Modelling basic perceptual functions

Dr Andrew P Paplinski, Assoc. Prof.
Faculty of Information Technology
Monash University
Melbourne, Australia

based on recent papers co-authored with

- Prof. Lennart Gustafsson , Luleå Tech. Uni, Sweden
- Dr. William M. Mount, Uni. NSW, Australia

ISDA 2014, Okinawa, Japan

Outline

- Introduction, inspiration, motivation
- Modelling hierarchical systems that can integrate Perceptual Objects with Spoken and Written Names
- Building blocks: functions, input and output signals
- Three versions of such systems:
  - binding concepts to spoken names,
  - binding written words to mental objects,
  - integrating visual and auditory stimuli.
- Working with signals on hyper-spheres.
- Incremental learning
- Transferring knowledge between perceptual systems.

How it started

- The work on modelling perception originates from our earlier involvement in modelling autism.
- Autism is considered to be a complex developmental disorder and one of its manifestations is the attentional deficit that we modelled.
- We have obtained some results related also to the problem of early intervention.
- At this stage we decided to model the “normal” brain first and to come back to the autistic brain.
- We have not finished the first part yet.

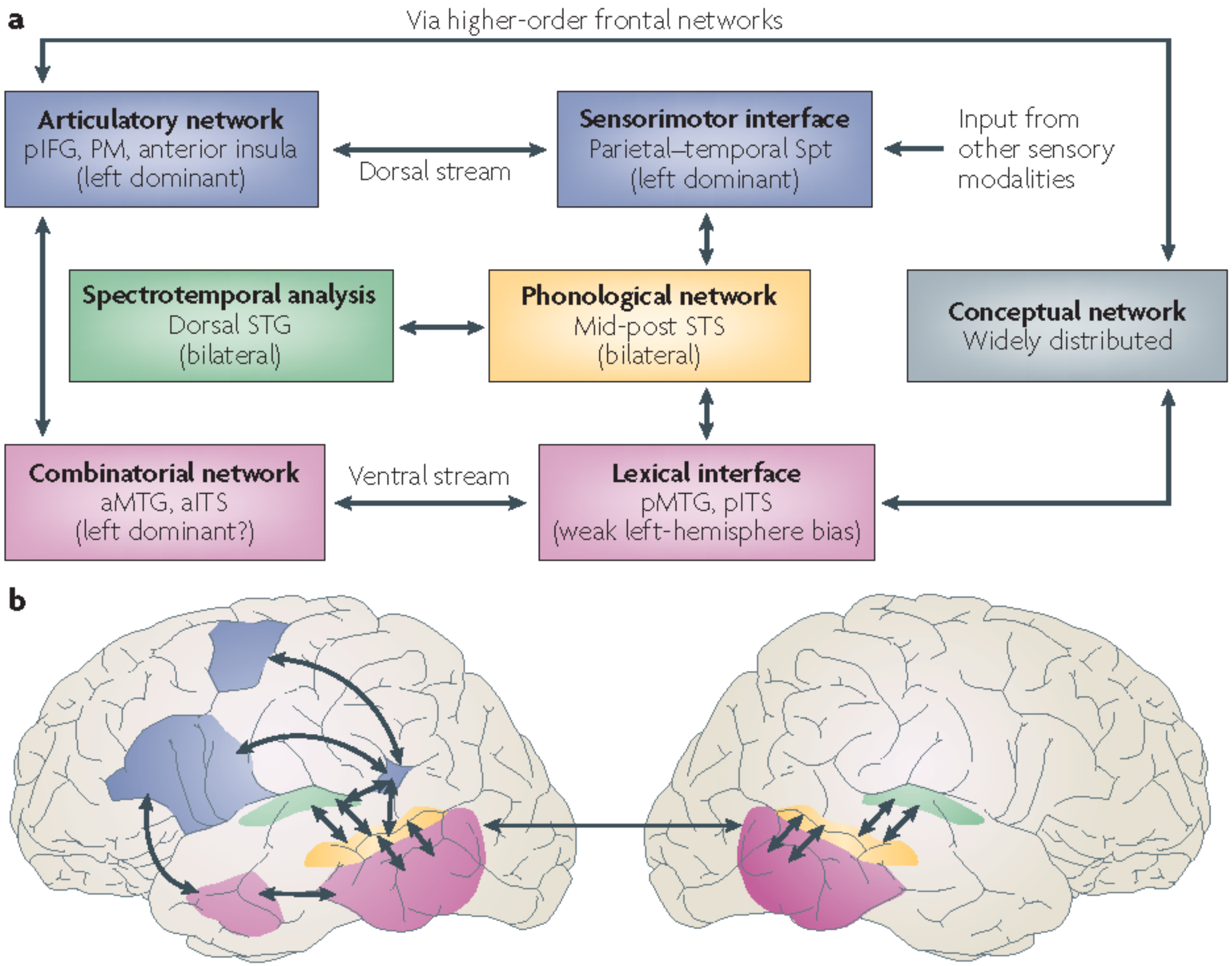
Perception

- Perception describes the way in which our brain interprets sensory information and creates the representation of the environment.
- We study systems that can integrate visual and auditory sensory information and bind it to the internal mental concepts.
- Two divergent objectives in studying how the brain works:
  – medical aspects
  – computational aspects

Inspiration 1: Speech Processing

Dual stream model:

G.Hickok & D.Poeppel: *The cortical organization of speech processing.* Nature Rev., Neurosci., vol.8, 2007



- Spectro-temporal analysis module
- phonological network

from which the processing diverges into two broad streams:

- the articulatory stream
- the lexical stream

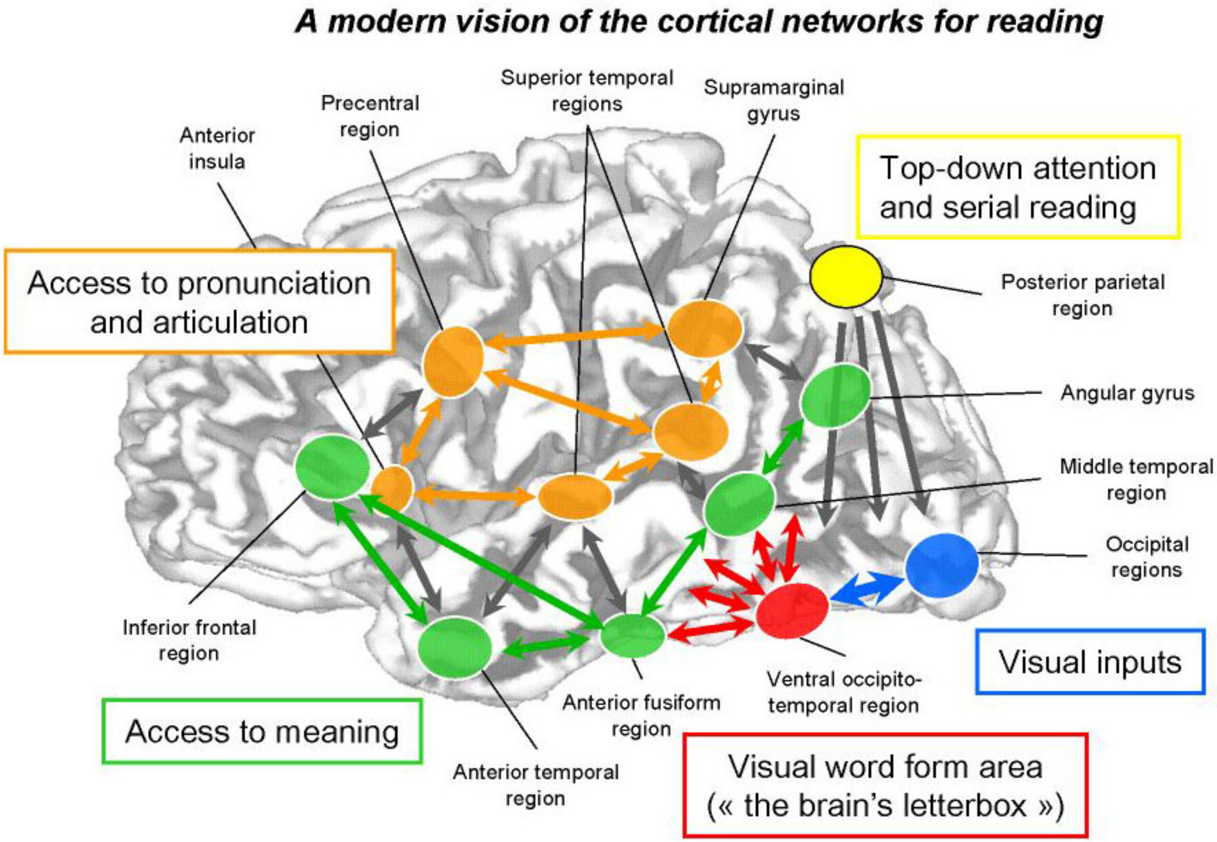
These two streams are interconnected by

- combinatorial network integrating lexical and articulatory processing,
- conceptual, higher-level network

Inspiration 2: Reading in the brain

S. Dehaene, *Reading in the Brain*, Viking 2009

Thirteen interconnected cortical areas, arranged in five groups:



- visual input,
- visual word form,
- access to meaning,
- access to pronunciation and articulation,
- top-down attention and serial reading.

