijeme procesiranja signala, čemu se treba dodati posebna važnost pri implementaciji u programska rješenja koja rade u stvarnom vremenu (eng. *realtime*). Primjerice, videokonferencijske platforme nastoje umanjiti kašnjenje, te time zahtijevaju brzu obradu signala. Ako su dostupni resursi te nije toliko važno kašnjenje, metode temeljene na strojnom učenju omogućuju efikasnije potiskivanje pozadinskih šumova. Razlog tome je što se kod takvih metoda generira potpuno novi signal, dok se kod tradicionalnih metoda od originalnog signala oduzima invertirani procijenjeni signal šuma. Ako nema dovoljno dostupnih resursa, tada tradicionalne metode imaju prednost iako su manje efikasne u potiskivanju šuma. Međutim, tu prepreku u današnje vrijeme je sve lakše i lakše preći, kako zbog razvoja sklopovlja tako i zbog same optimizacije metoda strojnog učenja [5].

3. OPIS TRADICIONALNIH METODA I METODA TEMELJENIH NA STROJNOM UČENJU

Skup podataka je skup svih materijala koji se koriste pri učenju neuronske mreže. Razdvajaju se na skup za učenje, skup za testiranje i skup za validaciju. U ovome radu odnosi se na skup zvučnih zapisa signala govora sa pozadinskim šumom

Funkcija gubitka je metoda evaluacije koja ukazuje na uspješnost učenja modela te time i procjenjivanju rezultata skupa podataka. Cilj modela je za funkciju gubitka dobiti što manji rezultat.

3.1 Tradicionalne metode potiskivanja šuma u zvučnom zapisu

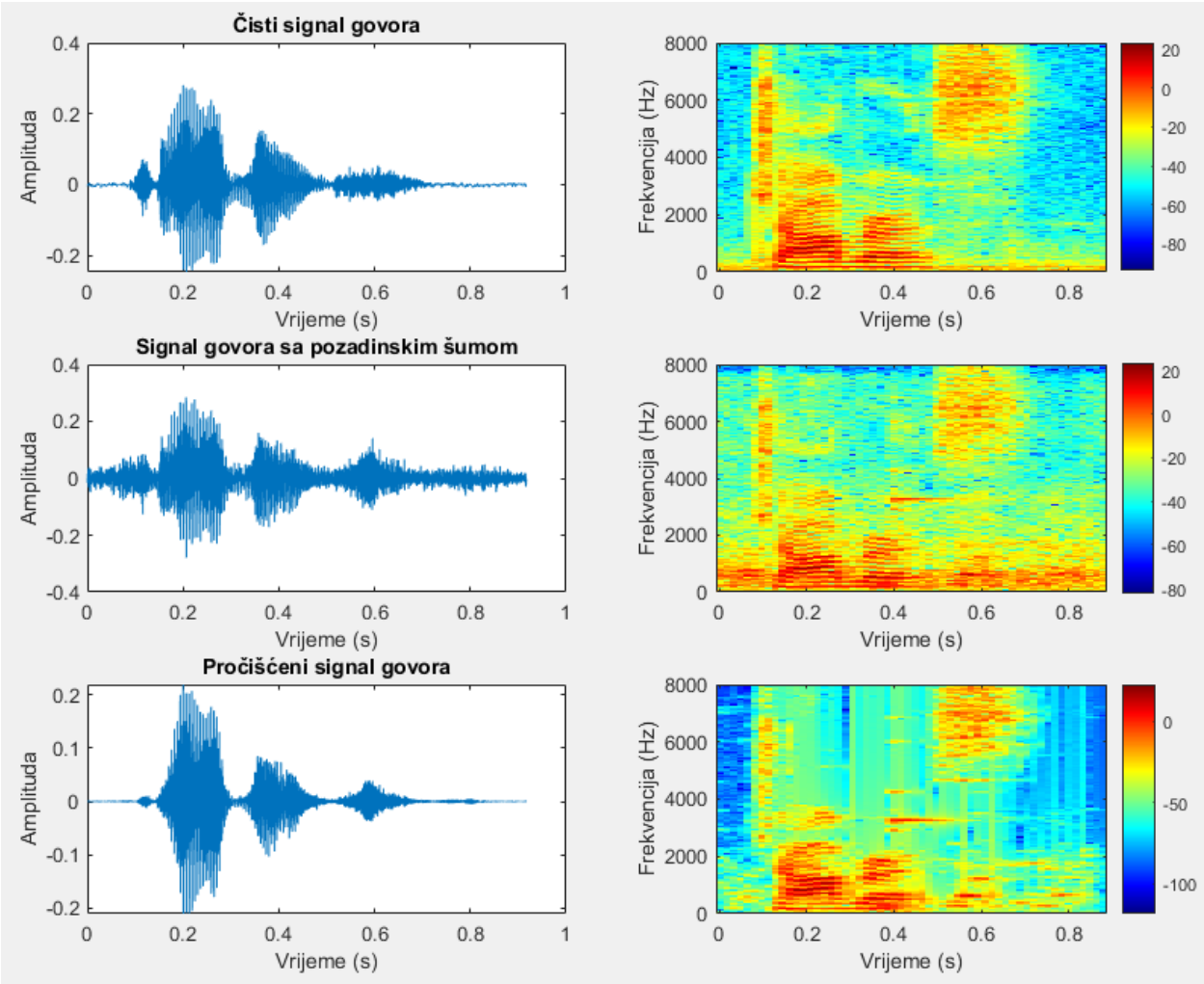
Tradicionalnim metodama smatraju se sve metode koje se temelje na prepoznavanju šuma pojedinih zvučnih zapisa te se potom invertirani signali šuma oduzima od početnog signala govora sa šumom kako bi se dobio procijenjeni čisti signal koji sadrži samo govor. U ovome poglavlju detaljnije su opisane dvije takve metode, a to su *Wiener* filtar te *spectral gating*.

3.1.1 Wiener filtar

Wiener filtar je industrijski standard za dinamičko procesiranje signala te se implementira u različite uređaje i primjene, poput slušnih aparata, mobitela i drugih raznih komunikacijskih uređaja. Vrsta je adaptivnog filtra te najbolje radi korištenjem dva signala – jedan koji sadrži signal i šum te drugi koji sadrži isključivo šum. Takvi signali mogu se snimiti postavljanjem dva mikrofona – jedan blizu izvora signala kojeg je potrebno izolirati te drugi na udaljenosti od izvora signala kako bi se odredili pozadinski šumovi. Za razliku od drugih metoda koje bi jednostavno oduzele ova dva signala, *Wiener* uzima u obzir činjenicu da se svaki mikrofon nalazi u sličnom, ali ne i istom okruženju te time je potrebno uzeti u obzir i druge faktore koji utječu na efikasnost filtra. *Wiener* filtar koristi značajke oba signala kako bi se generirala procjena čistog signala računajući srednju kvadratnu grešku te se potom greška pokušava minimizirati. Mane ovog pristupa su: potreba za dva različita signala kako bi se provela redukcija šuma te izobličenost signala u slučajevima gdje bi govor mogao sličiti pozadinskim šumovima – primjerice glas „s“ [5]. Primjena za situacije gdje ne postoji zaseban signal šuma se provodi putem rekurzivnog algoritma procjene s parametrom zaglađivanja između 0 i 1 [7].

Na slici 3.1. prikazan je zapis čistog govora u vremenskoj i frekvencijskoj domeni, zapis govora sa šumom u vremenskoj i frekvencijskoj domeni gdje se može primijetiti izobličenje signala u obje

domene, te pročišćeni signal govora u vremenskoj i frekvencijskoj domeni gdje se pokušava vratiti signal na početno stanje u kakvome se nalazi čisti zapis govora.



