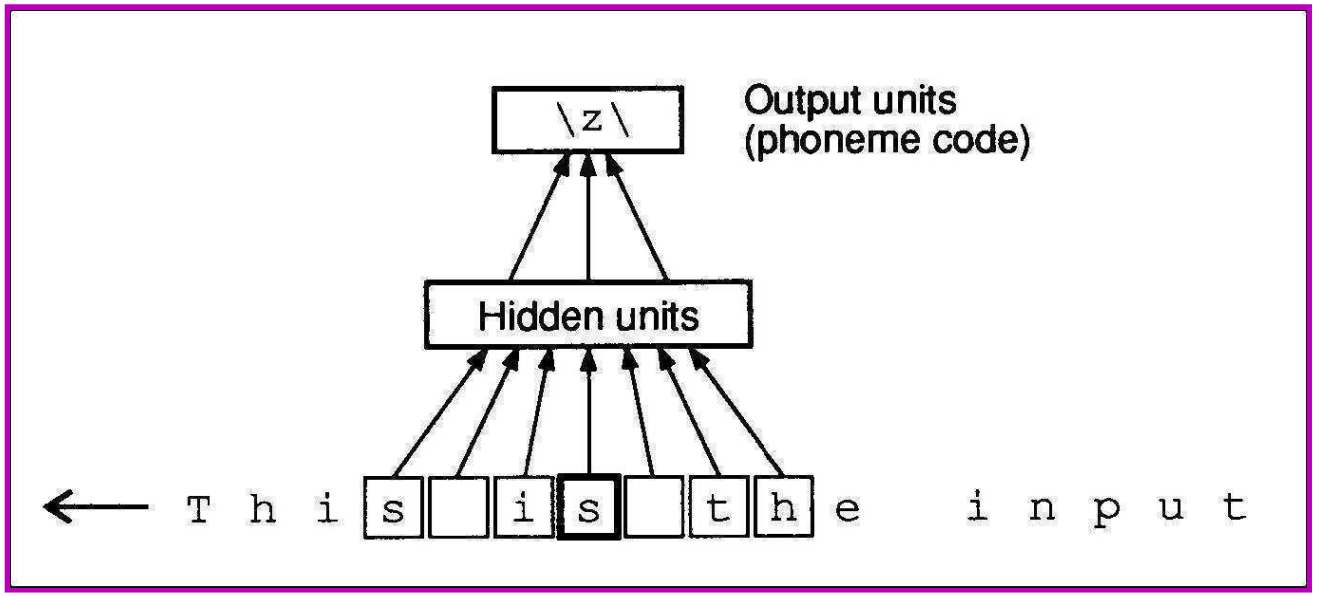
NETTalk

(Sejnowski & Rosenberg, 1987 "*Parallel Networks that Learn to Pronounce English Text*", Complex Systems 1, 145-168)

- Project for pronouncing English text: for each character, the network should give the code of the corresponding phoneme:
  - A stream of words is given to the network, along with the phoneme pronunciation of each in symbolic form
  - A speech generation device is used to convert the phonemes to sound
- The same character is pronounced differently in different contexts:
  - H<u>e</u>ad
  - B<u>e</u>ach
  - L<u>e</u>ech
  - Sk<u>e</u>tch

NETTalk – the architecture



- Input is rolling sequence of 7 characters
- 7 x 29 possible characters = 203 binary inputs
- 80 neurons in one hidden layer
- 26 output neurons (one for each phoneme code)
- 16,240 weights in the first layer; 2,080 in the second

➔ 203-80-26 two-layer network

NETTalk – Training results

- Training set: database of 1,024 words
- After 10 epochs the network obtains intelligible
  - generalization: 78% accuracy on continuation of train
  - Since three characters on each side are not always enough to determine the correct pronunciation, 100% accuracy cannot be obtained
-
  -
  -
  - damaging network produced graceful degradation, with rapid recovery on retraining
- Analysis of the hidden neurons reveals that some of them represent meaningful properties of the input (e.g., vowels vs. consonants)

NETTalk
Comparison to Rule-Based

- Generalization of NETTalk: only 78% accuracy
- Tools based on hand-coded linguistic rules (e.g., DECtalk) achieve much higher accuracy
- Hand-coded linguistic rules developed over a decade, and were worth thousands of $
- “Flagship” demonstration that converted many scientists, particularly psychologists, to neural network research
- The data for NETTalk used to be found at:

  http://homepages.cae.wisc.edu/~ece539/data/nettalk/