Neurobiologically plausible modeling of speech production and comprehension for improving our understanding of normal and disordered speech

Bernd J. Kröger

Department for Phoniatrics, Pedaudiology, and Communication Disorders, Medical School

RWTH Aachen University, Aachen, Germany

Neurobiologically plausible modeling of speech production and comprehension for improving our understanding of normal and disordered speech

UNIKLINIK RWTHAACHEN

Bernd J. Kröger

Department for Phoniatrics, Pedaudiology, and Communication Disorders, Medical School

RWTH Aachen University, Aachen, Germany

Neurobiologically plausible modeling of speech production and comprehension for improving our understanding of normal and disordered speech

A neural network model of the brain

in which we can insert specific neural dysfunctions at specific locations of the cortex

gives us a clear association of neural dysfunctions and symptoms of speech disorders (resulting from simulation studies)

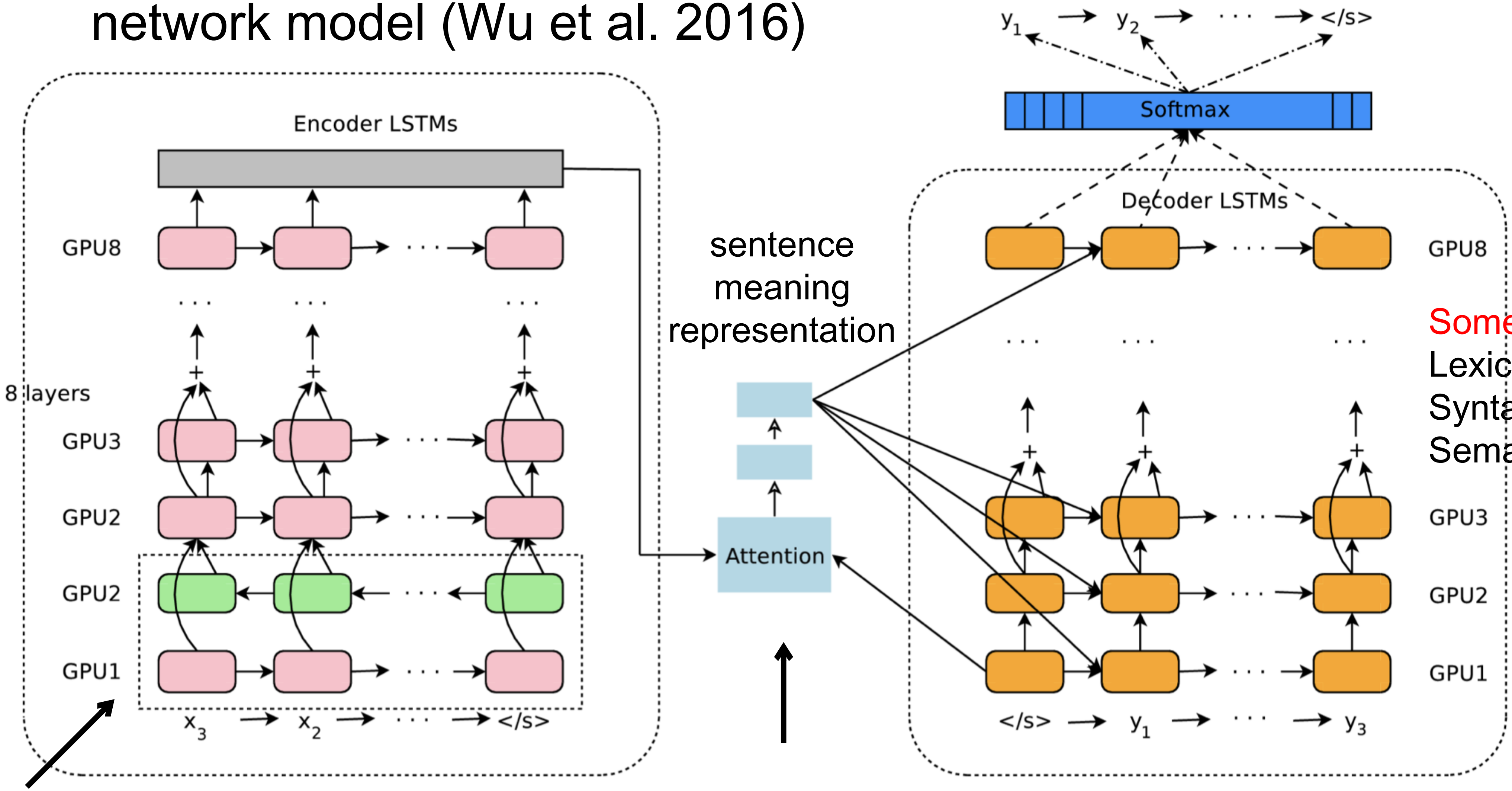
In practice: hard to recruit enough “well diagnosed” patients (concerning type and severity of speech disorder) willing to participate in clinical studies

We want a Neural Network Model like: Google's machine sentence translation network model (Wu et al. 2016)

High-level performance system

similar system:
part of chatGPT

language 2
word sequence output



sentence
meaning
representation

Somewhere inside:
Lexical processing
Syntactic processing
Semantic processing

Relatively equally structred / unstructured ensemble of modules

word sequence input
language 1

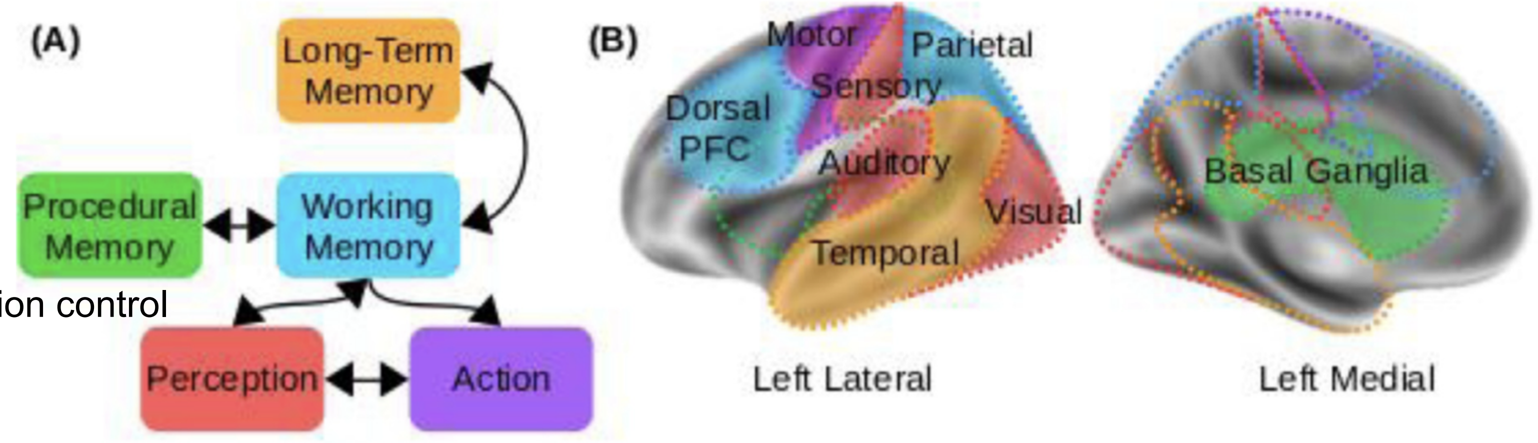
Interpretation of the neural activation patterns appearing here is difficult!!!

Goal of modeling: neurobiological plausibility

- Three points: (i) Model must in accordance with neurophysiological / neuroanatomical data (imaging, EEG, …), and behavioral data -> “box-and-arrow” models

functional modules

location of modules



action control

large-scale functional architecture of the brain:

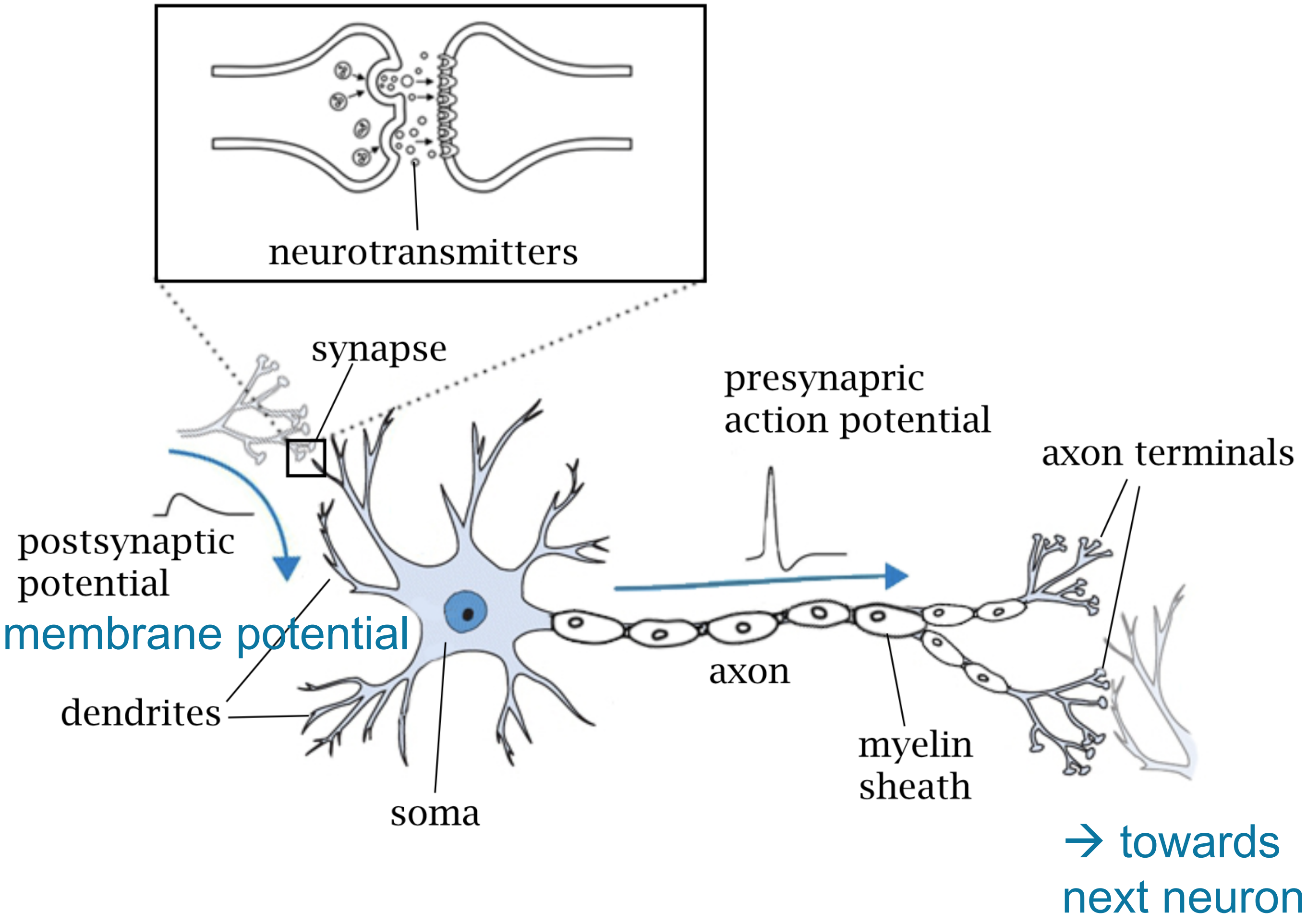
Common model of Cognition (CMC): Stocco et al. 2021

domain / task independent: speaking, gesturing, situational reasoning, relational reasoning, math task solving, solving any concrete or abstract task …

Goal of modeling: neurobiological plausibility

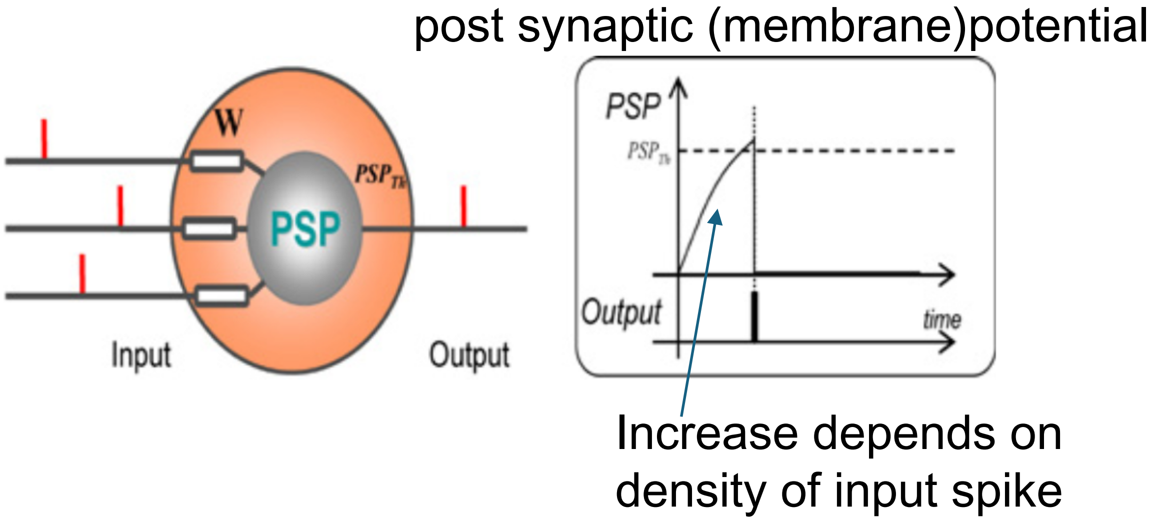
- (ii) realistic neuron model: spiking neuron model as “atomic unit” ->
  - leaky integrate and fire neuron model
  - plus: synapse model (exhibitory, inhibitory) (different degrees of strength for pule forwarding)
  - -> trained “link weights”

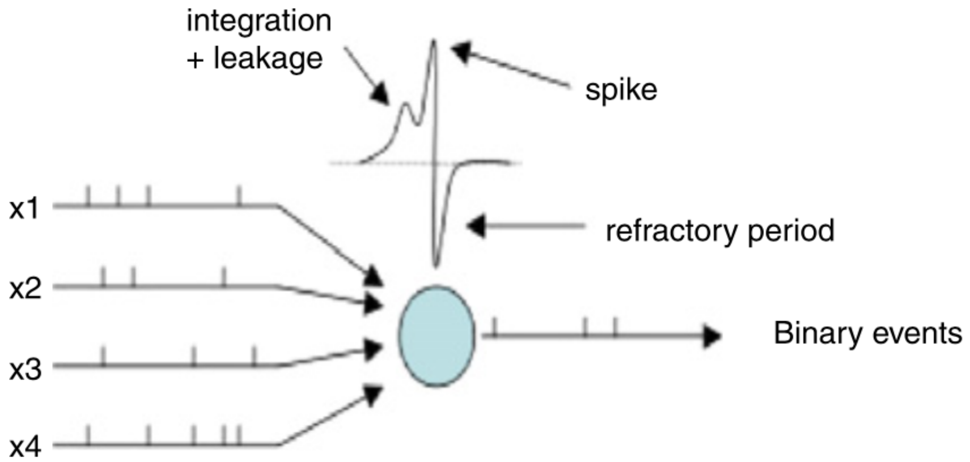
No longer: temporal-spatial averaging of “neural activation level”
(-> 2./3.generation - NNs)



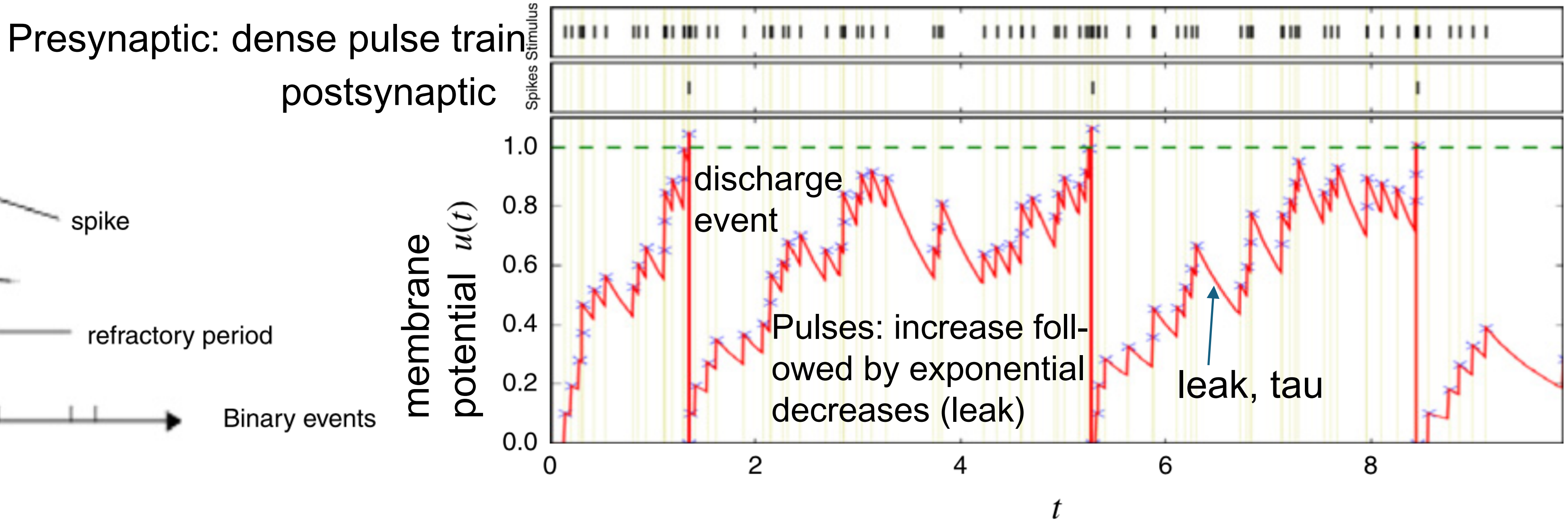
LIF neuron model

refractory period = about 10 msec





(a) The structure of a LIFM.



(b) Functionality of a LIFM.

