# Report on Fine tuning on Tacotron

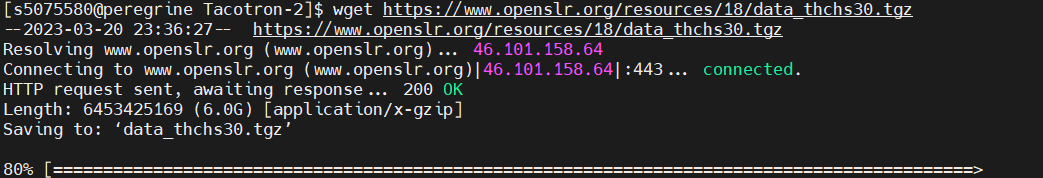
## Zhengkun mei

## Dataset

## 1.1 Data preparation

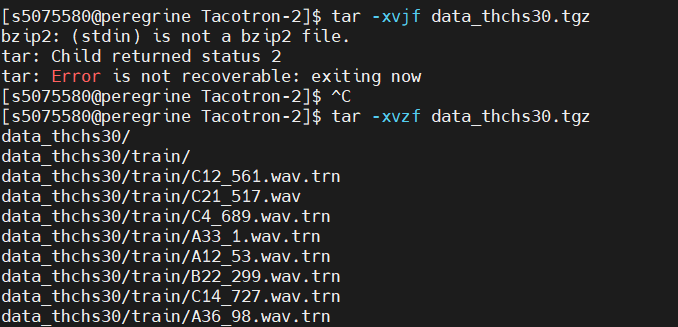
First I deleted the LJSpeech dataset to ensure that the steps of the LJSpeech dataset will not be repeated during the preprocessing and training phases.

At the first place, I considered to fine-tune the model based on the thchs30 dataset, which is a Chinese corpus dataset and which I am more familiar with. But in the end I changed my mind and turned to Russian corpus based on the fact that: 1. Russain corpus is far less in size compared with thchs30 dataset, it is good for preprocessing and training. 2. Vass promise me that she can evalutate the output and I also want to try a language I don’t know.

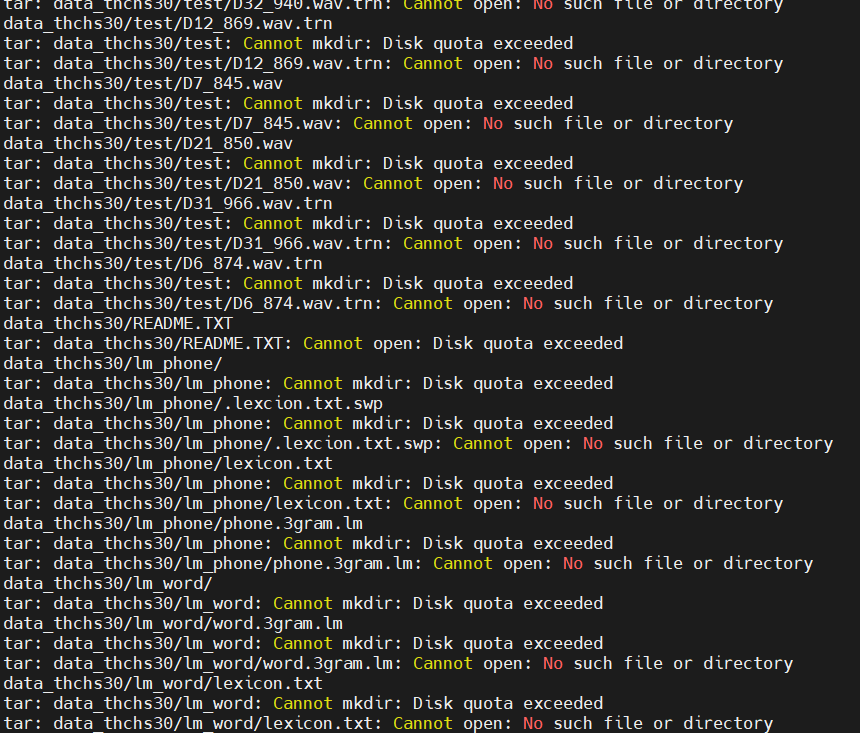


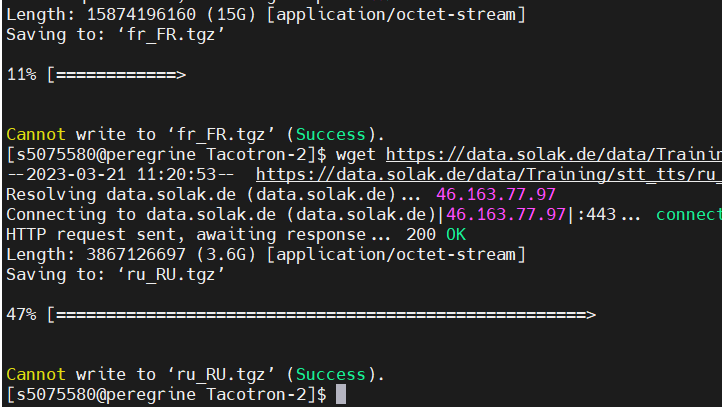
Although I finally abandon the plan to fine-tune with thchs30 dataset, there are still two interesting points I want to mention in the process.

