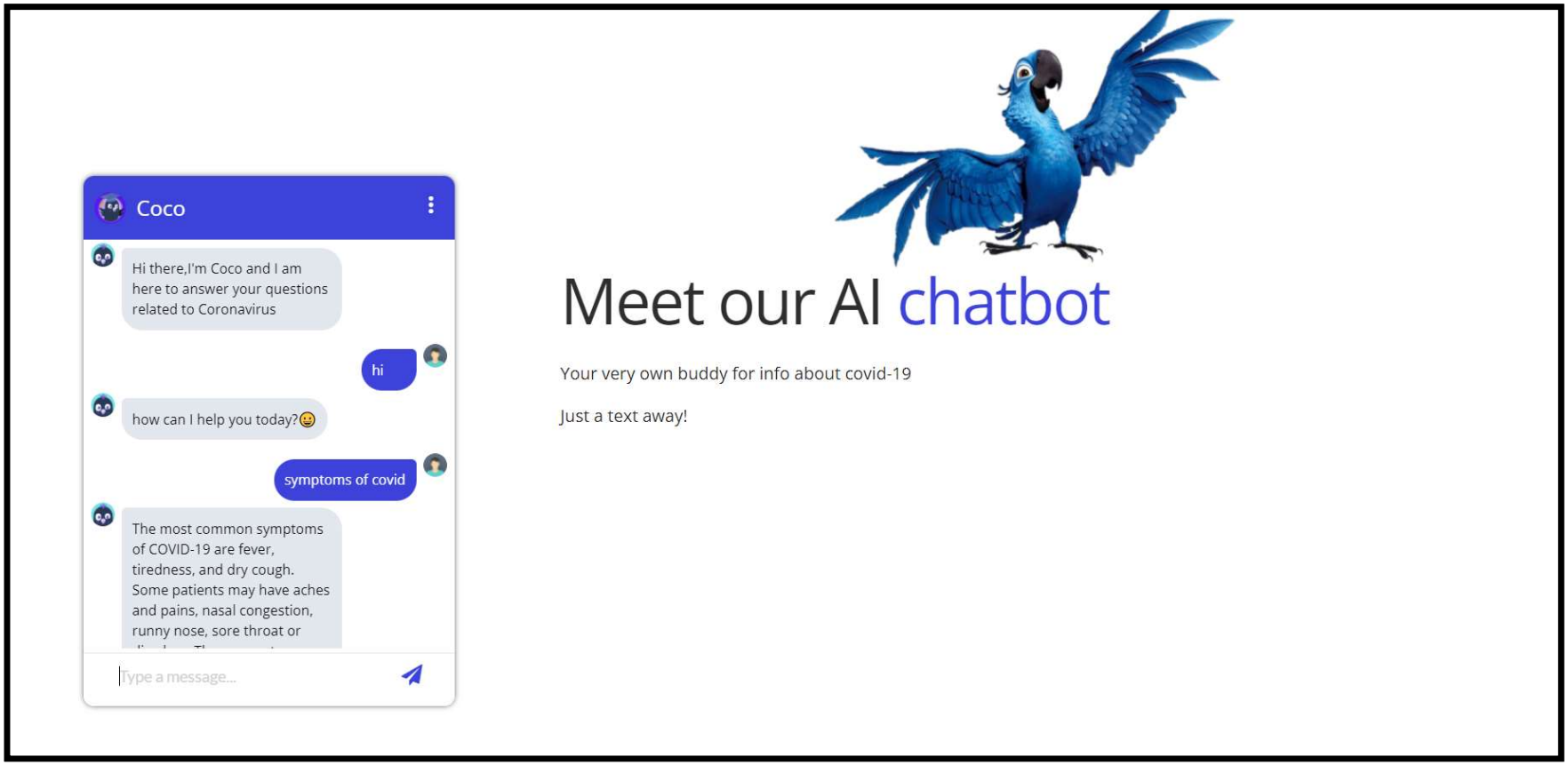
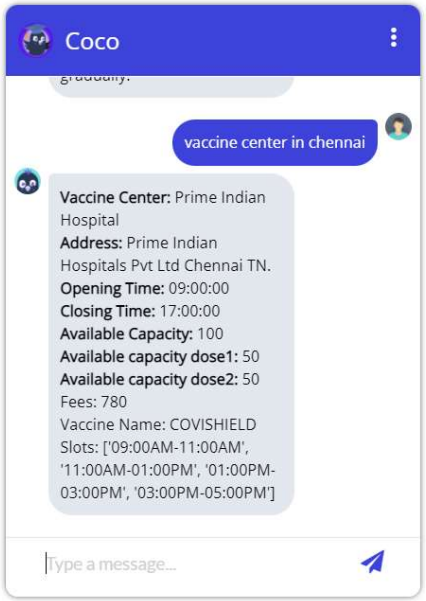
THUNDER BIRDS

