Computational Neuroscience in action

1 Introductory Concepts
   1.1 Computational Neuroscience ................................................................. 1
   1.2 Modelling objectives ................................................................................... 2
   1.3 Integration of letters and phonemes in the human brain — fMRI results ......... 3

2 Bottom-up: from an artificial neuron to Self-Organizing Maps/Moduls (SOMs)
   2.1 A Concept Neuron — A simplistic model of a biological neuron ................. 5
   2.2 Forming simple neural networks — Lateral inhibition ............................. 7
   2.3 Representation of percepts/stimuli as multidimensional points/vectors ........ 8

3 Structure of Self-Organizing Maps/Modules
   3.1 Similarity layer .......................................................................................... 9
   3.2 General structure of a Self-Organizing Map/Module ................................. 10
   3.3 Example of a Self-Organizing Map .......................................................... 11
   3.4 Learning Algorithm for Self-Organizing Maps ........................................ 12
   3.5 Example of a Self-Organizing Feature Map formation ............................. 13
   3.6 The Multimodal Self-Organizing Networks ............................................. 15
   3.7 A Simple Feedforward Multimodal Self-Organizing Network ................. 16
   3.8 Surfaces of post-synaptic activities ......................................................... 17
   3.9 The unimodal visual map for letters ....................................................... 18
   3.10 The auditory map for phonemes ........................................................... 19
   3.11 The bimodal map integrating phonemes and letters .............................. 20
   3.12 Robustness of the bimodal percepts against unimodal disturbances .......... 21
   3.13 Feedback and its significance for auditory perception ......................... 23
   3.14 Re-coded Phoneme Map ........................................................................ 25
   3.15 Robustness of the re-coded phoneme map .......................................... 26
1 Introductory Concepts

1.1 Computational Neuroscience

• Computational Neuroscience is a name given to the effort of modelling the brain using formal mathematical and computational tools

• We can formulate models on different levels that can describe
  – structures and/or
  – functions
  of neuronal cells, its components and aggregates, e.g., a synapse, dendrite, neuron, columns of neurons, cortical areas, and so on.

• The model can use differential equation like the Hodgkin-Huxley model, or describe interconnection of abstracted neurons (artificial neural networks or connectionist models).

• We can describe
  – low level functions, like a generation of action potential, or
  – high level perceptual or cognitive functions like integration of multi-modal stimuli, or attention in autism.

• Such models, if good, are then incorporated in other areas of neuroscience or psychology.

A.P. Papliński

1.2 Modelling objectives

We present a model of integration of auditory and visual information as in the human cortex.

• In essence, such an integration will be described as a mapping of multi-dimensional sensory stimuli on two-dimensional cortical areas.

• The model is based on Multimodal Self-organizing Networks (MuSoNs) also known as Artificial Cortical Networks (ACNs)

• Such networks are formed from interconnected Self-Organizing Modules/Maps (SOMs) that, in turn, are composed of artificial neurons

• We consider
  – structure and
  the learning procedure
  of Self-Organizing Networks.

• We concentrate on a network that forms bi-modal representation of phonemes and letters.
1.3 Integration of letters and phonemes in the human brain — fMRI results

- Bimodal and more generally multimodal integration of sensory information is important for perception since many phenomena have qualities in more than one modality.
- Multimodal integration has been studied extensively, for a comprehensive review, see:
- Recently, functional magnetic resonance imaging (fMRI) has made non-invasive studies of multimodal identification and recognition by human subjects increasingly tractable and important, see:
- After: N. van Atteveldt, E. Formisano, R. Goebel, and L. Blomert, “Integration of letters and speech sounds in the human brain,” *Neuron*, vol. 43, pp. 271–282, July 2004.: Nowadays most written languages are speech-based alphabetic scripts, in which speech sound units (phonemes) are represented by visual symbols (letters, or graphemes). Learning the correspondences between letters and speech sounds of a language is therefore a crucial step in reading acquisition, ...
2 Bottom-up: from an artificial neuron to Self-Organizing Maps/Moduls (SOMs)