Modelling action plan

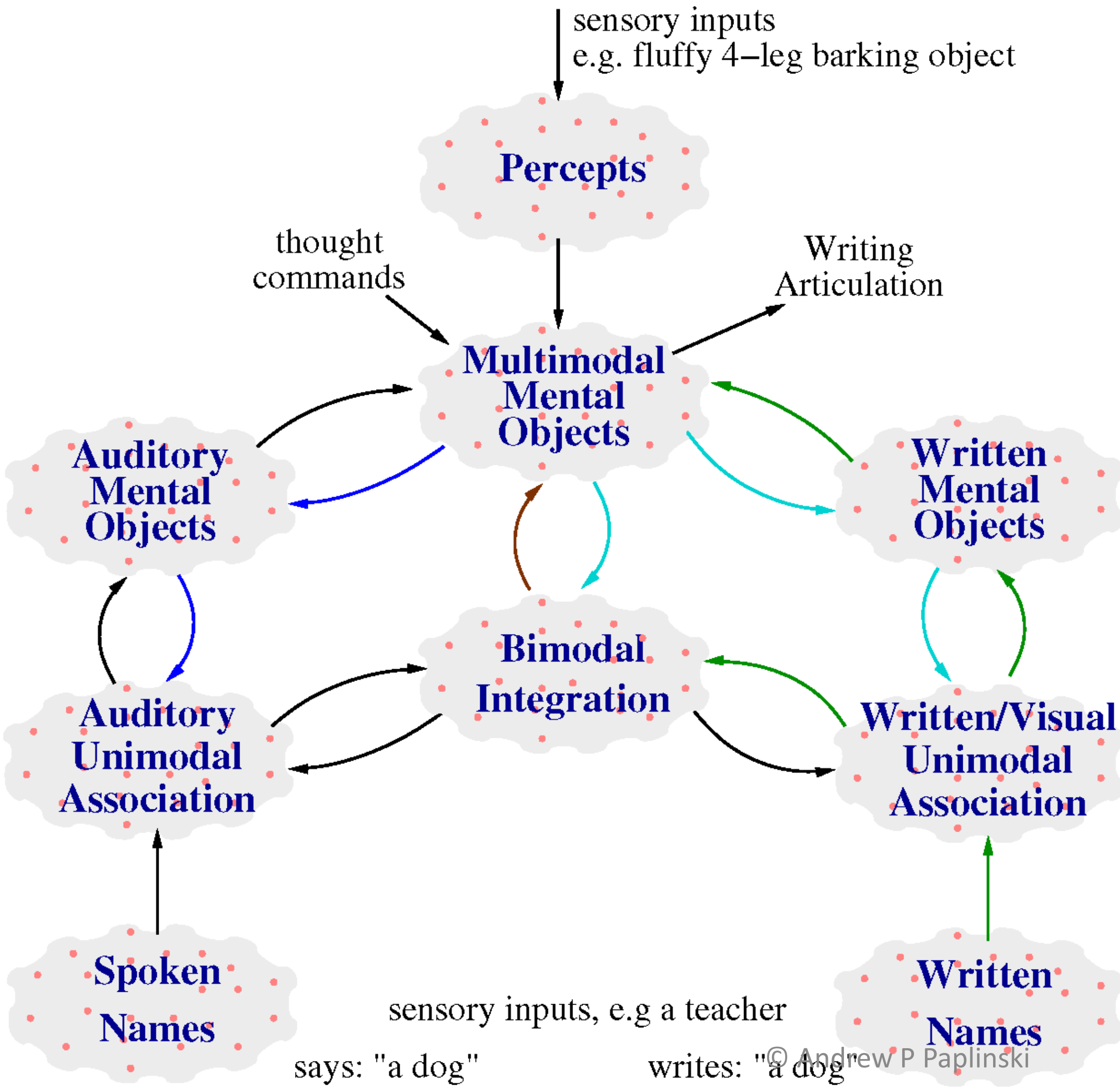
Common to both models is:

- phenomenological description of functions attributed to cortical areas,
- specification of interconnections between areas

Our action plan for modelling is to:

- formally specify functions/mappings of selected "cortical-like" areas.
- Specify signals between the areas in terms of a **uniform "neuronal code"**

Example: Integrating Perceptual Objects with Spoken and Written Names

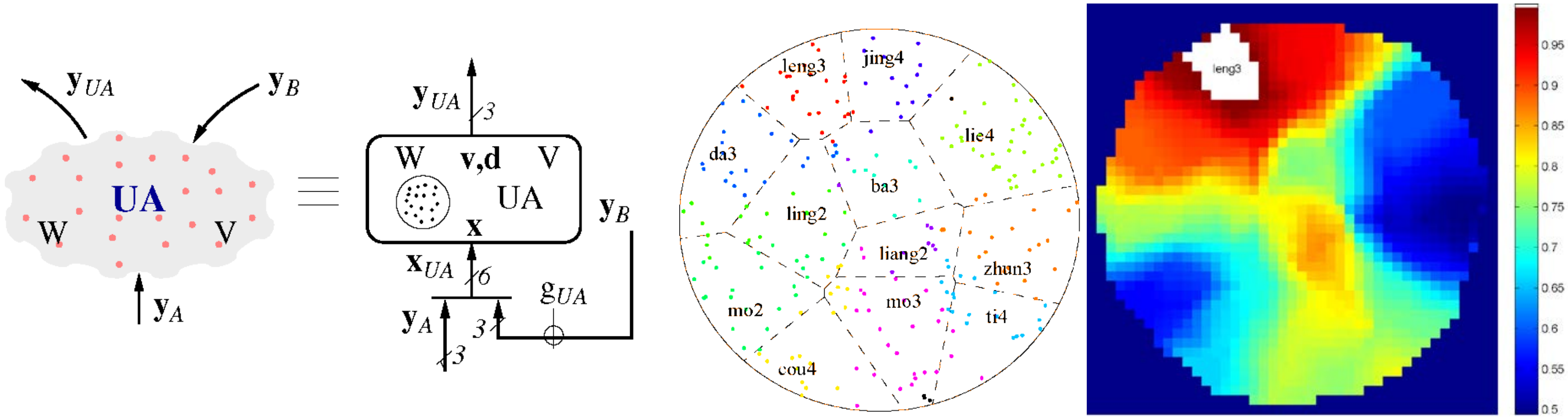


- Imagine a child learning about animals.
- Three types of sensory inputs and information processing path:
  - perceptual,
  - auditory (speech),
  - visual (written names)
- Sensory data is converted in a “neuronal code” also produced by all modules
- The codes are combined as the afferent signal to “cortical” modules
- Nine modules mapping input/output signals

Outline

- Introduction, inspiration, motivation
- Modelling hierarchical systems that can integrate Perceptual Objects with Spoken and Written Names
- Building blocks: functions, input and output signals
- Three versions of such systems:
  - binding concepts to spoken names,
  - binding written words to mental objects,
  - integrating visual and auditory stimuli.
- Working with signals on hyper-spheres.
- Incremental learning
- Transferring knowledge between perceptual systems.