Topic 2: ChatBot with Customizable Voice

Menaga M
Renuga Devi V
Krithika A
Nivethidha R V

2.1. Rasa Conversational AI chatbot

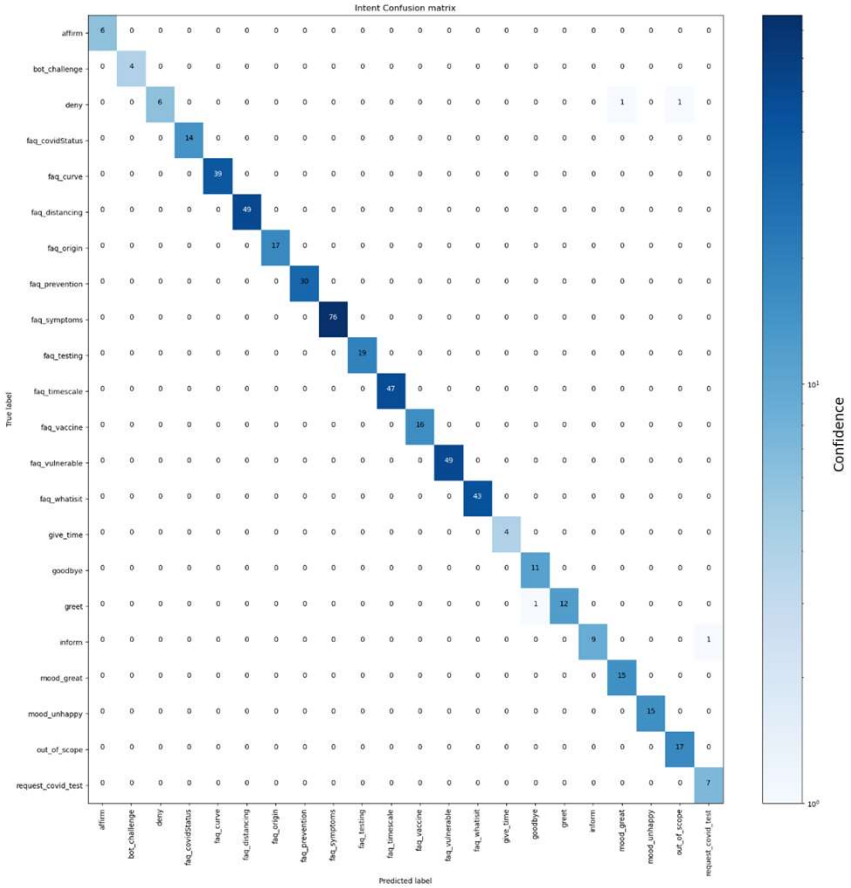