2.1 A Concept Neuron — A simplistic model of a biological neuron

Conceptual abstraction of biological neurons:

- Afferent Signals, $x$
- Excitatory Synapse
- Inhibitory Synapse
- Dendrite
- Post-synaptic activity
- Cell Body
- Axon
- Efferent Signal (Action potential)

Basic characteristics of a biological neuron for modelling:

- Information is coded in a form of instantaneous frequency of (action potential) pulses
- Synapses are either excitatory or inhibitory
- Signals are aggregated (“summed”) when travel along dendritic trees forming the post-synaptic activity
- The cell body generates at the axon the output pulse train of an average frequency proportional to the total (aggregated) post-synaptic activity.

The conceptual neuron of the above is still too complicated for most of the models and is further simplified to

- A single dendrite structure
- Operating with abstract values rather then with pulses of varying frequency:
  
  **Dendritic representation:**
  
  \[
  \begin{align*}
  \mathbf{x} &= [x_1, x_2, \ldots, x_n]^T \\
  \mathbf{w} &= [w_1, w_2, \ldots, w_n] \\
  d &= w_1 \cdot x_1 + \cdots + w_n \cdot x_n = \mathbf{w} \cdot \mathbf{x} \\
  y &= \sigma(d)
  \end{align*}
  \]

  - A single artificial neuron consists of $n$ synapses arranged along a linear dendrite.
  - The afferent signals $x_i$, arrive at synapses and are aggregated by the summing dendrite to form the post-synaptic activity $d$.
  - Each synapse is characterised by one parameter, the weight, or synaptic strength $w_i$.
  - The aggregation of signals is expressed as a linear combination or an inner product of afferent signals and synaptic weights:
  
  \[
  d = w_1 \cdot x_1 + \cdots + w_n \cdot x_n = \mathbf{w} \cdot \mathbf{x}
  \]
  
  - The post-synaptic activity, $d$, is passed through a limiting activation function, $\sigma(\cdot)$, which generates the axonal efferent signal $y$
2.2 Forming simple neural networks — Lateral inhibition

In the simple neural network that models lateral inhibition as in limulus vision (see http://www.esse.monash.edu.au/courseware/cse2330/2006/Lnts/L05D.pdf for details)

- Light intensity signals arrive from a number of visual receptors marked \( R \) to synapses of neurons,
- Neuronal axons form the local feedback excitatory connection, and inhibit the neighbouring neurons (lateral inhibition)
- Such connections form a spatial high-pass filter enhancing edges of the image received from the receptors.

A.P. Papliński

2.3 Representation of percepts/stimuli as multidimensional points/vectors

- Perceptual objects can be represented as points/vectors in multidimensional space.
- Consider an example of categorization of animals.
- The following 18 features are chosen to characterize each animal:

\[
\begin{align*}
x_1 & \; \log(\text{weight}) \\
x_2 & \; \{1, 2, 3\} \; \text{food (herbivores, omnivores, carnivores)} \\
x_3, x_4, x_5, x_6 & \; \{0, 1\} \; \text{locomotion (fins, wings, two legs, four legs)} \\
x_7 & \; \{0, 1\} \; \text{equipped with hooves (perissodactyls) or cloven hooves (artiodactyls)} \\
x_8 & \; \{0, 1\} \; \text{equipped with claws} \\
x_9 & \; \{0, 1\} \; \text{equipped with other feet} \\
x_{10} & \; \{1, 2, 3\} \; \text{cover (fur, feathers and scales)} \\
x_{11} & \; \{0, 1\} \; \text{colour black} \\
x_{12} & \; \{0, 1\} \; \text{colour white} \\
x_{13} & \; \{0, 1\} \; \text{even coloured} \\
x_{14} & \; \{0, 1\} \; \text{spotted} \\
x_{15} & \; \{0, 1/4\} \; \text{striped} \\
x_{16} & \; \{2, 4\} \; \text{facial feature (short faced, long faced)} \\
x_{17} & \; \{1, 2, 3\} \; \text{aquatic} \\
x_{18} & \; \{1, 2, 3\} \; \text{social behaviour (single living, pair living, group living)}.
\end{align*}
\]

