CS 786G *Cognitive Modeling*
Week 8 Discussion Paper
A connectionist success and its deep implication that .... ?

Steve Woods
89800462

May 26, 1995

Seidenberg & McClelland (1989) describes a model of *parallel distributed visual word recognition and pronunciation* restricted to short, monosyllabic words. This model is in the connectionist tradition, essentially constructing a mapping from textual or orthographic representation to a phonological description of pronunciation. A system of this type is perceived to be a crucial sub-portion of the larger context of lexical processing, and the even larger context of textual understanding.

The "network" is constituted in an orthographic level, a phonology level, and the *inter-level* of hidden units which provides the interconnection between orthography and phonology. The weightings of the network is formed by a "learning" process in which the network is trained based on pairings of orthographic strings (words) and their and their phonetic interpretation. An input word activates a subset of the hidden units, a word string feedback, feedforward to a phonetic output. The feedback and feedforward are compared to the "actual" values (word and correct phonological unit) respectively, a difference of each is computed, and the strengths of network connections are adjusted according to a measure of proportionality of effect of each change on total error.

After the adjustment in this fashion of the network through application of set of examples (in this case several thousand), the knowledge about "orthographic to phonological" mappings is essentially embedded or distributed in the network. No use is made of lexicons nor pronunciation rules, and subsequently there is no "lookup" process whatsoever. The pronunciation[1] is determined through application of the input word and simply utilizing the phonetic output.

[1]In fact we refer rather to the sub-process of *naming* rather than involving actual articulation plans and execution as well, but for simplicity I retain the term pronunciation.

While there are interesting aspects of this work from a learning point of view (in terms of what examples of a concept were seen and what behaviour is subsequently observed based on instances of this concept, for instance), it is interesting to focus on the authors intent of drawing comparisons between human behaviour and the model behaviour, particularly in cases of child-learning of the ortho-phono mapping, and observations of the effects of "brain damage" and subsequent ability in mapping. Similarities are evidenced through direct comparison of results between the model behaviour and at human behaviour detailed in several other works including the following items:

- *Regularity effects* in which people ascribe a regular pronunciation to a word that demonstrates physical similarity to a class of words pronounced in a particular way, but is in fact not a member of this class, it is an exception.

- *Frequency effects* in which naming of lower frequency words (those that occur less frequently) takes longer than high frequency words.

- *Frequency and Regularity effects* in which high frequency words seem to have no significant difference in naming time between regular and irregular words.

- *Strange words* that have very few if any similar similar spelling words in the entire range of words seen produce a large number of mispronunciation errors.

- *Neigborhood size* (where a neighbor is defined as a word that can be "reached" from a source word by changing one letter) is important for low frequency words, but less so as frequency increases.

- *Nonwords* were demonstrated to have a mean naming latency for both regular nonwords and exceptional nonwords that corresponds to human results.

Additionally, analogies are drawn between model and human behaviour in several other areas, including:

- Children's common mispronunciations seem to correspond to those evidenced in the model, particularly mentioned are errors related to words with multiple pronunciations. In addition, the model indicates a decreasing frequency of these types of errors with additional training.

- Impairments that affect ortho-phono capability such as is the case with some types of dyslexics is suggested by the authors to correspond to cases in which fewer hidden nodes were used in the model with identical training cases.

These correspondences are indeed impressive from the point of view of suggesting that this (or any) model is actually representative of a "how-to" for one aspect of human cognition. The important issue of scaling to a "real" human language is left untouched, however. One may wonder about what exactly is learned with this exercise of creating a connectionist model in this manner. What do the network nodes actually represent - neurons (surely not), some other knowledge structure ? What effect might a particular change in network structure (or even weighting) have ? What does this change "mean" or is it in some way so deeply embedded in a concept as to be unremovable from the whole ? If so, what does this tell us about knowledge representation methods in other types of symbolic AI ?

While it is certainly true that no claim is made to the effect that this simplistic model actually corresponds to any *physical* human structure, the "explanation" or at least correspondence between the model and human behaviour is elaborated on at *great* length in this paper, clearly with the intent of offering evidence of this correspondence as opposed to earlier models. The intention would clearly seem to be a modeling of human cognition towards duplicating the functionality of human cognition directly, rather than an attempt to duplicate the achievements of human cognition in some novel manner. Perhaps this is a fundamental difference between what has been described as "symbolic" AI and "connectionist" AI - attempting to duplicate achievements or to duplicate function. Certainly from a human standpoint it is difficult to meta-reason about one's reasoning process, at least to the point of explicitly detailing the arrival at a particular decision or at a lower level, the naming of a word. In contrast, a purely symbolic approach typically would offer this explanation ability. In this sense a critique of connectionist AI as lacking explainability would make little sense from the point of view of modeling human cognition, and such a critique will not be attempted. However, insofar as the process of actual human cognition is currently almost totally unexplained (as far as the writer is aware), a model such as the connectionist one seems very difficult to motivate, except from the "well, it works in this and this case" standpoint and consequently leaves little foothold for specific critique. Perhaps this inexplicability is actually an inherent property of intelligence, both physical and artificial, but until that is demonstrated or at least motivated with stronger human-model correspondence I remain uncertain about the meaning of these connectionist models, however successful in limited domains, even in the absence of symbolic analogs.

References

Seidenberg, M. & McClelland, J. (1989), 'A distributed, developmental model of word recognition and naming.', *Psychological Review* **96**, 523–568.