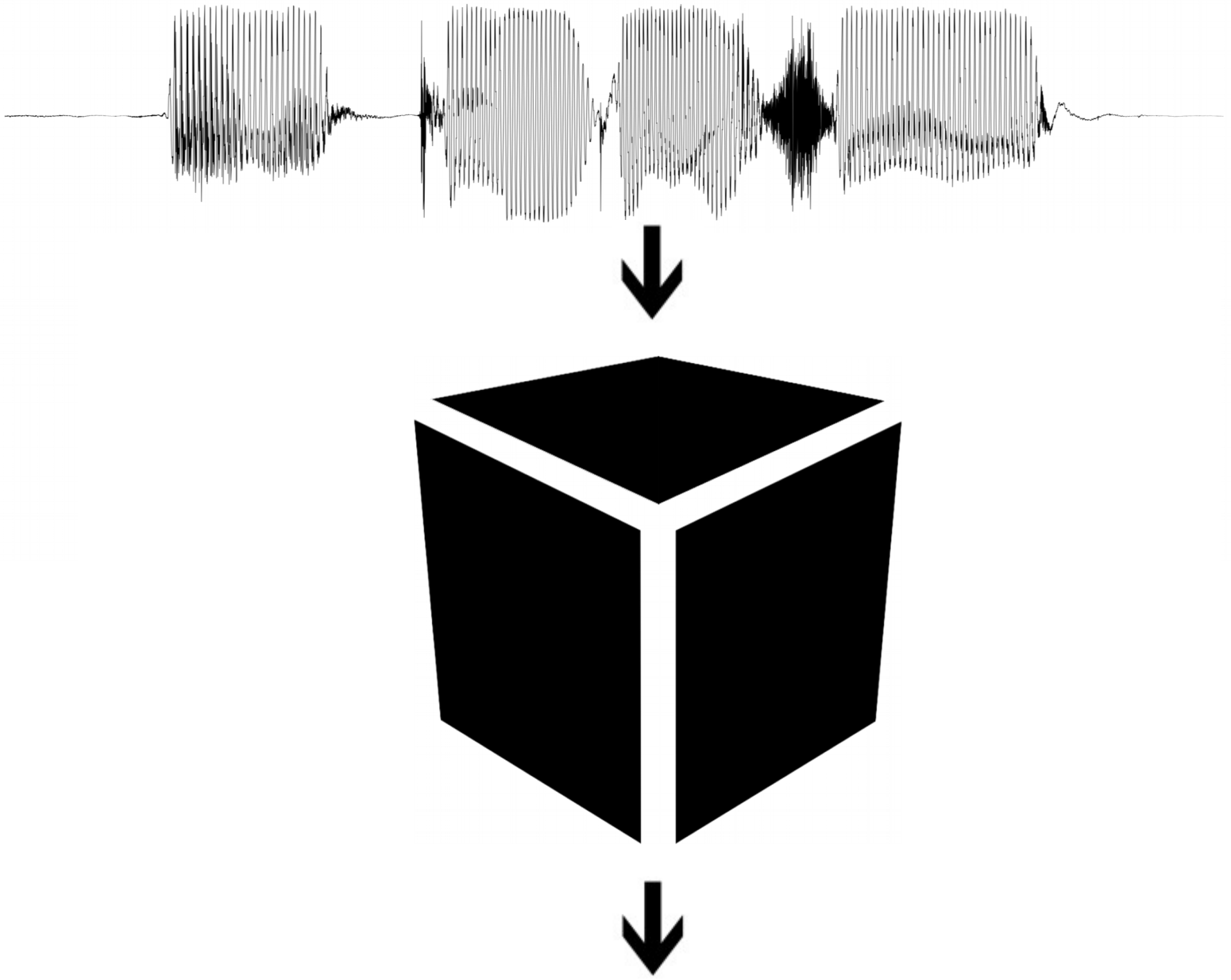
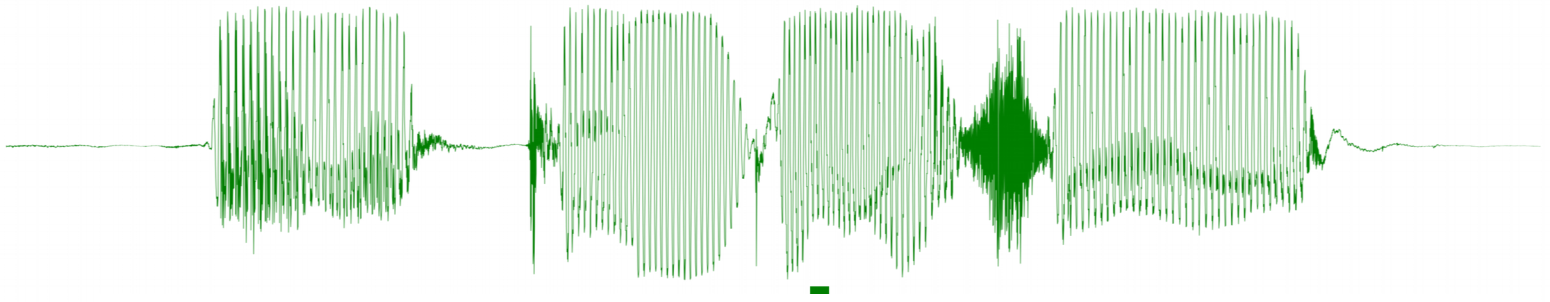
Automatic Speech Recognition:

A gentle introduction to Mozilla's DeepSpeech

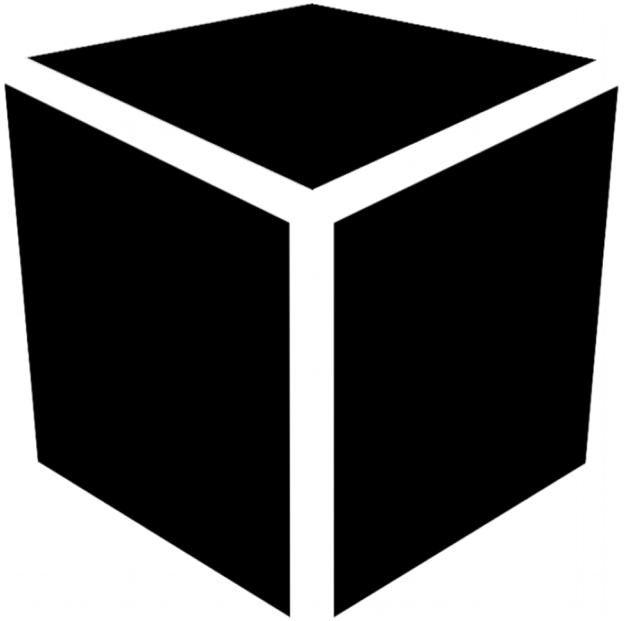
Josh Meyer + Fran Tyers
@_josh_meyer_ + @ezesanlasai