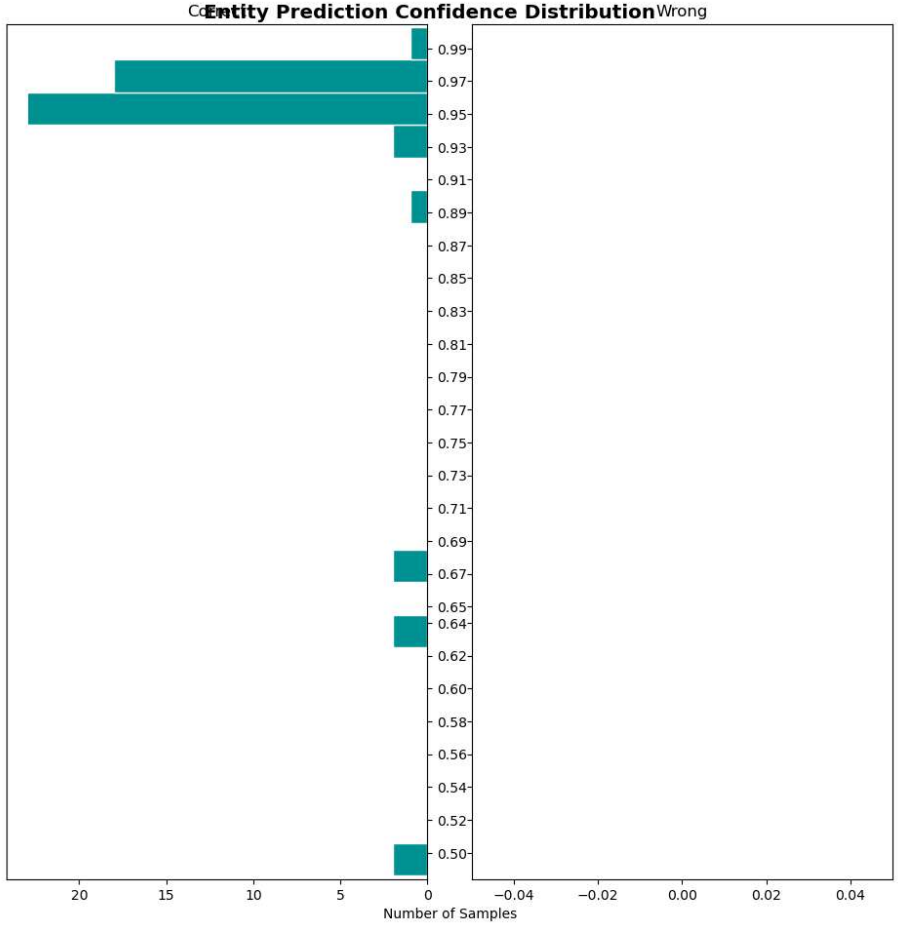


Meet our AI chatbot

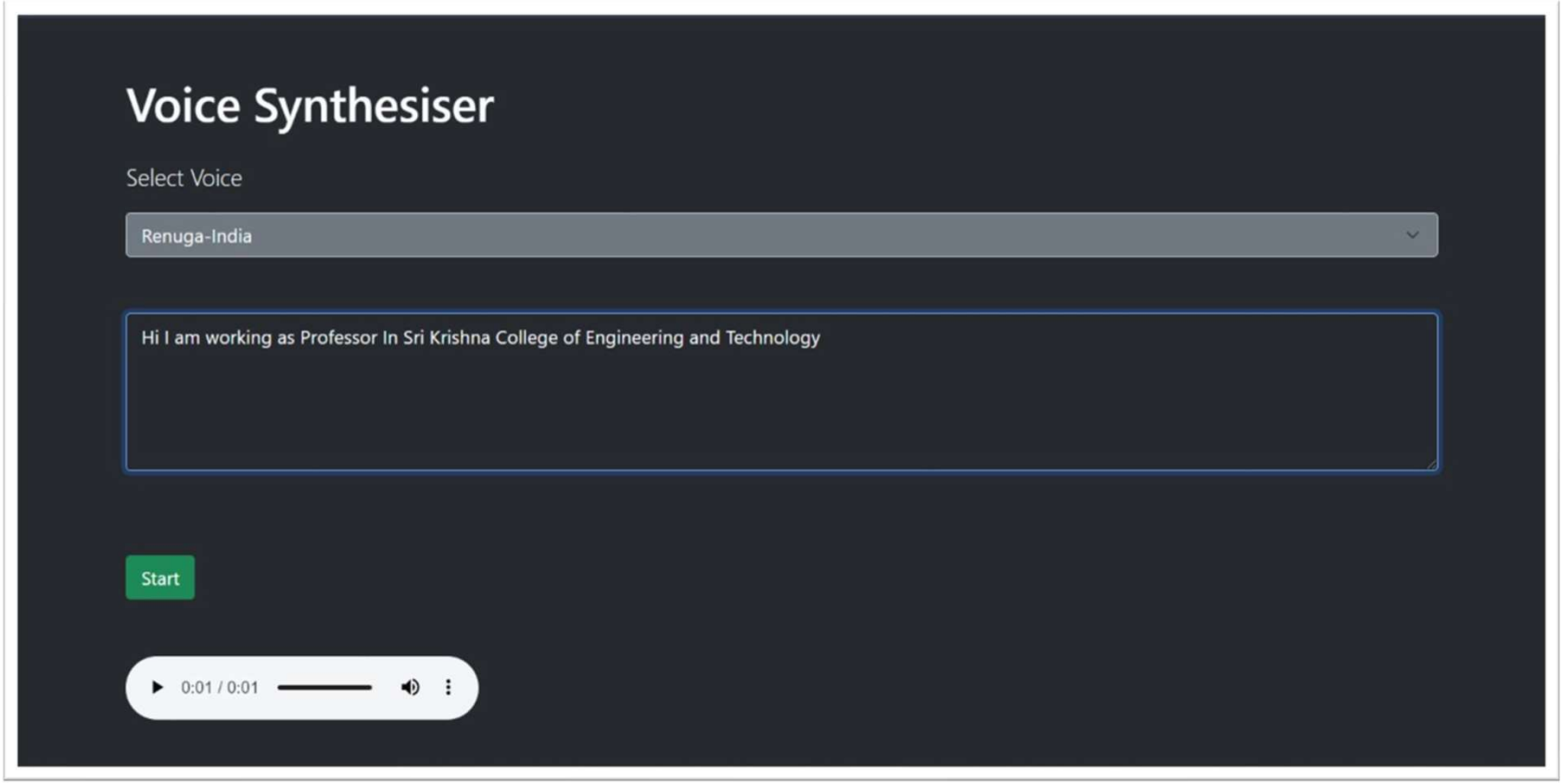
Your very own buddy for info about covid-19

Just a text away!

