UNIVERZITET U BEOGRADU

ELEKTROTEHNIČKI FAKULTET



**Algoritmi za prepoznavanje muzickih numera**

SEMINARSKI RAD

Mentor: Kandidat:

Dr Dragana Sumarac Pavlovic, profesor Uroš Stojanović, 3269/2023

Beograd, 2023.

Sadržaj

[Uvod -1st 1](#_Toc158041823)

[Shazam -1st 2](#_Toc158041824)

[Teorijski osnov -4st 3](#_Toc158041825)

[Zvuk 3](#_Toc158041826)

[Furijeova transformacija 4](#_Toc158041827)

[Spektrogram 4](#_Toc158041828)

[Performanse – 5+st 6](#_Toc158041829)

[Zaključak-1st 7](#_Toc158041830)

[Literatura 8](#_Toc158041831)

[Spisak slika 9](#_Toc158041832)

[Spisak tabela 10](#_Toc158041833)

[Prilog 11](#_Toc158041834)

# Uvod -1st

Potreba za prepoznavanjem numera i njihovih slicnosti je prisutna u nasim svakodnevnim zivotima, od razonodnih upotreba do komercijalnih aplikacija. Programi razvijani u ove svrhe postoje vec dugi niz godina u muzickoj industriji. Najkorisceniji programi ove vrste su Shazam, Soundhound i Adobe Audition. Ovi softveri kao osnovnu funkciju imaju kreiranje digitalnog otiska pesme i poredjenje dobijenih otisaka sa vec postojecim pesmama u svojim bazama podataka. Ovaj nacin upotrebe se pretezno koristi u razonodne svrhe radi pronalazenja pesme koju cujemo na javnom mestu i ne mozemo sami da je identifikujemo. Druga najcesca primena za koju se mogu koristiti neki od gorenavedenih softvera je prepoznavanje plagijata.

Ovaj rad se pretezno fokusira na teorijski princip funkcionisanja aplikacije *Shazam*, kao i testiranja granica softvera i mogucnosti njegovog zavaravanja.

Osnovni princip rada programa *Shazam* bazira se na kreiranjem digiralnog otiska muzicke numere. Otisak se formira pomocu podataka o vrhovima koje dobijamo sa spektrograma datog snimka. Pomocu odredjenog broja ovih vrhova, njihovih frekvenci i vremenskih udaljenosti relativno jednih u odnosu na druge u vremenu dobijamo dobru osnovu za podatke potrebne za formiranje digitalnog otiska. Odatle koriscenjem *hash* mapa svodimo dobijene podatke na znatno manje prostora sto omogucuje brze poredjenje sa bazom podataka.

Eksploatacijom poznavanja frekvencijskih opsega u kojima aplikacija pretezno formira digitalne otiske testiramo njene granice prepoznavanja numera koje su filtrirane ili imaju unetu distorziju na odredjenim frekvencama kao i prisustvo buke i suma u generalnom opsegu od 20Hz do 20kHz. Nakon ovoga ce isti snimci biti reprodukovati na slusalicama osobama koje su upoznate sa datim numerama radi poredjenja mogucnosti prepoznavanja softvera i ljudskog uha.

# Shazam -1st

*Shazam Entertainment Ltd.* je kompanija osnovana 2000-te sa ciljem da napravi mobilnu aplikaciju sa sposobnoscu da prepoznaje muzicke numere. Njihov algoritam je razvijan sa ciljem da moze da prepozna pesme na osnovu kratkih isecaka od svega nekoliko sekundi u prisustvu buke, kompresije i drugih izazova i neidealnih uslova. Originalni algoritam je uspesno radio ovaj posao sa bazom podataka od preko 2 miliona numera. Naknadno su uvedene dodatne nadogradnje na aplikaciju kao sto su mogucnost kupovine numera preko njihove aplikacije i monitoring za autorska prava.