A **building block** (module) maps signals from the input space to the latent/neuronal space



- A module, (e.g. Self-Organizing Map, SOM) performs mapping of input signals $\boldsymbol{x}_{UA}$ into the **latent/neuronal space** represented by colour dots located at points $\boldsymbol{v}_{UA}$
- The input signals $\mathbf{x}_{\mathrm{UA}}$ applied at the "synapses" of the module, and representing related objects, are combined with the synaptic weights $\mathbf{W}_{\mathrm{UA}}$ of all neuronal units into the **postsynaptic activity/strength** $d_{UA}(\boldsymbol{v}_{UA}) = \boldsymbol{W}_{UA} \cdot \mathbf{x}_{\mathrm{UA}}$
- Each object, e.g. *leng3* (a label) is mapped into a group of neuronal units, say, $\gamma = 20$.
- The neuron located at $\boldsymbol{v}_w$ with the highest postsynaptic strength d_w is call the **winner**.
- The output signal $\boldsymbol{y}_{UA} = [\boldsymbol{v}_w, d_w(\boldsymbol{v}_w)]$ aka **neuronal code**, combines the position of the winner with its postsynaptic activity/strength
- In other applications the number of neuronal nodes is smaller that the number of data points aka objects

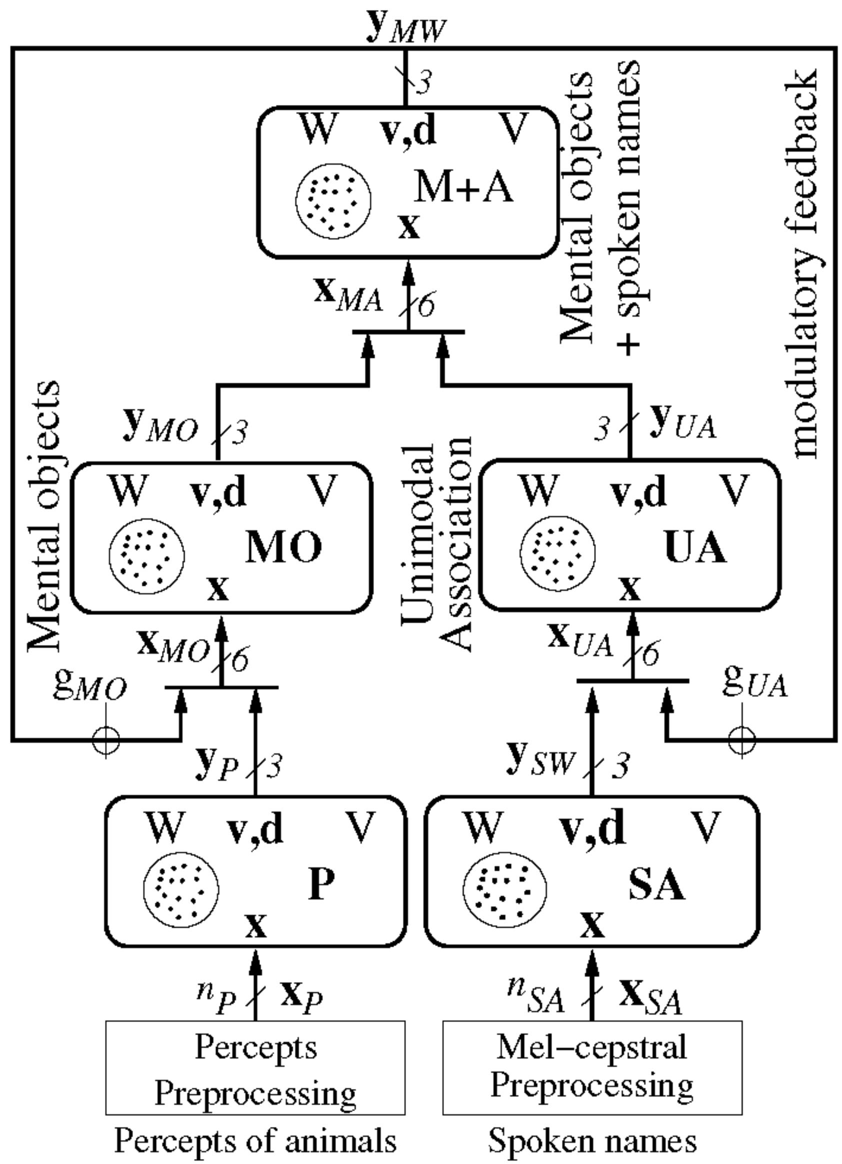
© Andrew P Paplinski

Outline

- Introduction, inspiration, motivation
- Modelling hierarchical systems that can integrate Perceptual Objects with Spoken and Written Names
- Building blocks: functions, input and output signals
- Three versions of such systems:
  - **binding concepts to spoken names,**
  - binding written words to mental objects,
  - integrating visual and auditory stimuli.
- Working with signals on hyper-spheres.
- Incremental learning
- Transferring knowledge between perceptual systems.

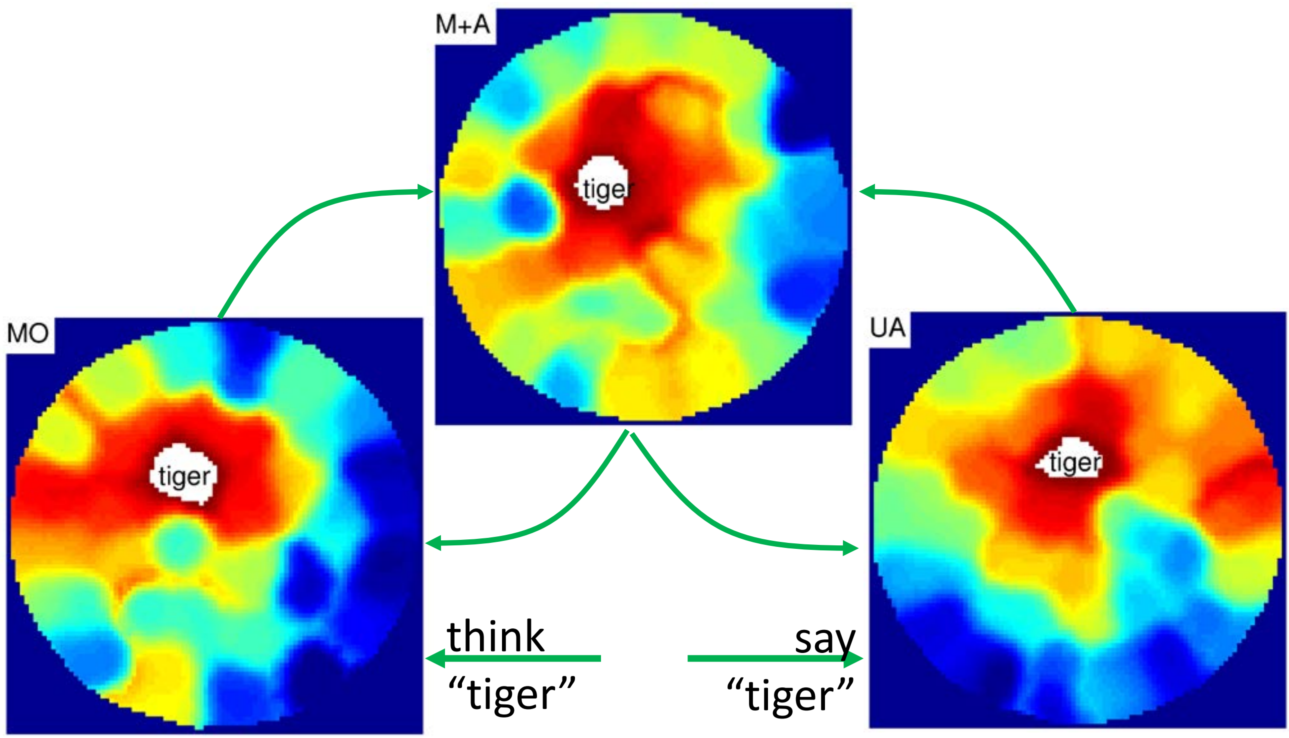
Binding **percepts** (of animals) with their **spoken** names

- Sensory observation/features of animals are converted into their semantic description or percepts
- The spoken names are coded in frequency domain: time samples are replaced by 36 mel-cepstral coefficients
- Two sensory level modules: **P** (storing percepts aka mental objects) and **SA** (storing internal representation of spoken words)
- At the top level, **M+A**, mental objects are bound with the spoken names
- Two intermediate level modules, **MO** and **UA**, accommodate the **modulatory feedback** from **M+A**



