This report compares the fine-tuned model from Task 3 against the pretrained model on the cv-valid-dev dataset in Task 2. Due to limited computational resources, I fine-tuned the model using only 10,000 samples from cv-valid-train.

To compare the pretrained model and the fine-tuned model results, we can use Word Error Rate (WER) and Character Error Rate (CER).

	WER (%)	CER (%)
Pretrained Model(facebook/wav2vec2-large-960h)	10.81	4.52
Fine-tuned Model (wav2vec2-large-960h-cv)	8.03	3.43

The fine-tuned model achieves a 2.78% reduction in WER and a 1.09% reduction in CER. This improvement is expected as the pretrained model is trained on generic speech data while the fine-tuned model has been adapted to our dataset.

Ways to improve model accuracy

1. Data Augmentation

Data augmentation enhances the model's generalization by exposing it to more diverse training conditions:

- Noise Injection: Adding background noise can improve robustness to real-world environment.
- Pitch & Speed Adjustment: Altering speed or pitch without changing the meaning can help the model learn invariance.
- Spectrogram: Applying transformations like time and frequency masking can improve robustness to missing or distorted signals.

2. Model Enhancements

Improving the model's architecture can enhance its ability to learn meaningful audio representations:

- Self-Attention Mechanisms: Transformer-based architecture can capture dependencies, improving contextual understanding.
- Regularization: Methods like dropout and batch normalization can prevent overfitting and stabilize training.
- Hyperparameter tuning/optimization. Given more time and resources, optimizing learning rates, batch sizes and optimizers can further enhance performance.

By implementing these strategies, the model can achieve better accuracy and robustness.