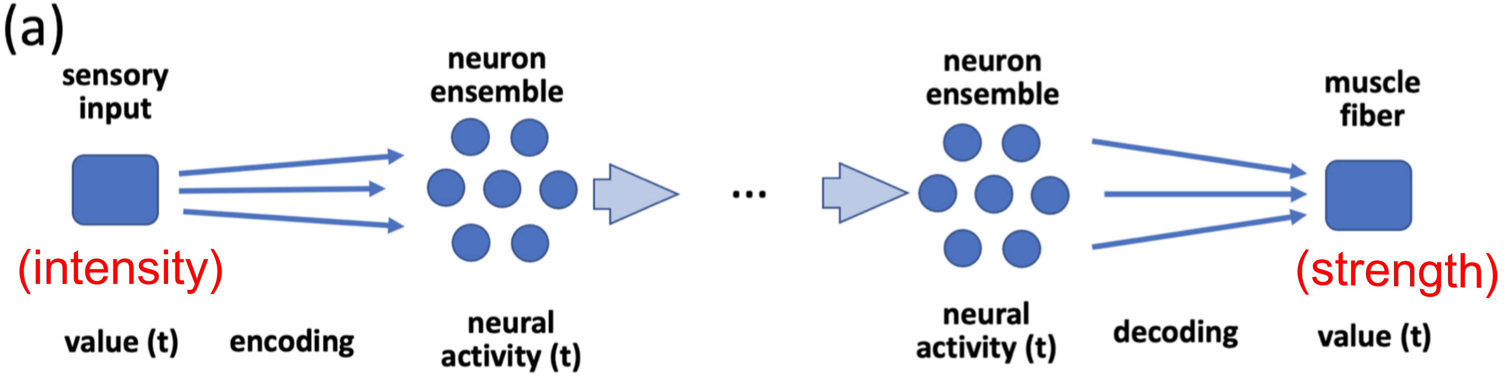
Leak: exponential decrease of membrane potential -> parameter tau describes the dynamics of the LIF neuron

Goal of modeling: neurobiological plausibility

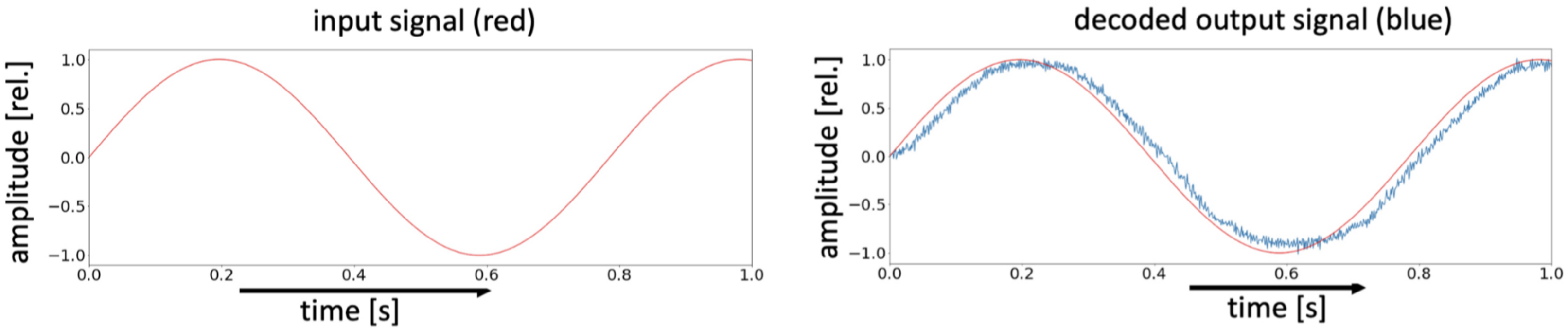
- (iii) build up the model using (probably genetically coded) **canonical neural network elements** (functional circuits)
  - Neural network units with a specific neural-level function, appearing in a large-scale neural network at many places
  - tiny building-blocks within all modules of the large-scale model (-> CMC model)
- Examples:
  - neural buffers (ensembles) for input encoding / output decoding of "values" (next slide)
  - forwarding and processing lower-level information (2nd next slide)
  - specific recurrent neural buffers (ensembles) for generating dynamics (oscillations, short-term memories) (3rd next slide)

defining **neural buffers** (neuron **ensembles**) for input encoding / output decoding

representing a “value”