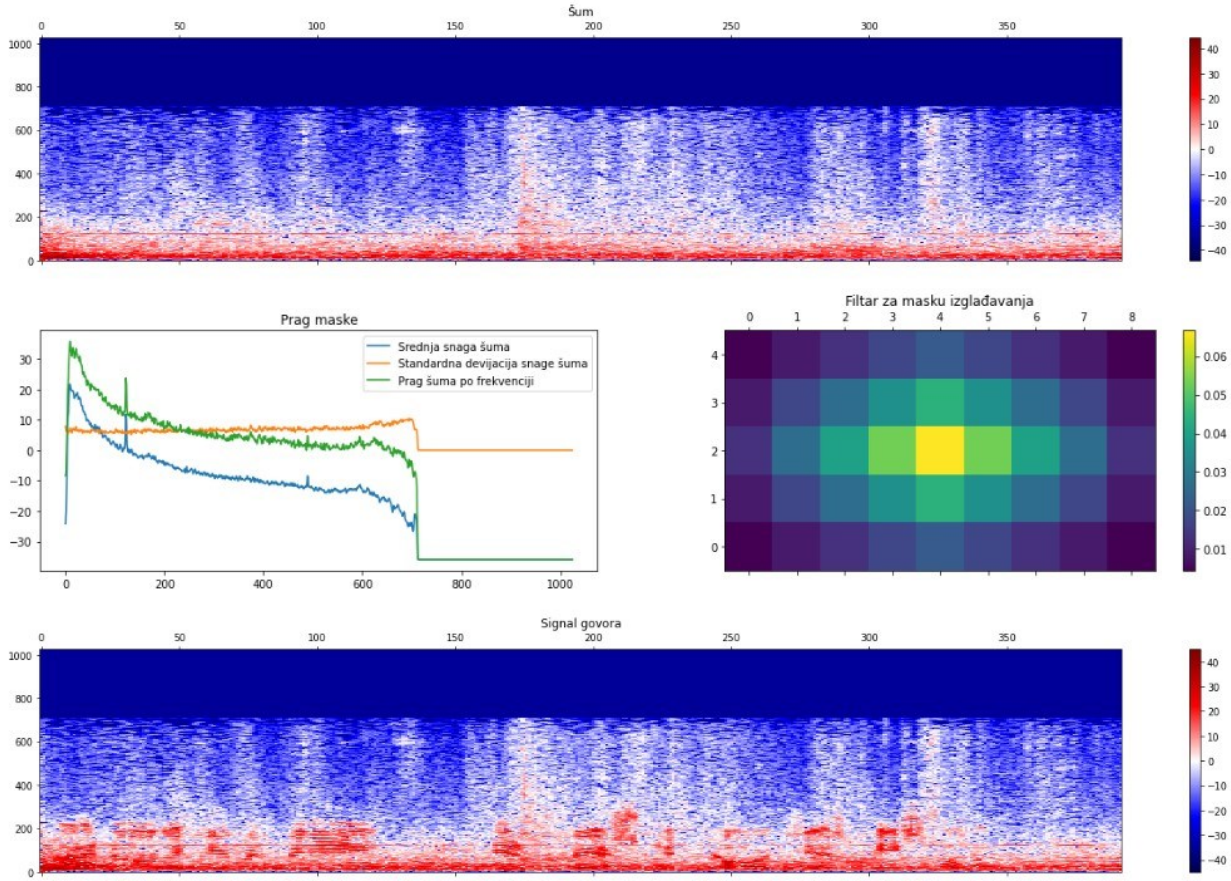
Slika 3.1. Vizualni prikaz rada *Wiener* filtra

3.1.2 Spectral gating

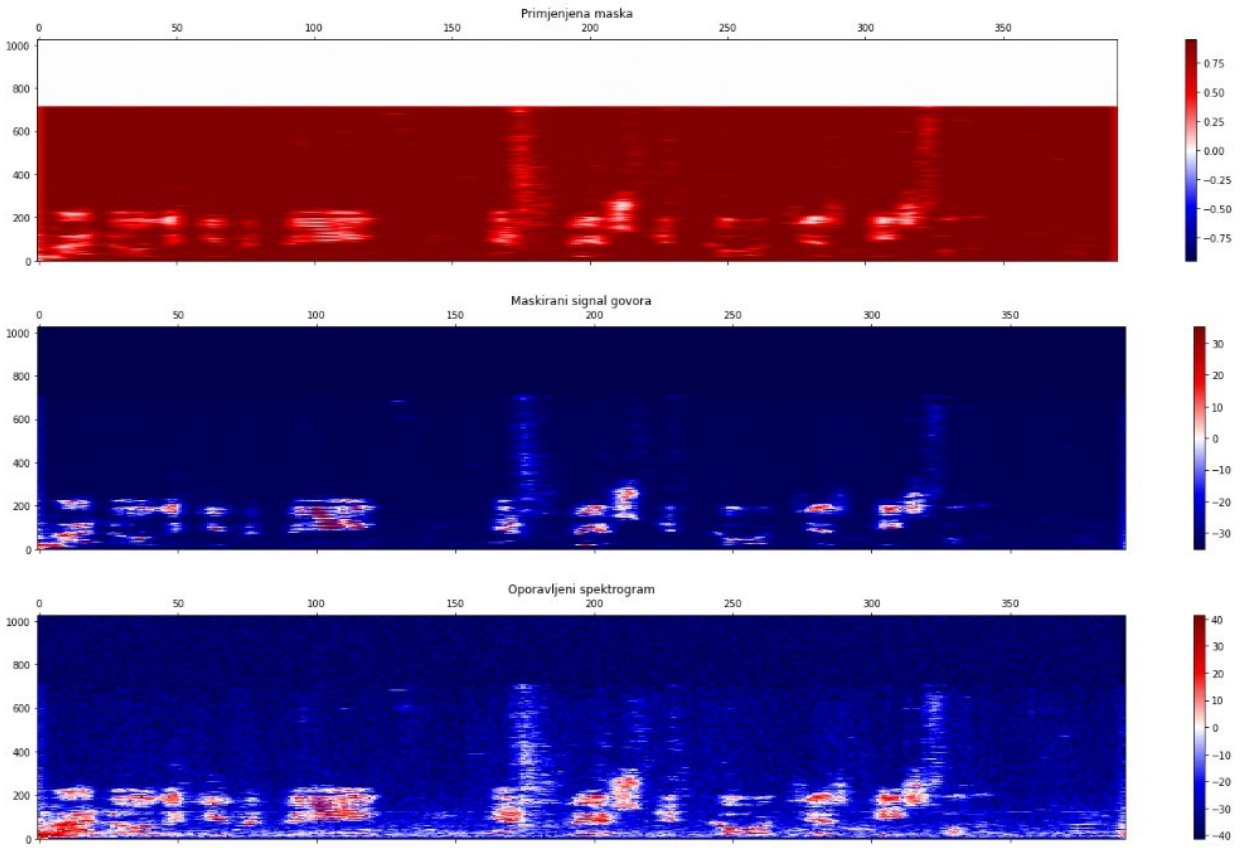
Spectral gating je verzija *Noise Gatea* koja radi na principu gdje se generira spektrogram signala (i proizvoljno spektrogram šuma) te se procjenjuje prag šuma (eng. *gate*) za pojedini frekvencijski pojas signala. Pomoću tih pragova računa se maska koja uklanja šum na temelju praga pojedinih frekvencijskih pojaseva. Najčešće se implementira jedna od dvije vrste algoritma – stacionarni (drži procijenjeni prag šuma istim kroz sve pojase) i ne-stacionarni (prag se računa zasebno za svaki pojas). Stacionarni pristup kreira spektrogram šuma zvučnog zapisa, računa se statistika u frekvencijskoj domeni, računa se prag s obzirom na statističke podatke šuma, kreira se spektrogram zvučnog zapisa, određuje se maska usporedbom spektrograma signala prema pragu koja se potom zagladi vremenskim i frekvencijskim filtrom. Na kraju se invertirana maska dodaje

na spektrogram signala kako bi se dobio pročišćeni signal. Ne-stacionarni pristup je proširenje stacionarnog pristupa gdje se računa spektrogram signala i vremenski zaglađena verzija spektrograma pomoću IIR filtra (eng. *Infinite Impulse Response*) koji se primjenjuje unaprijed i unazad na pojedinom frekvencijskom pojasu. Maska se generira nad vremenski zaglađenom spektrogramu te potom zagladi vremenskim i frekvencijskom filtrom. Takva maska se invertira te dodaje na izvorni signal [8].

Postupak uklanjanja šuma metodom *spectral gating*-a vizualno je prikazan slikama 3.2. i 3.3. Na slici 3.2. prikazan je spektrogram šuma, zatim računanje praga šuma, generira se maska, te prikazuje spektrogram signala. Na slici 3.3 može se vidjeti kako se dodaje maska na spektrogram signala, kako izgleda spektrogram signala nakon dodane maske, te spektrogram krajnjeg pročišćenog signala. Slike su generirane prema implementaciji [9].



