

# Vizualizacija glazbe

Ivan Vlahov

mentorica Željka Mihajlović, prof. dr. sc.

Fakultet elektrotehnike i računarstva, Zagreb

iv51747@fer.hr

**Sažetak**—U ovom radu prikazano je više metoda vizualizacije glazbe. Prikazuje se jednostavna vizualizacija amplitudom (glasnoćom), rastav zvučnih signala korištenjem algoritma brze Fourierove transformacije (FFT), te se daje pregled dostupne literature o vizualizaciji glazbe korištenjem generativnih suparničkih mreža (GAN). Prikazana je implementacija vizualizacije amplitudom i rastavom signala FFT-om u programskom paketu Processing.

**Ključne riječi**—vizualizacija, glazba, video, FFT, GAN

## I. UVOD

Vizualizacija glazbe je relativno nova grana umjetnosti koja se pojavila sedamdesetih godina prošlog stoljeća. Popularnost je stekla petnaestak godina kasnije, kada su se počeli pojavljivati programi za vizualizaciju glazbe za osobna računala. [3] Danas se vizualizacija glazbe koristi za razne stvari, od izrade vizualnih efekata na koncertima, glazbenih video spotova, pa sve do ukrasnih svjećica na božićnim drvcima. Neki od programa koji se koriste za vizualizaciju su Renderforest, Synesthesia i Videobolt.



Slika 1. Program za reprodukciju glazbe Winamp, Izvor: [5]

Iako već postoji puno programa za vizualizaciju, u ovom radu će biti predstavljene tri metode vizualizacije: vizualizacija amplitudom (glasnoćom) zvučnog signala, vizualizacija FFT-om, odnosno rastavom zvučnih valova na sinusoidne signale različitih frekvencija i amplituda, te se na kraju daje kratak uvod u generativne suparničke mreže (eng. *generative adversarial networks, GANs*) i uvid u dva rada u kojem se glazba vizualizira korištenjem GAN-ova.

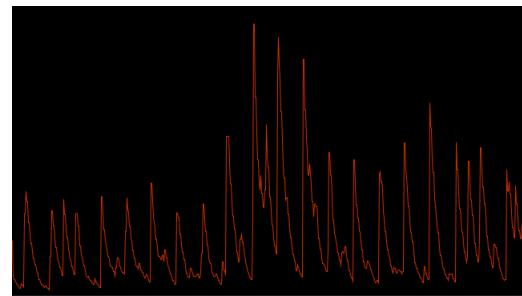
Za prve dvije metode dane su i snimke zaslona implementacije napravljene u programskom paketu Processing korištenjem biblioteke Sound. [1]

## II. JEDNOSTAVNE METODE VIZUALIZACIJE

### A. Vizualizacija amplitudom

Prva tehnika je vrlo jednostavna - varijabla koja se koristi za animaciju objekata jest amplituda zvučnog signala u određenom trenutku. Korištenje ove metode ne omogućuje veliku fleksibilnost, jer svi objekti koje animiramo ovise o istoj varijabli.

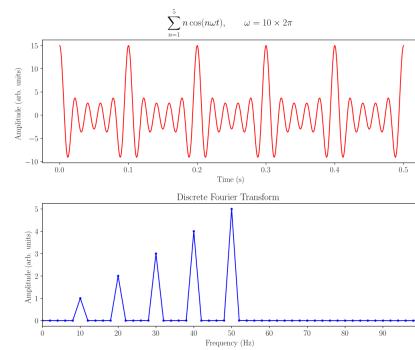
U programskom paketu Processing, ovaku vizualizaciju možemo izvesti korištenjem razreda Amplitude. Na Slici 2 prikazana je snimka zaslona vizualizacije amplitude (y-os) u odnosu na vrijeme (x-os).



Slika 2. Vizualizacija glasnoćom u Processingu

### B. Vizualizacija FFT-om

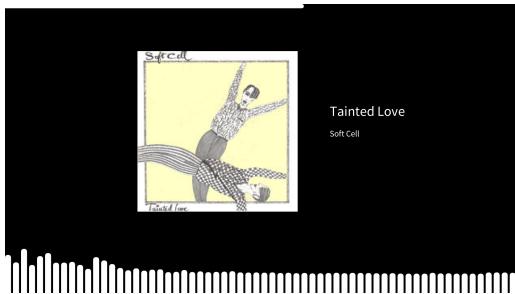
Druga tehnika je složenija, a radi se o rastavu zvučnih valova na zbroj sinusoida korištenjem algoritma brze Fourierove transformacije (eng. *Fast Fourier Transform, FFT*). Ovaj algoritam pretvara jedan uzorak zvučnog zapisa iz vremenske domene u frekvencijsku.



Slika 3. Primjer rastava korištenjem FFT-a, Izvor: [4]

Ovaj pristup omogućava puno veću prilagodbu, jer se objekti na zaslonu mogu animirati npr. ovisno o basu, vokalima i sl., tako što se filtriraju određene frekvencije, te se one postave kao argumenti funkcija kojima animiramo.

U *Sound* biblioteci programskog paketa Processing, ovaj algoritam dolazi ugrađen u razred FFT, te se na Slici 4 može vidjeti snimka zaslona implementacije jednostavnog programa za vizualizaciju, u kojem je prikazana slika, naslov i autor određene pjesme, a ispod tih informacija se nalazi stupčasti grafikon koji prikazuje koliko je u tom trenutku velika amplituda određenih frekvencija koje sačinjavaju uzorak zvučnih signala.