*Shazam* kao osnovni princip rada kreira digiralni otisak muzickih numera. Proces kreiranja ovih otisaka se bazira na pronalazenju vrhova u spektrogramu snimljenog isecka pesme i koriscenjem *hash* tokena dobijenih ovim procesom pronaci najslicniji token u bazi podataka vec poznatih numera. Otisvi treba svi da prate odredjene principe: treba da budu vremenski invarjantni, vremenski lokalizovani, robusni i sto priblizniji stanju entropije nakon *hash*-ovanja. Vremenska invarjantnost omogucuje podudaranje otisaka nezavisno od njegove pozicije unutar audio fajla. Lokalizovanost omogucuje da se svaki *hash* racuna uparivanjem blizskih audio isecaka. Robusnost osigurava da je moguce napraviti *hash* iako je snimak degradiran ili zasumljen. Glavni faktor koji osigurava nizak broj lazno pozitivnih rezultata je visok nivo entropije.

… JOS PONESS O SHAZAMU…

# Teorijski osnov -4st

## Zvuk

Zvuk je svaka vremenski promenljiva mehanička deformacija u elastičnoj sredini. Zvuk nam sluzi kao sredstvo komunikacije i medium za prenos informacija. Osnovne karakteristike zvučnih informacija su govor, muzika i ambijentalni zvuk. U akustici zvuk sagledamo u dva domena: vremenskom i frekvencijskom. Vremenski domen nam pretezno sluzi samo za snimanje i reprodukciju zvuka. U obradi signala najcesce koristimo frekvencijski domen zbog bolje preglednosti i lakse manipulacije nad signalom. Prelaz iz vremenskog domena u frekvencijski izgleda tako sto signal koji je analogni skup vrednosti u vremenu razlozimo u sumu sinusoida razlicitih frekvenci. Karakteristika da svaki signal mozemo da razlozimo u sumu sinusoida potice iz furijeovog reda. Karakteristike zvuka koje nas zanimaju gledano iz frekvencijskog domena su frekvence i njihove amplitude. Ove vrednosti dobijamo primenom Frujeove transformacije na signal.

Mikrofon je ulazni elektroakustički pretvarač koji je prilagođen radu u vazduhu kao mediju. Mikrofon pretvara zvučni pritisak, koji mu je ulazna veličina, u električni signal na njegovom izlazu. Ovi uredjaji rade u digitalnom domenu. Pri snimanju bilo kog zvuka, membrana mikrofona osciluje i pretvara mehanicku energiju u analogni elektricni signal. AD konvertori vrse konverziju iz analognog signala u digitalni kroz procese semplinga i kvantizacije. Sempling je proces kojim odabiramo da ubelezimo vrednost na svakih T sekundi. Ovim postupkom dobijamo 1/T odbiraka u sekundi, koji ljudskom uhu pri reprodukciji zvuce priblizno identicno kao originalni signal za 1/T > 44100Hz. Kvantizacija je proces diskretizacije snaga signala. Deljenjem skale jacine signala na konacno mnogo vrednosti koje mozemo dodeliti vrednosti amplitude mi ga kvantizujemo. Na slici ispod je prikazan signal pre i nakon konverzije iz analognog u digitalni.

A graph of a signal

Description automatically generated

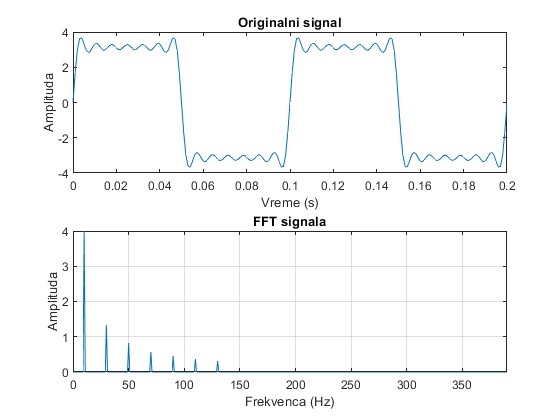
Slika 1. Analogni i digitalni signal

## Furijeova transformacija

Koriscenjem kvantizovanog i semplovanog vremenskog oblika signala mozemo da cuvamo podatke o signalu u digitalnoj formi, ali radi lakse obrade signala koristimo frekvencijski opseg. Signal konvertujemo iz vremenskog u frekvencijski oblik pomocu Furijeove transformacije. Furijeova transformacija signala racuna se na sledeci nacin:

za analogni signal. je kompleksna velicina. Njen moduo naziva sespektralna gustina amplituda, a argument spektralna gustina faza. Za digitalni signal Furijeova trasformacija glasi:

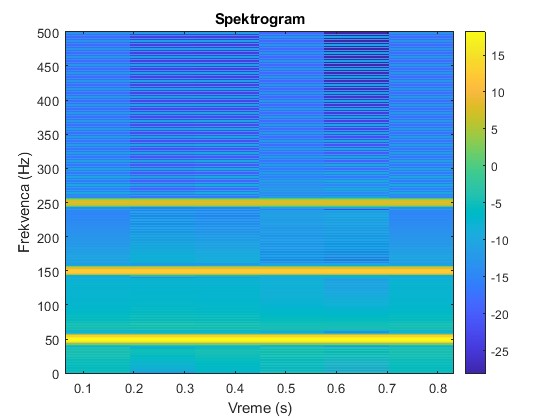
Furijeovom transformacijom dobijamo prikaz frekvenci koje su prisutne u nasem signalu i njihove amplitude.



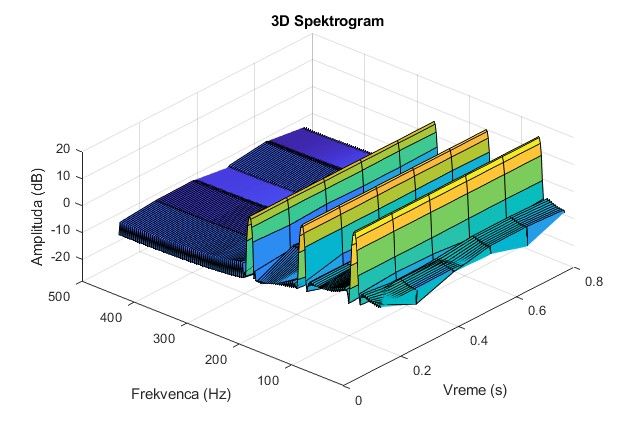
Slika 2. Signal u vremenskom i frekvencijskom obliku

## Spektrogram

Spektrogram je vizuelna reprezentacija frekvenci u signalu i njihovih promena u vremenu. Na X osi nam se nalaze frekvence, na Y osi vreme, a boje predstavljaju intenzitet ili ti amplitude na datim frekvencama. *Shazam* i drugi programi za prepoznavanje numera koriste mel-spektrogram, koji je kombinacija obicnog spektrograma, mel baze filtara i dB skale.



Slika 3. Spektrogram



Slika 4. 3D prikaz spektrograma

Sa slika 3 i 4 vidimo istaknute prisutne frekvence iz signala. Pozicije ovih vrhova su nam kljucni podaci koje *Shazam* algoritam koristi za kreiranje otiska jedne numere.

Zvuk, frekvence, mikrofon, digitalizacija, DFT - FFT, Spektrogram, Otisakivanje i algoritam, hash, menadzment bazom, sum i distorzija

# Autoenkoder – 3st

# Performanse – 5+st

Brzina, otpornost na sum i distorziju na razlicitim frekvenacama, false positive

# Zaključak-1st

# Literatura

1. Dragana Sumarac Pavlovic, Miomir Mijic, “Elektroakustika”, Beograd 2017
2. Avery Li-Chun Wang, “An Industrial-Strength Audio Search Algorithm”, Shazam Entertainment, Ltd., 2003.
3. Avery Wang, “The Shazam music recognition service”, 2006.
4. Conor Murphy, “Music-Recognition-Tab”, 2022
5. Pradeep Kumar Khanal, Achyut Kayastha, Ashish Khatakho, “Music recognition using deep learning”, *Tribhuvan University Institute of engineering*
6. Blog Ane Harris na sajtu Medium

<https://medium.com/@anaharris/how-does-shazam-work-d38f74e41359>

1. Clanak Jovana Jovanovica na stranici Toptal

<https://www.toptal.com/algorithms/shazam-it-music-processing-fingerprinting-and-recognition>

# Spisak slika

Slika 2.1 –

# Spisak tabela

Tabela 3.1 –

# Prilog