Binding Percepts (of animals) to Spoken Names

- The learning process develops the maps
- After learning we can test the behaviour of the maps for different percepts and names



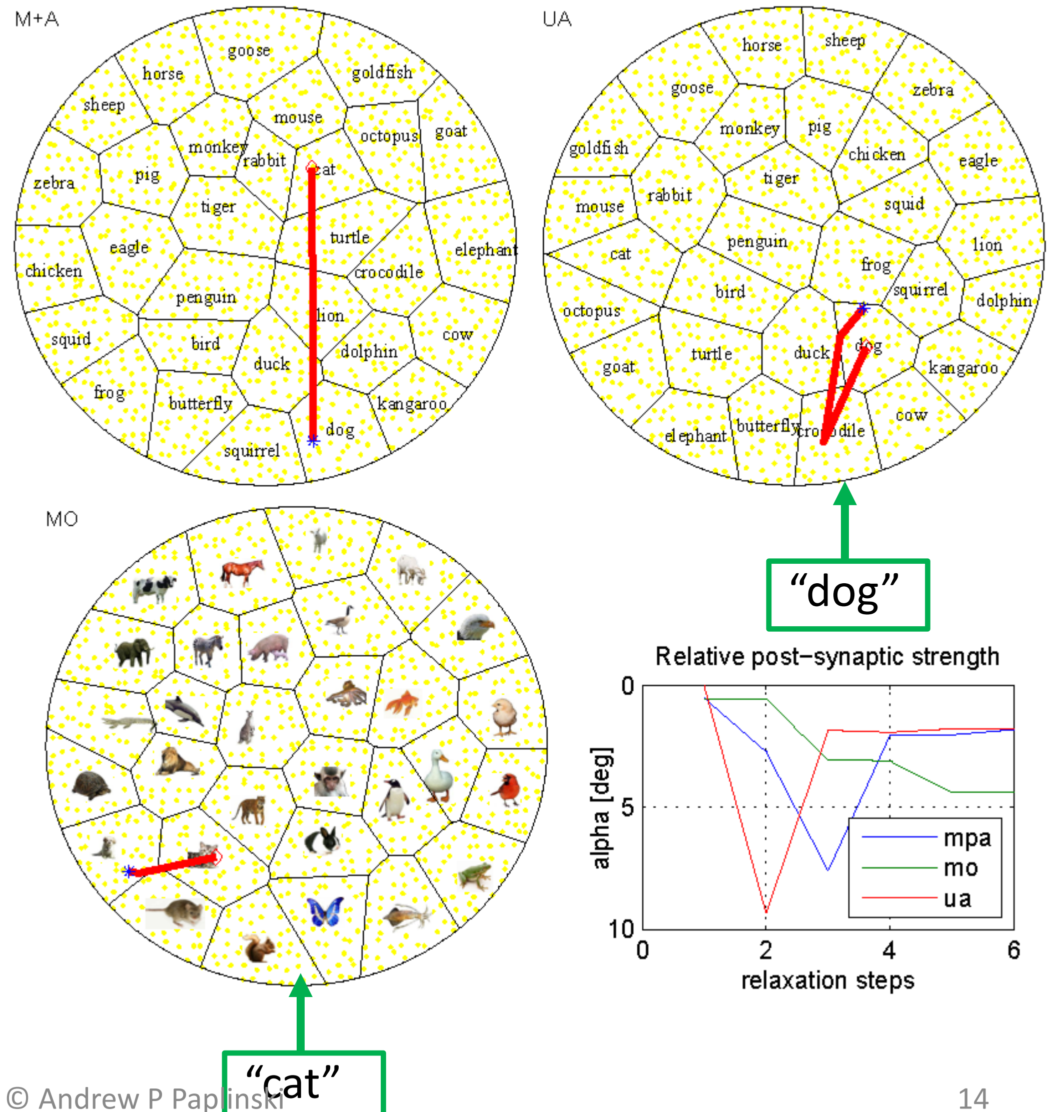
- During testing with congruent thought and spoken name the system quickly settles for the percept, e.g. "tiger"
- In the case of incongruent thoughts and names at least two cases can be considered: when either objects, or names are similar, e.g.

"cat – dog" or "frog – dog"

© Andrew P Paplinski

Similar percepts, dissimilar names

- The modules try to negotiate between the conflicting thoughts (think "cat", hear "dog")
- Initial values of postsynaptic strength is at the maximum and after six relaxation steps settle at the lower final values.
- Similar percepts make the auditory entry prevailing:
- all maps settle for "dog" with the varying degree of confidence measured by d

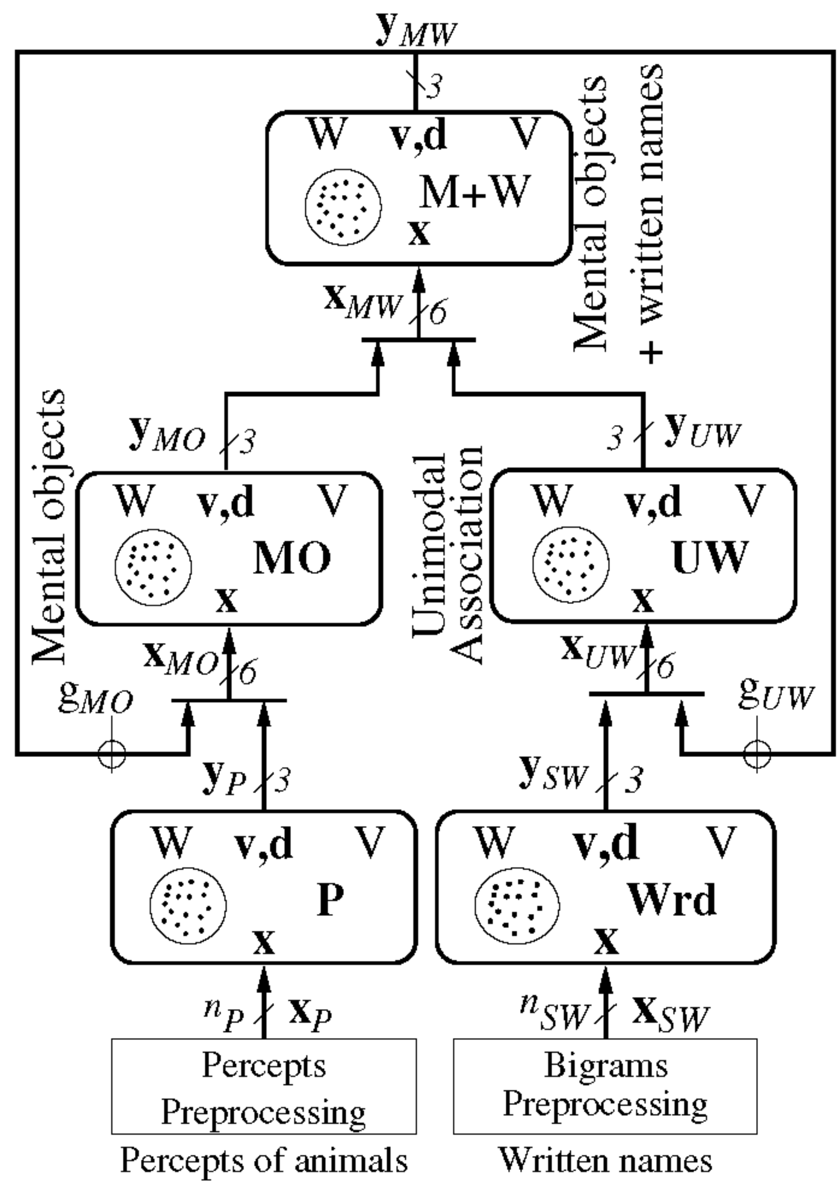


Outline

- Introduction, inspiration, motivation
- Modelling hierarchical systems that can integrate Perceptual Objects with Spoken and Written Names
- Building blocks: functions, input and output signals
- Three versions of such systems:
  - binding concepts to spoken names,
  - **binding written words to mental objects**,
  - integrating visual and auditory stimuli.
- Working with signals on hyper-spheres.
- Incremental learning
- Transferring knowledge between perceptual systems.