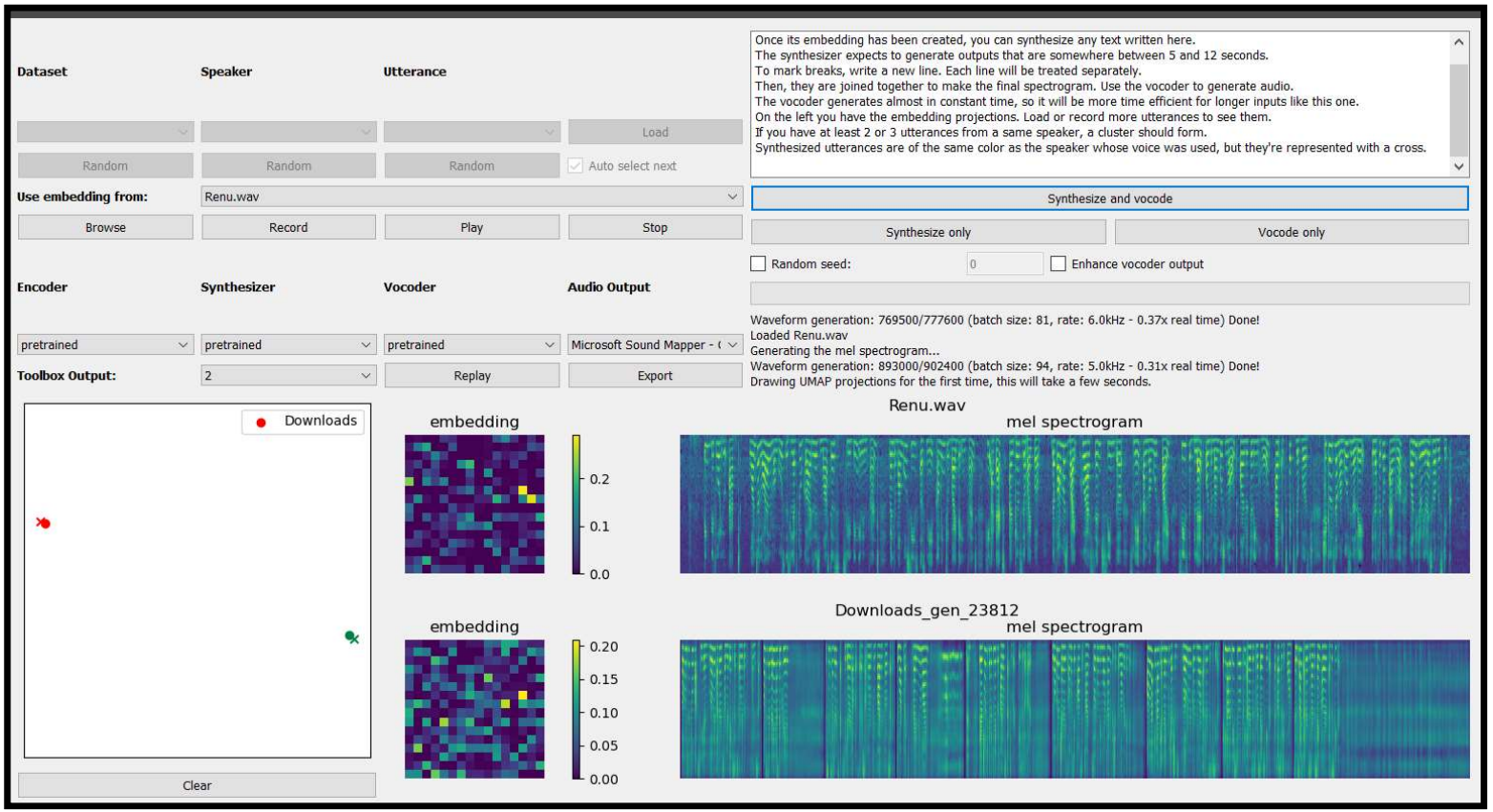

Intent Confusion matrix
affirm
bot_challenge
deny
faq_covidStatus
faq_curve
faq_distancing
faq_origin
faq_prevention
faq_symptoms
faq_testing
faq_timescale
faq_vaccine
faq_vulnerable
faq_whatisit
give_time
goodbye
greet
inform
mood_great
mood_unhappy
out_of_scope
request_covid_test
True label
Predicted label
Confidence


