

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

1. spk08_blendshapes.zip i spk14_blendshapes.zip

- **renamed_spk08/renamed_spk14**
 - .csv fajlovi koji sadrže koeficijente blendshape-ova po frejmu izdvojenih pomoću FaceCap aplikacije
 - Podrazumevani FPS je 60
 - .wav - audio snimci koji odgovaraju .csv fajlovima
 - Snimani simultano sa video-materijalom iz kog su izdvajani blendshape koeficijenti
 - Oko 180 rečenica/snimaka po govorniku (avatar je pravljen po crtama lica spk_08)
- **spk_08_transcripts.xlsx/spk_14_transcripts.xlsx**
 - Sadrže transkripte i redne brojeve snimaka u skladu sa nazivima audio i csv fajlova

2. labels_aligned.zip

- TXT fajlovi sa informacijom o vremenu početka i kraja svakog fonema, izraženom u sekundama

- Poravnanja su rađena automatski i mogu sadržati greške, što treba imati u vidu prilikom treninga i evaluacije modela

3. avatar.zip

Ovaj folder sadrži aplikaciju za testiranje avatara. Aplikacija se otvara jednostavnim dvoklikom na .exe fajl. Nakon što se pojavi lik avatara sa plavom pozadinom, klikom na strelicu u gornjem levom uglu otvara se mogućnost učitavanja fajlova. Učita se .csv fajl sa koeficijentima blendshape-ova, odgovarajući audio snimak, i postavi FPS na ono što je podrazumevano u .csv fajlu. Klikom na dugme *run*, avatar se pomera u skladu sa prosledjenim blendshape koeficijentima uz propratni audio koji se čuje u aplikaciji. U podfolderu *FTNFacialRig* nalazi se .txt fajl u kom je spisak blendshape koeficijenta podržan od strane avatara i u tačnom redosledu kako su i izdvojeni u .csv fajlovima. Ove .txt fajlove ne treba menjati. Ne treba ništa menjati u podfolderima da ne biste promenili funkcionisanje avatara.

4. audio_synth.zip

Ovaj folder sadrži nekoliko audio fajlova koji po tekstualnom sadržaju odgovaraju nekim od originalnih audio fajlova, ali su sintetizovani pomoću TTS. Ovakvi audio fajlovi biće dati i u test fazi samo sa drugačijim rečenicama.

Jasna definicija ulaza i izlaza modela

- *Ulaz u model nije striktno definisan.*
U test skupu biće obezbeđeni:
 - tekst koji avatar treba da izgovori i
 - audio signal (dobijen sintezom govora na osnovu tog teksta ili prirodni govor),a 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.
- Učesnici su slobodni da snime ili pronađu dodatni materijal za obuku ukoliko smatraju neophodnim, ali i da jasno navedu ako su koristili dodatni materijal i koji.
- *Očekivani izlaz modela je:*
 - CSV fajl koji sadrži predviđene blendshape koeficijente po frejmu, u unapred definisanom formatu i redosledu blendshape-ova (kao .csv fajlovi koji su dati za obuku).
 - Uz rezultate (CSV fajlove), takmičari predaju i fajl meta.json koji sadrži sledeća polja (po test fajlu):
 - inference_time_sec (po fajlu)
 - rtf (po fajlu)
 - lookahead_ms (jedna vrednost za sistem)
 - fps_out (FPS izlaza u CSV-u)

Organizator zadržava pravo da ponovo izmeri *inference_time_sec/RTF* na referentnom Colab okruženju radi fer poređenja.

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.

Važno je naglasiti da su učesnici slobodni da primene sopstveni preprocessing i filtriranje ili mapiranje postojećih koeficijenata. Cilj je postići što prirodniju animaciju avatara, a ne nužno reprodukovati originalne koeficijente jedan-na-jedan.

Napomena o kriterijumu brzine i kašnjenja (latency)

U ocenjivanju će se, pored kvaliteta animacije, uzimati u obzir i **brzina inferencije** i **algoritamsko kašnjenje** sistema.

1) Inference time / RTF (meren na Google Colab GPU)

Takmičari su dužni da obezbede skriptu ili notebook za inferenciju koji, za svaki test fajl, generiše izlazni CSV i ispisuje vreme potrebno za generisanje.

Za svaki test primer (audio + tekst) potrebno je prijaviti:

- **inference_time_sec**: vreme inferencije u sekundama, tj. vreme potrebno da model generiše sve blendshape koeficijente za ceo fajl (bez vremena koje je potrebno aplikaciji da učitava CSV i pusti animaciju).
- **RTF (Real-Time Factor)**: odnos vremena inferencije i trajanja audio signala:

$$RTF = \frac{\text{inference_time_sec}}{\text{audio_duration_sec}}$$

Napomena:

- Merenje se radi na **Google Colab-u sa uključenim GPU-om** (okruženje definiše organizator).
- U *inference_time_sec* ulazi samo vreme rada modela + priprema ulaza i postproces koji je nužan da se dobije CSV (npr. ekstrakcija feature-a ako je deo pipeline-a).
- Ne ulazi vreme učitavanja CSV u aplikaciju za animaciju, renderovanja videa, eksportovanja, itd.

2) Lookahead (algoritamsko kašnjenje) u milisekundama

Takmičari moraju prijaviti **lookahead_ms**: maksimalnu količinu “buduće” informacije (desni kontekst) koju sistem koristi da bi izračunao koeficijente za trenutni frejm.

- Ako je sistem **strogo kauzalan** (koristi samo prošlost i sadašnjost): *lookahead_ms* = 0.
- Ako koristi klizni prozor ili desni kontekst: *lookahead_ms* je trajanje tog desnog konteksta izraženo u ms.
- Ako model zahteva ceo signal unapred (npr. obrada cele rečenice pre generisanja): navesti realnu vrednost (praktično veliko kašnjenje u rangu trajanja fajla).

Za kraj

Generisani CSV fajlovi biće učitani u aplikaciju koja automatski generiše video animaciju avatara (pokreti usana i lica) – ista aplikacija koja je data učesnicima.

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

Spisak blendshape-ova na koje je rigovan avatar:

browInnerUp
browDownLeft
browDownRight
browOuterUpLeft
browOuterUpRight
eyeLookUpLeft
eyeLookUpRight
eyeLookDownLeft
eyeLookDownRight
eyeLookInLeft
eyeLookInRight
eyeLookOutLeft
eyeLookOutRight
eyeBlinkLeft
eyeBlinkRight
eyeSquintLeft
eyeSquintRight
eyeWideLeft
eyeWideRight
cheekPuff
cheekSquintLeft
cheekSquintRight
noseSneerLeft
noseSneerRight
jawOpen
jawForward

jawLeft
jawRight
mouthFunnel
mouthPucker
mouthLeft
mouthRight
mouthRollUpper
mouthRollLower
mouthShrugUpper
mouthShrugLower
mouthClose
mouthSmileLeft
mouthSmileRight
mouthFrownLeft
mouthFrownRight
mouthDimpleLeft
mouthDimpleRight
mouthUpperUpLeft
mouthUpperUpRight
mouthLowerDownLeft
mouthLowerDownRight
mouthPressLeft
mouthPressRight
mouthStretchLeft
mouthStretchRight
tongueOut