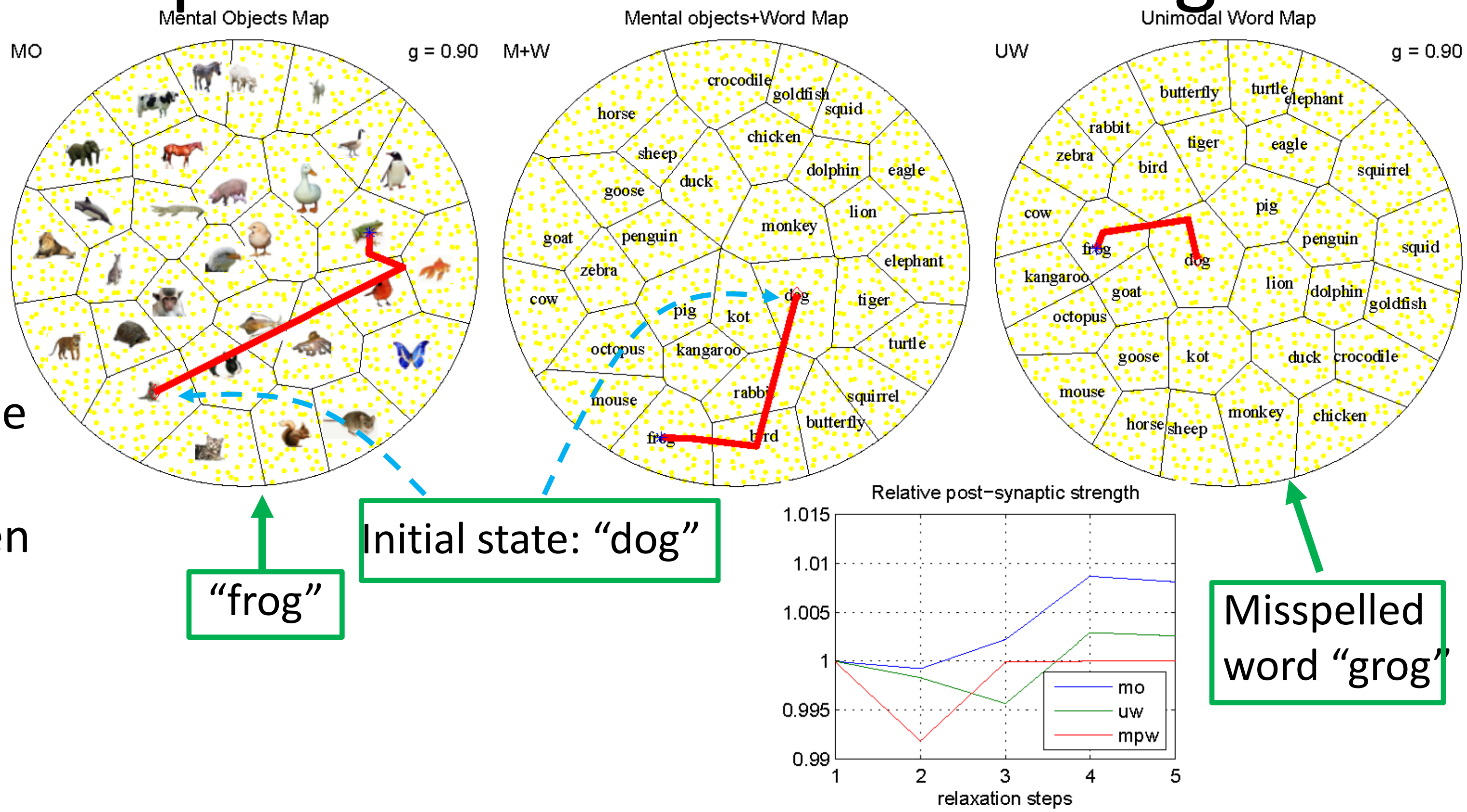
Binding **percepts** (of animals) with their **written** names

- Same five maps as before
- Written names of animals are pre-processed and converted into **bigrams**
- Two sensory level modules: **P** (storing percepts aka mental objects) and **Wrd** (storing internal representation of written words)
- At the top level, **M+W**, mental objects are bound with the written names
- Two intermediate level modules, **MO** and **UW**, accommodate the **modulatory feedback** from **M+W**



Percepts + written names: testing

- The learning process develops the maps
- After learning we can test the behaviour of the maps when percepts are incongruent with the written words



- Trajectories in the association maps go from initial "dog" to the percept "frog"
- The misspelled name "grog" is corrected in the Unimodal Word map **UW**
- The confidence of the proper guess is measured by the postsynaptic strength, $d = \boldsymbol{w} \cdot \boldsymbol{x}$ normalised to 1 for the "learned" object.
- The feedback loops settle in five relaxation steps.
- Note the values of the feedback gains.

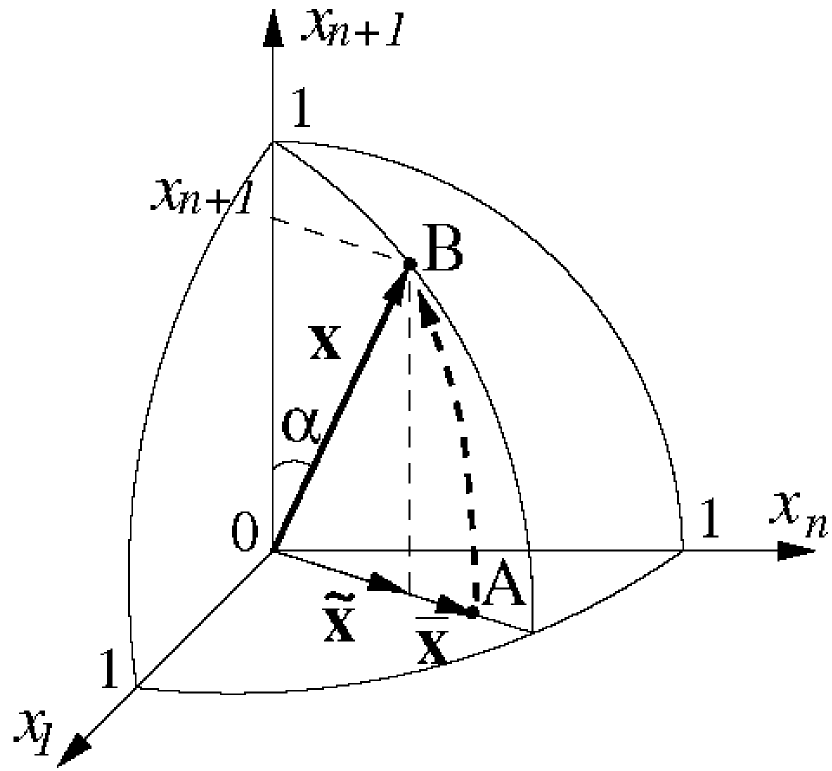
© Andrew P Paplinski

Outline

- Introduction, inspiration, motivation
- Modelling hierarchical systems that can integrate Perceptual Objects with Spoken and Written Names
- Building blocks: functions, input and output signals
- Three versions of such systems:
  - binding concepts to spoken names,
  - binding written words to mental objects,
  - integrating visual and auditory stimuli.
- Working with signals on hyper-spheres.
- Incremental learning
- Transferring knowledge between perceptual systems.

Technicality: putting vectors on unity hyper-spheres.

