**CENG 796 - Peer-review form**

**Reviewed project ID:** Group **01**

**Reviewed project's title (title of the paper): It’s Raw! Audio Generation with State-Space Models**

**Reviewer name(s)**: Ege Demir, Alparslan Sertel

Instructions:

* Answer = *Yes*, *No* or *Partial*.
* You may expand sections as necessary.
* For most questions, you do not need to add comments, unless the instructions tell you otherwise.
* "Notebook" refers to "Jupyter Notebook" file that is expected to be named as main.ipynb

|  |  |  |
| --- | --- | --- |
| **Question** | **Answer** | **Comments** |
| Contains a jupyter notebook file | Yes |  |
| Notebook is located at <project\_root>/main.ipynb | Yes |  |
| Notebook's first section contains paper information (paper title, paper authors, and project group members' name & contact information)  Some good examples: see group03, group10, group11 (and a couple of other groups). | Partial | Contact information is in README.md, but not in the notebook |
| Notebook contains a section for hyper-parameters of the model. | Yes |  |
| Notebook contains a section for training & saving the model. | Yes |  |
| Notebook contains a section (or a few sections) for loading a pre-trained model & computing qualitative samples/outputs. | Yes |  |
| Notebook contains reproduced plots and/or tables, as declared. | Yes |  |
| Notebook contains pre-computed outputs. | Yes |  |
| Data is included and/or a proper download script is provided. | Yes |  |
| Notebook contains a section describing the difficulties encountered. | Yes |  |
| The paper has achieved its goals and/or explained what is missing. | Yes | The reproduction of the goals has failed but the simple overfits seem convincing and the problem seems to be inadequate computing resources. |
| The notebook contains a section that reproduces the figure(s) and table(s) declared in the goals. | Yes | Although the results are far off. |
| The notebook also reports the original values of the targeted quantitative results, for comparison. | No |  |
| MIT License is included. | Yes |  |
| As the reviewer(s), you have read the paper & understood it. | Yes |  |
| Implementation of the model seems correct. | Yes | Fig 1 S4 Block is straightforward and seems correct- <https://github.com/gcinbis/DeepGenerativeModels-2023-Spring-Projects-Version1/blob/2286a4bd9b933e3466c3ff78d57281fbcab8d1a2/group01_version1/sashimi.py#L6>  <https://github.com/gcinbis/DeepGenerativeModels-2023-Spring-Projects-Version1/blob/main/group01_version1/S4/layer.py>  Since we are not familiar with state space models we cannot make healthy comments on the implementation of the actual S4 unit in the S4 folder, however we were able to observe the details reported in A.1 and 4.1 which describe how the S4 here is different from the original one. Also their implementation is consistent with their equations in the notebook.  There is no explanation given for the line: “p = (p / 255.0) \* 2.0 - 1.0” in sampling (cell 16 of main.ipynb) but we guess it has got to do with the audio domain. Other than this the autoregressive sampling seems correct.  The training loop is standard and seems correct.  Their hyperparameters are correct.  They train the timescale (delta, dt) although the paper recommends against this because of instability but the code authors rightfully say that they are not given initial values. Maybe they could initialize these by training for once and then retraining the rest of the network after freezing dt. The change of gradients of dt could also shine a light on instability. |
| Notebook looks professional (in terms of notation, readability, etc.) | Yes | We would like to praise our classmates on this. |
| Source code looks professional (in terms of coding style, comments, etc.) | Yes | We would like to praise our classmates on this. |

**Additional comments:**

|  |
| --- |
| The training seems unstable for the toy models and the actual models. Maybe longer training, higher context size and lower learning rate could help. The number of epochs they train for is ~1/15 of the paper.  The losses are reported manually, and no reason is given.  Overall we would like to congratulate the code authors for their excellent code and comments. We would recommend them to work on training rather than the model which seems correct. |