The dog

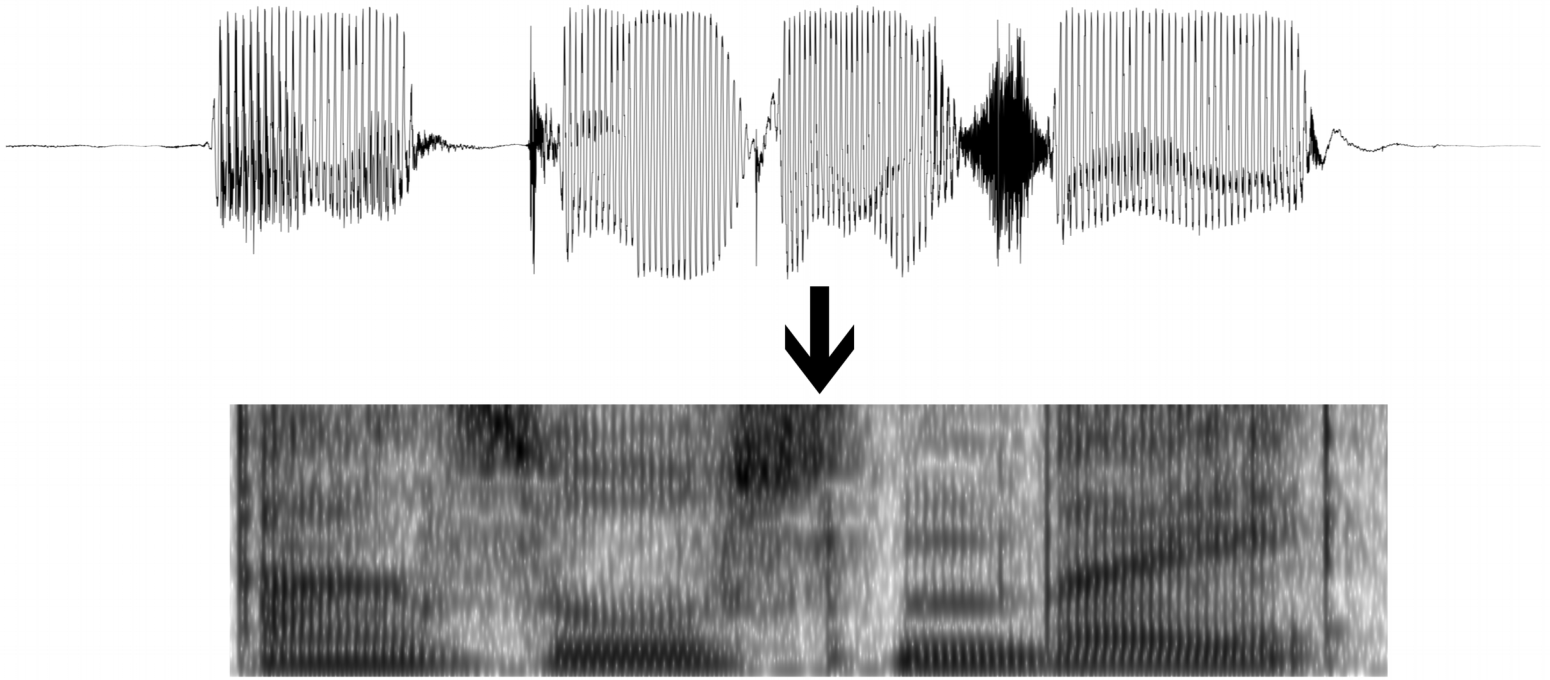


DATA



MODEL

"The dog"