- All multidimensional data:
  - Sensory data, $\boldsymbol{x}_{\mathrm{S}}$
  - Internal neuronal codes $\boldsymbol{x}$
  - Weight vectors $\boldsymbol{w}$
  - Neuronal position vectors $\boldsymbol{v}$
- are projected on a unity hyper-sphere
- Hence, we work with unity vectors.
- The distance between vectors is calculated as inner product



Comments re. Learning

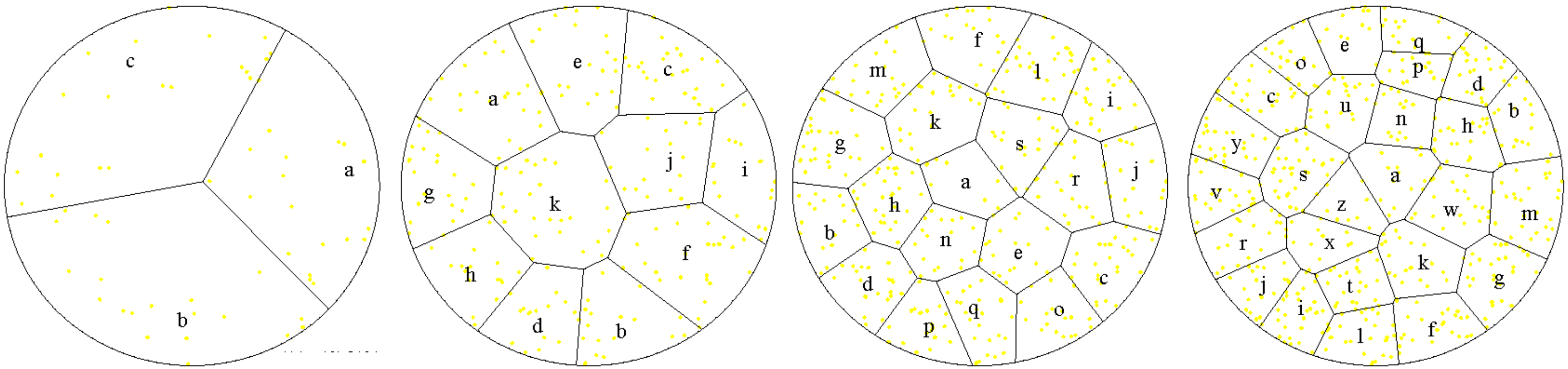
- The objective of learning is
  - to map multidimensional input objects/vectors into neuronal/latent space in such a way that
  - vectors close to each other in the input space remains such in the latent space
- In addition, in our case, we aim at maintaining stochastically constant ratio of neuronal units to the objects, e.g. $\gamma = 20$
- The motivation comes from the redundancy required in biological systems and ability to place noisy signals within the neurons allocated to the given objects
- Two learning systems are considered:
  - Kohonen SOMs with dot-product learning law,
  - Elastic Nets, ENs, implementing Gaussian Mixture Models (GMM) with the Expectation Maximization learning law

Outline

- Introduction, inspiration, motivation
- Modelling hierarchical systems that can integrate Perceptual Objects with Spoken and Written Names
- Building blocks: functions, input and output signals
- Three versions of such systems:
  - binding concepts to spoken names,
  - binding written words to mental objects,
  - integrating visual and auditory stimuli.
- Working with signals on hyper-spheres.
- Incremental learning
- Transferring knowledge between perceptual systems.

Incremental learning

Demo



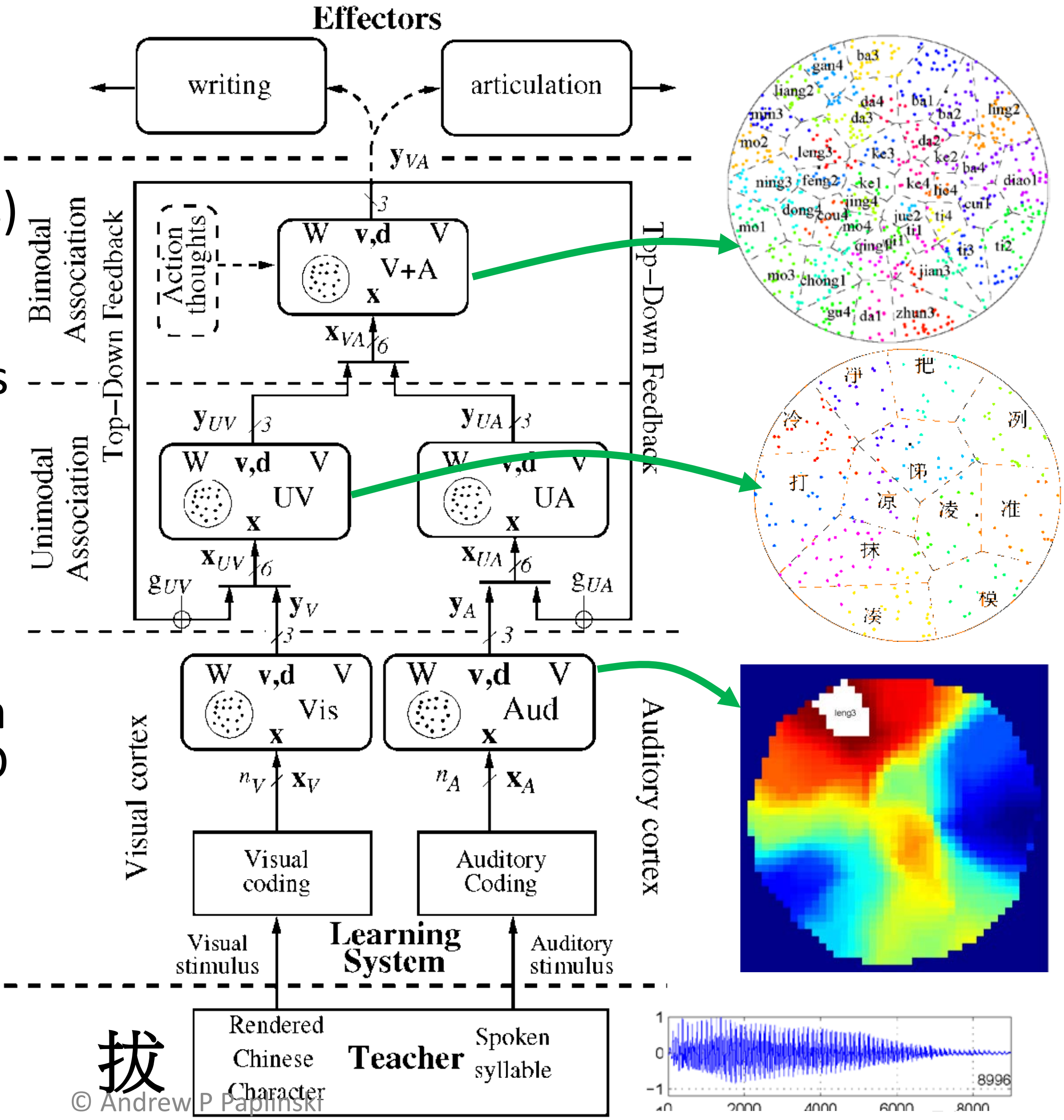
- Start with some initial number of stimuli (three in the example) and nodes ($3\gamma = 60$)
- Apply the selected learning law.
- For n added new objects we generate additional $n\gamma$ neuronal units randomly distributed in the neuronal space.
- The selected learning law is applied again
- As expected, at each stage the map organizes the stimuli according to their visual features, e.g., keeping \`f’, \`l’, and \`i‘ together.

Outline

- Introduction, inspiration, motivation
- Modelling hierarchical systems that can integrate Perceptual Objects with Spoken and Written Names
- Building blocks: functions, input and output signals
- Three versions of such systems:
  - binding concepts to spoken names,
  - binding written words to mental objects,
  - integrating visual and auditory stimuli.
- Working with signals on hyper-spheres.
- Incremental learning
- Transferring knowledge between perceptual systems.

