First peer review of the OSM: One-Shot Multi-speaker project.
Reviewed by MMDF team.
23.04.2021

1. Are project objectives clearly explained?
If not write what was not clear. Provide your opinion on the problem.

The general problem statement is well-explained, the team is planning to provide a framework that would transform text to speech using a small sample of voice and make it more flexible and user-friendly. In my opinion the goal of making a convenient API is great in itself and especially valuable for such a demanding topic as speech generation.

2. *Is the proposed baseline solution relevant to the problem? Could you recommend something to the team?*

The baseline solution is implementation of the paper “Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis”. This is a good start since this solution contains a whole implemented pipeline of text to voice generation, namely Speaker Encoder, Synthesizer and Vocoder.
Before diving into the project I would recommend to explore more solutions in this area. If there are no other solutions that combine all three stages, that probably it will be helpful to search for each stage separately. This will expand the horizon of understanding how each part can be implemented in terms of API.

3. Is it clear from the report how the team is going to test / evaluate the results?
(not only metrics but code and bugs). Could you recommend something here?

The team is going to provide API, therefore, it will be possible to evaluate their solution from the point of view of usability by an external user. I advise the team to pay attention to the solutions and libraries that were mentioned in the course lectures. This can be useful from the point of view of code development and from the point of view of reference for user-friendly interfaces.

4. Is it possible to guess next project development steps from the report and github repository?
What a team is going to do next, provide with 3 next possible steps?
e.g.
- code baseline solution
- run tests and evaluate
- try several variants to improve baseline solution

This is not clearly spelled out in the report, but presumably the team will primarily focus on improving the existing baseline framework. Then they will add customization to the existing pipeline and finally implement various speaker encoders. It is a good idea to write exact steps and milestones for this project.

5. Would you recommend to improve the report and how? Or it is all good.

From the report it is not completely clear how the team plans to use the original implementation. Will it be improved in terms of flexibility or it have to be rewritten? How mentioned new Speaker Encoders will affect the baseline solution? I think a couple sentences on this topic may improve the report.

6. Is project github repository easy to follow?

So far the repository is well organized. Links and images in README are helpful and good looking. Baseline solution is yet not forked from the original repo, so probably the team will add necessary files manually.