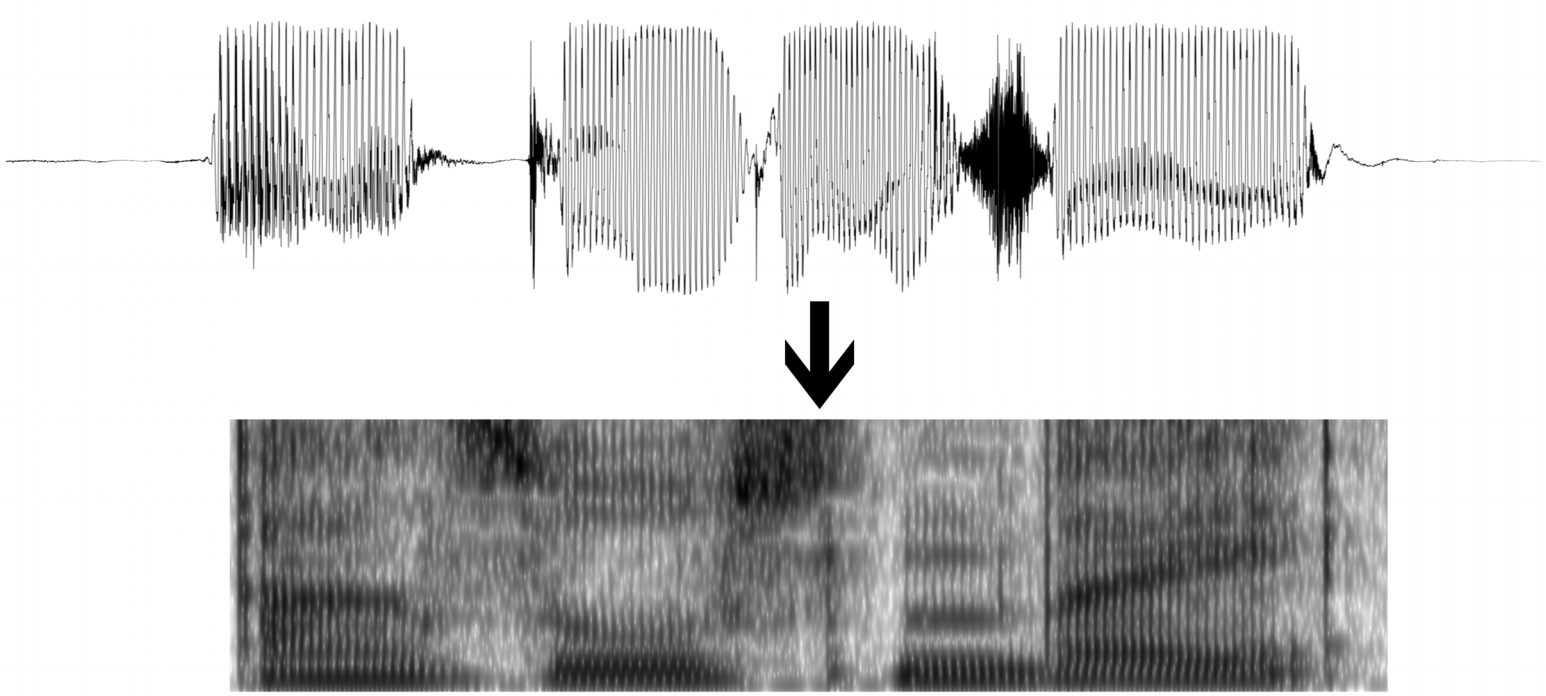


t h e d o g

The dog

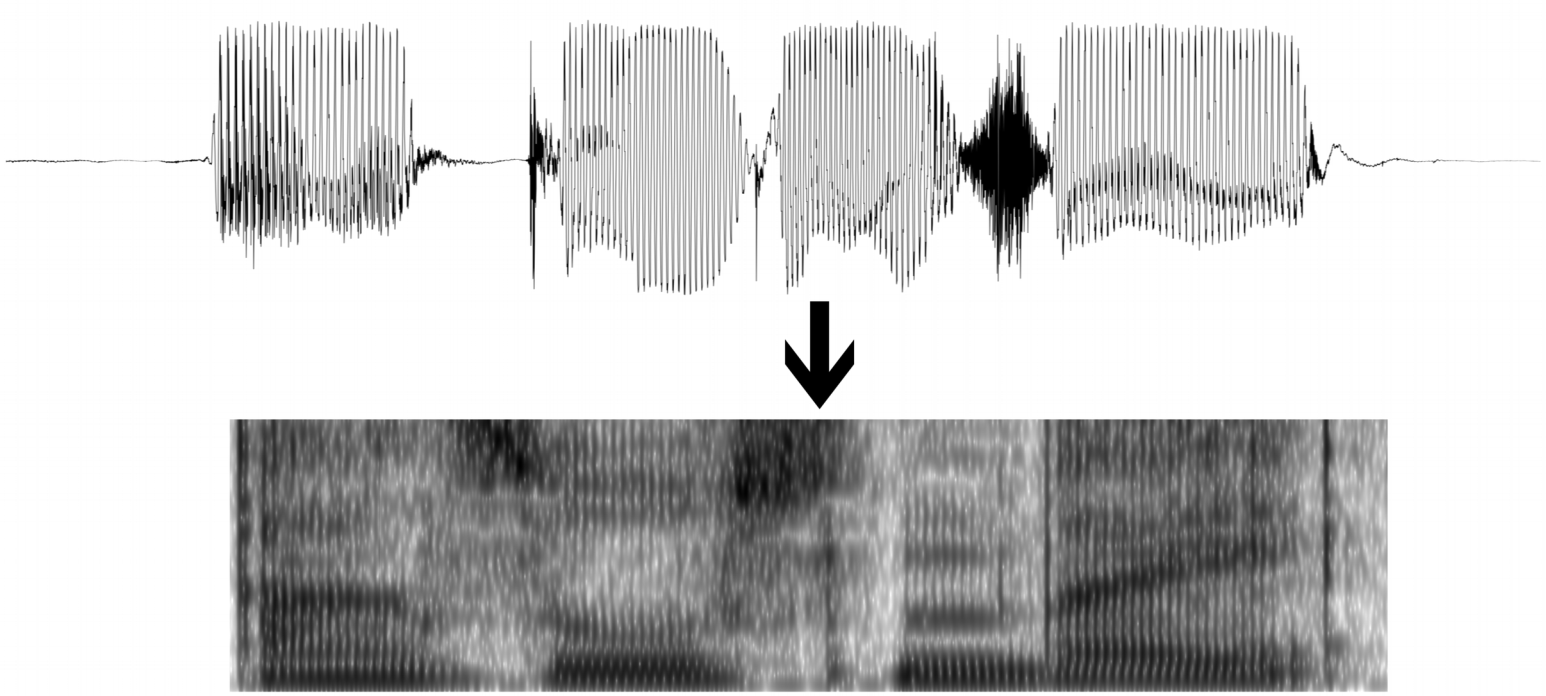
FEATURE EXTRACTION

- Sliding window feature extraction



FEATURE EXTRACTION

- Sliding window **feature** extraction



* a **feature** is measurable characteristic of the data
  (e.g. loudness of a certain frequency)