Integrating written and spoken Chinese

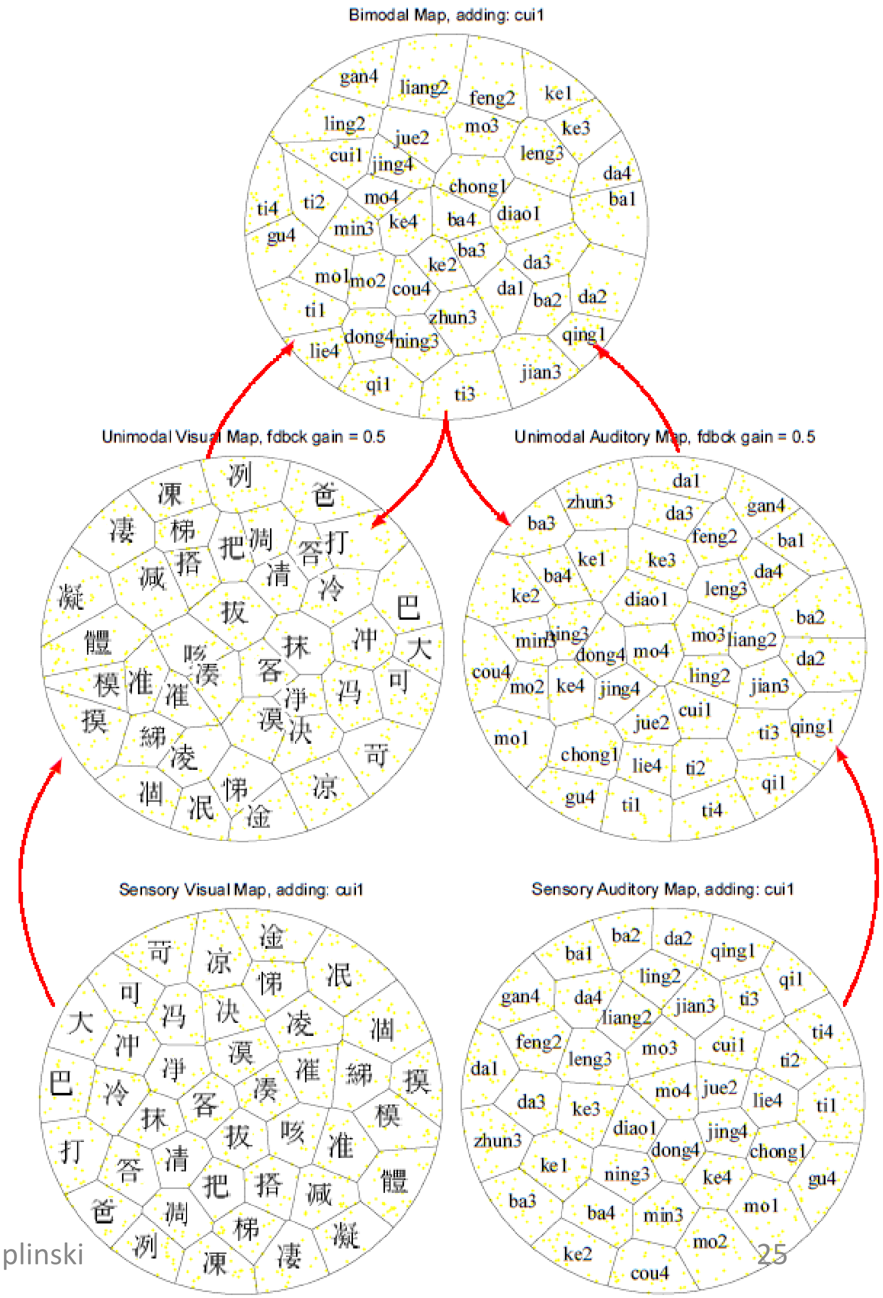
- A learning system **incrementally** maps stimuli of different modalities (Chinese characters and related Mandarin utterances) into the latent spaces.
- Note a number of hierarchical processing levels and modulated feedback
- Each afferent signal at each module excites the group of neuronal unit
- Location of the highest excited unit and the strength of the excitation form the 3D “neuronal code”
- Bimodal association module store the accumulated knowledge and can drive writing and articulation effectors



© Andrew P Paplinski

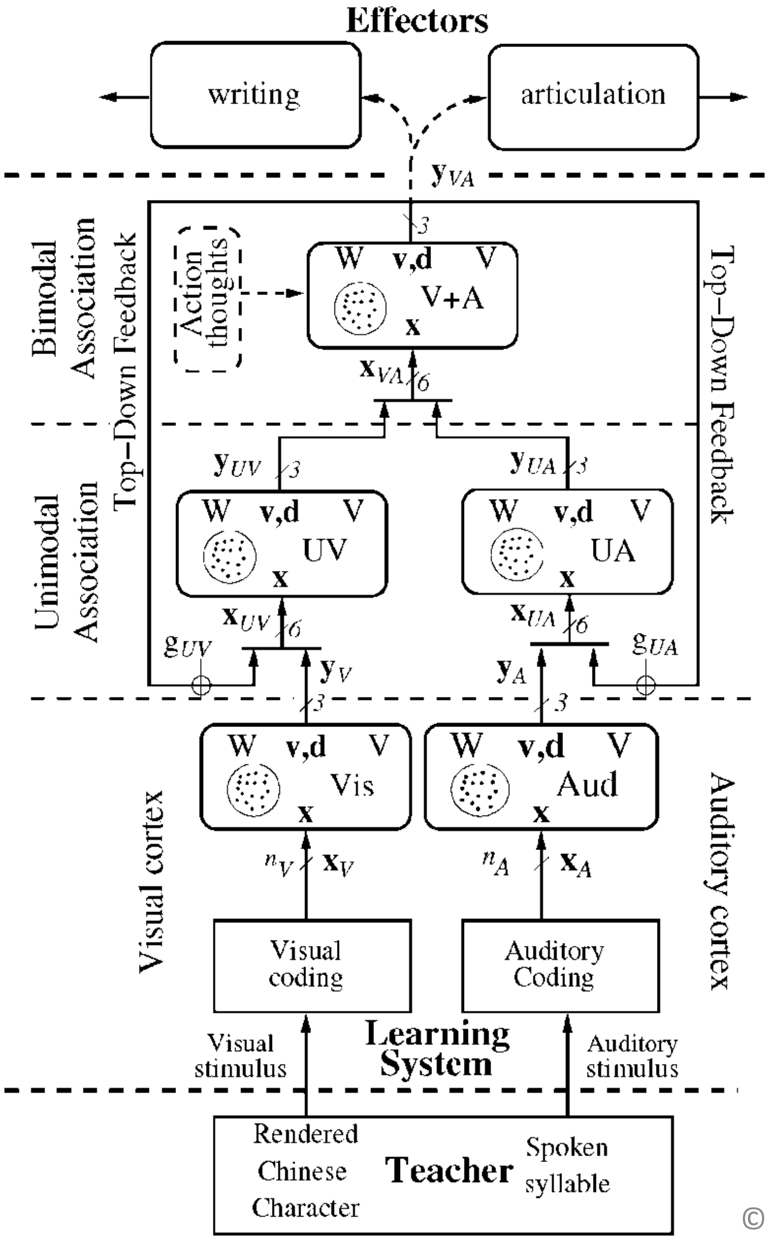
Integrating written and spoken Chinese

- Showing mapping in all modules after incremental learning
- Chinese characters are converted into vectors using the angular integral of Radon transform (aniRT)
- Mandarin utterances are coded using 36 mel-cepstral coefficients
- Sensory maps show similarities based on the respective coding vectors.
- Unimodal association maps combine information from the sensory modules and from the bimodal module



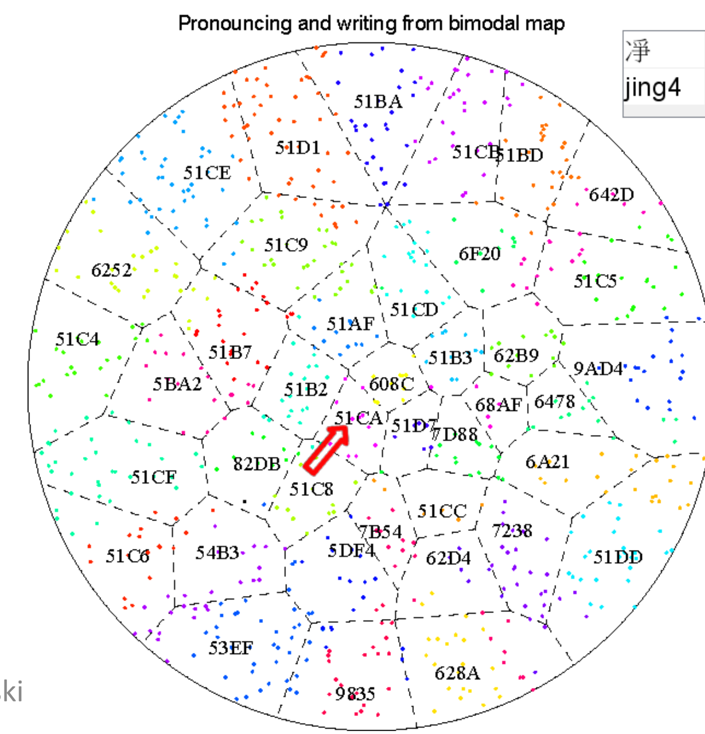
© Andrew P Paplinski

Writing and Articulation



- The bimodal module accumulates the body of the system knowledge
- We assume that an endogenous "action thought" applied to the bimodal module can induce the writing and/or articulation action, e.g.

