# Designing networks of spiking neurons for practical use: the case of reading

Szymon Rutkowski

University of Warsaw

*szymon@szymonrutkowski.pl*

October 18, 2018

- Spelling correction (for Polish)
    - non-dictionary errors (for theoretical simplicity)
    - ranking candidate corrections from dictionary by edit or vector distance
    - generating corrections directly using LSTM recurrent network
        - with multi-layer ELMo embeddings representing phonology, morphology, whole words
- What would some more humanlike process look like?
- We seemlessly "correct" spelling errors by reading (eg. *yu can porbalby reed tihs*).

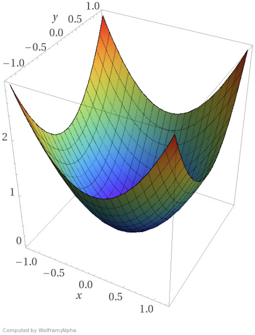
# Ways to (try to) build intelligence

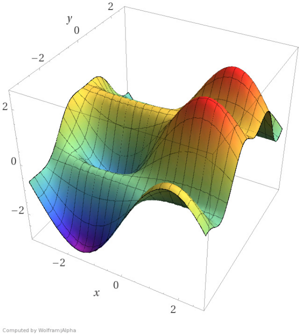
# From logic to “curve fitting”

Some of the ways to construct intelligent systems where we need them:

- thinking hard about how they should work: “Good Old-Fashioned AI” (GOFAI)
  - building intelligent systems from philosophical principles
  - formal logic & logic programming
  - encoding rules manually, expert systems
- fitting mathematical models to data: machine learning and statistics
- empirical research, reverse engineering: back to biology?

# Fitting complex functions




In ML, we take a model (such as a multi-layered neural network) and search for a set of parameters that minimizes some cost function on data.

# Possible limitations of machine learning

- In general, we require the math to be differentiable
  - or at least need some evaluating function (of numerical value) to be there: reinforcement learning, evolutionary algorithms
- We prefer matrix operations in practice (easier to implement), but this encourages neural models to be simpler
  - for example, connections *within* a neural layer are somewhat troublesome to implement
  - no (pseudo)continuous time – discrete epochs isolated from each other – "turn-based" operation
- Although in theory, this shouldn't matter? – we should learn things anyway

# Intepretability and abstraction

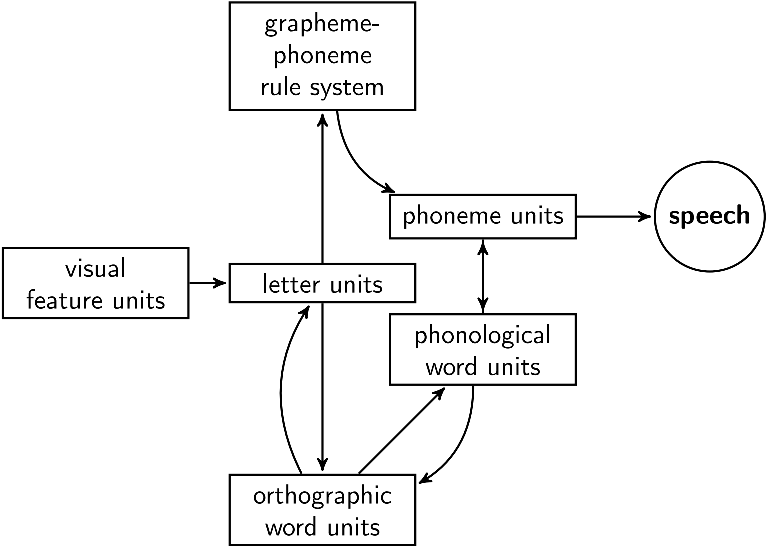
- The experience from fields such as natural language processing shows that algorithmic learning tends to beat human-collected knowledge
  - *Every time I fire a linguist, the performance of the speech recognizer goes up* – Frederick Jelinek, head of IBM speech recognition in 1980s
- But abstract models hit their limits eventually – as Markov chains most widely used to model language before deep learning
- Properties of biological systems may directly represent important features of information processing
  - such as Weber-Fechner law: neurons getting increasingly "bored" with arriving information

# Dual-route models of reading

# The original dual-route model of reading

- The model was developed and implemented in 1990s and 2000s among debates on the neuroscience of reading
- The dual-route approach maintains that in order to read, we **1)** read the individual letters and, in parallel, **2)** recognize words from our vocabulary, then agree the results
- An opposite stance is that there is no hard separation between those

# Organization of neuronal units in the model



# Interpretation of neuronal units

- Letters are excited by visual features that form them, graphemes and words by letters contained by them etc.
- Model parameters (weight strengths) are hand-tuned
- Most important is the balance between evidence from graphemic and lexical path

# The original dual-route model of reading

- Recent empirical research is skeptical of its strict biological truthfulness, but it has some advantages:
  - clear, understandable structure
  - there is a computational implementation
- The implementation can read a corpus of monosyllable English words, replicating some known phenomena
  - frequent and regular words are read faster than rare and irregular ones
  - non-existing words sounding like existing ones are also read faster