FEATURE EXTRACTION

- **Sliding window** feature extraction

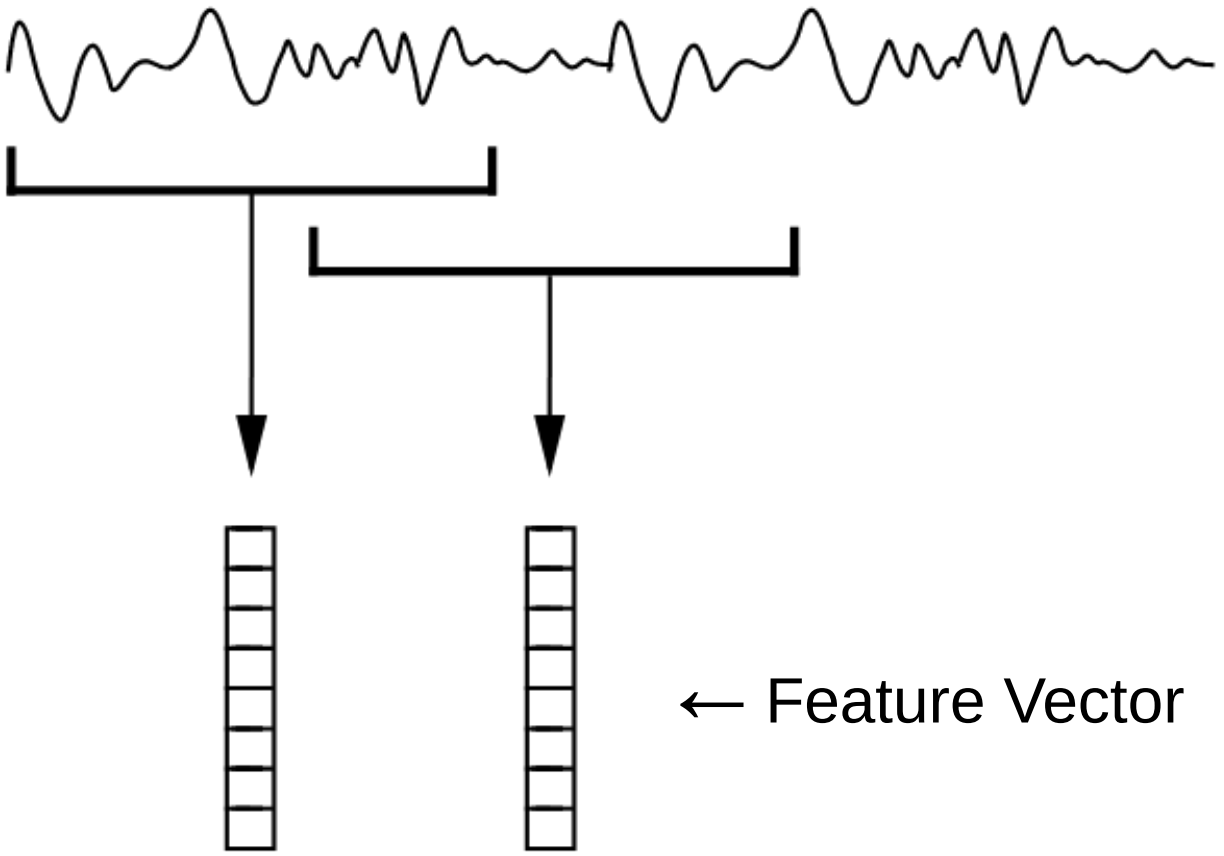
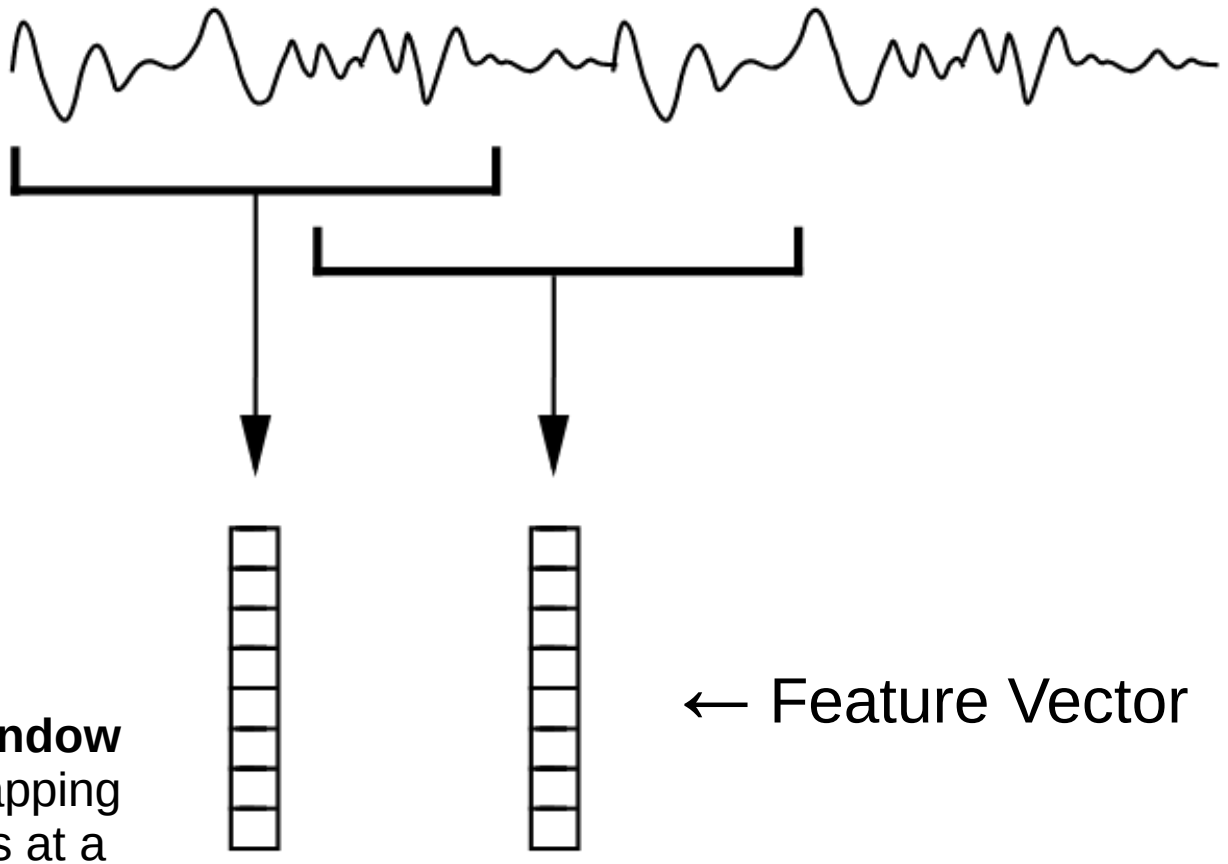


Figure credit
HTK Book

FEATURE EXTRACTION

- **Sliding window** feature extraction



← Feature Vector

* the **sliding window** will make overlapping audio snapshots at a moment in time