Slika 3.2. Prikaz rada *spectral gatinga* – 1.dio



Slika 3.3. Prikaz rada *spectral gatinga* – 2.dio

3.2 Metode potiskivanja šuma u zvučnom zapisu zasnovane na strojnom učenju

Metode temeljene na strojnom učenju su značajno pripomogle sustavima za poboljšavanje govornih sadržaja. Takve mreže su većinom napravljene nadgledanim metodama učenja uz korištenje sintetičkih kombinacija čistog govora s poznatim zapisima šuma. Općenito, metode temeljene na strojnom učenju rade na princip da se od skupa podataka za učenje uzme određeni broj uzoraka koji se potom provedu kroz mrežu, vrši se evaluacija iteracije putem funkcije gubitka te se ovisno o rezultatu promijene težinske veze među neuronima. Time se kroz svaku iteraciju modela mijenjaju težinski parametri te se na kraju dobije gotovi model. Takav model se zatim evaluira pomoću različitih metrika ovisno o zadatku te se rade dodatne promjene u svrhu poboljšanja procesa učenja modela te se na osnovu tih metrika procjenjuje efikasnost modela u praksi. Za zadatak potiskivanja šuma prikladne su povratne neuronske mreže zbog mogućnosti prepoznavanja uzorka sekvencijskih tipa podatka kao što su zvučni zapisi [6].

U svrhu potiskivanja pozadinskog šuma zvučnih zapisa, većinom se koriste modeli za procjenu dodane veličine u pojasevima u vremensko-frekvencijskoj reprezentativnoj domeni signala sa šumom. Primjer takvog modela je *Stacked denoising Auto-encoder* koji koristi funkciju gubitka

zvanu *weighted reconstruction*, te procjenjuje spektar snage čistog govora [10]. Modeli koji su *phase-aware* (svjesni su utjecaja pomaka faze) koriste kompleksne matrice omjera poput duboke neuronske mreže temeljene na cIRM procjeni (eng. *Complex Ideal Ratio Mask and Estimation*) koja koristi funkciju gubitka temeljene na srednjoj kvadratnoj grešci za imaginarni i realni dio [11], dok drugi pristupi rade direktno nad valnim oblikom poput *Wave-U-Net* koji koristi funkciju gubitka *LeakyReLU* [12].

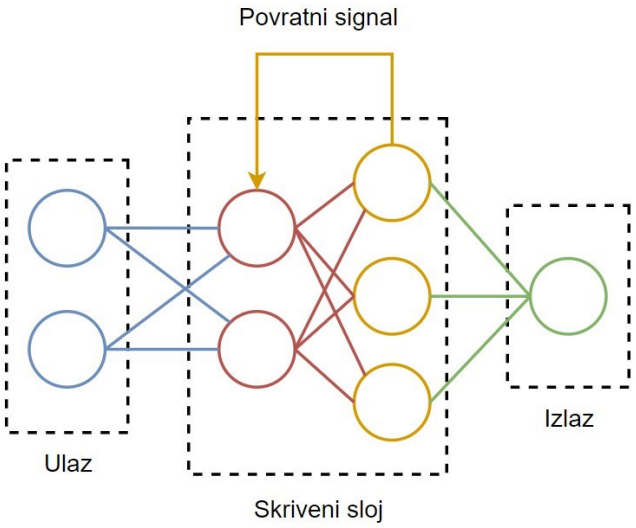
Postoje više problema koji su vezani uz zadatak potiskivanja šuma. Prvi se odnosi na potrebu za robusnosti modela pri različitim uvjetima i okolnostima rješavanja zadatka koji su prisutni u stvarnome svijetu. Drugi problem je dostupnost čistih govornih zapisa, koji su trenutno ograničeni u javnoj domeni. Trenutno najviše skupa podataka dolazi od čitanog materijala, kao što je i Libri Speech Noise Dataset [13] koji je korišten u ovome radu. Kao treći problem, zadatak postaje sve teži za obaviti kako se odnos signal/šum zvučnog zapisa za obradu smanjuje. To se može zaobići učenjem modela nad većim skupom podataka no to povećava šansu da će model imati veći pomak prema jednoj vrsti šuma, čime se smanjuje robusnost modela u stvarnim primjenama. I kao zadnji problem, nesuglasica između ljudske percepcije kvalitete zapisa i funkcije za smanjenje gubitka može rezultirati time da rad dobro optimiziranih modela bude lošije percipiran od strane čovjeka [14].

3.2.1 Povratne neuronske mreže

Povratne neuronske mreže se sastoje od ulaznog sloja, skrivenog sloja, povratne petlje za skriveno stanje iz skrivenog sloja, izlaznog sloja. Povratna petlja omogućuje da se skriveno stanje mijenja svakom iteracijom kroz model. Primjerice, svaki zvučni zapis se može podijeliti na vremenske intervale. Prolaskom pojedinog dijela zapisa, skriveno stanje se mijenja svakom iteracijom, pohranjujući stanje prošle iteracije. Na kraju svake iteracije, izlaz se šalje naprijed kroz mrežu kako bi se generirao potpuno novi zvučni zapis bez šuma. Problem koji se pojavljuje kod ovakvog tipa mreža je pohrana informacije o stanju na duži vremenski period te ili nestajući gradijent (smanjenje gradijenta iteracijama zbog kojih nadopuna težinske vrijednosti postane zanemariva ukoliko gradijent postane zanemariv) ili „eksplodirajući“ gradijent (zbrajanje velikih gradijenata uzrokuje nadopunu težinskih vrijednosti sa većim parametrima) što uzrokuje sporo učenje u usporedbi s ostalim mrežama [15]. Zbog toga su nastale varijante RNN koje koriste pragove (eng. *gates)*. To su operacije koje uče koje informacije treba dodati ili izdvojiti iz skrivenog stanja mreže. Dvije varijante unaprijeđenih RNN-a su LSTM (eng. *Long Short-Term Memory*) i GRU (*eng.*

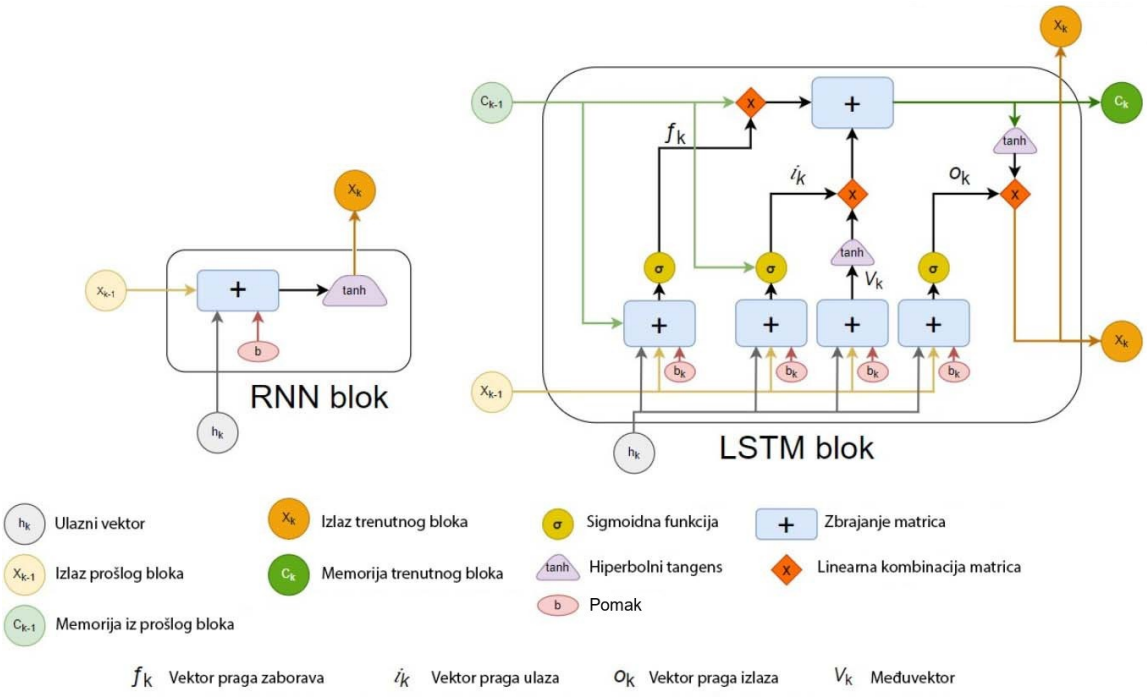
Gated Recurrent Unit). Obje varijante su računski zahtjevnije nego li tradicionalne RNN, ali su pogodnije za uklanjanje pozadinskih šumova zvučnih zapisa [5].