# The original dual-route model of reading

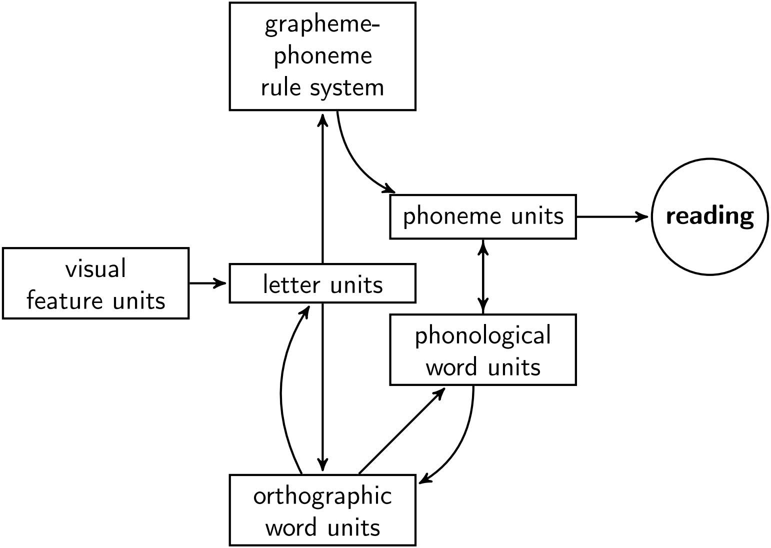
> *[...] we have specified the architecture of the model ourselves, rather than relying on backpropagation to do this. Our computational model is hand-wired rather than learned. [...] One disadvantage of the learning algorithm approach is that the range of possible model architectures that can be developed through backpropagation is constrained, and the correct model may not be within that range.*

Max Coltheart et al. (2001), *DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud*

# A spiking dual-route model for Polish

- It is fed letters directly, instead of assembling from visual features
- The output is graphemes, not phonemes: silent reading instead of reading aloud
- Grapheme reading is neuron-, not rule-based
    - Note that digraphs (and *n*-graphs) – such as *sh*, *wh*, *ee* ... in English – form single graphemes
    - The reader assembles words from graphemes, not letters, so in **nohwere* it can see *hw* as *wh* (supported by recognizing the whole word)
- For Polish, we try to recognize suffixes (for inflection) in parallel to word stems
- The model was moved to a biological neuronal simulator (NEST), so it could be implemented and expanded easier
    - (the original employed linear, not spiking neurons)

# Organization of neuronal units in the model



# Readings

English – toy lexicon and error list

| Network input | Reading output | Correct form |
|---|---|---|
| lese | **lease** | lease |
| lxte | **lute** | lute |
| ltue | **lute** | lute |
| elsee | **else** | else |

Polish – national corpus (NKJP) spelling errors, SGJP lexicon

| Network input | Reading output | Correct form |
|---|---|---|
| sytacjach | **sytuacjach** | sytuacjach |
| smieszny | **śmiesznn** | śmieszny |
| dziękujem | **dzięckujeee** | dziękujemy |

# Problems of constructing spiking nets

# Maintaining meaningful state

- We preserve information in the long term in connections, but what about the working memory?
- For example, we've just read *e*, so it should be less likely to appear again (as in *dzięckujeee*)
- (We do have lateral inhibition and inhibitor populations!)
- But biological(-like) neurons revert to their resting potential after firing – the information is erased
- It can be tackled with **active synapses** – `tsodyks_synapse` model in NEST – their excitability can decrease or increase with spiking events

# Methods of coding

- Implementations of DRC model used the local method of coding, where one neuron (or population) encodes one “concept” (such a word or a letter)
- This makes the number of units scale linearly with the number of concepts – 400k neurons for 400k word stems
- We think that the brain uses sparse coding: one codeword is represented by synchronous activation of many neurons
- ...in a way, our model has that – a word is “encoded” by its letter neurons etc.

# The role of noise

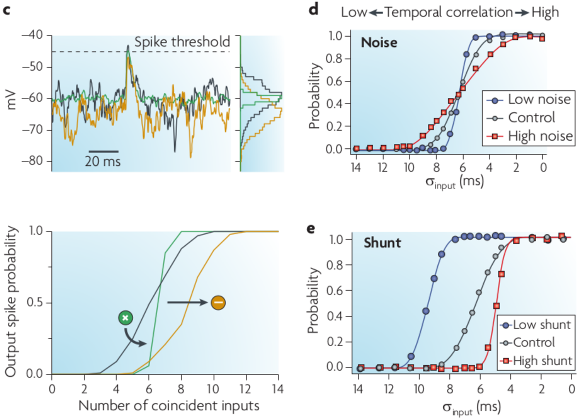



Figure: from Silver (2010), *Neuronal arithmetic*

Thank you for your attention!