- Therefore, each animal can be perceived as a point in 18-dimensional space:

\[
\mathbf{x} = [x_1, x_2, \ldots, x_{18}]
\]

- The animal kingdom will be represented by a number of points in such a space
- The 18 signals/number form the feature vector that is applied as afferent signals to our neural networks

3 Structure of Self-Organizing Maps/Modules

Self-Organizing Maps/Modules (SOMs) also known as Kohonen maps or topographic maps were first introduced by von der Malsburg (1973) and in its present form by Kohonen (1982).

3.1 Similarity layer

- In self-organizing maps/modules, the synaptic weight vector, \( w_i \), associated with each neuron approximates some aspects of feature vectors/stimuli \( x \) applied to its synapses.
- It can be shown that the post-synaptic activity \( d_i \) is a measure of similarity between the synaptic weight vector \( w_i \) and the applied stimulus \( x \).
- the more similar the vectors, the larger postsynaptic activity \( d_i \).

3.2 General structure of a Self-Organizing Map/Module

A Self-Organizing Map/Module (SOM) consists of three main parts/layers:

- A similarity layer that measure the similarity of the current stimulus to each synaptic weight vector.
- A competition layer that identifies the winning neuron, the one that has the synaptic weight vector most similar to the current stimulus.
- Positional information. Each neuron is located with respect to each other on a two-dimensional grid known as the feature space. Alternatively, each neuron can store its position relative to other neurons on the grid.
- Mathematically we say that for a given matrix of neuronal weights, \( W \), the network maps \( k \)th stimulus \( x(k) \) onto a position in the neuronal grid of the winning neuron \( v(k) \), that is, the neuron with the weight vector most similar to the stimulus.

\[
v(k) = g(x(k); W) ; \quad v \in \mathcal{R}^l
\]  

In other words, for each stimulus, one neuron, the winner of the competition, is most activated.
3.3 Example of a Self-Organizing Map

Consider a Self-Organizing Map/Module consisting of \( m = 12 \) neurons located on a 2-D \( 3 \times 4 = 12 \) neuronal grid:

- The postsynaptic activity in the similarity layer indicates the distance of the current stimulus from the weight vector.
- The competition layer is based on the lateral inhibition and local self-excitatory connections and enhances the response of the winning neuron suppressing the activities of other neurons.

A.P. Papliński

3.4 Learning Algorithm for Self-Organizing Maps

- The objective of the learning algorithm for a SOM is to capture in the synaptic weights essential characteristics of the stimuli.
- The learning algorithm consists of two basic steps, namely, competition and cooperation between neurons of the output grid.

**Competition:** the current stimulus \( x(k) \) is compared with each weight vector \( w_j(k) \) and the winning neuron \( j(k) \) and its position on the neuronal grid is established.

**Cooperation:** The winning neurons and its neighbours have their weight vectors modified pulling them in the direction of the current stimulus.

\[
\Delta W = \eta(k) \cdot \Lambda(j, i) \cdot (x^T(k) - w_j(k))
\]

with the strength \( \Lambda(j, i) \) related to their distance \( \rho(j, i) \) from the winning neuron.

- The **neighbourhood function**, \( \Lambda(j, i) \), is usually a Gaussian function:
- The neighbourhood function should initially include all neurons and shrink when the learning progresses so that finally the neighbourhood function includes only the winning neuron.
- The learning gain \( \eta \) is gradually reduced to stabilize the algorithm.

A.P. Papliński
3.5 Example of a Self-Organizing Feature Map formation

- In this example we consider two-dimensional stimuli generated by two sources marked by by o and + respectively.
- We use three groups of data each containing 10 stimuli (i.e. points in a two-dimensional space) in each source, sixty points all together.
- The sources can be thought of as producing, e.g., two dialects of a very limited protolanguage, each with three protophonemes.