Na slici 3.4. prikazana je arhitektura osnovnog tipa povratne mreže te se može primijetiti u skrivenom sloju jedan blok mreže vraća informaciju o stanju na prijašnji blok mreže, time se omogućuje „razumijevanje“ sekvencijalnih tipova podataka.



Slika 3.4 Prikaz strukture RNN tipa mreže

Na slici 3.5. prikazani su RNN i LSTM blokovi uz popratnu legendu. Kompleksnost LSTM bloka je primjetna s obzirom na RNN blok. S obzirom na RNN blok koji prima samo izlaz prošlog bloka, LSTM blok također prima i memoriju prošlog bloka.



Slika 3.5. Prikaz strukture RNN i LSTM bloka

3.2.2 Poconet arhitektura

Unatoč prednostima RNN s obzirom na konvolucijske neuronske mreže (eng. *Convolutional Neural Network* – CNN), PoCoNet arhitektura donosi nove inovacije kako bi se model bolje prilagodio zadatku uklanjanja šuma iz zvučnih zapisa. Korištenjem frekvencijsko pozicijskih značajki modelu se omogućuje efikasnije raspoznavanje frekvencijsko ovisnih karakteristika. Polu-nadgledana metoda učenja proširuje skup za treniranje poboljšavajući uzorke prije treniranja te se transformacijama sam skup proširuje. Transformacije koje su odrađene su:

- **EQ** – nasumični niski i visoki filtri za ujednačavanje, središnja frekvencija odabrana ujednačeno iz logaritamske domene između 40 i 8000 Hz, pojačanje između ±10 dB. Dvije nasumične EQ Gaussove krivulje po podatku, simetrične u logaritamskoj domeni, s Q vrijednostima između 0.5 i 1.5, frekvencija odabrana iz istog intervala kao i središnja frekvencija. Nasumično dodano na signal govora i signal šuma.
- **Izmjena visine tona** – nasumično uzorkovanje s ±10% originalnog broja uzoraka.
- **Rezanje (eng. *clipping*)** – nasumično rezanje vršne vrijednosti signala između 0.5 i 1 s primijenjene 10% vremena.
- **Simulacija praznog međuspremnika** – nasumično brisanje prvih 0.5 do 1 s za simulaciju napola punog međuspremnika u evaluaciji situacije niske latencije.

- **Razina i tišina** – zanemarivanje podataka signala sa RMS (eng. *Root Mean Square*) vrijednosti dominantnog dijela manjim od -38dBFS (eng. *dB relative to full-scale of 1.0*) i normalizacija svakog signala kako bi imali vrijednost RMS-a od -20 dBFS. Zatim se primijeni nasumično smanjene glasnoće između -30 i 0 dB za pozadinski dio, normalizira se cijeli signal na -20 dBFS RMS, te se potom primijeni nasumično pojačanje između -25 do 5 dB na cijeli signal. Dodatno se koristi tišina kao dominantni dio signala 3% cjelokupnog vremena.
- **Pojasno ograničavanje** – kako bi model bio robustan u slučajevima gdje je ulazni signal pojasno ograničen, primjenjuje se niskopropusni pojasni filtar frekvencije između 4 i 7 kHz, 2.5% vremena samo za pozadinski dio, 2.5% vremena samo za dominantni dio i 5% vremena za cijeli signal [14].

Funkcija gubitka dana je izrazom (3-1)

$$\mathcal{L}(y,\hat{y}) = \lambda_{audio}\mathcal{L}_{audio}(y,\hat{y}) + \lambda_{spectral}\mathcal{L}_{spectral}(Y,\hat{Y})$$

(Pogreška! U dokumentu nema teksta navedenog stila.-2)

gdje je funkcija audio gubitka L1 gubitak dan izrazom (3-2)

$$\mathcal{L}_{audio}(y,\hat{y}) = |y-\hat{y}|$$

(Pogreška! U dokumentu nema teksta navedenog stila.-2)

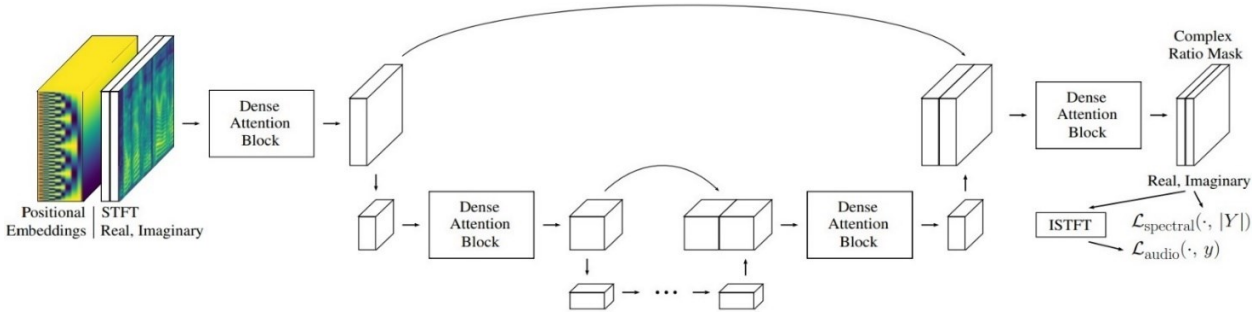
Za spektralnu funkciju gubitka $\mathcal{L}_{spectral}$, neka su $Y_{t,f} = |STFT(y)_{t,f}|$ i $\hat{Y}_{t,f} = |STFT(\hat{y})_{t,f}|$, STFT (eng. *Short Time Fourier Transformation*) veličine pojasa, te je funkcija dana izrazom (3-3)

$$\mathcal{L}_{spectral} = \sum_{t,f} w(f)\left(\lambda_{over}1_{\hat{Y}\geq Y,t,f} + \lambda_{under}1_{\hat{Y}<Y,t,f}\right)|Y_{t,f} - \hat{Y}_{t,f}|$$

(Pogreška! U dokumentu nema teksta navedenog stila.-3)

gdje je *w* frekvencijski težinska funkcija, y je dominanti dio originalnog signala bez šuma, $\hat{y}$ je procjena dominantnog djela pročišćenog signala sa šumom nakon potiskivanja šuma, $1_{\hat{Y} \geq Y,t,f}$ je karakteristična funkcija s vrijednost 1 ako je $\hat{Y} \geq Y, t, f$ ili 0 ako je suprotno. Varijable λ_{under} i λ_{over} predaju težinsku vrijednost za podcjenjivanje i precjenjivanje veličine govora. Ta funkcija gubitka omogućuje pomak (eng. *bias*) za održavanje kvalitete govora kako bi se poboljšala ljudska percepcija kvalitete [14].

Model se sastoji od 2D U-Net arhitekture sa *self-attention* slojevima i 4-slojnim DenseNet blokovima na svakoj razini. Modelu se kao ulaz daju zvučni zapisi sa šumom koji su transformirani gore navedenim transformacijama pretprocesiranje. Učenje je provedeno sa 700 tisuća iteracija, veličinom serije zvučnih zapisa od 122, korištena je *Adam* funkcija za optimizaciju, stopa učenja je bila $1e^{-4}$ te je za svakih 100 tisuća iteracija prepolovljena [14]. Na slici 3.6. može se vidjeti detaljnija struktura mreže te pojedini blokovi.



