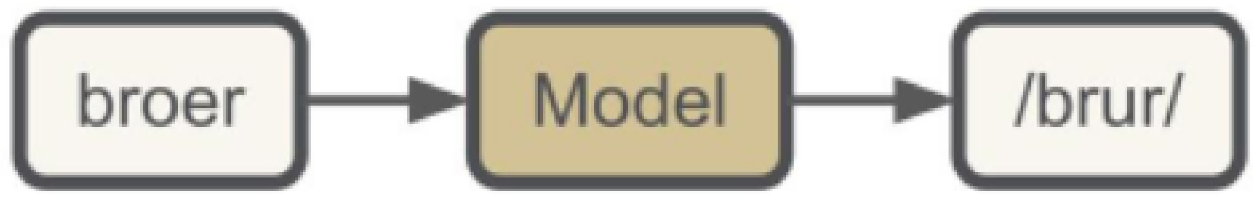
Grapheme to Phoneme Conversion for Dutch

Brian Hicks, Enze Chen, Minjia Zhong
CS 229 Fall 2017, Stanford University

Introduction

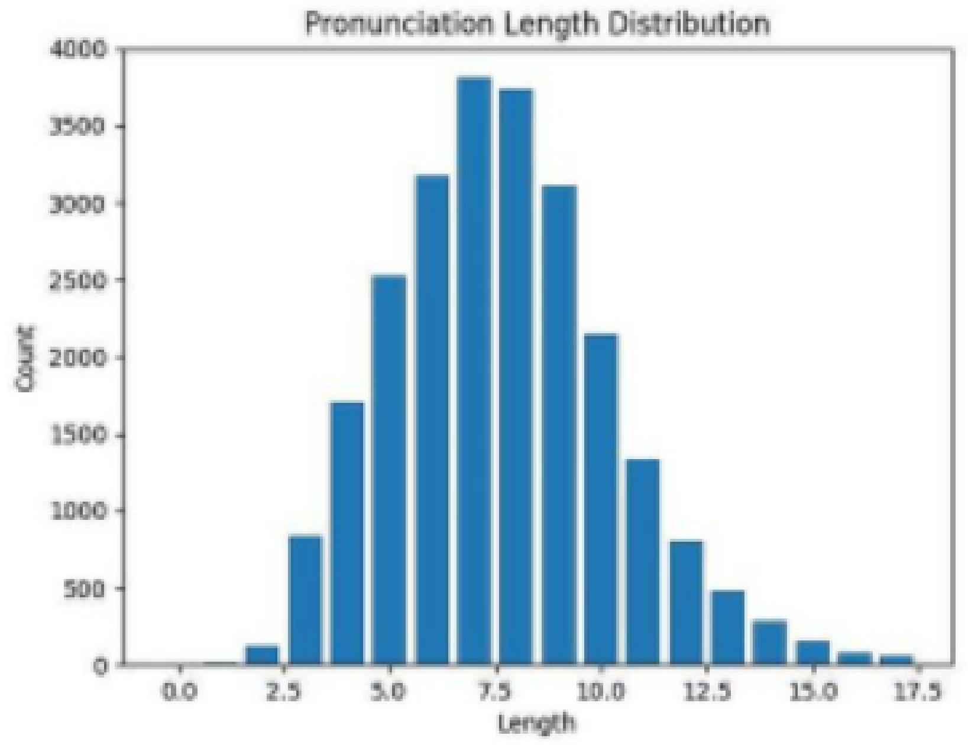
Grapheme-to-phoneme conversion (G2P) converts a written word to its pronunciation.


broer → Model → /brur/

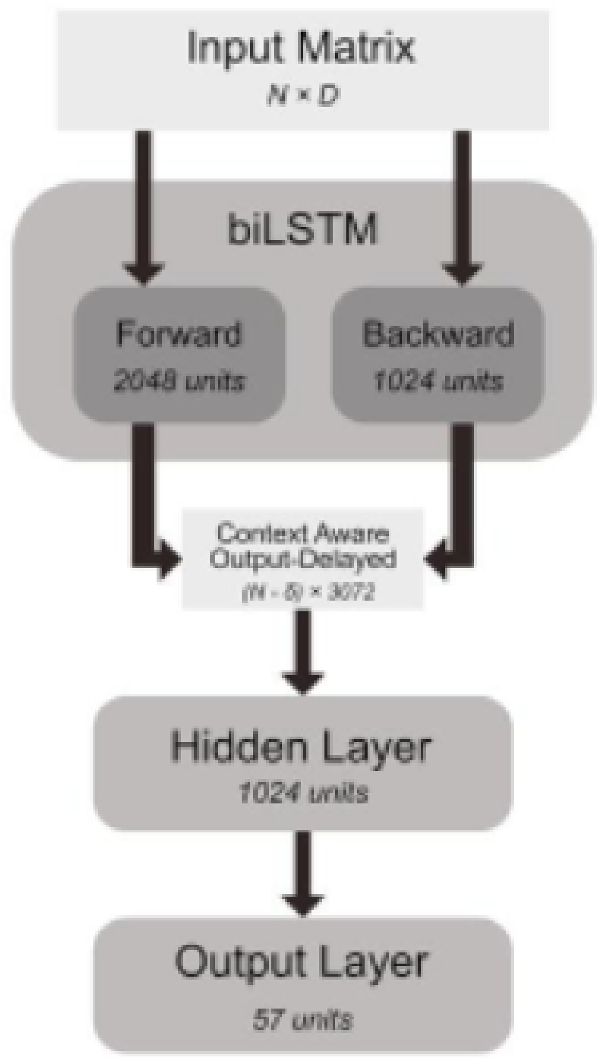
Applications include automatic speech recognition and text-to-speech systems.