Figure credit
HTK Book

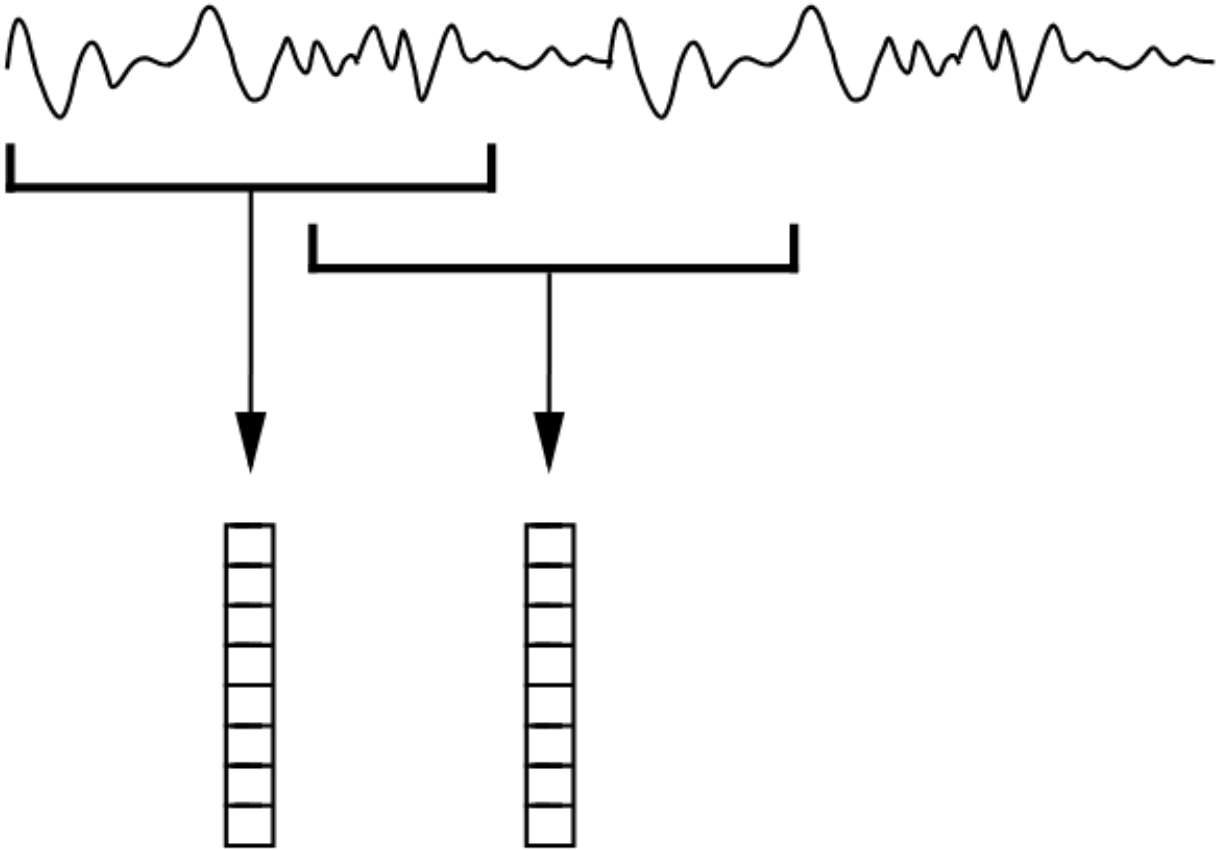


Figure credit
HTK Book

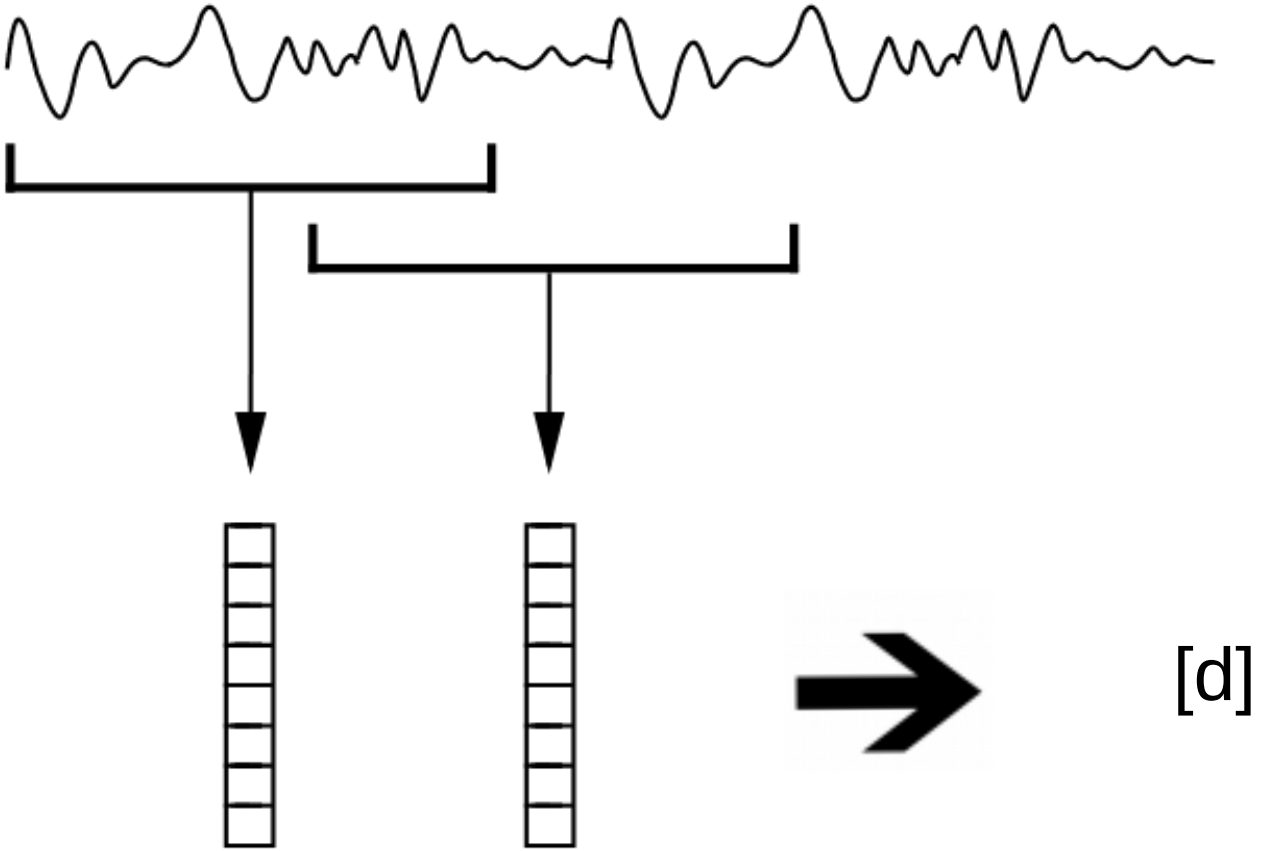


Figure credit
HTK Book

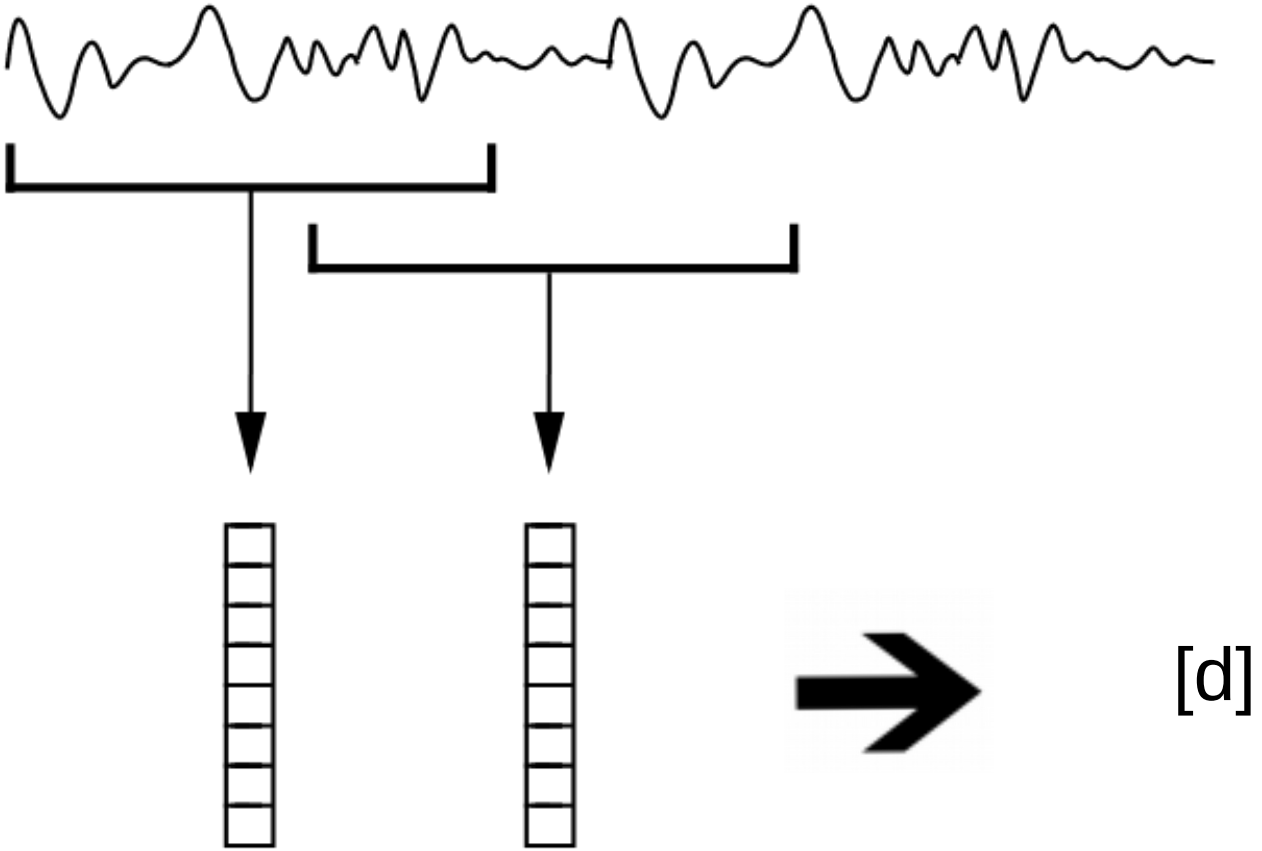


Figure credit