- We can imagine that a source is a representation of a parent of a child pronouncing three phonemes in ten slightly different ways.
- The parallel is far from perfect but might be helpful for a conceptual understanding of the map formation and interpretation.
- Real sensory stimuli, like the phonemes of speech are of course larger in number and dimension.

Let our Self-Organizing Map/Module have 9 neurons organized on a $3 \times 3$ grid
- The $3 \times 3$ neuronal grid is to approximate 60 stimuli organised in three pairs of clusters.
- Initially, the 2-D weight vectors marked with ‘*' in the plot, are assigned random values.
- During each learning step, the weight vector closest to the stimulus is pulled towards the stimulus together with its closed neighbours.

- The resulting feature map assigns nine neurons to three pairs of stimuli clusters (60 stimuli),
- Note that neurons are assigned either to data clusters, or to groups of clusters to have the best approximation of data distribution.
- Three neurons in the centre of the grid seem to be not assigned to any group.

- Note that the synaptic weight vectors form an elastic net the is shaped to approximate the stimuli distribution.
- Finally observe that if the number of neurons is significantly higher than the number of stimuli, then each stimuli will be assigned to a patch of neurons.
3.6 The Multimodal Self-Organizing Networks

- A Multimodal Self-Organizing Network (MuSON) is built from several Kohonen Self-Organizing Modules (SOMs).

- For the purpose of modelling integration of phonemes and letters as in the cortex we consider a two-level feedforward structure and a three-level feedback network.

- Bimodal integration takes place in SOM_{bm} module.

- Re-coded phoneme module SOM_{lt} provides better interpretation of noisy stimuli thanks to inclusion of the visual information.

- The sensory level consists of a visual module processing letters, SOM_{lt} and an auditory module processing phonemes, SOM_{ph}.

3.7 A Simple Feedforward Multimodal Self-Organizing Network

- Unisensory maps, SOM_{lt} and SOM_{ph}, receive sensory stimuli, \( x_{lt} \) and \( x_{ph} \).

- The bimodal map, SOM_{bm}, receives combined pair of three-dimensional outputs from sensory maps, \( y_{lt} \).

- The learning process can take place concurrently for each SOM, according to the Kohonen learning law.

- After self-organization each SOM performs the mapping of the form:

  \[ y(k) = g(x(k); W, V) \]

  where \( x(k) \) represents the \( k^{th} \) stimulus for a given map, \( W \) is the weight matrix, and \( V \) describes the structure of the neuronal grid.

- The 3-D output signal \( y(k) = [d(k) \ v(k)] \) combines the 2-D position of the winner \( v(k) \) with its 1-D post-synaptic activity \( d(k) \).
3.8 Surfaces of post-synaptic activities

The trained map post-synaptic activity is:

\[ d(k) = W \cdot x(k) \]

- Example of the postsynaptic activity in the trained bimodal map consisting of 36 × 36 neurons when the stimuli representing letter/phoneme ő are presented to:
  - the visual letter map and
  - the auditory phoneme map

- The activity in the map for one phoneme/letter combination shows one winning patch with activity descending away from this patch.

36 × 36 neuronal grid

- The patch of neurons representing ő is clearly distinguished from other patches.
- Collections of such patches forms the final maps.

3.9 The unimodal visual map for letters

- The stimulus \( x_{\text{lt}} \) to the visual letter map is a 22-element vector obtained by a principal component analysis of scanned \((21 \times 25)\)-element binary character images.
- After self-organization the visual letter map SOM\(_{\text{lt}}\) is represented by patches of highest post-synaptic activities of a neuronal population consisting of 36×36 neurons

- The map shows the expected similarity properties — symbols that look alike are placed close to each other in the map.
- For example, the cluster of characters f, t, r, i, l, k, b and p based predominantly on a vertical stroke.
- The neuronal populations within the patches of the letter map constitute the detectors of the respective letter.
3.10 The auditory map for phonemes

- The auditory material consists of twenty-three phonemes as spoken by ten native Swedish speakers.
- Each phoneme spoken by each speaker is represented by thirty-six melcepstral coefficients.
- These feature vectors were averaged over the speakers, yielding one thirty-six element feature vector for each phoneme.
- This averaged set of vectors constitutes the inputs $x_{ph}$ to the auditory map SOM$_{ph}$.