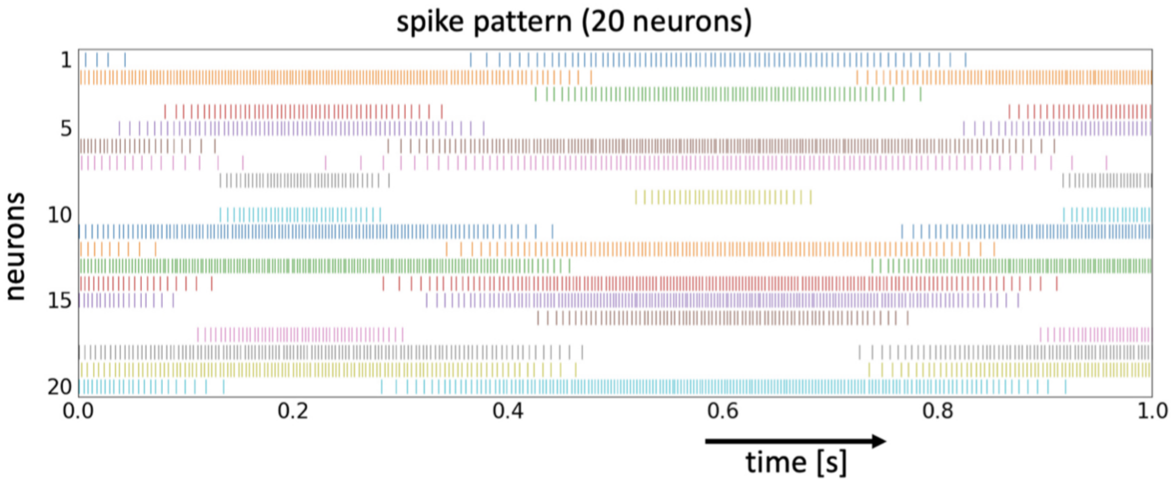


representing a “value” over time



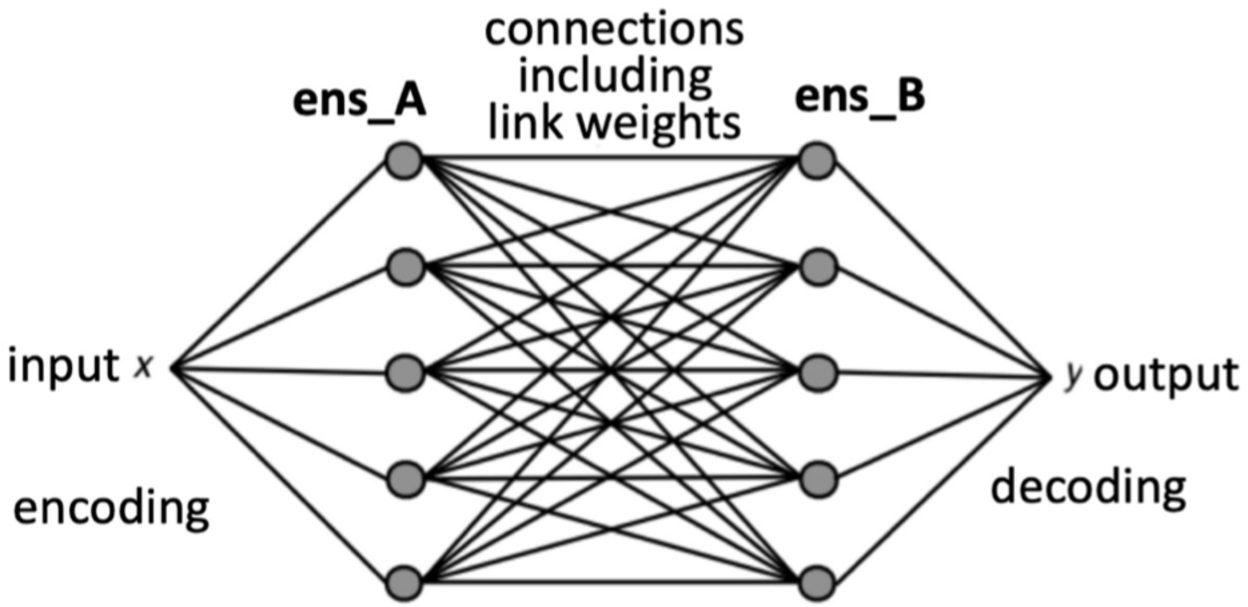
NEF
Eliasmith 2013

N = 20…100 neurons
depending on accuracy needed

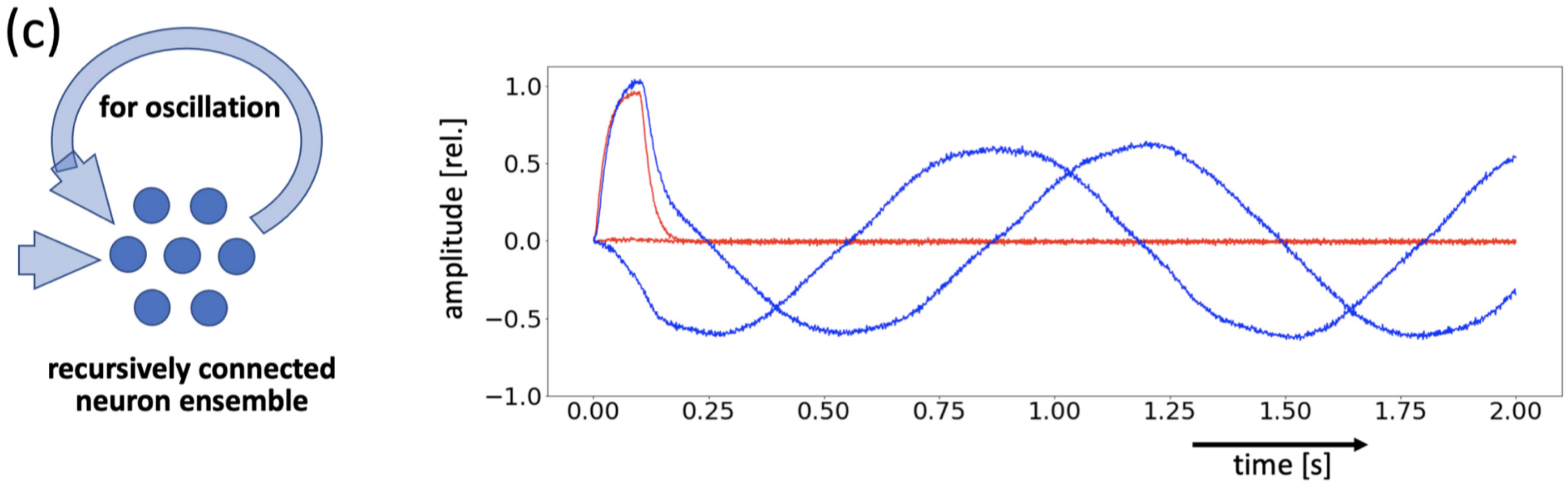


Kröger 2023, JIN

forwarding and processing of lower-level information: **neural connections**

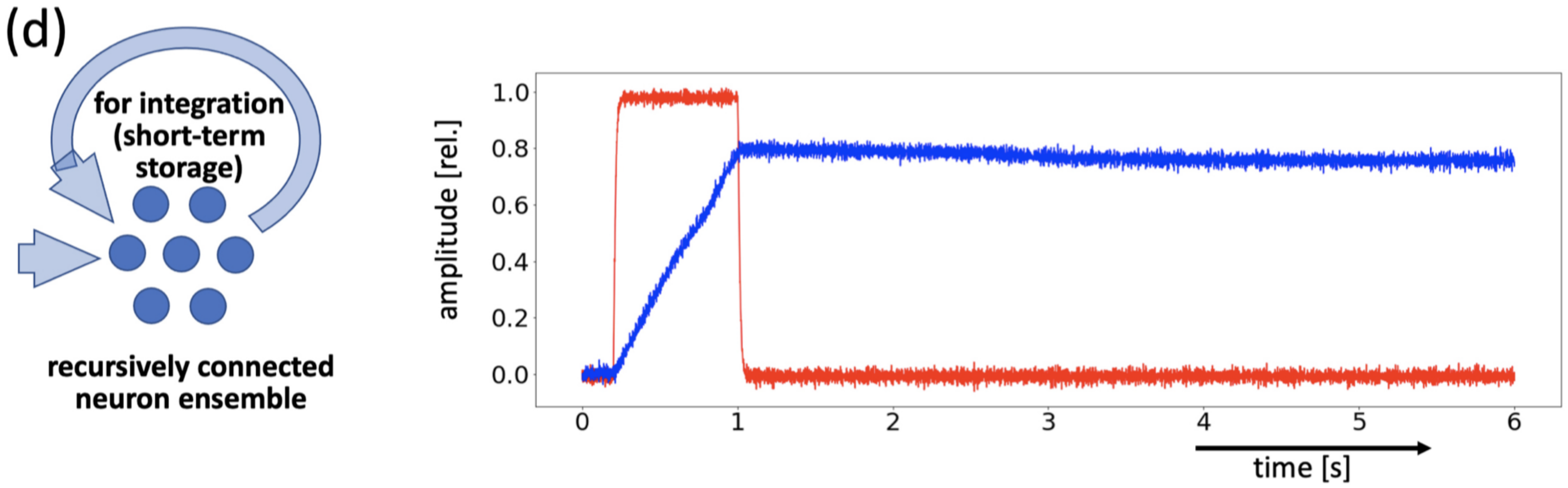


recurrent neural ensembles for generating dynamics -> **oscillations**



for defining intrinsic time scales

recurrent neural ensembles for enabling **short term storage**



for holding a value over a time period

Beyond coding of “values”: Coding of “items”

→ moving forward towards cognition

Eliasmith 2013: How to build a brain
Stewart & Eliasmith 2014: IEEE review on semantic pointer architecture

Values (NEF) =

- loudness, frequency, muscle strength
- -> directly coded by neuron ensembles as a specific “value”

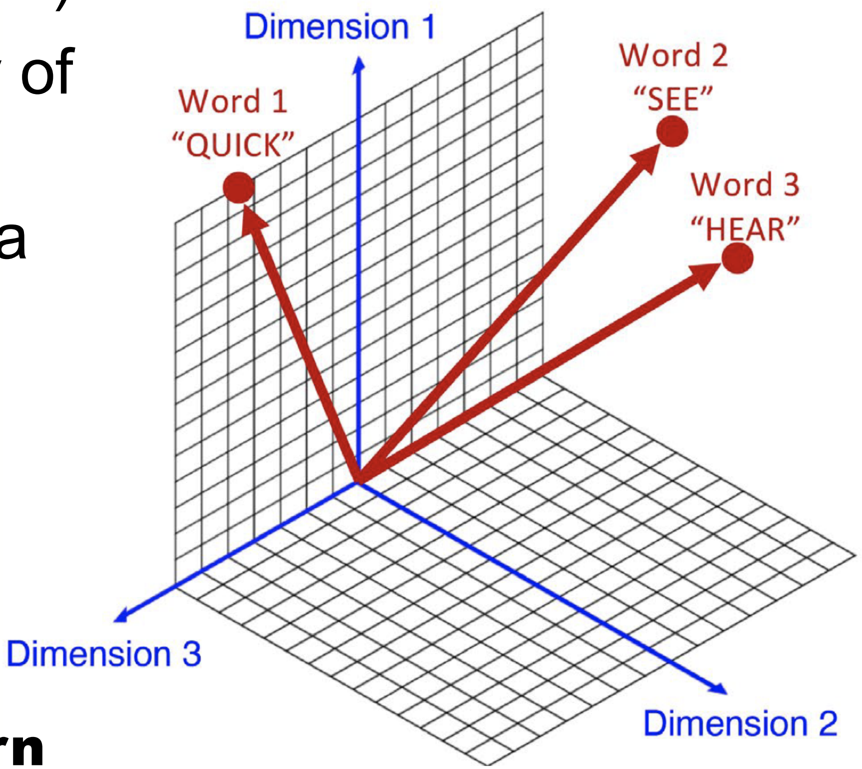
Items (SPA) =

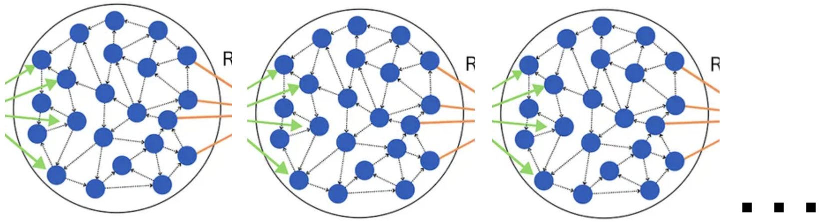
- a word (concept, lemma, phonol. form) ,
- a syllable (phonol. form, gesture score, higher-level auditory form, pre-motor pattern)
- a sentence meaning, a thought, a decision, … (abstract cognitive)

Beside NEF-SPA system (Eliasmith et al. 2014) ← includes a concept for cognitive modeling
NEST simulator for building up complex models (Gewaltig et al. 2012)
NEURON tool box (Hines Carnevale 2001)

Beyond coding of “values”: Coding of “items”

