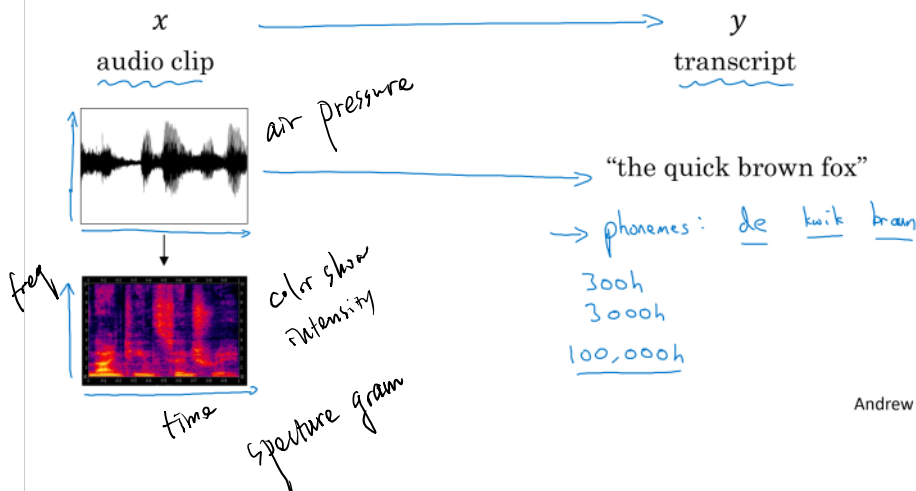




Audio data

Speech recognition

## Speech recognition problem

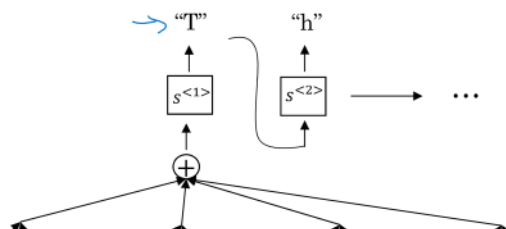


Andrew Ng

older generation speech recognition  
Before: use phonemes for the task.  
manually created, in linguistics

Must use large dataset.  
2000 hours — research hours  
10,000 — 100,000 hours commercial systems

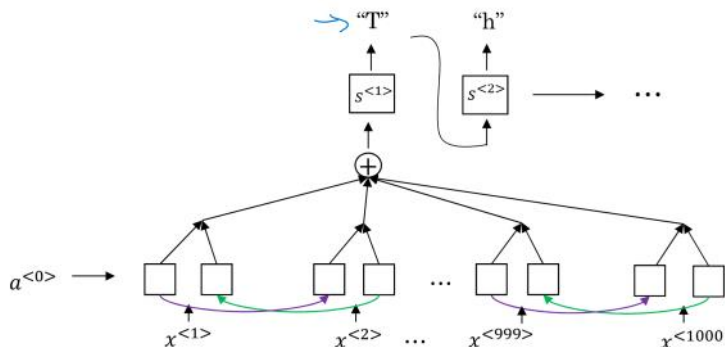
## Attention model for speech recognition



use the attention model for speech recog.

CTC cost.

## Attention model for speech recognition



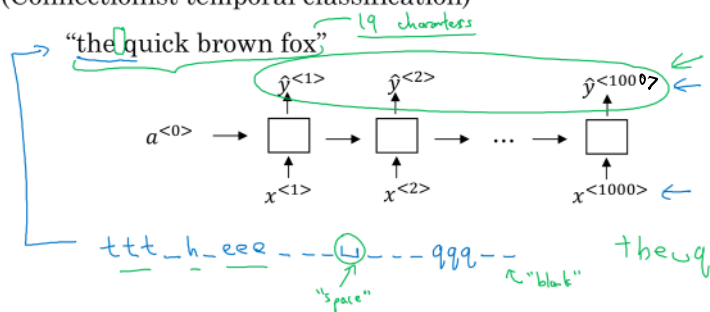
Andrew Ng

use the attention model for speech recog.

CTC cost.

## CTC cost for speech recognition

(Connectionist temporal classification)



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

① 这啥。

... 没听懂

Input 是 continuous 的，  
其输出还是 discrete。  
但有时 segmentation 的 input，  
和中文字符是一样的。  
没听懂啥意思

trigger word



Audio data

## Trigger word detection

What is trigger word detection?



Amazon Echo  
(Alexa)



Baidu DuerOS  
(xiaodunihao)



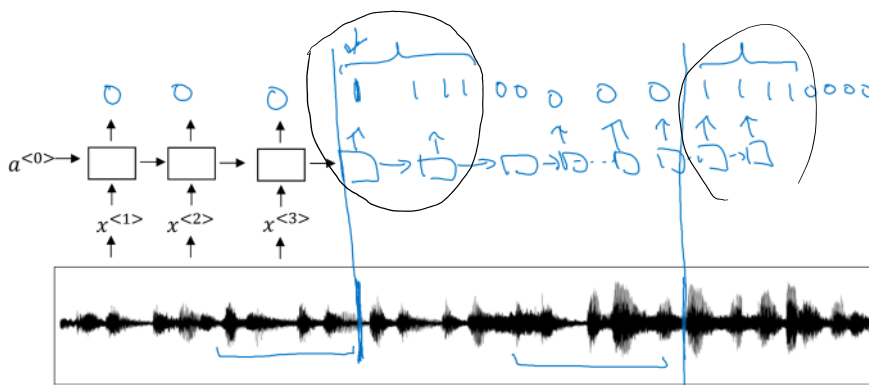
Apple Siri  
(Hey Siri)



Google Home  
(Okay Google)

Andrew Ng

## Trigger word detection algorithm



Problem  
imbalanced training set!

lots of 0. few 1

trick: can make it output a few 1.  
(e.g. - trigger any 1 or 2).

will appear in problem set.



Andrew Ng