HTK Book

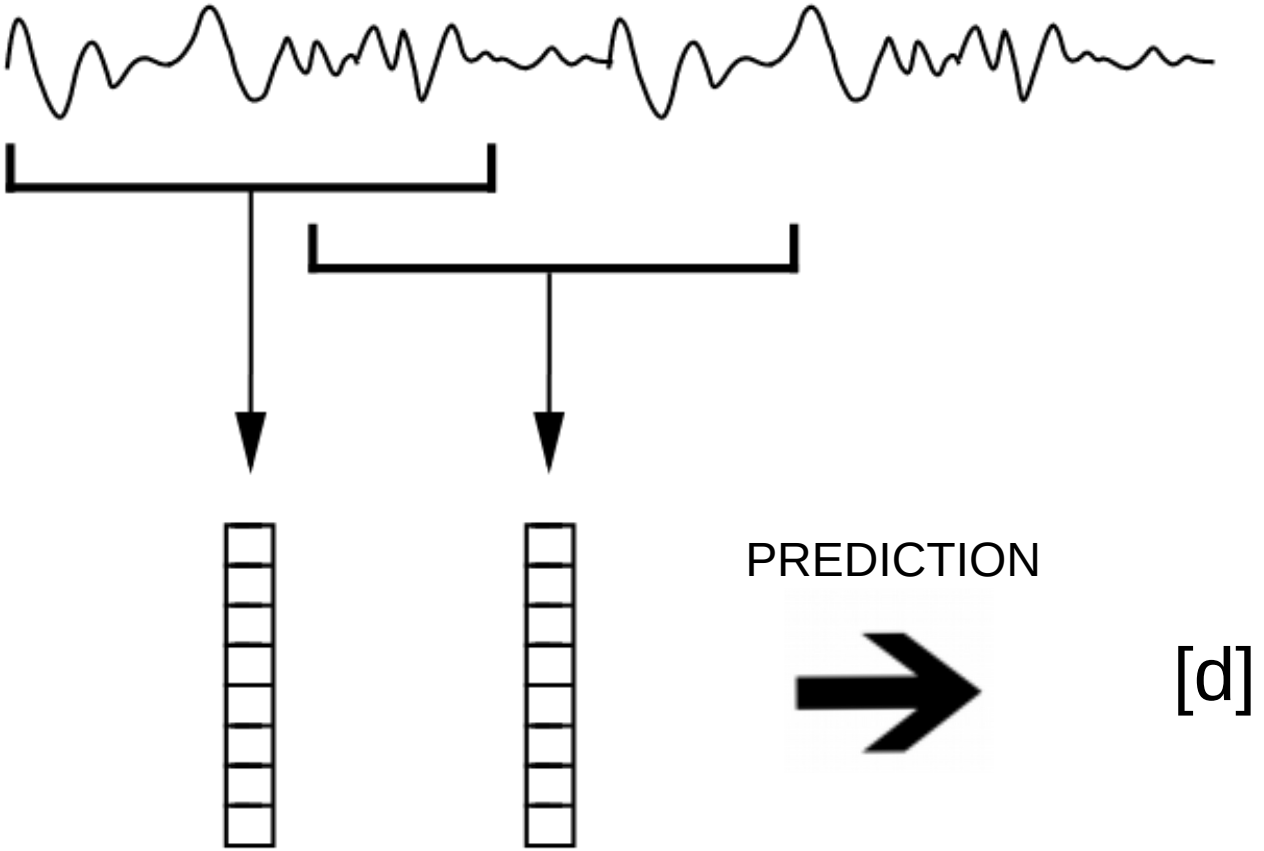


Figure credit HTK Book

DeepSpeech TRAINING

- Labeled Data
- Parameter Search

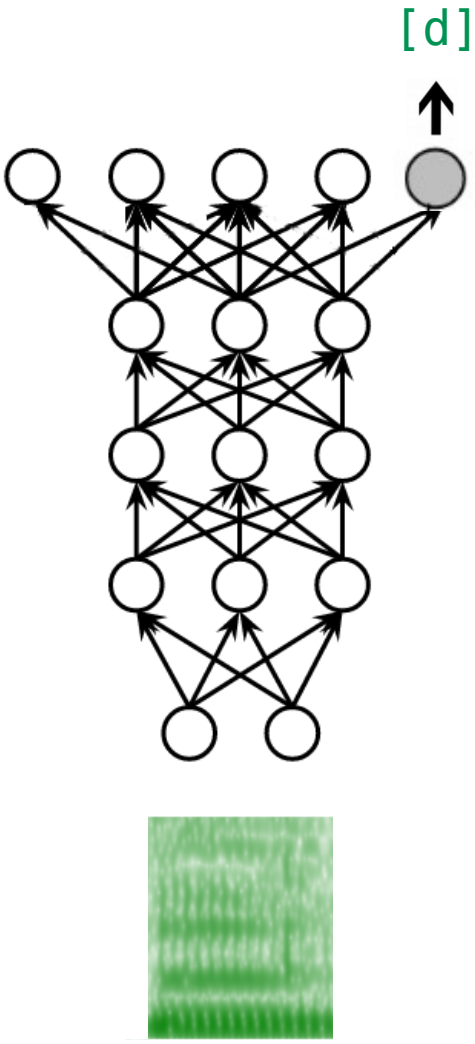
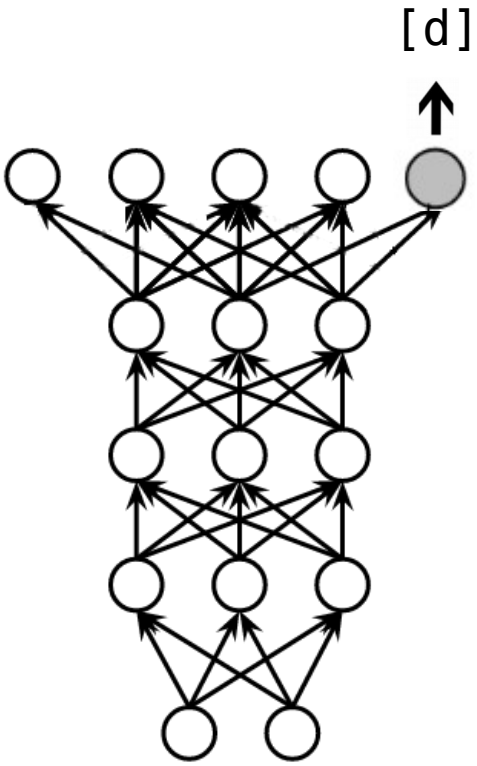