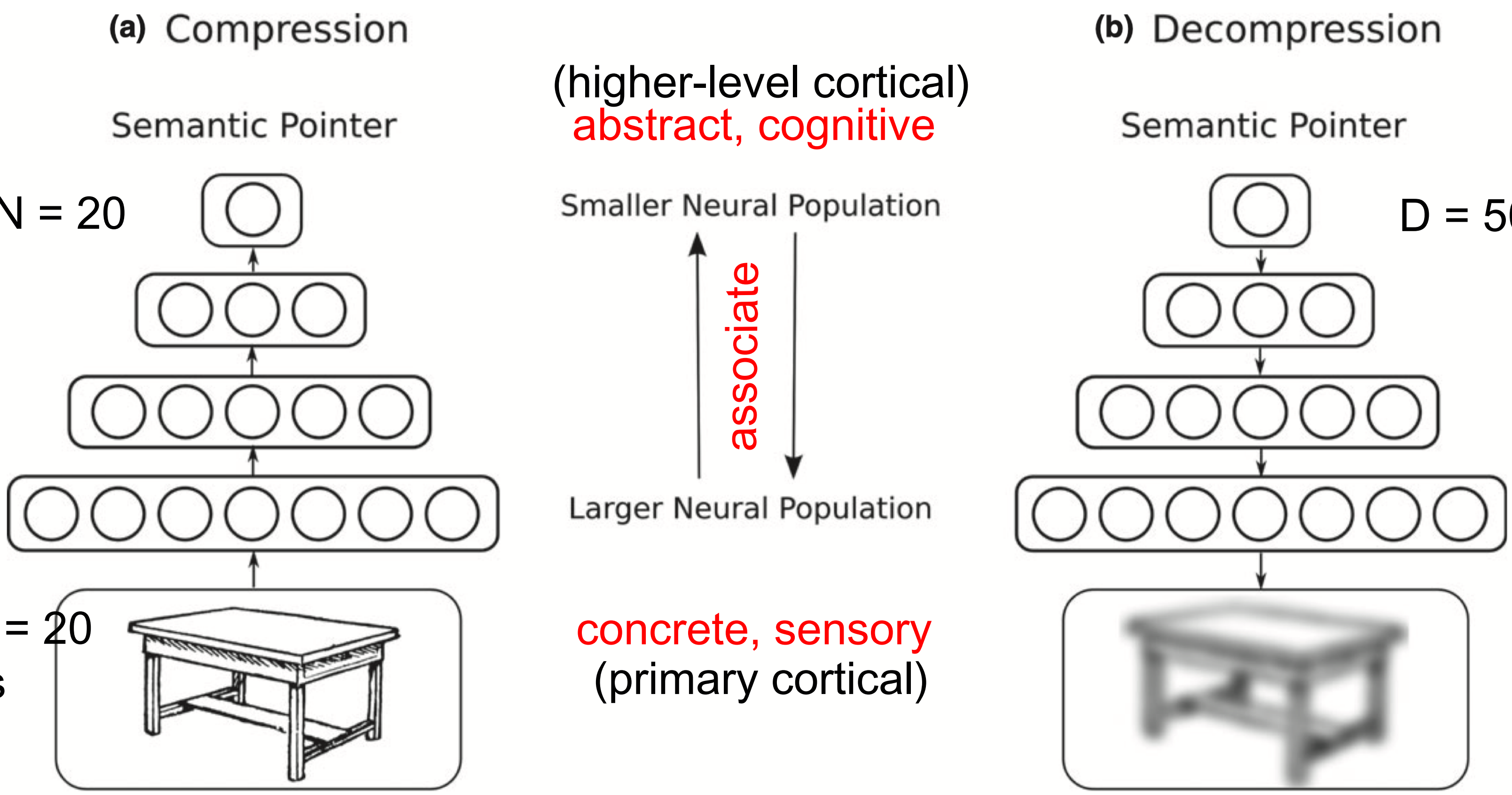
- the idea: items = represented by **vectors** (**S-pointers**)
- need: up to D=500 dimensions in case of a vocabulary of 60.000-100.000 items (mental lexicon, each level)
- the vector (mathematics) is only in the background as a valuable helper:
  - behavioral level:
    - vector points on items: cat / dog
    - quantifies similarity / dissimilarity of items = distance in vector space
  - neuronal level:
    - each value of the vector is coded as **activation pattern** in one of D neuron ensembles (each with N neurons)
    - **neuron buffer**: hosts states = neural representations of items
    - typically: D=64 (1000 items vocab) N = 50…100 -> 3200…6400 neurons





S-pointer and Semantic Pointer Architecture

D = 500, N = 20



(higher-level cortical)
abstract, cognitive

associate

concrete, sensory
(primary cortical)

(words)

D = 500, N = 20

visual input:
Many thousand ensembles (NEF)
each ensemble: N = 20
representing pixels of picture

mental image:
Inner activation

(or for syllables:
abstract higher level: phonological form /pla:/
concrete lower level: sensory: auditory, somatosensory;
motor: pre-motor gesture plan; motor: detailed muscular activation)

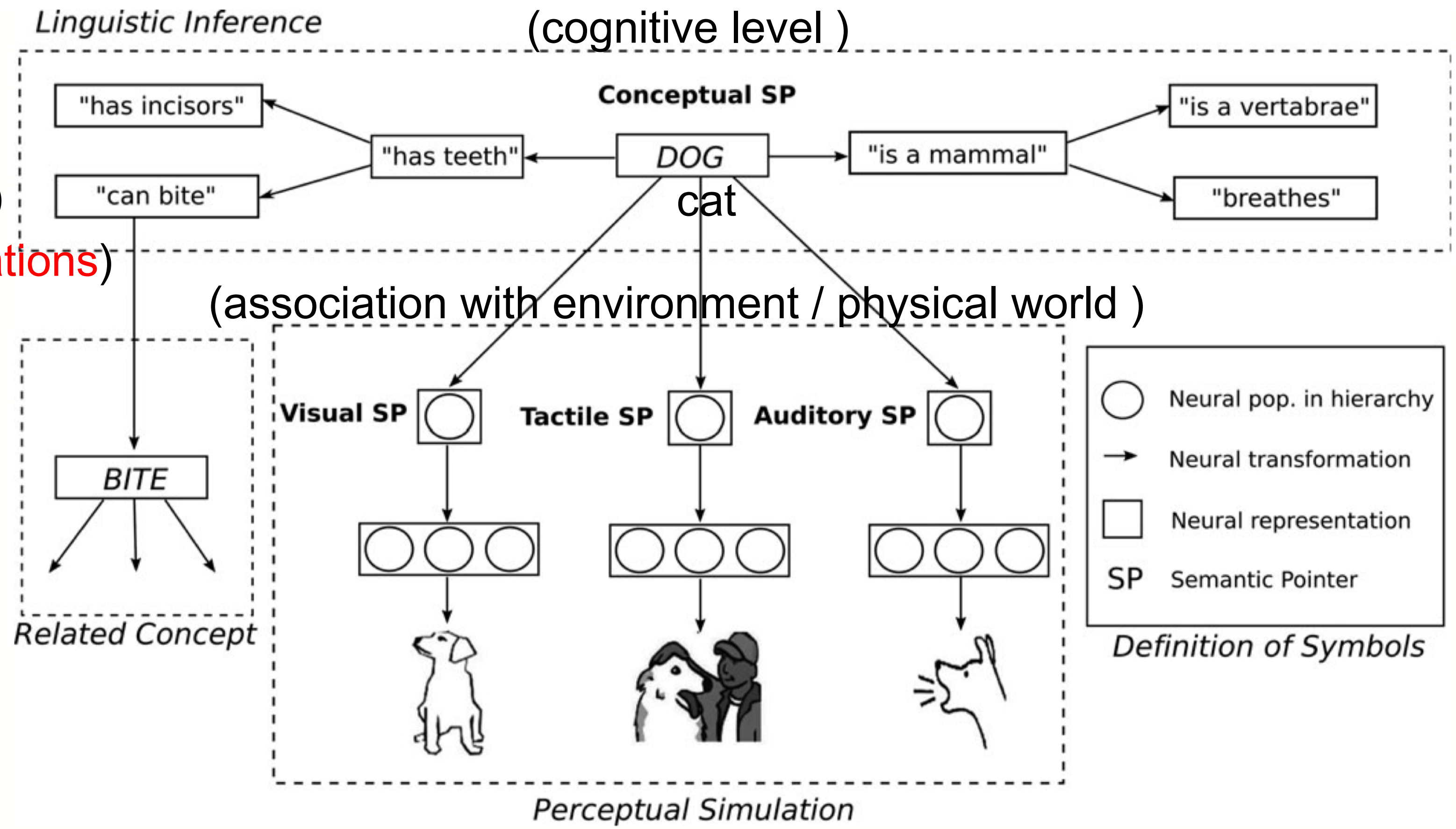
S-pointer and Semantic Pointer Architecture

Defining similarity
Example: word concepts

S-pointer networks:
(Coding of associations)
(Coding of similarity relations)

(cognitive level)

(association with environment / physical world)