Dataset and Challenges

Our dataset contains 24,404 orthography-pronunciation pairs.



The lack of a one-to-one correspondence of graphemes and phonemes make alignment- based approaches difficult.



Methodology

We leverage a bidirectional long-short-term-memory recurrent neural network (biLSTM) to encode the words using both past and future context.

Our model then enforces an output delay δ (i.e., ignores the first δ characters), which gives the model "time" to read the first few characters before having to make a prediction.

We then construct various ensembles, utilizing different voting and averaging techniques to account for model noise.

Results

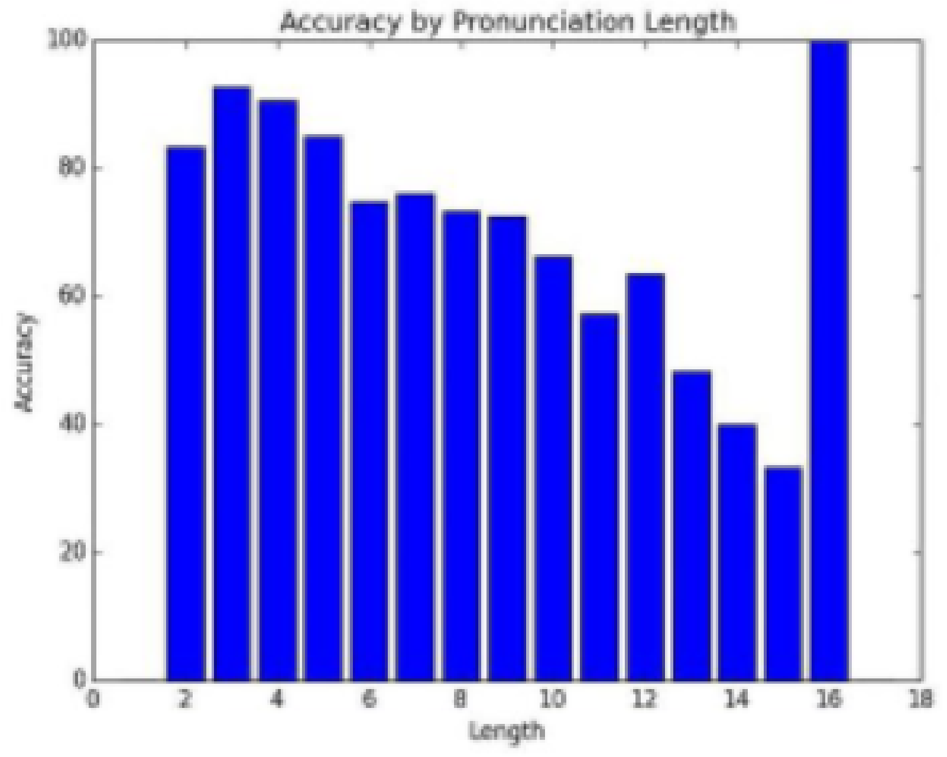
We define accuracy as completely correct predictions, and ED ratio as the ratio of the edit distance to the pronunciation length.

INDIVIDUAL MODEL PERFORMANCE

δ	Accuracy (%)	ED	ED ratio
0	59.9	1.734	0.194
1	64.7	1.658	0.188
2	66.2	1.656	0.188
4	64.7	1.639	0.185
-	63.9	1.672	0.189

ENSEMBLE PERFORMANCE

Model	Accuracy (%)	ED	ED ratio
Average	75.5	1.535	0.174
Weighted Average	75.4	1.527	0.173
Voting	74.9	1.583	0.179
Weighted Voting	74.9	1.573	0.179



Ensembles produced a significant increase in performance (11.6%). In general, the longer a word was, the harder it was to make accurate predictions, very short and very long words excepted.

References

[1] Paardekooper, P.C. *ABN-uitspraakgids*, 1978.
[2] Rao, *et al.* *IEEE ICASSP*, 2015.