Figure credit
Heigold 2013

DeepSpeech TRAINING

- Labeled Data
- **Parameter** Search



* **parameters** are the connection weights

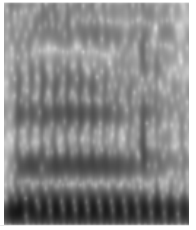
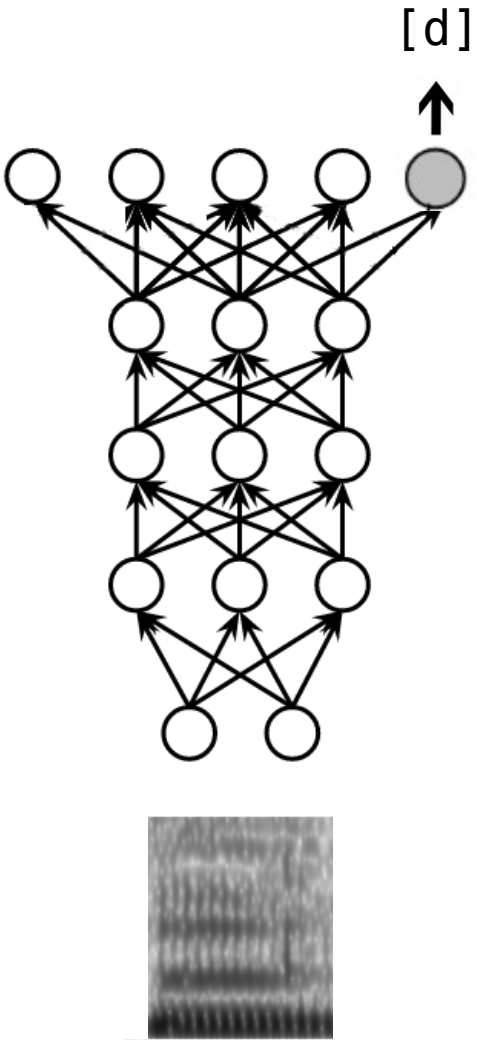


