Number of incorrectly aligned phones vs utterance length, model, and decoding algorithm. Softmax with $\beta = 2$ Plain Keep 50 frames Window ± 150 phones of the state of the stat Mixed Utt. Dataset Same Utt. aligned - 200 -600 -J 400 − Number 0 -200 400 600 200 600 200 400 600 200 400 600 400 Utterance length [phones]