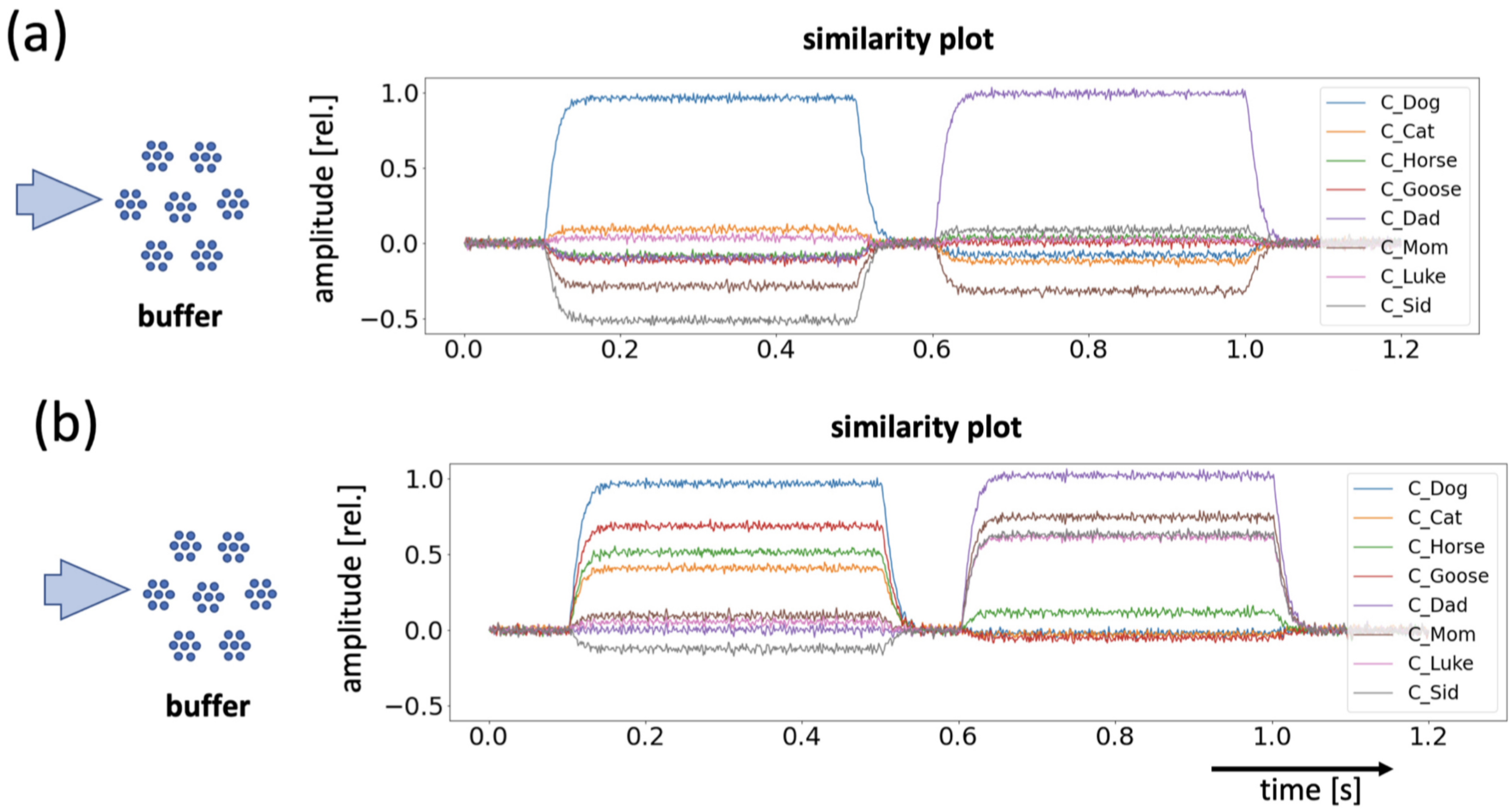


Mental images, auditory simulations, …
barking, meowing, …

Summary: neurobiological plausibility

- Build up a complex (large-scale) model using **canonical neural network elements (SPA)**
  - defining basic motor, sensory, or cognitive functions
  - Basic network elements within different modules of main functions (-> CMC model)
- Further examples:
  - Represents items with or without similarity relations (next slide -> state buffers)
  - Holds items in short term memory (2nd next slide -> recurrent buffers)
  - Associating items: phono forms -> semantic concepts (mental lexicon) (3rd next slide -> associative memories)
  - Binding of items (states) allows reasoning, allows representation of sentence meaning, etc. (4th next slide: binding buffers)

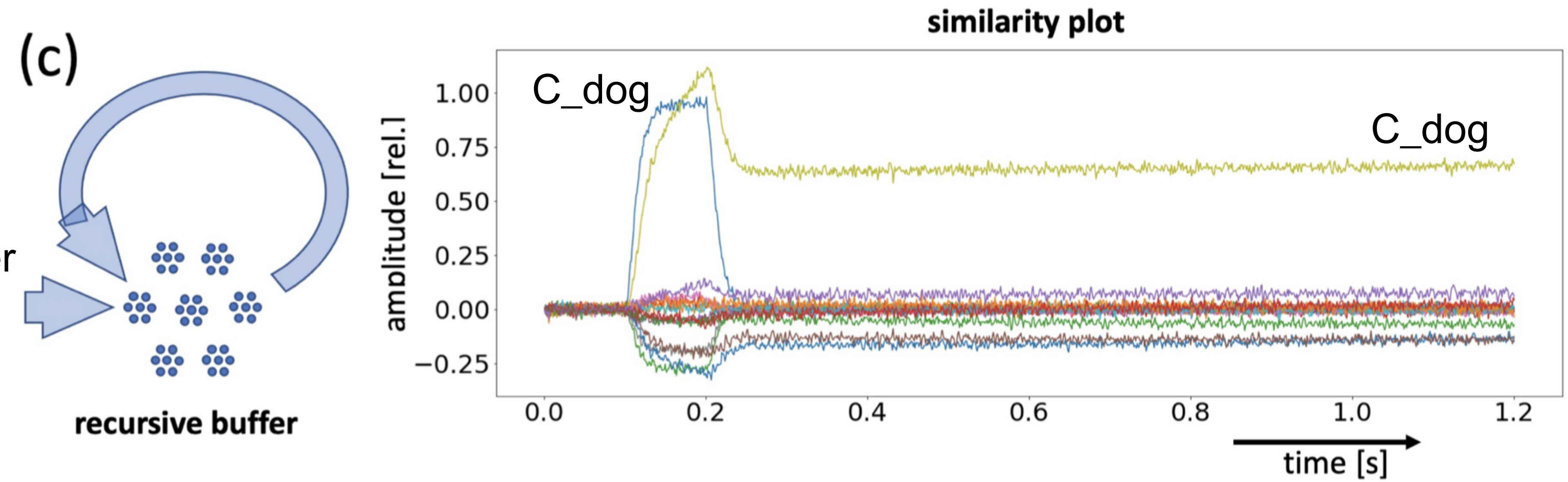
Represents items with or without / with similarity relations: **state buffers**



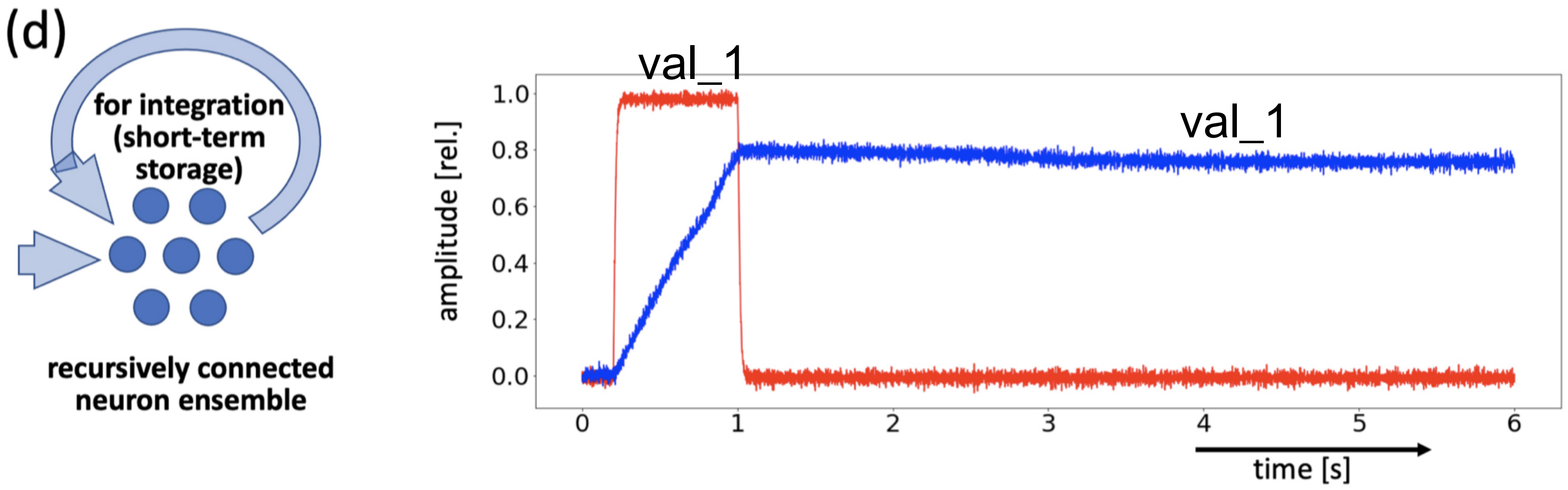
Probing: activation (t) -> similarity
plots: decoding activation (t) as S-pointer activity

Holds item representations (S-pointer) in short term memory: **recursive buffers**

