the notebook performs speech-to-text transcription using a pretrained Wav2Vec2 model from torchaudio. It includes noise addition, denoising, and evaluation using WER (Word Error Rate).

# Speech2Text with Wav2Vec2

This project provides an end-to-end speech-to-text (STT) system using Facebook's pretrained Wav2Vec2 model via `torchaudio`. It allows users to upload audio files, optionally apply noise or denoising, and transcribe speech to text using a custom CTC decoder.

## Features

- Upload `.wav` files for transcription

- Uses `torchaudio`'s pretrained Wav2Vec2 model

- Includes preprocessing options:

- Add Gaussian noise

- Apply denoising using `librosa`

- Visualize waveform and spectrogram

- Evaluate transcription accuracy using Word Error Rate (WER)

## Requirements

Install the required packages:

```bash

pip install torchaudio librosa jiwer matplotlib

Note: Ensure your Python environment supports PyTorch with appropriate CUDA or CPU configuration.

**How to Use**

1. Open Speech2text.ipynb in Jupyter Notebook or Google Colab.
2. Run the cells step-by-step:
   * The notebook installs libraries, loads a pretrained model, and defines helper functions.
   * You can upload .wav audio files when prompted.
3. You can apply preprocessing (noise or denoising) optionally.
4. The model will predict the transcription, and WER will be calculated if a reference is provided.

**Pretrained Model**

* **Model**: WAV2VEC2\_ASR\_BASE\_960H (from torchaudio)
* **Decoding**: Custom CTC (Connectionist Temporal Classification) decoder

**Example Output**

* Input: Audio of someone saying *"Hello, how are you?"*
* Output: "Hello how are you"

**License**

MIT License

**Acknowledgments**

* [Facebook AI - Wav2Vec2](https://ai.facebook.com/blog/wav2vec-20-a-framework-for-self-supervised-learning-of-speech-representations/)
* [torchaudio documentation](https://pytorch.org/audio/stable/index.html)
* [librosa](https://librosa.org/)
* [jiwer](https://github.com/jitsi/jiwer) for WER computation