The first thing to do is unzip the command. At first I still used the -xvjf to unzip the dataset, but there is an error, so I check what is the meaning for -xvjf. Both "xvjf" and "xvzf" are options of the tar command to decompress files, and they are used for different types of compressed files. "xvjf" is used to decompress files ending in .tar.bz2, where: ".tar" means a file packaged into tar format; ".bz2" indicates a file compressed using the bzip2 algorithm. "xvzf" is used to decompress files ending in .tar.gz, where: ".tar" means a file packaged into tar format; ".gz" indicates a file compressed using the gzip algorithm. So I changed it to the -xvzf

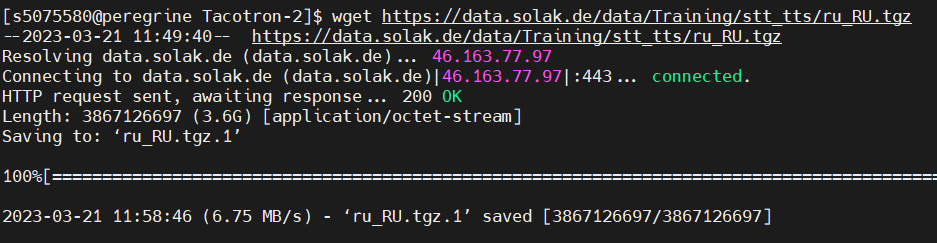


The second thing I feel like I'm a bit stupid. I found that I have been using the tacotron-2 model in the home2 folder, so that when I unzip the data set, I found that there is not enough space, so I realized the mistake and switched to the data folder to start all the steps again.



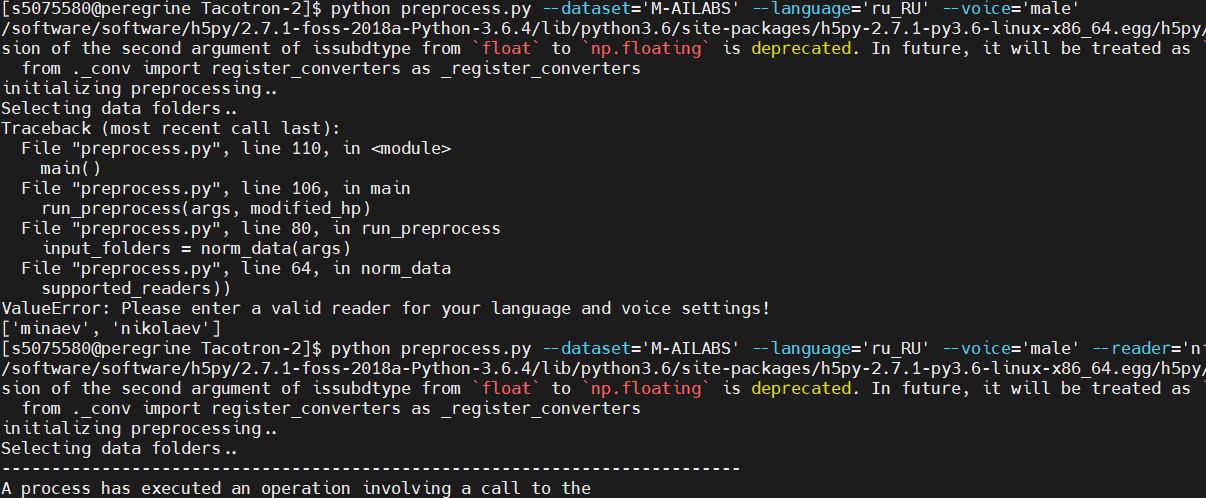
and the same problem also appears in the other datasets, for example: 