Slika 3.6. Prikaz arhitekture PoCoNet

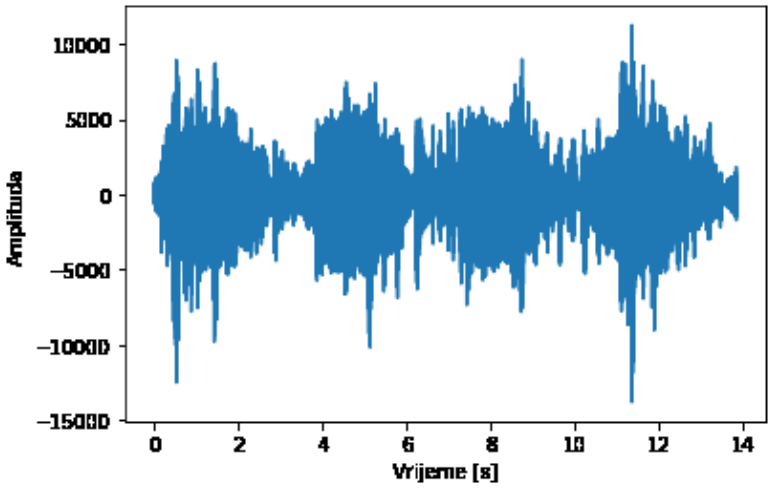
4.OPIS UZORAKA, TEHNOLOGIJA I EVALUACIJA RJEŠENJA ZA POTISKIVANJE POZADINSKIH ŠUMOVA

U ovom poglavlju dan je najprije opis korištenih skupova podatka uz popratni tehnički opis zvučnih zapisa, dan je popis korištenih implementacija metoda i popratne skripte korištene u radu. Zatim je dan kratki opis načina objektivne i subjektivne evaluacije te postignuti rezultati.

4.1 Tehnički opis uzoraka, korištene implementacije i tehnologije

Programski kod za obradu zvučnih zapisa u svrhu potiskivanja šuma pisan je u Python programskom jeziku pri čemu su korištene sljedeće biblioteke: Librosa [16], Pedalboard [17], Noisereduce [8], Scikit-learn [18], TensorFlow [19], Seaborn [20], Scipy [21], Numpy [22], MatPlotLib [23], WADA-SNR [24].

Za odabir gotovih uzoraka zvučnih zapisa sa šumom, koristio se Libri Speech Noise Dataset kao izvor. Pojedini uzorak se sastoji od govornika koji čita nekakav sadržaj te je u pozadini dodan šum. Skup podataka sadrži različite kombinacije kako govornika tako i vrsta šumova, no za evaluaciju je odabrano 5 uzoraka iz skupa podataka. Uzorci su odabrani po kompatibilnosti s WADA-SNR algoritmom zbog limitacije algoritma. Algoritam ima poteškoće pri procjeni uzoraka kojima je stvarni SNR jednak ili manji od -20 dB te uzoraka kojima je stvarna vrijednost SNR veća od 100 dB [25]. Govornici pričaju na engleskom jeziku. Trajanje pojedinog zapisa može varirati te postoje zapisi od 4 sekunde sve do zapisa koji traju i 17 sekundi. Svaki zapis ima frekvenciju uzorkovanja od 16.000 Hz. Zapisi su u .wav mono formatu. Na slici 4.1. možemo vidjeti primjer signala sa šumom u vremenskoj domeni, govor ženske osobe na engleskom jeziku pri čemu je u pozadini zvuk sirene.



Slika 4.1. Prikaz uzorka signala sa šumom u vremenskoj domeni

Za uzorak na hrvatskome jeziku snimljen je vlastiti uzorak sa muškim govornikom hrvatskog jezika uz pozadinski šum žamora glasova u trajanju od 7 sekundi. Uzorak također ima frekvenciju uzorkovanja od 16.000 Hz te je zapisan u .wav mono formatu.

U eksperimentalnom dijelu rada za tradicionalnu metodu potiskivanja šuma koristila se metoda za *spectral gating* koje se nalazi u biblioteci *noisereduce* [8] dok je za metodu strojno učenja korišten Intelov OpenVINO model pod nazivom noise-suppression-poconetlike-0001. Model je temeljen na PoCoNet arhitekturi i treniran na DNS-Challenge datasetu. Model radi s mono zvučnim zapisima s frekvencijom uzorkovanja od 16.000 Hz. Audio je iterativno procesiran u dijelovima veličine od 2048 [26]. Korištenje modela se odvija putem gotove skripte koja se može preuzeti s OpenVINO repozitorija [27].

4.2 Metrike za evaluaciju

U radu će se provesti evaluacija efikasnosti potiskivanja šuma iz zvučnog zapisa na dva načina. Prvo je izvršena objektivna evaluacija računanjem SNR procjene te subjektivna evaluacija metodom srednje vrijednosti mišljenja (eng. *Mean Opinion Score* - MOS).

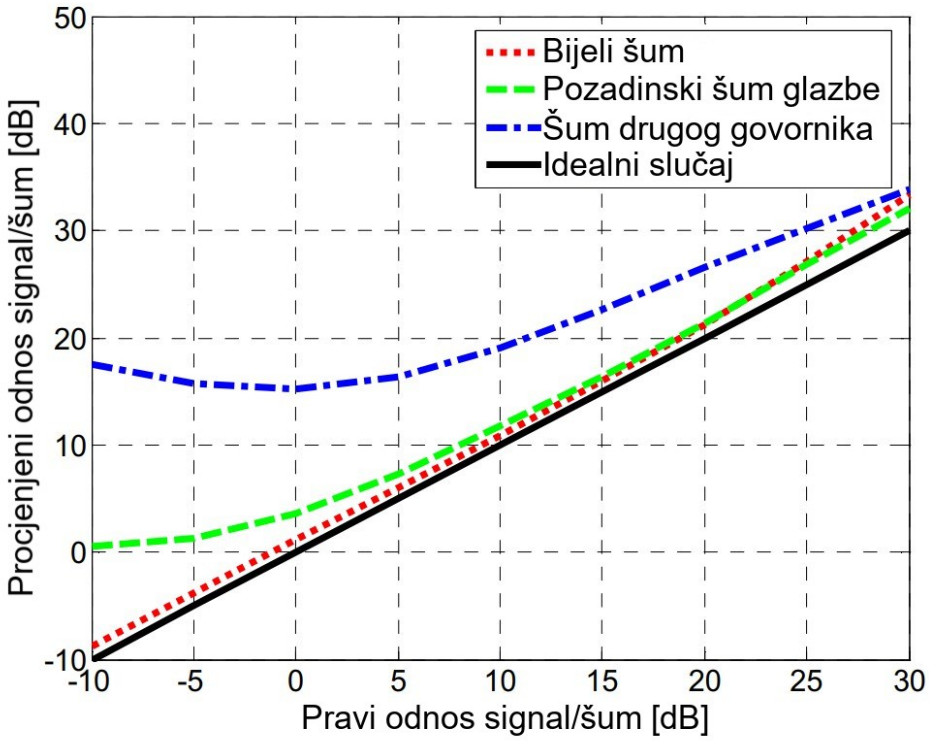
4.2.1 Objektivna metrika (procjena odnosa signal/šum)

Dobra procjena odnosa signal/šum tj. SNR može značajno pomoći pri optimizaciji algoritama u području potiskivanja šumova zvučnih zapisa. Procjena SNR-a se može svrstati u više kategorija. Jedan od pristupa se temelji na razlikovanju spektra signala od spektra šuma. Pod tu kategoriju pripadaju tehnike *Noise spectrum estimation* i *Spectral subtraction*. Drugi pristup se temelji na mjerenju energije signala. *NIST STNR* algoritam pripada ovoj skupini. Time se kreira histogram

energije kratkog vremena (eng. *short-time energy*). Međutim, neki od drugih pristupa se ne bave ni energijom niti spektralnim koeficijentima, već statističkom analizom [28].

U ovome radu se SNR procjenjivao putem WADA-SNR (eng. *Waveform Amplitude Distribution Analysis*) koji se temelji na pretpostavci da se distribucija amplituda većinom može karakterizirati preko gama distribucije s parametrom između 0.4 i 0.5. Pretpostavlja se da su signal i šum nezavisni, da čisti signal prati gama distribuciju s fiksnim parametrom, te da pozadinski šum ima Gaussovu distribuciju [28].

