



deeplearning.ai

Audio data

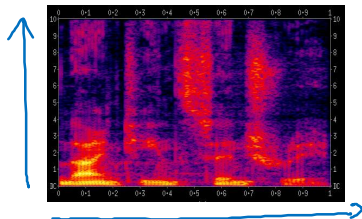
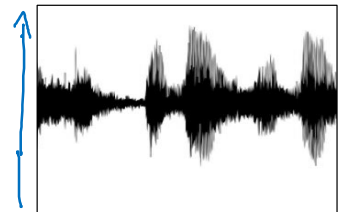
---

Speech recognition

# Speech recognition problem

$x$

audio clip



$y$

transcript

“the quick brown fox”

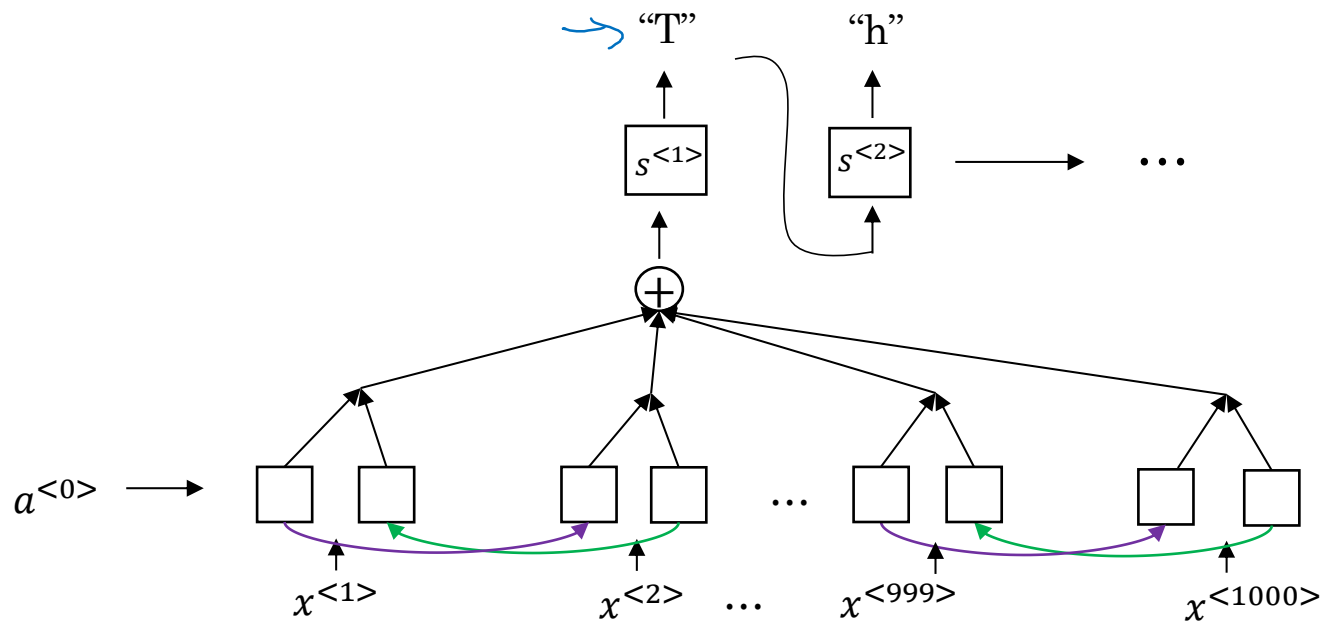
→ phonemes: de kwi br au n

300h

3000h

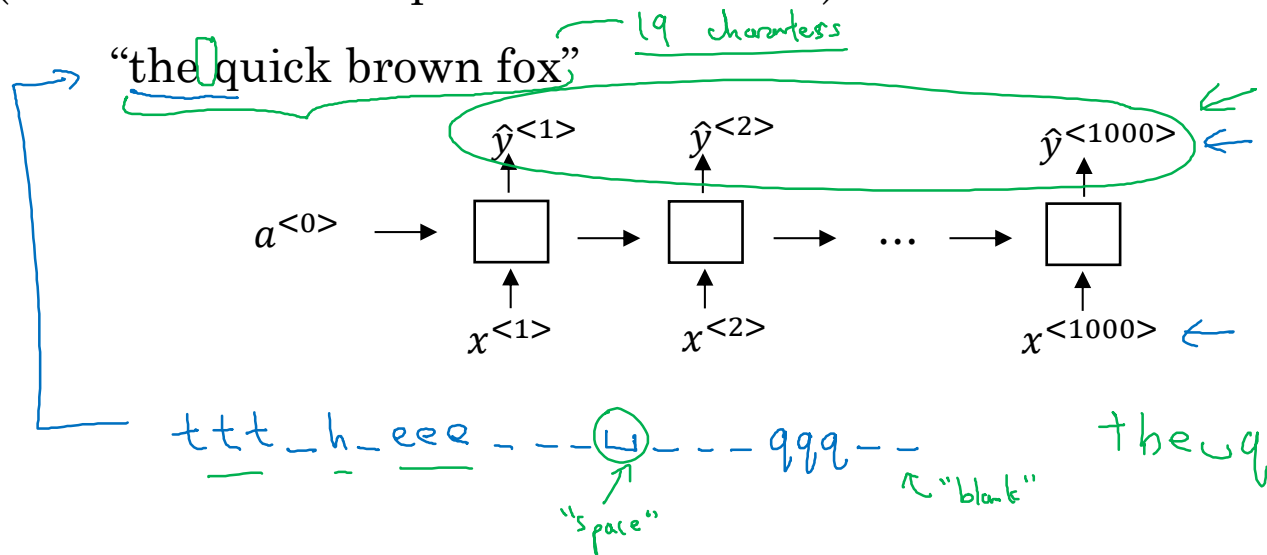
100,000h

# Attention model for speech recognition



# CTC cost for speech recognition

(Connectionist temporal classification)



Basic rule: collapse repeated characters not separated by "blank"