**Speech to Text in Python with Deep Learning**

## Revolutionary Model Wav2Vec

Here I use the  [Wave2Vec](https://ai.facebook.com/blog/wav2vec-20-learning-the-structure-of-speech-from-raw-audio/) — a state-of-the-art speech recognition approach by Facebook.

My standalone Machine learning and deep learning project which I initialed the idea during pursuing the online course from IBM, Google, Sololearn etc. Now a day there is no interface for our documentation, and note preparation - key on word processors, emails, social media’s. So there is huge time wastage for typing the document to a file .Its time consuming and take more error on the document by typing. In this AI world we don’t have a documentation tools to many purposes.

The Project was done at the time of pursuing the online courses .The Courses provides a step towards the path of Machine learning and deep learning in great extent and promote me to attend the goal in knowledge acquiring and execute the idea in workable form.

There are thousands of languages spoken around the world, many with several different dialects, which presents a huge challenge for building high-quality speech recognition technology. It’s simply not feasible to obtain resources for each dialect and every language across the many possible domains (read speech, telephone speech, etc.). Our new model, wav2vec 2.0 , uses self-supervision to push the boundaries by learning from unlabeled training data to enable speech recognition systems for many more languages, dialects, and domains. With just one hour of labeled training data, wav2vec 2.0 outperforms the previous state of the art on the 100-hour subset of the LibriSpeech benchmark — using 100 times less labeled data.

The Wav2Vec2 model was proposed in [wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations](https://arxiv.org/abs/2006.11477) by Alexei Baevski.

*We show for the first time that learning powerful representations from speech audio alone followed by fine-tuning on transcribed speech can outperform the best semi-supervised methods while being conceptually simpler. wav2vec 2.0 masks the speech input in the latent space and solves a contrastive task defined over a quantization of the latent representations which are jointly learned. Experiments using all labeled data of Librispeech achieve 1.8/3.3 WER on the clean/other test sets. When lowering the amount of labeled data to one hour, wav2vec 2.0 outperforms the previous state of the art on the 100 hour subset while using 100 times less labeled data. Using just ten minutes of labeled data and pre-training on 53k hours of unlabeled data still achieves 4.8/8.2 WER. This demonstrates the feasibility of speech recognition with limited amounts of labeled data.*

This model was contributed by [patrickvonplaten](https://huggingface.co/patrickvonplaten).

**Usage tips**

* Wav2Vec2 is a speech model that accepts a float array corresponding to the raw waveform of the speech signal.
* Wav2Vec2 model was trained using connectionist temporal classification (CTC) so the model output has to be decoded using [Wav2Vec2CTCTokenizer](https://huggingface.co/docs/transformers/v4.38.2/en/model_doc/wav2vec2#transformers.Wav2Vec2CTCTokenizer).

A list of official Hugging Face and community (indicated by 🌎) resources to help you get started with Wav2Vec2. If you’re interested in submitting a resource to be included here, please feel free to open a Pull Request and we’ll review it! The resource should ideally demonstrate something new instead of duplicating an existing resource.  A pre-trained model is a model that has already been trained by someone else which we can reuse in our system. The model we are going to import is trained by Face book.