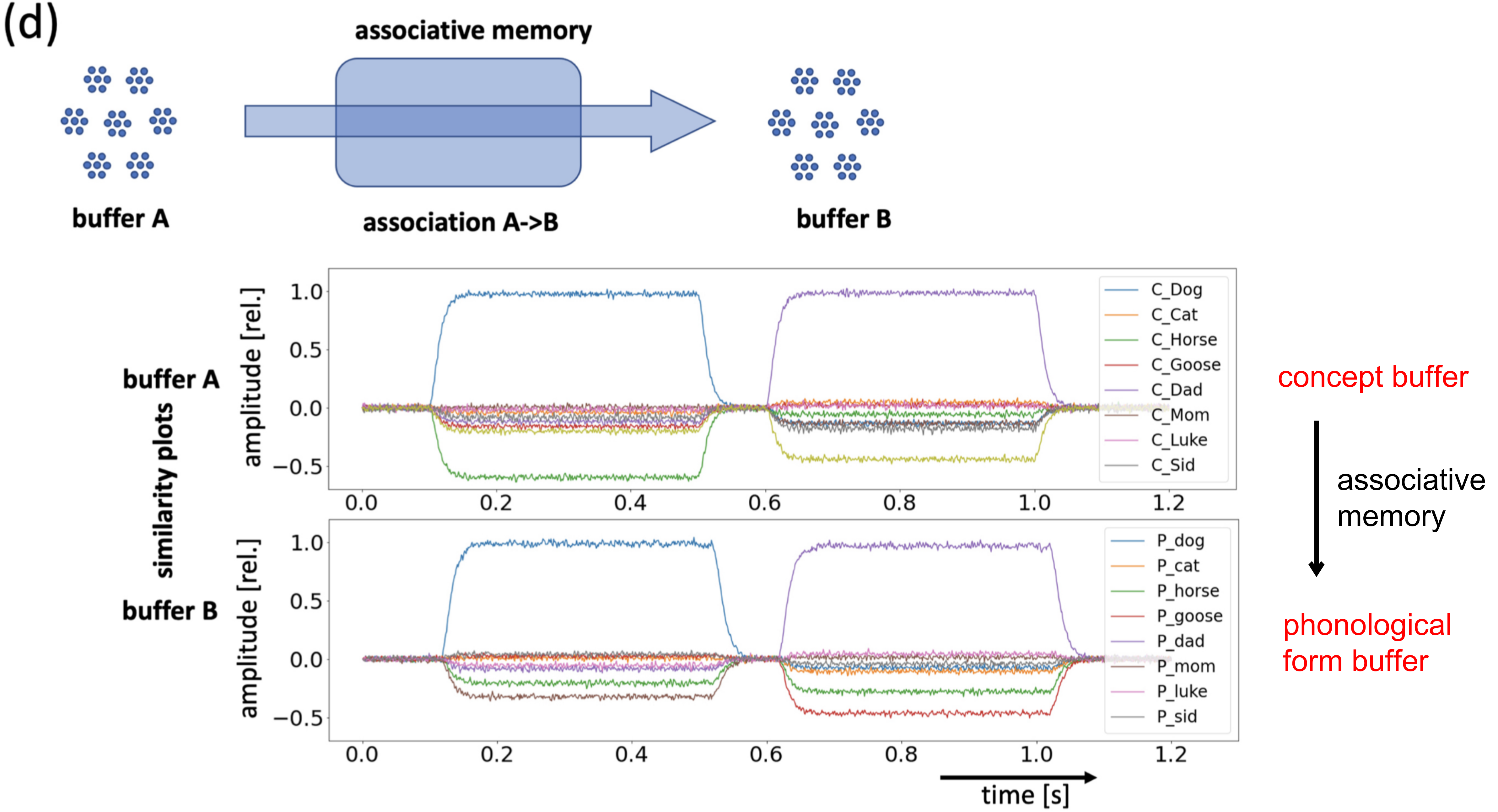
neuron buffer
state buffer
-> items



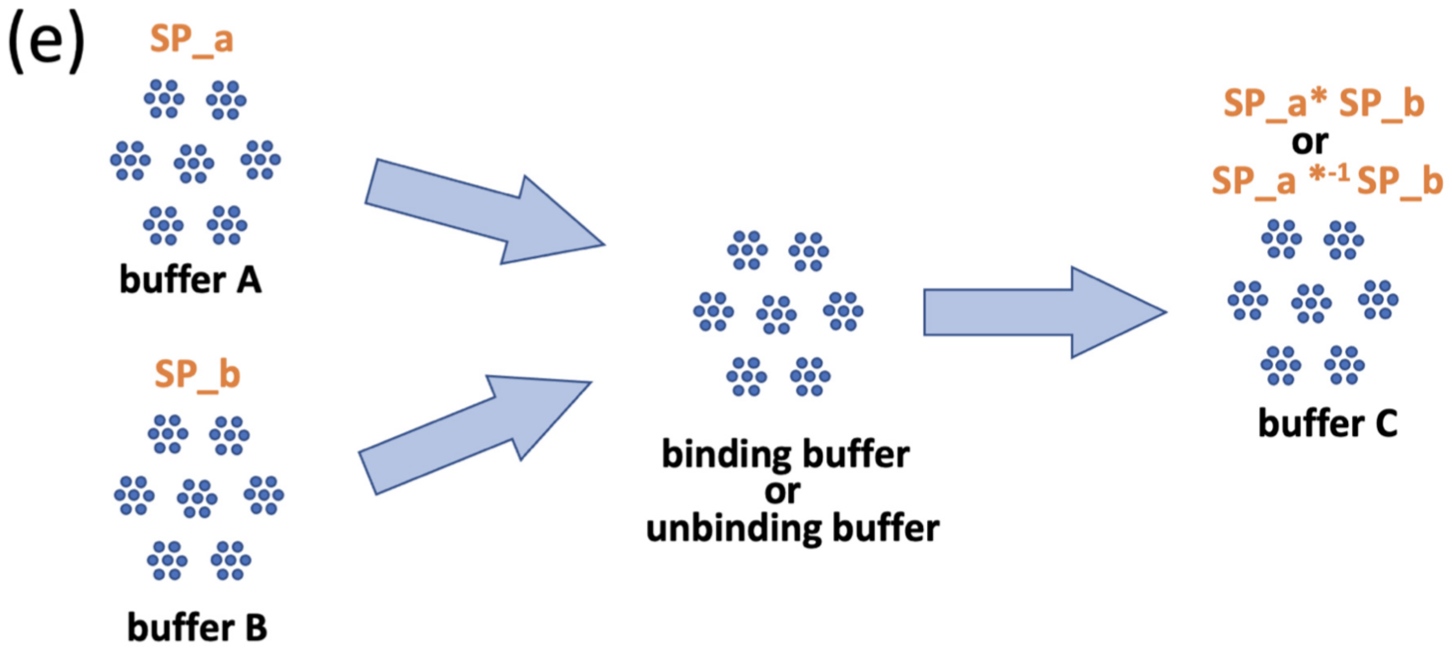
neuron ensemble
-> values



Associating items: transformation -> **associative memories**



binding of items (of states): **binding buffer** and binding network



Sentence meaning: R_action * C_drinking + R_agent * C_Benno + R_patient * C_coffee → E_sentence1

R_ : roles
C_ : concepts
E_ : event meaning

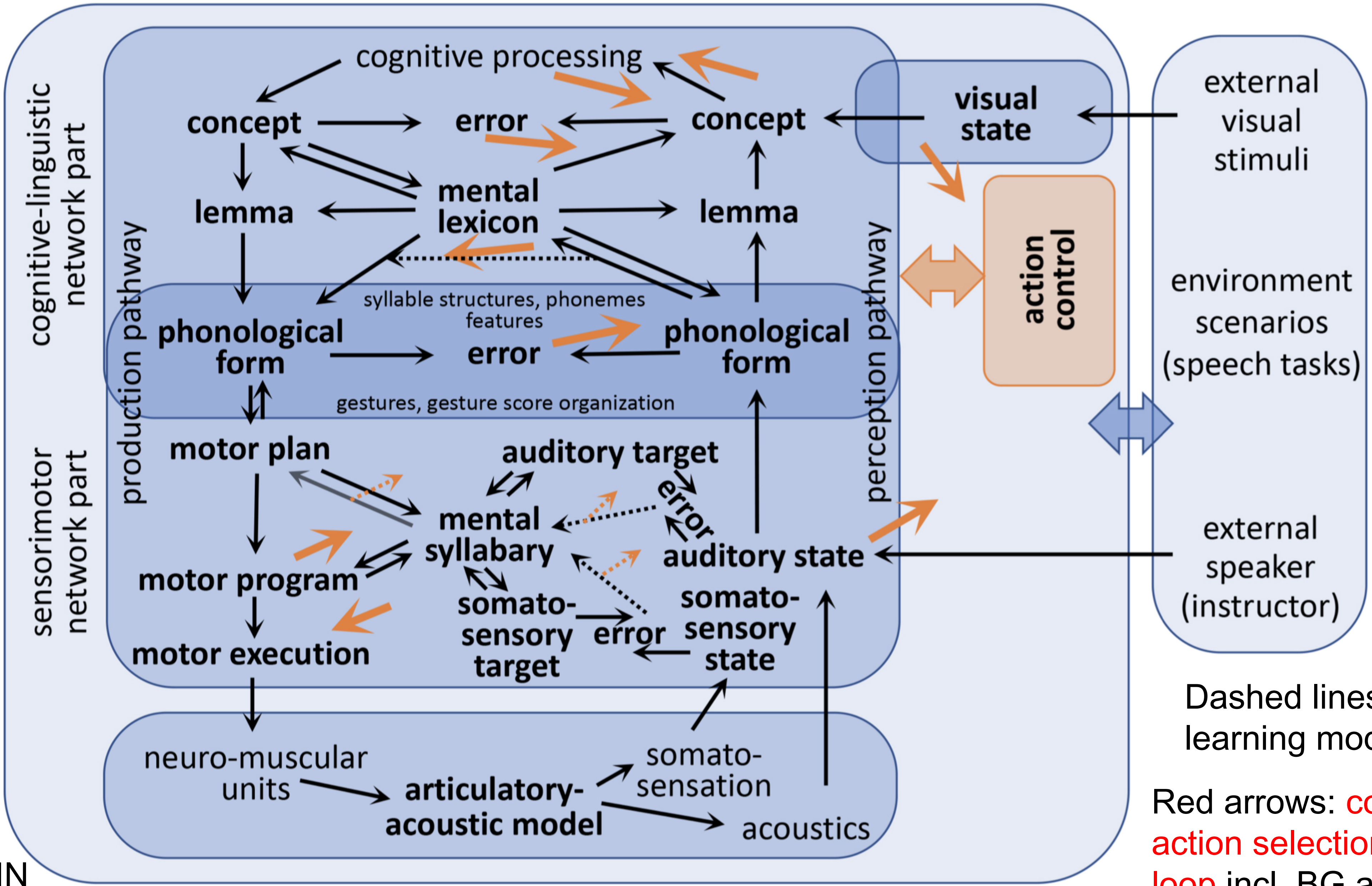
One binding buffer: inputs:
roles (buff_A) and concepts (buff_B)
come in synchronously over time
last step: additive integration (?)