Slika 4. Vizualizacija FFT-om u Processingu

### III. VIZUALIZACIJA GAN-OM

Generativne suparničke mreže, ili kraće GAN-ovi, su modeli neuronskih mreža koji se sastoje od dva dijela: prvi dio je generator, koji stvara izlazne podatke pokušavajući ih učiniti što sličnijima podacima iz skupa za učenje, a drugi dio je diskriminator, koji za određene podatke pokušava prepoznati radi li se o pravim podacima iz skupa za učenje ili o generiranim podacima. Ova dva modela djeluju kao suparnici - generator pokušava "prevariti" diskriminatorsku mrežu, a diskriminator pokušava "otkriti" razlike između stvarnih i generiranih podataka.

GAN-ovi su relativno nova metoda, objavljeni u radu Iana Goodfellowa iz 2014. godine [12]. Na temelju tog rada, kroz sljedećih nekoliko godina pojavili su se modeli koji generiraju jako realistične slike ljudi, životinja, umjetničkih djela itd. Primjer takvog modela je StyleGAN iz 2019., koji se može i isprobati u aplikaciji This person does not exist [6].



Slika 5. Slike koje je generirao StyleGAN; ni čovjek ni mačka na slici ne postoje. Izvor: [6] [7]

#### A. Vizualizacija StyleGAN-om

U radu *Stylizing Audio Reactive Visuals* [8], autori predlažu metodu vizualizacije glazbe korištenjem GAN-ova, konkretno modela StyleGAN. Ideja rada jest da se vizualizacija provodi u tri dijela: izdvajanje značajki, šetnja po ulaznom prostoru i mapiranje značajki u StyleGAN.

Značajke se izdvajaju tako što se dijele na niske, srednje i visoke pojaseve korištenjem audio-filtera. Za svaki uzorak zvučnog signala računa se razlika ovih triju značajki u odnosu na prošli uzorak te se pomoću njih provodi šetnja po ulaznom prostoru. Za svaku značajku frekvencije se generira slučajni vektor čije dimenzije odgovaraju dimenzijama ulaznog prostora, a magnituda mu se postavlja na magnitudu odgovarajuće značajke frekvencije. Zatim se ti novogenerirani vektori dodaju ulaznim vektorima iz prošlog koraka. Ovaj pristup omogućava gladak prijelaz između slika.



Slika 6. Primjer nekoliko slika iz vizualizacije StyleGAN-om, Izvor: [8]

#### B. Vizualizacija BigGAN-om

Drugi primjer modela korištenog za vizualizaciju glazbe je BigGAN, u radu *The Deep Music Visualizer* [9]. Ulagani podaci ovog modela sastoje se od vektora s 1000 stupaca koji odgovara jednom od 1000 ImageNet razreda različitih objekata [10], te vektora šuma sa 128 stupaca pomoću kojeg se postavljaju različite izlazne značajke poput boje, dimenzije objekata, položaja, orijentacije i sl.

Autor ovog rada također provodi izdvajanje značajki i mapiranje na GAN. Na početku se odabire do 12 različitih ImageNet razreda za koje se želi generirati vizualizacija. Visina tona je prva izdvojena značajka, kojom se upravlja koji će se od 12 razreda objekata prikazati. S druge strane, tempo i glasnoća upravljaju vektorom šuma. Slično kao i u prethodnom primjeru, ovaj pristup omogućava gladak prijelaz između slika.



Slika 7. Primjer nekoliko slika iz vizualizacije BigGAN-om, Izvor: [13]

#### IV. DISKUSIJA, DALJNJI RAD I ZAKLJUČAK

Vizualizacija glazbe je zanimljivo, još uvijek novo područje u kojem i dalje ima sadržaja za istraživanje. S obzirom na svoj status u popularnoj kulturi i svakodnevnom životu, za očekivati je da će se i dalje razvijati, i da ćemo sve češće viđati glazbene spotove koje je u potpunosti generirao model umjetne inteligencije.

Ovaj rad pokriva samo uvod u jednostavne metode vizualizacije, kao i jednostavne implementacije navedenih metoda. Kombiniranjem ovih metoda, korištenjem filtera različitih frekvencijskih pojaseva i dalnjim istraživanjem različitih generativnih modela, ostavljen je velik prostor za dalnje istraživanje metoda vizualizacije glazbe.

Programski kod korišten u ovom radu može se pronaći na GitHub stranici autora [11].

#### LITERATURA

- [1] Processing Sound Library,  
<https://processing.org/reference/libraries/sound/index.html>
- [2] Let's learn about waveforms  
<https://pudding.cool/2018/02/waveforms/>
- [3] The basics of Music Visualizers  
<https://www.ipr.edu/blogs/audio-production/the-basics-of-music-visualizers/>
- [4] Fast Fourier Transform  
[https://en.wikipedia.org/wiki/Fast\\_Fourier\\_transform](https://en.wikipedia.org/wiki/Fast_Fourier_transform)
- [5] Webamp  
<https://webamp.org/>
- [6] This person does not exist  
<https://www.thispersondoesnotexist.com/>
- [7] This cat does not exist  
<https://thiscatdoesnotexist.com/>
- [8] Stylizing Audio Reactive Visuals, Han-Hung Lee et al.  
<https://neurips2019creativity.github.io/doc/Stylizing%20Audio%20Reactive%20Visuals.pdf>
- [9] The Deep Music Visualizer, Matt Siegelman  
<https://github.com/msieg/deep-music-visualizer>
- [10] IMAGENET 1000 Class List  
<https://deeplearning.cms.waikato.ac.nz/user-guide/class-maps/IMAGENET/>
- [11] music-visualisation, I. Vlahov  
<https://github.com/vlahovivan/music-visualisation>
- [12] Generative Adversarial Nets, I. Goodfellow et al.  
<https://arxiv.org/pdf/1406.2661.pdf>
- [13] Deep Music Visualizer: Where is my mind - Maxence Cyrin  
<https://vimeo.com/365200802?signup=true>