But in the end I got the dataset



## preprocessing

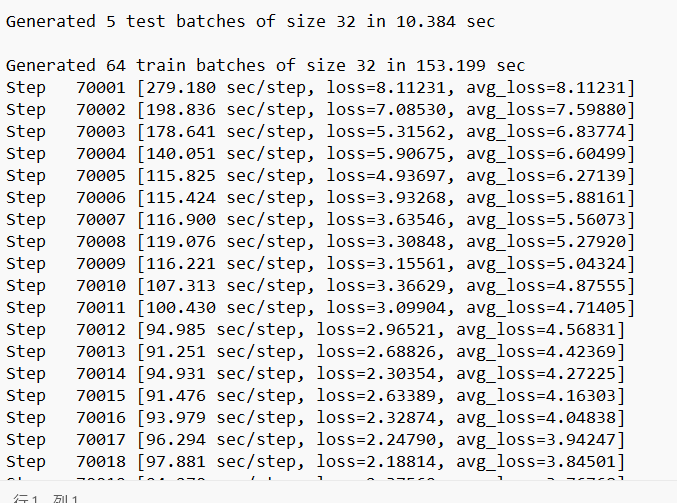
There is only one thing I found that should be mentioned, is that I need to use flag to point certain parameters. It can be better explained by my screenshot:



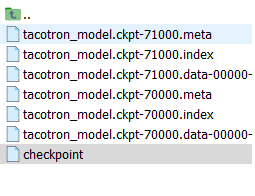
## Train

## 2.1 General process

Put the trained 70000steps model into the tacotron taco-pretrain folder, and modify the number of checkpoint steps accordingly. In the training.sh file, change the model type to tacotron, and modify the step to 71000, so that the training On the basis of the original 70000 steps model, 1000 steps of training will be carried out based on the Russian dataset. Normally, after the training is completed, the model will also become a 71000 steps model.



As can be seen from the screenshot, the model starts training from step 70001, which also confirms what I just said. And when I completed the training process, it had file like this:



The checkpoint also changed accordingly.

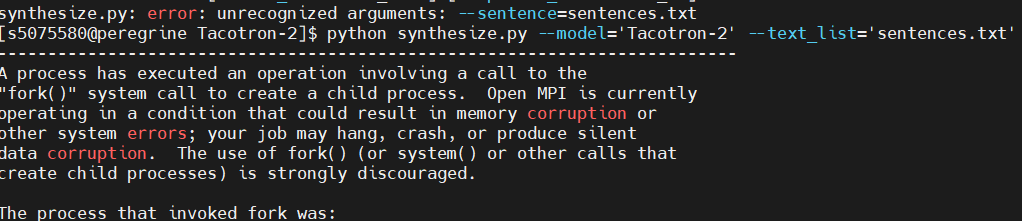
## Mistakes in training

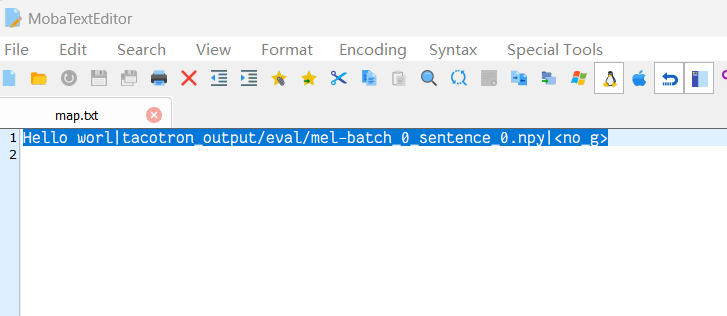
Firstly, still the problem of space, since I originally used home2 file as the working space,

Second, since my previous work is based on tacotron-2, so I firstly uploaded model in tacotron-2 pretrain file and also used the tacotron-2 to train, but it is wrong. For the provided models were trained based on tacotron and griffin-lim, so I guess when scanning the tacotron-2 files, it found the wavenet file also, which should not be included in tacotron model. So I followed the suggestion of frank, submit training.sh in one step for creating the complete structure of tacotron, and repeated the training process based on tacotron and in the end solved this mistake.

## Synthesizing

1. In the process of synthesis, I found an interesting thing. When I want to synthesize sentences different from those in sentences.txt, I modify the content at the beginning, but the synthesized ones are still the original ones in sentences.txt. content, so I thought, the txt file should be specified for synthesis during the synthesis process. Or directly modify the text content in hparam.





Although I have synthesized some voices, I can't judge whether these voices are of higher quality or have better quality than the model before fine-tuning. I asked Vass for help and hoped that she would evaluate these voices, but Vass is very busy recently, and I will have time after April, and I have not used the model before fine-tuning to synthesize Russian speech, so there is no comparison experiment, and I have no idea what the evaluation criteria should be, so in this In this report, these contents will not be improved. I plan to try my best to synthesize and compare voices and formulate evaluation standards in April, and complete the entire content of this report with the help of Vass.