**Learning rate Comparisons:**

Alpha = .01

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|100.0|0.0|511.2|611.2|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.2|99.8|0.0|518.3|618.1|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|46.8|52.8|0.3|528.4|581.6|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|38.7|60.9|0.4|528.3|589.6|100.0|

Alpha = 1.0

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|69.7|30.3|2.0|102.0|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|33.6|66.4|0.0|100.0|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|23.6|32.0|44.4|1.6|78.0|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|11.9|29.8|58.3|0.1|88.3|100.0|

Alpha = 2.0

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|71.7|28.3|0.0|100.0|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|62.1|37.9|3.6|103.6|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|24.1|26.8|49.1|0.3|76.3|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|20.5|25.7|53.8|1.3|80.8|100.0|

Alpha = 3.0

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|67.8|32.2|0.0|100.0|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|66.2|33.8|4.3|104.3|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|23.1|23.1|53.8|0.3|77.2|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|21.6|25.3|53.0|1.6|79.9|100.0|

**Alpha = 4.0 (Baseline)**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|63.2|36.8|0.0|100.0|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|62.1|37.9|3.6|103.6|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|22.1|19.9|58.0|0.5|78.4|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|20.6|22.9|56.6|1.2|80.7|100.0|

Alpha = 5.0

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|67.8|32.2|0.0|100.0|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|66.4|33.6|4.3|104.3|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|23.1|23.1|53.8|0.3|77.2|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|21.7|25.4|52.8|1.6|79.9|100.0|

Alpha = 6.0

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|44.7|55.3|0.0|100.0|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.2|50.5|49.3|1.9|101.7|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|17.0|20.4|62.7|0.2|83.2|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|18.6|25.0|56.4|2.3|83.7|100.0|

—---------------------------------------------------------------------------------

**BPE Comparisons: Conclusions: Not much improvement with toggling different BPE values**

**BPE = 100**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|100.0|0.0|143.4|243.4|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|100.0|0.0|140.0|240.0|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|30.9|63.3|5.8|21.3|90.5|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|30.5|58.3|11.2|24.0|93.6|100.0|

**BPE = 125**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|90.8|9.2|6.6|106.6|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.5|82.4|17.1|10.7|110.2|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|29.1|37.3|33.6|3.6|74.5|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|26.0|37.2|36.8|4.4|78.4|100.0|

**BPE = 150 (BASELINE)**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|63.2|36.8|0.0|100.0|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|62.1|37.9|3.6|103.6|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|22.1|19.9|58.0|0.5|78.4|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|20.6|22.9|56.6|1.2|80.7|100.0|

**BPE = 175**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|80.9|19.1|0.7|100.7|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|71.7|28.3|5.2|105.2|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|26.7|29.4|43.9|1.1|74.5|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|23.0|28.6|48.5|1.7|78.8|100.0|

**BPE = 200**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|47.4|52.6|0.0|100.0|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|42.9|57.1|0.2|100.2|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|15.5|9.2|75.3|0.0|84.5|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|14.0|9.8|76.1|0.2|86.1|100.0|

—---------------------------------------------------------------------------------

**Using Char instead of BPE for token**

**Alpha = 4.0 (Baseline), Token = char**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|100.0|0.0|118.4|218.4|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|100.0|0.0|154.5|254.5|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|32.8|58.6|8.6|23.6|90.8|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|33.3|60.2|6.4|39.7|106.3|100.0|

**Using Word instead of BPE for token**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|2.0|5.3|92.8|0.0|98.0|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.7|9.8|89.5|0.0|99.3|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|3.7|2.3|94.0|0.0|96.3|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|4.5|3.9|91.5|0.3|95.7|100.0|

—---------------------------------------------------------------------------------

**Config: Baseline but with Encoder Number of Blocks = 10, Decoder Number of Blocks = 5**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|42.8|57.2|0.0|100.0|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|32.6|67.4|0.0|100.0|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|23.4|28.4|48.1|3.7|80.3|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|18.2|27.3|54.5|2.5|84.3|100.0|

**Config: Baseline but with Encoder Number of Blocks = 14, Decoder Number of Blocks = 7**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|71.7|28.3|0.7|100.7|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|0.0|73.3|26.7|4.3|104.3|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|24.2|27.0|48.8|0.8|76.6|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|23.5|31.0|45.5|2.0|78.5|100.0|

—---------------------------------------------------------------------------------

**Parameters for best WER**

**Alpha = 1.0**

**BPE = 200**

**Encoder, Decoder Number Blocks: 10, 5**

**Token type = word**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.7|7.2|92.1|0.0|99.3|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|1.0|13.8|85.2|0.0|99.0|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|2.1|2.9|95.0|0.0|97.9|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|4.8|5.7|89.5|0.1|95.2|100.0|

**Parameters for best CER**

**Alpha = 5.0**

**BPE = 125**

**Encoder, Decoder Number Blocks: 14, 7**

**Toke type = bpe**

### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|152|0.0|78.3|21.7|1.3|101.3|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|420|1.7|75.7|22.6|7.6|106.0|100.0|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|

|---|---|---|---|---|---|---|---|---|

|inference\_asr\_model\_valid.acc.ave/dev\_vi|25|619|22.9|30.4|46.7|1.3|78.4|100.0|

|inference\_asr\_model\_valid.acc.ave/test\_vi|57|1737|22.5|33.8|43.8|2.1|79.7|100.0|