```
Fine-tuned WER Score: 1.0266326158336374
Original WER Score: 1.0130557148069006
```

The fine-tuned model and the original model exhibit similar Word Error Rate (WER) scores, which is to be expected since the fine-tuned model was trained on only a subset of the full training dataset. As a result, the model may not have reached its full potential yet.

However, if I were to train on the complete dataset, I would anticipate the fine-tuned model to outperform the original pre-trained model used in Task 2a, as fine-tuning on domain-specific data should improve performance.

There is still potential to enhance the accuracy and robustness of the model. One approach is to expand the training dataset with more diverse audio samples, incorporating a variety of accents, alterations of speaking speeds, and having different background environments. Additionally, data augmentation techniques—such as introducing background noise, adding overlapping music, adjusting the speed of the speech, pitch shifting, and volume adjustments could help the model become more robust to real-world audio variations.

Moreover, since the model was trained on a smaller subset of the full training set, it likely limited its performance. The fine-tuned model also wasn’t optimised for hyperparameters, which can have a significant impact on its results. To improve performance, systematic hyperparameter tuning could be applied, experimenting with parameters like learning rate, batch size, gradient accumulation steps, and dropout rates. Techniques such as grid search, random search, or more advanced methods like Bayesian optimisation could be used to identify the best settings. Additionally, implementing curriculum learning starting with simpler audio data and progressively introducing more complex examples could improve training efficiency and model accuracy.

Lastly, integrating a language model during decoding (e.g., KenLM or Transformer based models) could enhance transcription quality by correcting grammatical errors and improving contextual understanding. Post-processing techniques like spell checking and grammar correction could also address any residual errors in the transcriptions. With these strategies, we can build a more accurate and robust speech recognition system capable of performing well across a variety of scenarios and datasets.