Figure credit Heigold 2013

DeepSpeech TRAINING

- Labeled Data
- Parameter **Search**

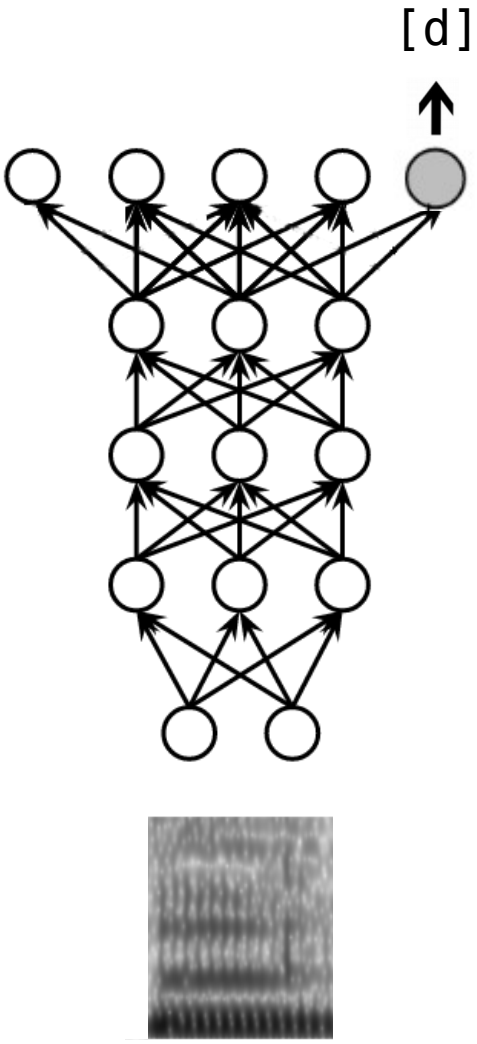


* **search** is finding best connection weights for the given data

Figure credit
Heigold 2013

DeepSpeech TRAINING

- Labeled Data
- **Iterative** Parameter Search

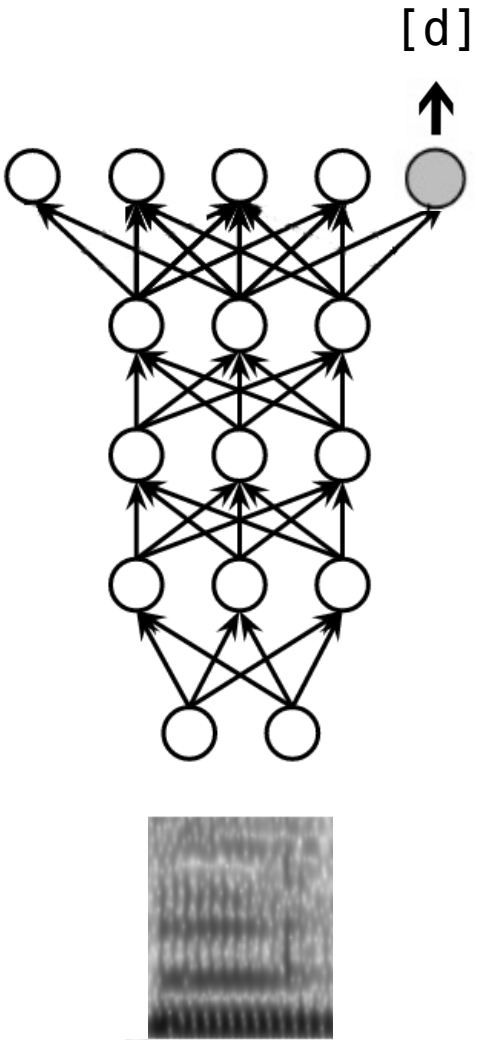


* we will **iterate** over the data, updating parameters incrementally

Figure credit
Heigold 2013

DeepSpeech TRAINING

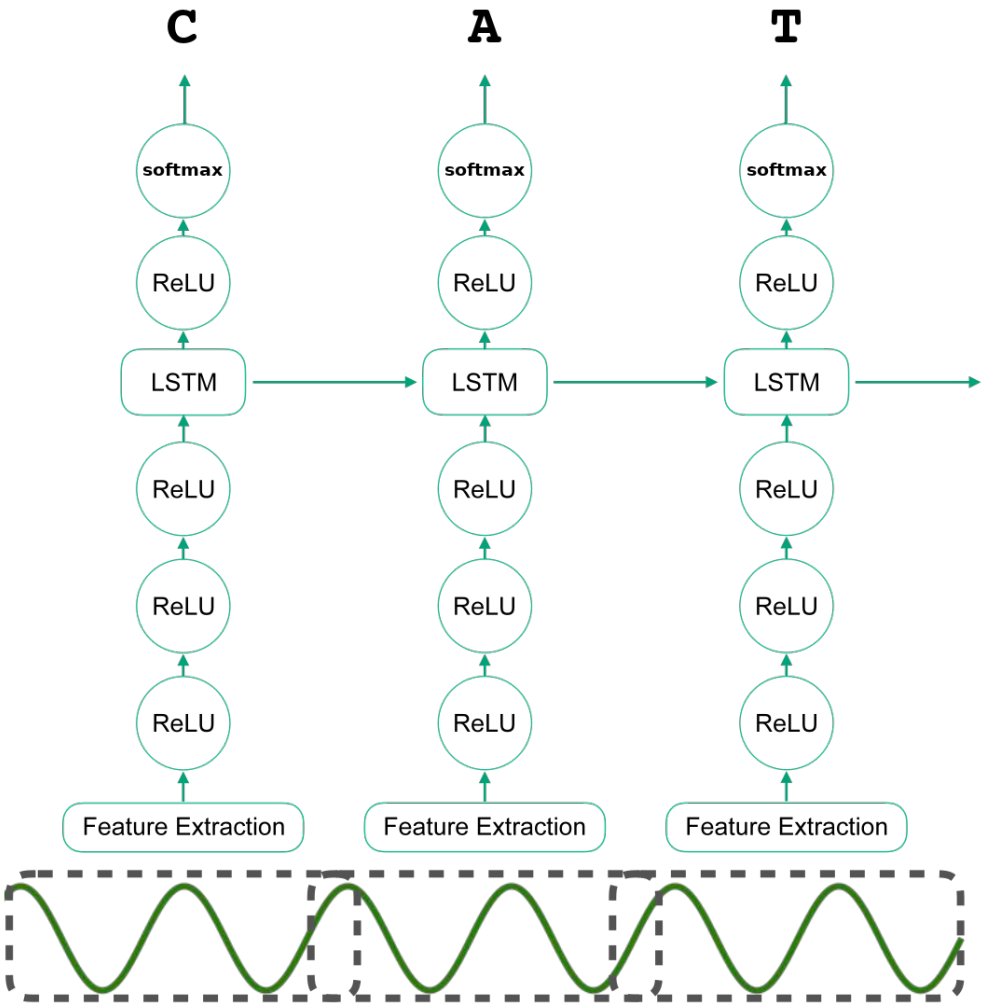
- Labeled Data
- **Good** Parameter Search



* **good** parameters will maximize the probability of predicting letters found in unseen data

Figure credit
Heigold 2013

For more information* on DeepSpeech architecture...



* check out Reuben's blog post for more
www.hacks.mozilla.org/2018/09/speech-recognition-deepspeech

Thanks for your attention:)

Рахмат)))