Entity Prediction Confidence Distribution
Wrong
Number of Samples

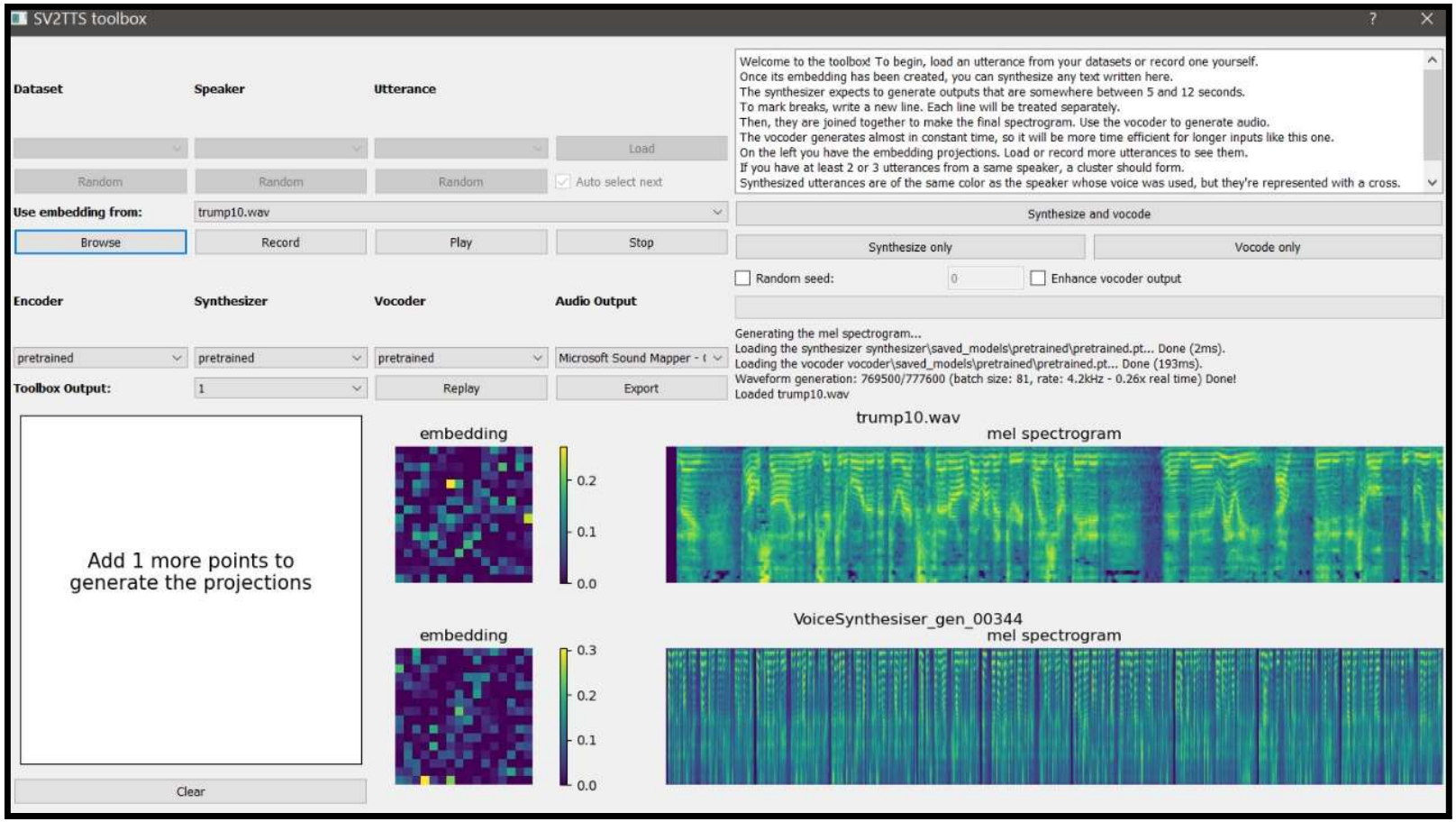
2.2. Synthetic Voice generation model



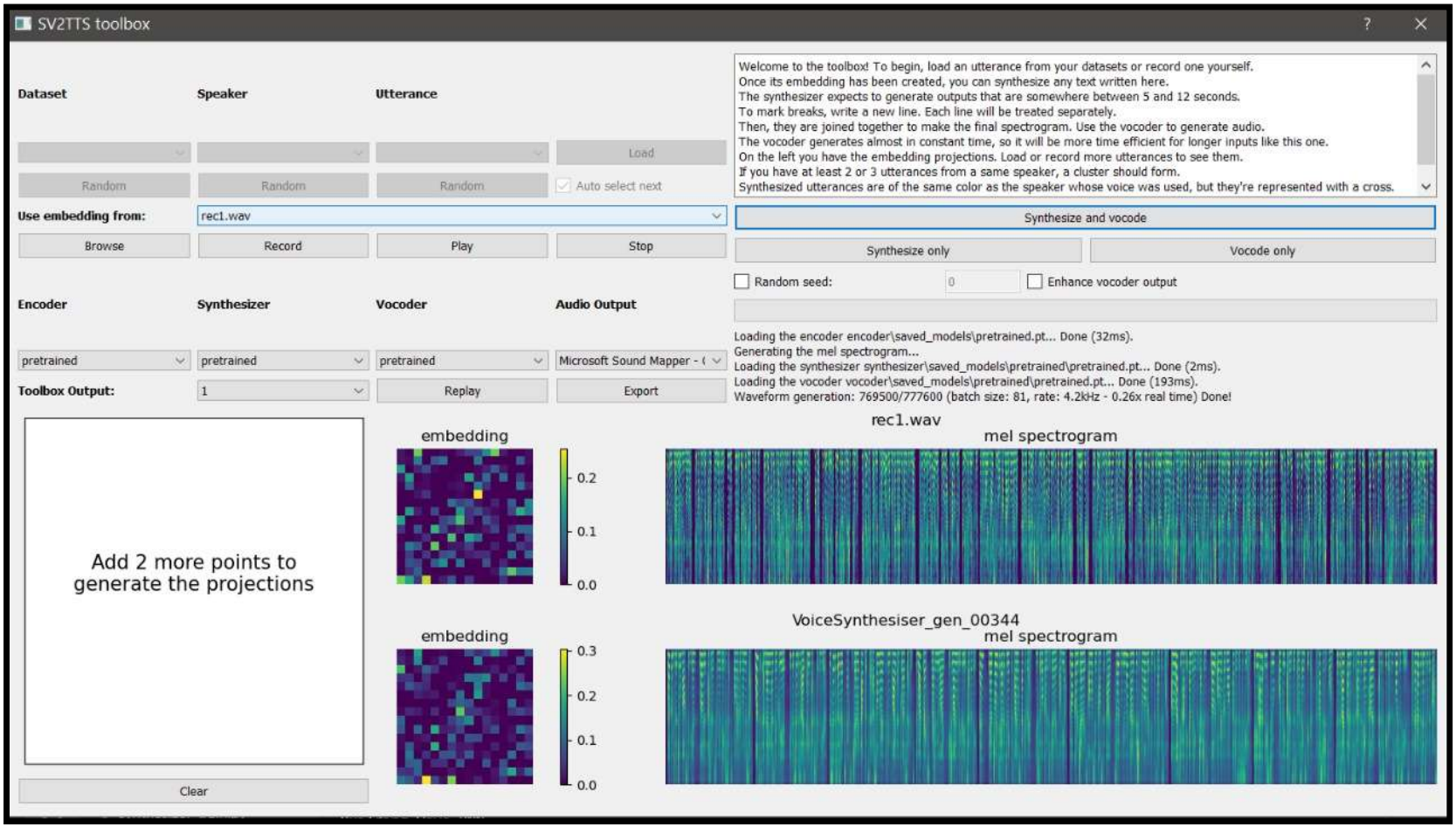
VOICE 1


Dataset
Speaker
Utterance
Load
Random
Random
Random
Auto select next
Use embedding from:
Renu.wav
Browse
Record
Play
Stop
Encoder
Synthesizer
Vocoder
Audio Output
pretrained
pretrained
pretrained
Microsoft Sound Mapper - (
Toolbox Output:
2
Replay
Export
Once its embedding has been created, you can synthesize any text written here.
The synthesizer expects to generate outputs that are somewhere between 5 and 12 seconds.
To mark breaks, write a new line. Each line will be treated separately.
Then, they are joined together to make the final spectrogram. Use the vocoder to generate audio.
The vocoder generates almost in constant time, so it will be more time efficient for longer inputs like this one.
On the left you have the embedding projections. Load or record more utterances to see them.
If you have at least 2 or 3 utterances from a same speaker, a cluster should form.
Synthesized utterances are of the same color as the speaker whose voice was used, but they're represented with a cross.
Synthesize and vocode
Synthesize only
Vocode only
Random seed:
0
Enhance vocoder output
Waveform generation: 769500/777600 (batch size: 81, rate: 6.0kHz - 0.37x real time) Done!
Loaded Renu.wav
Generating the mel spectrogram...
Waveform generation: 893000/902400 (batch size: 94, rate: 5.0kHz - 0.31x real time) Done!
Drawing UMAP projections for the first time, this will take a few seconds.
Downloads
Clear
embedding
0.2
0.1
0.0
Renu.wav
mel spectrogram
embedding
0.20
0.15
0.10
0.05
0.00
Downloads_gen_23812
mel spectrogram

