deeplearning.ai

Audio data
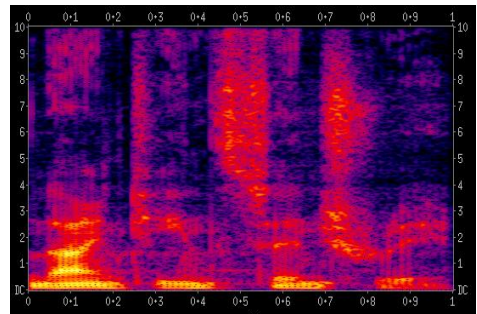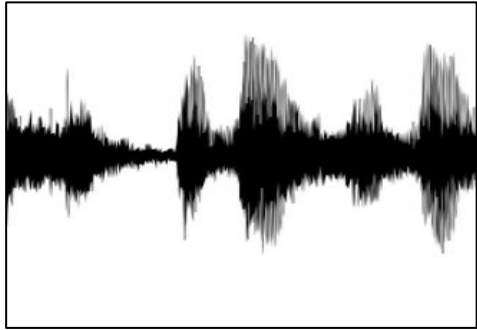
Speech recognition

Speech recognition problem

x $\longrightarrow$ y

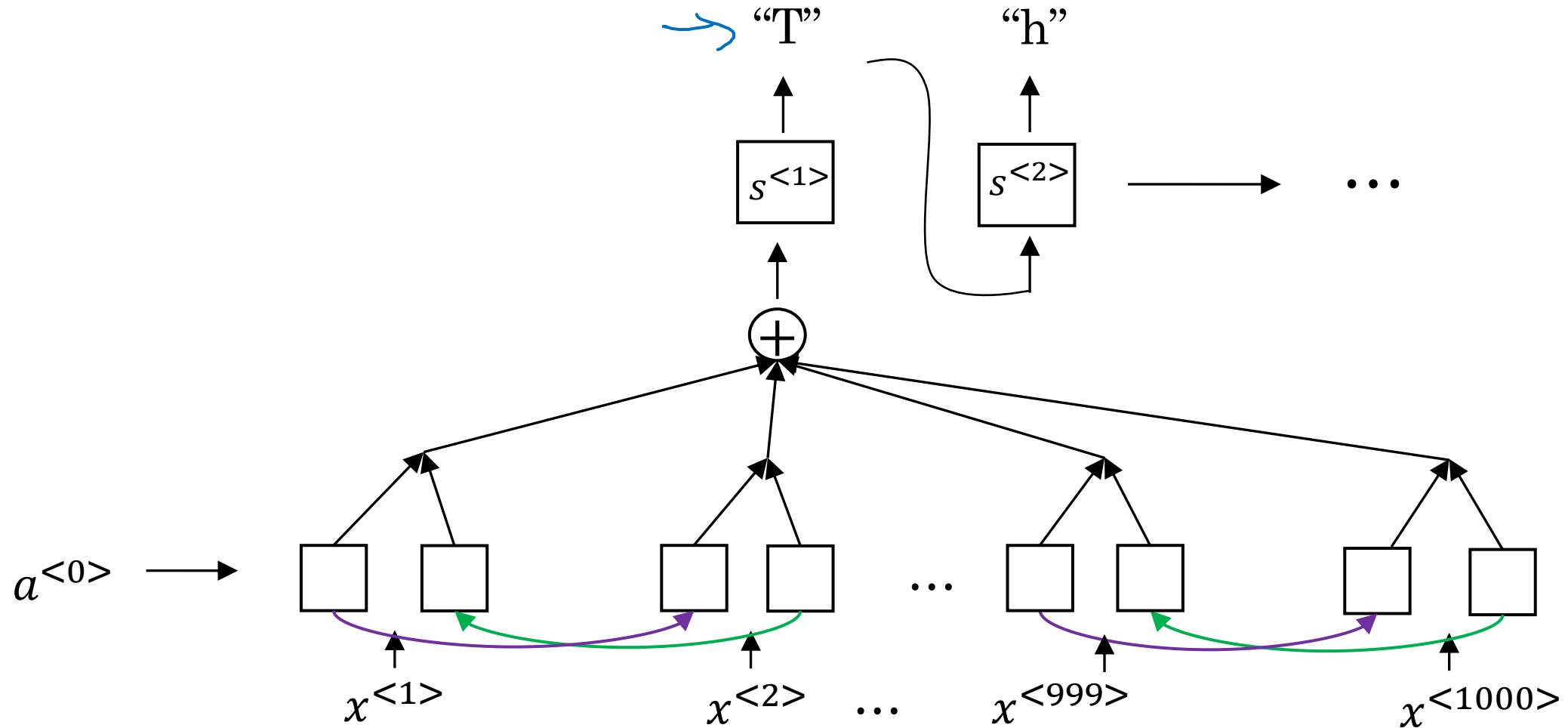audio clip   transcript

"the quick brown fox"
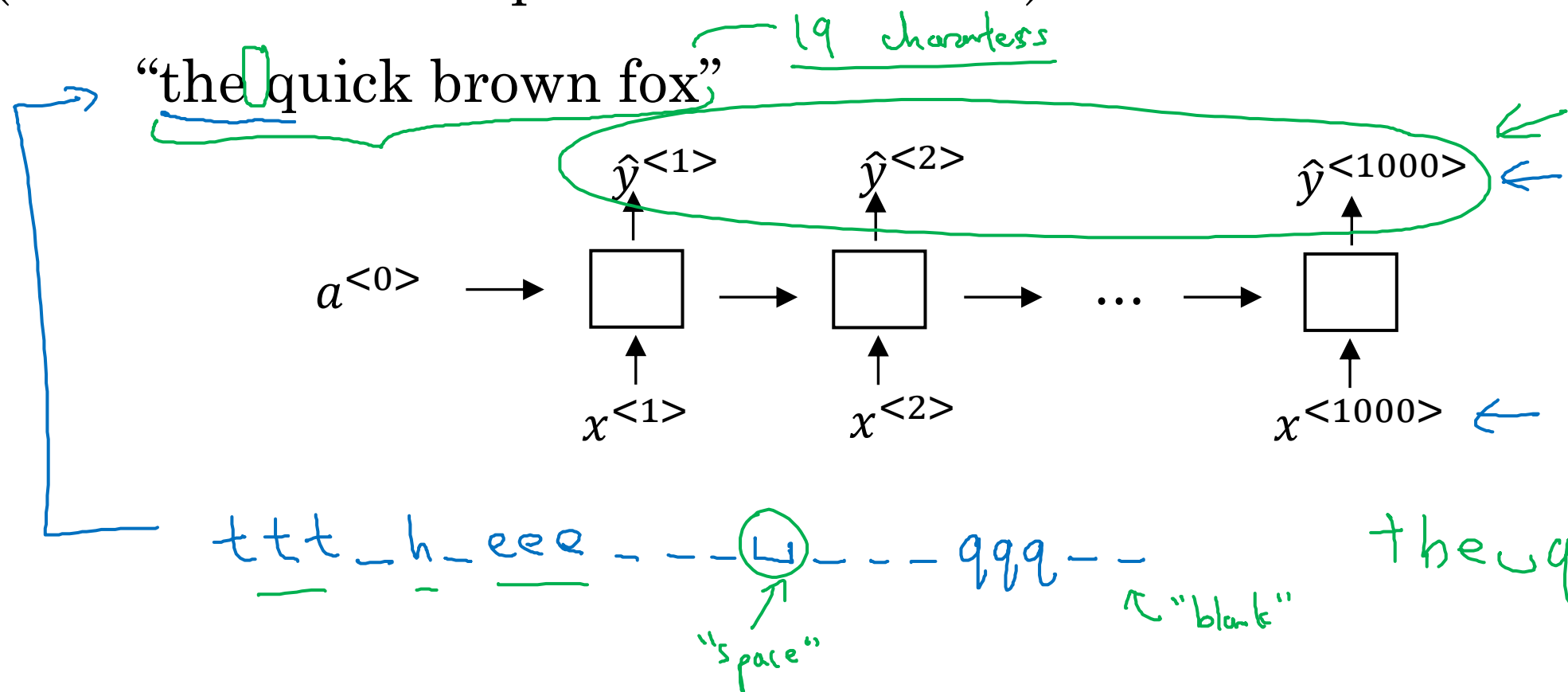
phonemes: de kwik braun

300h

3000h

100,000h

Andrew Ng

Attention model for speech recognition

Andrew Ng

CTC cost for speech recognition

(Connectionist temporal classification)

19 characters

"the quick brown fox"

$a^{<0>}$ → $\hat{y}^{<1>}$ $\hat{y}^{<2>}$ $\hat{y}^{<1000>}$

...

$x^{<1>}$ $x^{<2>}$ $x^{<1000>}$

ttt_h_eee - - - - ⊔ - - - - qqq - - theᵤq

"space"

↖ "blank"

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

[Graves et al., 2006. Connectionist Temporal Classification: Labeling unsegmented sequence data with recurrent neural networks]   Andrew Ng