[demo]

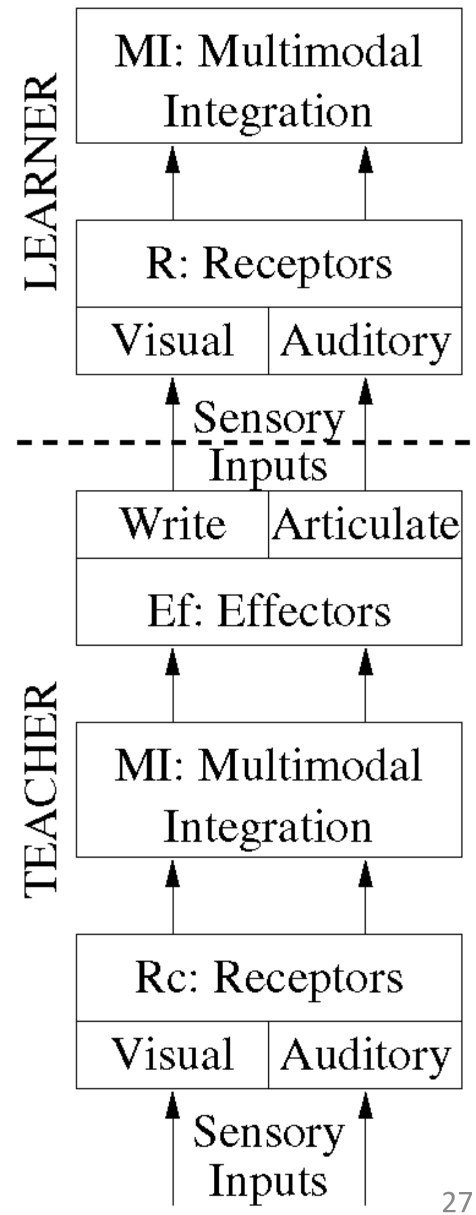


Closing the loop: from teacher to learner

- The output from the articulation and writing effectors can be used as an input to another learning system.
- We consider this issue in the paper presented in this conference

Each system has three main parts:

- Rc — Receptors that receive the external sensory information, auditory and visual in our case,
- MI — Multimodal Integration part that interprets the sensory information and incorporates it within the internal knowledge structure of self-organizing modules
- Ef — Effectors that produce an external representation of knowledge, articulation and writing effectors in our case.



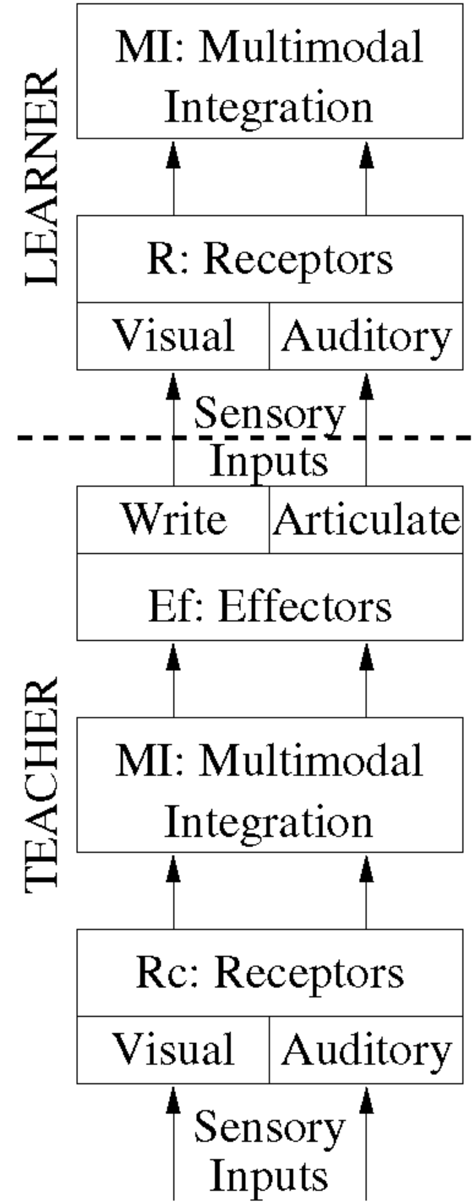
© Andrew P Paplinski

Outline

- Introduction, inspiration, motivation
- Modelling hierarchical systems that can integrate Perceptual Objects with Spoken and Written Names
- Building blocks: functions, input and output signals
- Three versions of such systems:
  - binding concepts to spoken names,
  - binding written words to mental objects,
  - integrating visual and auditory stimuli.
- Working with signals on hyper-spheres.
- Incremental learning
- Transferring knowledge between perceptual systems.

Transferring the knowledge from teacher to learner

- The teacher has its knowledge stored in the three modules:
  - two unimodal association modules, *UV* and *UA*
  - the bimodal map representing the top level of the system hierarchy.
- The transfer of knowledge between the teacher and the learner can occur in one of the following three modes:
  - Incrementally from the "fully learned" teacher.
  - Concurrently with the teacher in the incremental way,
  - All in one step (batch mode)

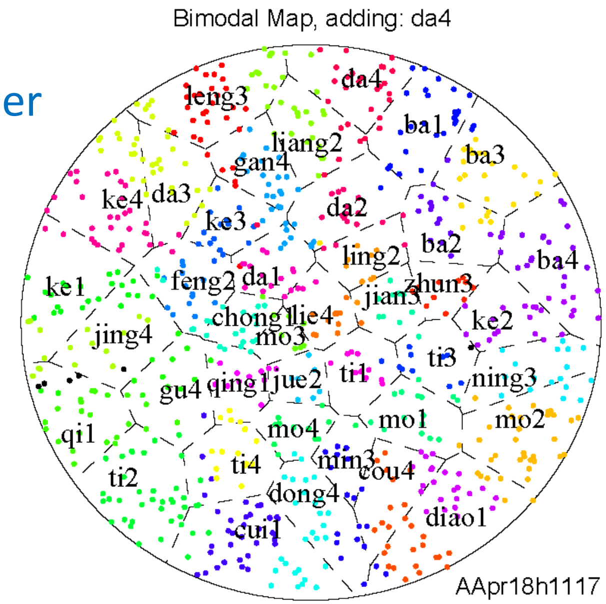


© Andrew P Paplinski

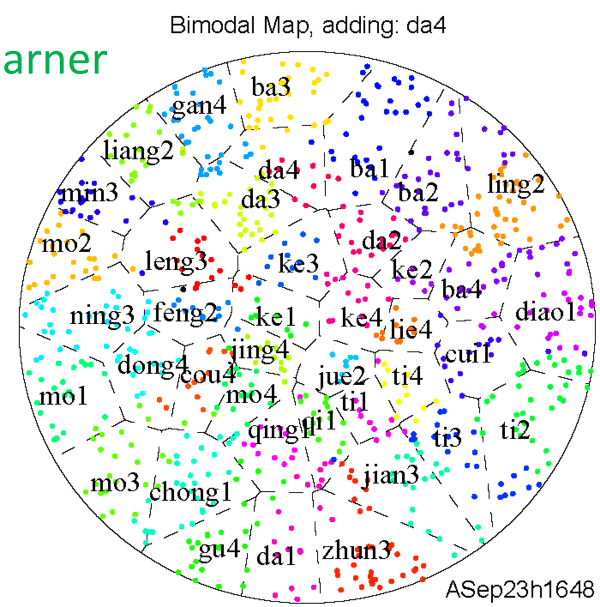
Example of the incremental learning

- Example of bimodal maps for the teacher and the learner.
- The teacher and the learner maps are different
- the teacher and the learner are different individuals in the sense that they have formed different bimodal associations between the written and spoken language components,
- More generally: they created different views of their limited "worlds" due to the history of the learning process.

teacher



learner



© Andrew P Paplinski

Summary

- We model aspects of perception showing how
  - meaning,
  - speech,
  - reading,
  - writing,

  can be integrated together inside learning systems.
- Building blocks of the systems are self-organizing modules (Kohonen SOMs or Elastic Nets)
- The blocks generate a universal "neuronal code" which combines the position of the winner in the neuronal/latent space with its post-synaptic strength.
- We show how knowledge can be transferred between the teacher and the learner systems.