VOICE 2


SV2TTS toolbox
Dataset
Speaker
Utterance
Load
Random
Random
Random
Auto select next
Use embedding from:
trump10.wav
Browse
Record
Play
Stop
Encoder
Synthesizer
Vocoder
Audio Output
pretrained
pretrained
pretrained
Microsoft Sound Mapper - (
Toolbox Output:
1
Replay
Export
Welcome to the toolbox! To begin, load an utterance from your datasets or record one yourself.
Once its embedding has been created, you can synthesize any text written here.
The synthesizer expects to generate outputs that are somewhere between 5 and 12 seconds.
To mark breaks, write a new line. Each line will be treated separately.
Then, they are joined together to make the final spectrogram. Use the vocoder to generate audio.
The vocoder generates almost in constant time, so it will be more time efficient for longer inputs like this one.
On the left you have the embedding projections. Load or record more utterances to see them.
If you have at least 2 or 3 utterances from a same speaker, a cluster should form.
Synthesized utterances are of the same color as the speaker whose voice was used, but they're represented with a cross.
Synthesize and vocode
Synthesize only
Vocode only
Random seed:
0
Enhance vocoder output
Generating the mel spectrogram...
Loading the synthesizer synthesizer\saved_models\pretrained\pretrained.pt... Done (2ms).
Loading the vocoder vocoder\saved_models\pretrained\pretrained.pt... Done (193ms).
Waveform generation: 769500/777600 (batch size: 81, rate: 4.2kHz - 0.26x real time) Done!
Loaded trump10.wav
Add 1 more points to generate the projections
Clear
embedding
0.2
0.1
0.0
trump10.wav
mel spectrogram
embedding
0.3
0.2
0.1
0.0
VoiceSynthesiser_gen_00344
mel spectrogram

VOICE 3


SV2TTS toolbox
Dataset
Speaker
Utterance
Load
Random
Random
Random
Auto select next
Use embedding from:
rec1.wav
Browse
Record
Play
Stop
Welcome to the toolbox! To begin, load an utterance from your datasets or record one yourself.
Once its embedding has been created, you can synthesize any text written here.
The synthesizer expects to generate outputs that are somewhere between 5 and 12 seconds.
To mark breaks, write a new line. Each line will be treated separately.
Then, they are joined together to make the final spectrogram. Use the vocoder to generate audio.
The vocoder generates almost in constant time, so it will be more time efficient for longer inputs like this one.
On the left you have the embedding projections. Load or record more utterances to see them.
If you have at least 2 or 3 utterances from a same speaker, a cluster should form.
Synthesized utterances are of the same color as the speaker whose voice was used, but they're represented with a cross.
Synthesize and vocode
Synthesize only
Vocode only
Random seed:
0
Enhance vocoder output
Encoder
Synthesizer
Vocoder
Audio Output
pretrained
pretrained
pretrained
Microsoft Sound Mapper - (
Toolbox Output:
1
Replay
Export
Loading the encoder encoder\saved_models\pretrained.pt... Done (32ms).
Generating the mel spectrogram...
Loading the synthesizer synthesizer\saved_models\pretrained\pretrained.pt... Done (2ms).
Loading the vocoder vocoder\saved_models\pretrained\pretrained.pt... Done (193ms).
Waveform generation: 769500/777600 (batch size: 81, rate: 4.2kHz - 0.26x real time) Done!
Add 2 more points to generate the projections
Clear
embedding
0.2
0.1
0.0
rec1.wav
mel spectrogram
embedding
0.3
0.2
0.1
0.0
VoiceSynthesiser_gen_00344
mel spectrogram