Na slici 4.2. prikazan je graf odstupanja procjene odnosa signal/šum korištenjem metode WADA-SNR. Graf prikazuje estimacije SNR ovisno o slučaju, primjerice idealni slučaj (crna boja) je slučaj gdje nema pozadinskog šuma, drugi slučaj je kada imamo pozadinsku glazbu (zelena boja), treći gdje postoji glas drugog govornika (plava boja), te četvrti gdje je dodan bijeli šum (eng. white noise) (crvena boja). Prema grafu može se primijetiti da procjena SNR-a za bijeli šum ne odstupa značajno pri različitim vrijednostima stvarnog SNR-a, ali se ta razlika povećava nakon 25 dB stvarnog SNR-a. Procjena za šum vrste pozadinske glazbe značajno odstupa od stvarnog SNR-a kada je vrijednost stvarnog SNR-a manja od 5 dB i veća od 25 dB. Procjena za šum vrste drugog govornika u pozadini značajno odstupa za različite vrijednosti stvarnog SNR-a, no ta razlika se značajno smanjuje kako stvarni SNR raste od 25 dB. Graf je preuzet iz rada [28].



Slika 4.2. Prikaz stvarnog i procijenjenog SNR-a WADA algoritma

4.2.2 Subjektivna metrika (MOS)

Kao mjerilo subjektivne ocjene metoda, koristila se MOS metoda nad 12 ispitanika. Ispitanici su dobrog sluha te su intervalu dobi od 18 do 25 godina. Zvučni zapisi su poslani ispitanicima digitalnim putem te su preslušavanje zvučnih zapisa odradili na vlastitoj opremi i u vlastitom okuženju. Pojedini ispitanici su prvo preslušali originalni zapis sa šumom, zatim zapis pročišćen metodom temeljenoj na strojnom učenju (PoCoNet arhitektura) i na kraju zapis pročišćen tradicionalnom metodom (*spectral gating*). Ocjene su dodijeljene zasebno za signal (SIG) i zasebno za šum (BAK) pojedinog zapisa te su dane putem tablice 4.2. napravljene u skladu s načinom ocjenjivanja predstavljenom u radu [29].

Tablica 4.2. Prikaz tablice ocjena i opisi dani ispitanicima prilikom evaluacije

Ocjena	Skala dominantnog dijela (SIG)	Skala pozadine (BAK)
5	U potpunosti prirodno, nema degradacije	Neprimjetno
4	Prilično prirodno, blaga degradacija	Donekle primjetno
3	Donekle prirodno, srednja degradacija	Primjetno, ali neupadljivo
2	Prilično neprirodno, velika degradacija	Prilično primjetno, donekle upadljivo
1	Iznimno ne prirodno, iznimna degradacija	Iznimno primjetno, iznimno upadljivo

4.3 Postignuti rezultati

4.3.1 Rezultati objektivne evaluacije

Na uzroke je najprije primijenjena metoda *spectral gating* te je kao rezultat dobiven pročišćeni signal, potom je iz istih uzorka dobiveni pročišćeni signal kao rezultat modela metode strojnog učenja (PoCoNet arhitektura). Tako su dobivena dva pročišćena signala nad kojima je provedena daljnja procjena SNR-a. U tablici 4.1. original se odnosi na zvučni zapis sa šumom, PoCoNet se odnosi na zvučni zapis pročišćen metodom strojnog učenja (PoCoNet arhitektura) te se *spectral gating* odnosi na zvučni zapis pročišćen metodom *spectral gating*. Iz rezultata prema tablici 4.1. može se zaključiti da su oba načina poboljšali procijenjeni SNR s obzirom na originalni zapis. U prvome slučaju oba pristupa značajno poboljšaju SNR. Kod drugog slučaja SNR navodi ka tome da je metoda *spectral gating* ostvarila veći procijenjeni SNR nego li PoCoNet. U trećem slučaju

PoCoNet ima približno dvostruko bolji SNR nego metoda *spectral gating*. U četvrtom slučaju nailazimo na veliko odstupanje kod SNR, gdje je metoda *spectral gating*, prema SNR vrijednosti, značajno bolja nego li PoCoNet. U petome slučaju SNR vrijednosti PoCoNet i metode *spectral gating* su međusobno različite za približno 7 dB. Za šesti slučaj kada se govori hrvatskim jezikom obje metode su ostvarile podjednaki SNR s razlikom od otprilike 1 dB manje za PoCoNet.

Tablica 4.1. Objektivna ocjena metoda u odnosu na tip šuma u usporedbi s originalom.

TIP ŠUMA	ORIGINAL	POCONET	SPECTRAL GATING
1. Zaključana vrata	8.86 dB	60.1 dB	46 dB
2. Sirena	5.68 dB	26.46 dB	36.36 dB
3. Tok vode	0.02 dB	33.68 dB	16.35 dB
4. Plač djeteta	11.03 dB	27.28 dB	43.15 dB
5. Usisavač	4.17 dB	31.31 dB	24.15 dB
6. Žamor glasova (HR)	6.45 dB	33.3 dB	34.36 dB

4.3.2 Rezultati subjektivne evaluacije

Provedena je subjektivna evaluacija uzoraka zvučnih zapisa te se tablice ocjena nalaze se u prilozima P4.1, P4.2, P4.3 i P4.4. Prema subjektivnim rezultatima iz tablice 4.3., model metode strojnog učenja (PoCoNet arhitektura) ima bolje performanse nego li tradicionalna metoda (*spectral gating*). SIG PoCoNet kroz sve testove MOS ima srednju vrijednost od 4.32, BAK PoCoNet kroz sve testove MOS ima 4.89. SIG *spectral gating* kroz sve testove MOS ima srednju vrijednost od 2.38, BAK *spectral gating* kroz sve testove MOS ima 1.74. Kroz sve testove MOS, ispitanici smatraju da PoCoNet gotovo u potpunosti ukloni pozadinski šum uz blagu degradaciju originalnog signala. Također, PoCoNet uspješno potiskuje pozadinski šum na primjeru vlastitog snimljenog uzorka gdje je jezik govornika na hrvatskom, za razliku od rezultata dobivenim *spectral gating-om* za koje ispitanici smatraju da loše uklanja pozadinski šum, a da kvaliteta signala opadne. Oba načina su najveće ocjene dobili prilikom uklanjanja šuma usisavača. PoCoNet i *spectral gating* su najslabije ocijenjeni prilikom uklanjanja šuma protoka vode. Prema slici 4.2. može se zaključiti da WADA-SNR najbolje procijeni SNR kada je tip šuma sličan bijelom šumu.

Tablica 4.3. Subjektivna ocjena metoda u odnosu na tip šuma.

TIP ŠUMA	SIG - POCONET	BAK - POCONET	SIG - SPECTRAL GATING	BAK - SPECTRAL GATING
1. Zaključana vrata	4.5	5	2.67	1.92
2. Sirena	4.17	4.83	2.25	1.58
3.Tok vode	3.5	4.75	1.83	1.25
4. Plač djeteta	4.58	4.92	2.42	1.67
5. Usisavač	4.67	5	2.75	2.25
6. Žamor glasova (HR)	4.5	4.83	2.33	1.75

4.3.3 Vrijeme procesiranja

U radu se koristilo osobno računalo s Windows 10 operacijskim sustavom, Intel-ov i5-8300H CPU @ 2.30GHz sa 16 GB 2666 MHz. Tablica 4.4 prikazuje vrijeme potrebno za obradu pojedinog uzorka zvučnog zapisa ovisno o metodi obrade. Za obje metode se koristila *CPU* implementacija. Prema tablici 4.4 može se primijetiti da PoCoNet duže obrađuje zvučne zapise nego li *spectral gating.* Prosječno vrijeme obrade za PoCoNet je 1.87 s dok za *spectral gating* je 0.57 s. Prosječno, za obradu 1s zvučnog zapisa PoCoNet treba 154 ms, dok *spectral gating* treba 48.5 ms.

