

I performed testing of OSM demonstration by running the code in demonstration.ipynb and following their instructions:

1. I downloaded the pre-trained models and put them into corresponding directories
2. Next, I used default .wav and .txt file with the config provided and launched the inference
3. Finally, I listened to the generated .wav file

The demonstration presented by the team reproduces stated baseline results. Namely, the phrase “Hello, world! My favourite game is playing football. I live in Moscow. I like watching films.” generated by the network is clearly distinguishable although it lacks an intonation and is somewhat noisy.

I assume that the project is in its final stage of completion as all the required components for speech generation are present. Though it is hard to evaluate how easy it will be to introduce new methods to the project without spending some time working with it.