The two files were evaluated in three areas including spelling and contrations errors by visual checking the csv files.

| Metric | cv-valid-dev (base model) | cv-valid-test (finetuned model) |
|---|---|---|
| Spelling | Lower proportion of spelling errors. | Higher proportion of spelling errors. |
| Contractions | During transcription of complicated contractors such as “a’int”, the model appends other alphabetical characters that sound like the word. For example, the transcription of “i ain't gonna squeal” is “aink on a squeal” | Performs better in distinguishing complicated contractors it could be due to the dataset that it was finetuned on contained a substantial proportion of contractors. For example, “aren't you going to tell me” gets transcribed to “ain't you going to tell me”. |

Proposed Improvement (Text)

1. Perform a similarity dataset check between cv-valid-dev and cv-valid test. The dataset characteristic (at word level and character level) Identify if there is some form of distribution between the two dataset.
2. Compare the distribution of whitespaces
3. Compare the significance of contractions in the dataset.
4. Compare n-gram character sequences between both models to see if the fine-tuned model has adapted to specific contractions or colloquialisms.

Proposed Improvement (Audio)

1. This dataset that the base model is trained on is derived from has high-quality audio well-articulated speech with minimal background noise, recorded in mostly controlled environments. However, the common voice dataset is sourced from public contributords.
2. Using STFT and spectrogram check the noise difference between the two dataset.