

Opis baze podataka i zadatka takmičenja

U okviru dostavljenog foldera nalazi se baza podataka namenjena obuci modela za **predviđanje blendshape koeficijenata** kojima se animira avatar u realnom vremenu. Cilj takmičenja je razvoj modela koji, na osnovu ulaznih informacija (tekst i/ili audio), predviđa sekvencu blendshape koeficijenata kompatibilnih sa avатарom rigovanim po **ARKit 52 standardu**, i to na način pogodan za real-time upotrebu.

Tokom **testiranja**, učesnicima će biti data rečenica koju avatar treba da izgovori. Za tu rečenicu biće obezbeđeni:

- tekstualni zapis i
- audio signal.

Model treba da generiše blendshape koeficijente koji će omogućiti prirodnu i sinhronizovanu animaciju avatara u realnom vremenu.

Struktura dostavljenih podataka

Baza sadrži više foldera sa sledećim sadržajem:

1. Spk08.zip

- **video_a_ser_spk08.zip**

- Video snimci govornika po kome je avatar napravljen
- Ukupno 80 MP4 fajlova, po jedna rečenica na srpskom jeziku
- FPS: 100
- Svaki fajl sadrži video i pripadajući audio zapis
- Snimci su frontalni i namenjeni analizi pokreta lica

- **audio.zip**

- Visokokvalitetni audio snimci
 - Snimani simultano sa video-materijalom, pomoću profesionalnog mikrofona
 - Ovi snimci su korišćeni za polu-manuelna poravnanja frejmova po fonemima
-

2. labels_08_aligned.zip

- TXT fajlovi sa informacijom o vremenu početka i kraja svakog fonema, izraženom u sekundama
 - Poravnanja su delimično manuelna i mogu sadržati greške, što treba imati u vidu prilikom treninga i evaluacije modela
-

3. spk08_ser_blendshapes.zip

Ovaj folder sadrži sve informacije vezane za izdvojene blendshape koeficijente:

- .json fajlove sa metapodacima za svaki snimak:
 - FPS
 - ukupan broj frejmova
 - broj blendshape koeficijenata po frejmu
 - spisak blendshape-ova i njihov redosled
 - .npy fajlove koji sadrže numeričke vrednosti blendshape koeficijenata po frejmu
-

4. spk08_audio_synth.zip

Ovaj folder sadrži audio fajlove koji po tekstualnom sadržaju odgovaraju originalnim audio fajlovima, ali su sintetizovani pomoću TTS. Ovakvi audio fajlovi biće dati i u test fazi uz snimak prirodnog govora osobe iz baze za obuku, samo sa drugačijim rečenicama.

Napomena o blendshape standardu, ekstrakciji i mogućnostima

Avatar je rigovan prema ARKit 52 blendshape standardu, a spisak podržanih blendshape-ova dat je u nastavku dokumentacije.

Kako su za trening dostupni gotovi video snimci, ciljni blendshape koeficijenti su inicijalno izdvojeni pomoću MediaPipe Blendshape V2 modela (<https://storage.googleapis.com/mediapipe-assets/Model%20Card%20Blendshape%20V2.pdf>).

Važno je naglasiti sledeće:

- Izdvojeni blendshape koeficijenti nisu idealni, jer su dobijeni iz 2D video-snimaka, dok je avatar rigovan za 3D animaciju (ARKit 52).
- Učesnici su slobodni da:
 - koriste druge alate za ekstrakciju blendshape-ova,
 - ponovo izvuku koeficijente iz dostupnih video snimaka, ili
 - primene sopstveni preprocessing, filtriranje ili mapiranje postojećih koeficijenata
- Cilj je postići što prirodniju animaciju avatara, a ne nužno reprodukovati originalne koeficijente jedan-na-jedan

Jasna definicija ulaza i izlaza modela

- Ulaz u model nije striktno definisan.
U test skupu biće obezbeđeni:
 - tekst koji avatar treba da izgovori
 - audio snimak u kom osoba iz trening skupa izgovara taj tekst ili
 - audio signal dobijen sintezom govora na osnovu tog teksta.

Učesnici su slobodni da kao ulaz koriste:

- samo tekst,
- samo audio,
- kombinaciju teksta i audija, ili
- bilo koja dodatna obeležja koja mogu biti izvedena iz navedenih podataka (npr. fonemi, prozodijska obeležja, embeddings i sl.).
- FPS izlaza nije striktno propisan.
Učesnici su dužni da jasno navedu frekvenciju (FPS) po kojoj generišu blendshape koeficijente, kao i eventualne postupke interpolacije ili resamplovanja.
- Latency nije striktno ograničen, ali će modeli sa manjim kašnjenjem biti pozitivno vrednovani, posebno u kontekstu rada u realnom vremenu.
- Očekivani izlaz modela je:
 - CSV fajl koji sadrži predviđene blendshape koeficijente po frejmu, u unapred definisanom redosledu blendshape-ova.

Generisani CSV fajlovi biće učitani u aplikaciju koja automatski generiše video animaciju avatara (pokreti usana i lica). Rezultate iz aplikacije za nekoliko primera učesnici će moći da dobiju do 5 puta u toku takmičenja o čemu će biti informisani putem veb sajta takmičenja.

NAPOMENA: Očekuju se drugačija akustička obeležja iz sintetizovanog govora u odnosu na obeležja koja se mogu izdvojiti iz prirodnog govora. Na učesnicima je da omoguće da ove razlike ne degradiraju kvalitet animacije.

Spisak blendshape-ova na koje je rigovan avatar:

browInnerUp	jawLeft
browDownLeft	jawRight
browDownRight	mouthFunnel
browOuterUpLeft	mouthPucker
browOuterUpRight	mouthLeft
eyeLookUpLeft	mouthRight
eyeLookUpRight	mouthRollUpper
eyeLookDownLeft	mouthRollLower
eyeLookDownRight	mouthShrugUpper
eyeLookInLeft	mouthShrugLower
eyeLookInRight	mouthClose
eyeLookOutLeft	mouthSmileLeft
eyeLookOutRight	mouthSmileRight
eyeBlinkLeft	mouthFrownLeft
eyeBlinkRight	mouthFrownRight
eyeSquintLeft	mouthDimpleLeft
eyeSquintRight	mouthDimpleRight
eyeWideLeft	mouthUpperUpLeft
eyeWideRight	mouthUpperUpRight
cheekPuff	mouthLowerDownLeft
cheekSquintLeft	mouthLowerDownRight
cheekSquintRight	mouthPressLeft
noseSneerLeft	mouthPressRight
noseSneerRight	mouthStretchLeft
jawOpen	mouthStretchRight
jawForward	tongueOut