Tablica 4.4 Vrijeme obrade zvučnog zapisa za pojedinu metodu

TIP ŠUMA	POCONET	SPECTRAL GATING
1. Zaključana vrata	2.5 s	0.7 s
2. Sirena	2 s	0.6 s
3 Tok vode	2.5 s	0.7 s
4 Plač djeteta	1.1 s	0.4 s
5. Usisavač	2 s	0.6 s
6. Žamor glasova (HR)	1.1 s	0.4 s

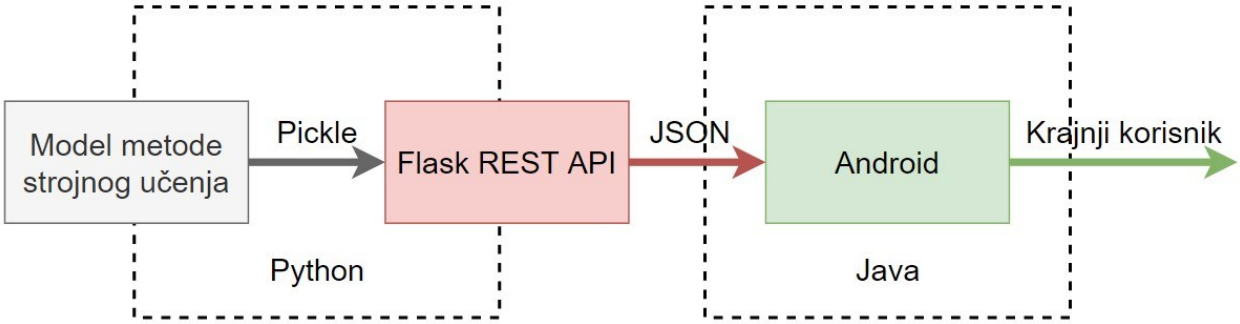
5. MOGUĆNOST IMPLEMENTACIJE MODELA METODE STROJNOG UČENJA

Za implementaciju modela metode strojnog učenja na Android sustav, postoje već gotova rješenja kao što su ML Kit – SDK *(*eng.*Software Development Kit*) od strane Googlea, TF Hub za pretragu već postojećih modela, TF Lite Model Maker za treniranje postojećih modela na novi set podataka.

Što se tiče predikcija modela, postoje dvije opcije – raditi predikcije direktno na mobilnom uređaju ili pak na udaljenom servisu te potom dohvatiti rezultate. Obje metode imaju svoje prednosti i mane. Kod obrade na uređaju prednosti su što gotovo pa ni nema latencije u obradi zahtjeva, aplikacija može raditi *offline*, te korisnički podaci i uzorci ne napuštaju uređaj. Mane su što su performanse ograničene ovisno o resursima uređaja, može se očekivati veći trošak baterije i skidanje modela na uređaj može dugo trajati. Kod obrade na udaljenom servisu, prednosti su te da udaljeni servis ima znatno jače i dostupnije računalne resurse potrebne za obradu podataka, mogućnost asinkronog rada (korisnik ne mora čekati na završetak obrade) te zbog male ili nepostojeće potrebe za obradom na uređaju, baterija može duže trajati. Mane ovog pristupa su te što bi aplikacija zahtijevala konekciju s mrežom, povećana latencija i veća potrošnja podatkovnog prijenosa te, uz to, podaci napuštaju uređaj te se treba osigurati dobra sigurnost pri razmjeni potencijalno osjetljivih podataka.

Neki od udaljenih servisa s takvom namjenom su: Firebase ML – rješenje prvotno namijenjeno za mobilne uređaje, Azure ML – Microsoftovo rješenje šire namjene s pristupnom točkom na API te nije nužno vezano za jednu vrstu platforme [30].

Na slici 5.1. prikazuje se potencijalna arhitektura i navedene tehnologije s kojima bi se moglo kreirati rješenje pogodno za mobilne uređaje. Model metode strojnog učenja se postavi na poslužitelj (eng. *backend*) u obliku *web* aplikacije (*Pickle* omogućuje serijalizaciju objekta za postavljanje na poslužitelj) te se implementira API kao komunikacija između aplikacije na korisnikovom uređaju i *web* aplikacije pomoću JSON-a (eng. *JavaScript Object Notation*) koji serijalizira poruke između API i aplikacije na korisnikovom uređaju.



Slika 5.1. Prikaz arhitekture potencijalnog rješenja s navedenim tehnologijama

6. ZAKLJUČAK

Zadaka ovog rada bilo je istražiti te usporediti rad metoda temeljenih na strojnom učenju u usporedbi s tradicionalnim metodama u svrhu potiskivanja pozadinskih šumova zvučnih zapisa. Od tradicionalnih metoda istražene su *Wiener* filtar te *spectral gating*. *Wiener* filtar se smatra industrijskim standardom te je adaptivna vrsta filtra. Iako postoji implementacija za rad s jednim signalom koji sadrži dominantni dio sa šumom, najbolje radi uz dva signala gdje je jedan govornik sa šumom te drugi samo šum. Zbog toga je u radu istražen i *spectral gating* kao alternativni pristup obrade jednog signala sa šumom. Od metoda temeljenih na strojnom učenju istražene su povratne neuronske mreže te PoCoNet arhitektura mreže. U radu je odabran već gotovi *Intelov OpenVINO* model temeljen na PoCoNet arhitekturi. Oba načina su evaluirana objektivnom i subjektivnom metodom. Kao objektivna metoda evaluacije korišten je WADA-SNR algoritam za procjenu odnosa signala/šuma dok za subjektivnu je odabrana metoda srednje vrijednosti mišljenja (MOS). Rezultati obje metode evaluacije ukazuju na veću efikasnost metode temeljene na strojnom učenju (PoCoNet arhitektura). Vrijedno je spomenuti također da model metode strojnog učenja (PoCoNet arhitektura) održava iste performanse na primjeru vlastitog snimljenog uzorka gdje je jezik govornika na hrvatskom.

LITERATURA

[1] *Noise Cancellation Basics, Types, Headphones and More* [online], T. Abdulgafar, 2021. dostupno na: https://krisp.ai/blog/noise-cancellation-types/ [pristupljeno u srpnju 2022.]

[2] *Krisp* [online], dostupno na: https://krisp.ai/ [pristupljeno u kolovozu 2022.]

[3] *Audo* [online], dostupno na: https://audo.ai/ [pristupljeno u kolovozu 2022.]

[4] *Audacity* [online], dostupno na: https://www.audacityteam.org/ [pristupljeno u kolovozu 2022.]

[6] *Recurrent Neural Networks for Noise Reduction in Robust ASR* [online], A. Maas, Q. Le, T. O'Neil, O. Vinyals, P. Nguyen, A. Ng, 2012., dostupno na: http://ai.stanford.edu/~amaas/papers/drnn_intrspch2012_final.pdf [pristupljeno u kolovozu 2022.]

[5] *Background Noise Removal: Traditional vs AI Algorithms* [online], P. Guduguntla, 2021. dostupno na: https://audo.ai/blog/background-noise-removal-traditional-vs-ai-algorithms [pristupljeno u srpnju 2022.]

[7] *Single Channel Speech Enhancement: Using Wiener Filtering with Recursive Noise Estimation* [online] , N. Upadhyay, R. Jaiswal, 2016, dostupno na: https://www.sciencedirect.com/science/article/pii/S1877050916300758?ref=pdf_download&fr=RR-2&rr=741e5440dda7788b [pristupljeno u kolovozu 2022.]

[8] *Noise reduction in python using spectral gating* [online], T. Sainburg, 2022. dostupno na: https://pypi.org/project/noisereduce/#description [pristupljeno u srpnju 2022.]

[9] *Noise reduction using spectral gating in python* (vizualizacija koraka) [online], T. Sainburg, 2018, dostupno na : https://timsainburg.com/noise-reduction-python.html [pristupljeno u kolovozu 2022.]

[10] *Speech Enhancement with Weighted Denoising Auto-Encoder* [online], B. Xia, C. Bao, 2013, dostupno na: https://www.isca-speech.org/archive_v0/archive_papers/interspeech_2013/i13_3444.pdf [pristupljeno u kolovozu 2022.]

[11] *Complex Ratio Masking for Monaural Speech Separation* [online], D. Williamson, Y. Wang, D. Wang, 2016, dostupno na:

https://aspire.sice.indiana.edu/publication_files/williamsonetal.cRM.2016.pdf [pristupljeno u kolovozu 2022.]

[12] *A Wavenet for Speech Denoising* [online], D. Rethage, J. Pons, X. Serra, 2018, dostupno na: https://arxiv.org/pdf/1706.07162.pdf [pristupljeno u kolovozu 2022.]

[13] *Libri Speech Noise Dataset* [online], dostupno na: https://www.kaggle.com/datasets/earth16/libri-speech-noise-dataset [pristupljeno u srpnju 2022.]

[14] *PoCoNet: Better Speech Enhancement with Frequency-Positional Embeddings, Semi-Supervised Conversational Data, and Biased Loss* [online], U. Isik, R. Giri, N. Phansalkar, J.-M. Valin, K. Helwani, A. Krishnaswamy, 2008. dostupno na: https://arxiv.org/pdf/2008.04470.pdf [pristupljeno u kolovozu 2022.]