![Phoneme map](image)

- After the learning process we obtain a phoneme map.
- As for the letter map the patches of neuronal populations constitute the detectors of the respective phonemes.
- Note the plosives g, k, t, and p, the fricatives s, S (sh-sound) and f, and the nasal consonants m and n form three close groups on the map.
- Also vowels with similar spectral properties are placed close to each other.
- The exact placing of the groups vary from one self-organization to another, but the existence of these groups is certain.

3.11 The bimodal map integrating phonemes and letters

- The outputs from the auditory phoneme map and the visual letter map are combined as 6-dimensional inputs $[y_{lt} y_{ph}]$ to the bimodal map SOM$_{bm}$.

![Bimodal map](image)

- The patches of highest post-synaptic activity combine characteristics of letters and phonemes.
- Note, for example, groups of the fricative consonants s, S and f, and the nasal consonants m and n.
- Most, but not all, vowels form a group and those who are isolated have obviously been placed under influence from the visual letter map.
3.12 Robustness of the bimodal percepts against unimodal disturbances

- An important advantage of integration of stimuli from sensory-specific cortices into multimodal percepts in multimodal association cortices is that even large disturbances in the stimuli may be eliminated in the multimodal percepts:

- We excite the network with the three letters i, â and m which are all uncorrupted.
- The corresponding phonemes i, â and m are heavily corrupted as shown.
- These corrupted phonemes cause the activity on the phoneme map to move significantly from the original positions.

- In the bimodal map the activities have moved very little.
- The recognition of these bimodal percepts is much less influenced by the auditory corruptions than the recognition in the phoneme map.
- This holds for all other letter/phoneme combinations as well.
- The above testing was done for the feedforward network.
3.13 Feedback and its significance for auditory perception

The robustness of the bimodal percepts, demonstrated previously can be employed to benefit through feedback to enhance auditory perception.

- We introduce feedback in our MuSON through the re-coded phoneme map SOM$_{rph}$.
- The 6-dimensional input stimuli to the re-coded phoneme map $[y_{ph}, y_{bm}]$ is formed from the feedforward connection from the sensory phoneme map SOM$_{ph}$ and the top-down feedback connection from the bimodal map SOM$_{bm}$.

• The first phase of self-organization is identical as for the feedforward network.
• In the second phase of self-organization the re-coded phoneme map SOM$_{rph}$ is initialized to have the 23 winners in the same positions as in the phoneme map SOM$_{ph}$.
• The weight vectors are then trained by the Kohonen rule.
• Relaxation is included between two maps in the feedback loop.
3.14 Re-coded Phoneme Map

- After self-organization the re-coded phoneme map $SOM_{rph}$ is seen to be similar to the phoneme map $SOM_{ph}$.

- The bimodal map is similar to that in the feedforward case.

3.15 Robustness of the re-coded phoneme map

- The re-coded phoneme map through its feedback connections has a dramatically different property when phonemes are corrupted.

- The map shows post-synaptic activities for the three letters and phonemes, i, ā and m.

- Solid lines show the results when the auditory and visual inputs to the sensory-specific maps are perfect.
Dotted lines show the results when the auditory inputs to the unisensory phoneme map are heavily corrupted.

Note that when phonemes are corrupted the activities in the re-coded phoneme map change insignificantly compared to the activities in the phoneme map.

This holds for all other letter/phoneme combinations as well.

Conclusion

With a Multimodal Self-Organizing Network we have simulated bimodal integration of audiovisual speech elements, phonemes and letters.

We have demonstrated that the bimodal percepts are robust against corrupted phonemes, and that when these robust bimodal percepts are fed back to the auditory stream the auditory perception is greatly enhanced.

These results agree with known results from psychology and neuroscience.