

# Final Plan GSoC'21

By Vasudev Gupta

<b>Details</b>	<b>1</b>
Before Phase-1	1
After Phase-1 Evaluation	1
<b>Timeline</b>	<b>2</b>

## Details

### Before Phase-1

Currently, Wave2Vec2 is open-sourced only in PyTorch. I will be implementing this model in TensorFlow 2. Since Wave2Vec2 is a pre-trained model, I will be converting its pre-trained checkpoints into TensorFlow compatible format using mapping recipes (from this [blog post](#)). This will allow us to get the pre-trained TensorFlow checkpoints without any costly pre-training & obtained checkpoints will give similar performance when compared to official wave2vec2.

I will then try to reproduce the results mentioned in the paper by fine-tuning the TensorFlow pre-trained model on the [TIMIT](#) & [LibriSpeech](#) dataset. I will add all the fine-tuned TensorFlow SavedModel to [TensorFlow Hub](#) & will add code-dependent checkpoints to my [HuggingFace Hub](#). For fine-tuning, I will be relying on GCP credits while for obtaining pre-trained checkpoints, no compute is required.

Finally, I will work on a blog post that will include a discussion on the paper and my learnings during the project.

### After Phase-1 Evaluation

I will work on converting fine-tuned Wave2Vec2 SavedModel to TensorFlow Lite and will work on a blog post focusing on TensorFlow Lite & my conversion.

## Timeline

Timeline	Work description
May 28 - 12 June	<ul style="list-style-type: none"><li>• Paper reading &amp; finish Wave2Vec2 coding in <a href="#">this</a> GitHub repository.</li></ul>
June 13 - June 21	<ul style="list-style-type: none"><li>• Finish conversion script to convert PyTorch pre-trained checkpoints to TensorFlow format.</li><li>• Finish training script, data pipelines for fine-tuning the pre-trained model.</li><li>• Setup GCP instance &amp; GCS bucket for storing data and checkpoints.</li></ul>
June 22 - June 25	<ul style="list-style-type: none"><li>• Time for code review.</li><li>• Make changes as per suggestions.</li></ul>
June 26 - July 5	<ul style="list-style-type: none"><li>• Fine-tune Wave2Vec2 on TIMIT &amp; LibriSpeech dataset and inspect the training.</li></ul>
July 6 - July 10	<ul style="list-style-type: none"><li>• Prepare Colab notebooks showing demos of using fine-tuned models.</li><li>• Time for notebooks review.</li><li>• Export code-dependent checkpoints to HuggingFace Hub.</li></ul>
July 11 - July 13	<ul style="list-style-type: none"><li>• Export TensorFlow SavedModel to TensorFlow Hub &amp; include Colab notebooks while publishing model to Hub.</li></ul>
<b>Phase-1 Evaluation</b>	
July 14 - July 28	<ul style="list-style-type: none"><li>• Prepare a Blog post explaining Wave2Vec2 paper and my learnings</li><li>• Work on TensorFlow SavedModel conversion to TensorFlow Lite</li></ul>
July 29 - Aug 6	<ul style="list-style-type: none"><li>• Work on a blog post focussing Wave2Vec2 conversion to TensorFlow Lite.</li></ul>
Aug 7 - Aug 23	<ul style="list-style-type: none"><li>• Final submission + Buffer time</li></ul>

*Buffer time is kept for compensating for an unpredictable delay during the project.*