[15] *Recurrent Neural Network (RNN) Tutorial: Types, Examples, LSTM and More* [online], A Biswal, 2022, dostupno na : https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn [pristupljeno u kolovozu 2022.]

[16] *Librosa* [online], dostupno na : http://librosa.org/doc/main/index.html [pristupljeno u srpnju 2022.]

[17] *Pedalboard* [online], dostupno na: https://github.com/spotify/pedalboard [pristupljeno u srpnju 2022.]

[18] *Scikit-learn* [online], dostupno na: https://scikit-learn.org/stable/ [pristupljeno u srpnju 2022.]

[19] *TensorFlow* [online], dostupno na: https://www.tensorflow.org/ [pristupljeno u srpnju 2022.]

[20] *Seabron* [online], dostupno na: https://seaborn.pydata.org/ [pristupljeno u srpnju 2022.]

[21] *Scipy* [online], dostupno na: https://scipy.org/ [pristupljeno u srpnju 2022.]

[22] *Numpy* [online], dostupno na: https://numpy.org/ [pristupljeno u srpnju 2022.]

[23] *MatPlotLib* [online], dostupno na: https://matplotlib.org/ [pristupljeno u srpnju 2022.]

[24] *WADA SNR* [online], dostupno na : https://gist.github.com/johnmeade/d8d2c67b87cda95cd253f55c21387e75 [pristupljeno u kolovozu 2022.]

[25] *WADA-SNR* (limitacije) [online], dostupno na: https://gist.github.com/johnmeade/d8d2c67b87cda95cd253f55c21387e75?permalink_comment_id=3545389%23gistcomment-3545389 [pristupljeno u kolovozu 2022.]

[26] noise-suppression-poconetlike-0001 [online], dostupno na : https://docs.openvino.ai/latest/omz_models_model_noise_suppression_poconetlike_0001.html [pristupljeno u srpnju 2022.]

[27] OpenVINO noise suppresion [online], dostupno na: https://github.com/openvinotoolkit/open_model_zoo/blob/master/demos/noise_suppression_demo/python/noise_suppression_demo.py [pristupljeno u srpnju 2022.]

[28] *Robust Signal-to-Noise Ratio Estimation Based on Waveform Amplitude Distribution Analysis* [online], C. Kim, R. M. Stern, 2008. dostupno na: http://www.cs.cmu.edu/~robust/Papers/KimSternIS08.pdf [pristupljeno u kolovozu 2022.]

[29] *Noise Cancellation Method for Robust Speech Recognition* [online], K. U. Shajeesh, K. P. Soman, 2012. dostupno na: https://research.ijcaonline.org/volume45/number11/pxc3879438.pdf [pristupljeno u kolovozu 2022.]

[30] *Build smarter apps with machine learning* [online], Android Developers, dostupno na: https://developer.android.com/ml [pristupljeno u kolovozu 2022.]

SAŽETAK

Naslov: Izdvajanje signala govornika potiskivanjem pozadinskog šuma pomoću metoda temeljnih na strojnom učenju

U ovome radu, u prvome poglavlju napravljen je kratki uvod na tematiku, opisani su osnovni pojmovi i definicije. U drugome poglavlju detaljnije su opisani opći načini potiskivanja pozadinskih šumova, dan je pregled gotovih programskih rješenja za potiskivanje pozadinskih šumova te je dan kratki uvod u tradicionalne metode i metode temeljene na strojnom učenju. U trećem poglavlju, detaljnije su opisane tradicionalne metode i metode temeljene na strojnom učenju. Kod tradicionalnih metoda detaljno su opisani principi rada Wiener filtra te *spectral gating* pristupa – *spectral gating* je odabran kao tradicionalna metoda u eksperimentalnom dijelu. Kod metoda temeljenih na strojnom učenju detaljnije je opisan opći postupak rada metoda te su detaljno opisane povratne neuronske mreže i PoCoNet arhitektura koja je bila odabrana za metodu temeljenu na strojnom učenju u eksperimentalnom dijelu rada. Četvrto poglavlje daje pregled tehnologija danih u radu, tehnički opis uzoraka i skupa podataka, detaljan opis metrika za evaluaciju kako za objektivnu tako i za subjektivnu metriku, te su na kraju prikazani rezultati objektivne i subjektivne metode. U petome poglavlju dan je kratki uvod u mogućnost implementacije modela metode strojnog učenja na mobilnim uređajima.

Ključne riječi: potiskivanje šuma, pozadinski šum, spectral gating, strojno učenje

ABSTRACT

Title: Background noise suppresion of speech signals based on machine learning methods

In this paper, in the first chapter, a short introduction to the topic is made, the basic terms and definitions used in the paper are described, as well as the presentation of sound recordings in both the time and frequency domains. In the second chapter, general methods of background noise suppression are described in more detail, an overview of ready-made software solutions for background noise suppression is given, and a brief introduction to traditional methods and methods based on machine learning is given. In the third chapter, traditional methods and methods based on machine learning are described in more detail. With traditional methods, the working principles of the Wiener filter and the spectral gating approach are described in detail - spectral gating was chosen as the traditional method in the experimental part. In the case of methods based on machine learning, the general procedure of the methods is described in more detail, details of recurrent neural networks and the PoCoNet architecture which was chosen for the method based on machine learning in the experimental part of the work. The fourth chapter provides an overview of the technologies presented in the paper, a technical description of the samples and data set, a detailed description of the evaluation metrics for both objective and subjective metrics, and finally the results of the objective and subjective methods are presented.. In the fifth chapter, a short introduction to the possibility of implementation on mobile devices is given.

Keywords: noise suppresion, background noise, spectral gating, machine learning

PRILOZI

P4.1. Tablica SIG ocjena za metodu strojnog učenja

Tip šuma \ Redni broj ispitanika	1	2	3	4	5	6	7	8	9	10	11	12
Zaključana vrata	5	5	4	3	5	5	3	5	5	5	4	5
Sirena	4	4	3	4	5	4	5	4	4	4	4	5
Tok vode	3	4	2	3	4	4	3	4	4	4	3	4
Plač djeteta	5	5	4	4	4	5	4	4	5	5	5	5
Usisavač	5	5	4	4	5	4	5	4	5	5	5	5
Žamor glasova (hrvatski jezik govornika)	5	5	4	4	5	5	4	4	5	4	4	5

P4.2. Tablica BAK ocjena za metodu strojnog učenja

Tip šuma \ Redni broj ispitanika	1	2	3	4	5	6	7	8	9	10	11	12
Zaključana vrata	5	5	5	5	5	5	5	5	5	5	5	5
Sirena	5	4	5	5	5	5	5	5	4	5	5	5
Tok vode	5	3	5	5	5	5	5	5	5	5	4	5
Plač djeteta	5	5	5	5	5	5	5	5	4	5	5	5
Usisavač	5	5	5	5	5	5	5	5	5	5	5	5
Žamor glasova (hrvatski jezik govornika)	5	4	5	5	5	5	5	5	4	5	5	5

P4.3. Tablica SIG ocjena za tradicionalnu metodu

Tip šuma \ Redni broj ispitanika	1	2	3	4	5	6	7	8	9	10	11	12
Zaključana vrata	3	4	2	1	3	3	1	2	2	4	5	2
Sirena	2	3	2	1	4	2	1	2	1	3	5	1
Tok vode	3	3	1	1	2	1	1	2	2	2	3	1
Plač djeteta	3	1	1	3	4	4	1	2	2	3	3	2
Usisavač	3	3	2	3	4	3	2	2	2	4	4	1
Žamor glasova (hrvatski jezik govornika)	3	1	2	3	3	3	1	2	2	3	3	2

P4.4. Tablica BAK ocjena za tradicionalnu metodu

Tip šuma \ Redni broj ispitanika	1	2	3	4	5	6	7	8	9	10	11	12
Zaključana vrata	2	1	3	1	1	2	2	2	2	3	1	3
Sirena	2	1	2	1	2	2	2	1	1	2	1	2
Tok vode	2	1	1	2	1	1	1	1	1	2	1	1
Plač djeteta	2	1	2	1	2	3	1	2	1	2	1	2
Usisavač	3	3	2	2	2	2	1	1	3	3	3	2
Žamor glasova (hrvatski jezik govornika)	2	1	2	2	1	2	1	2	2	3	1	2