The perception-production model

- just using **canonical neural network elements**: -> large-scale network for speech production and speech perception (language processing)
- have **pre-defined form** for representation of items:
  - mental lexicon: phono-forms, concepts, lemmata: noun, verb, determiner, adjective, preposition
  - syntax: dependency arc names:
  - sentence level semantics: event description in terms or role-concept pairs
- **pre-defined structure for network modules**: Three levels of representations in the mental lexicon (strongly associated / connected)
  - Semantic level: cat-dog, car-bike, eat-drink, …
  - Phonological-level: /kEt/-/ka:/, /dOg/-/drank/, …
  - Lemma level: determiners (the-a) vs. adjectives (god-bad), vs. nouns vs. verbs, …

Neural model of mental lexicon

including word processing pathways



Dashed lines: learning mode

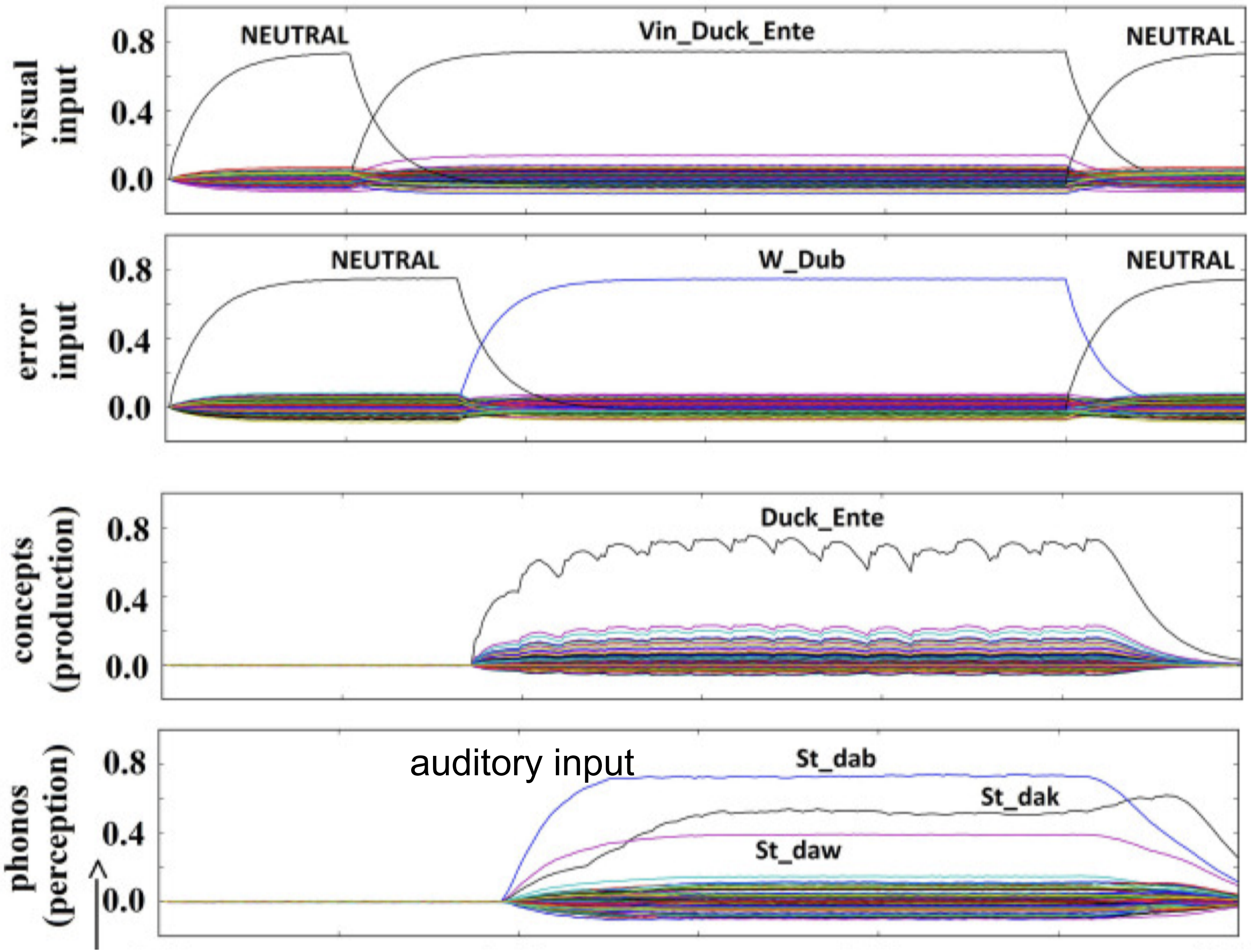
Red arrows: cortico-cortical action selection / control loop incl. BG and Thal.

Kröger 2023 JIN

Simulation example: production

Picture naming with distractor word

auditory distractor word

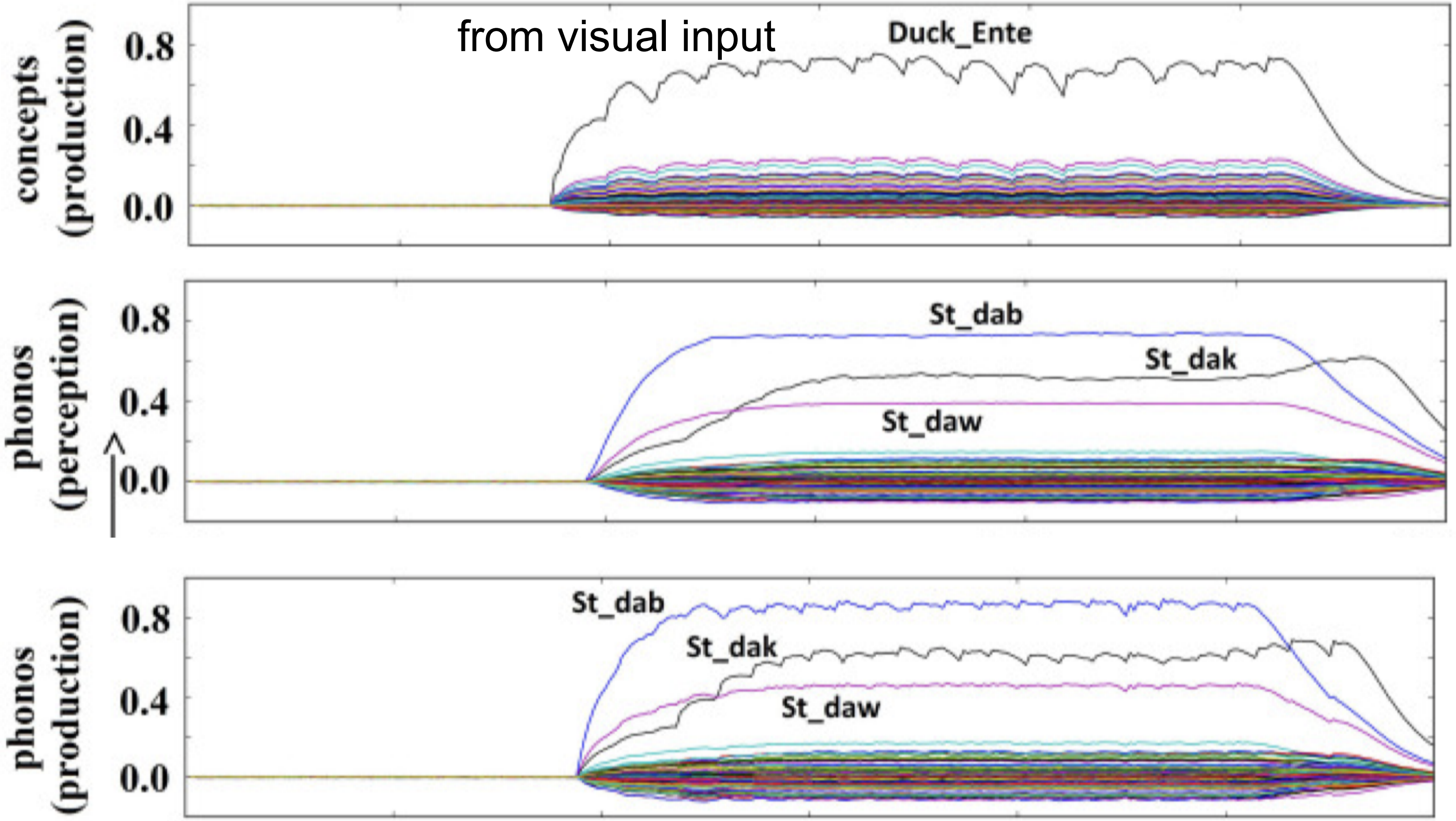


visual perception process not displayed

mental auditory co-activation resulting from visual input (-> temporally delayed)

Simulation example: production

Picture naming with distractor word

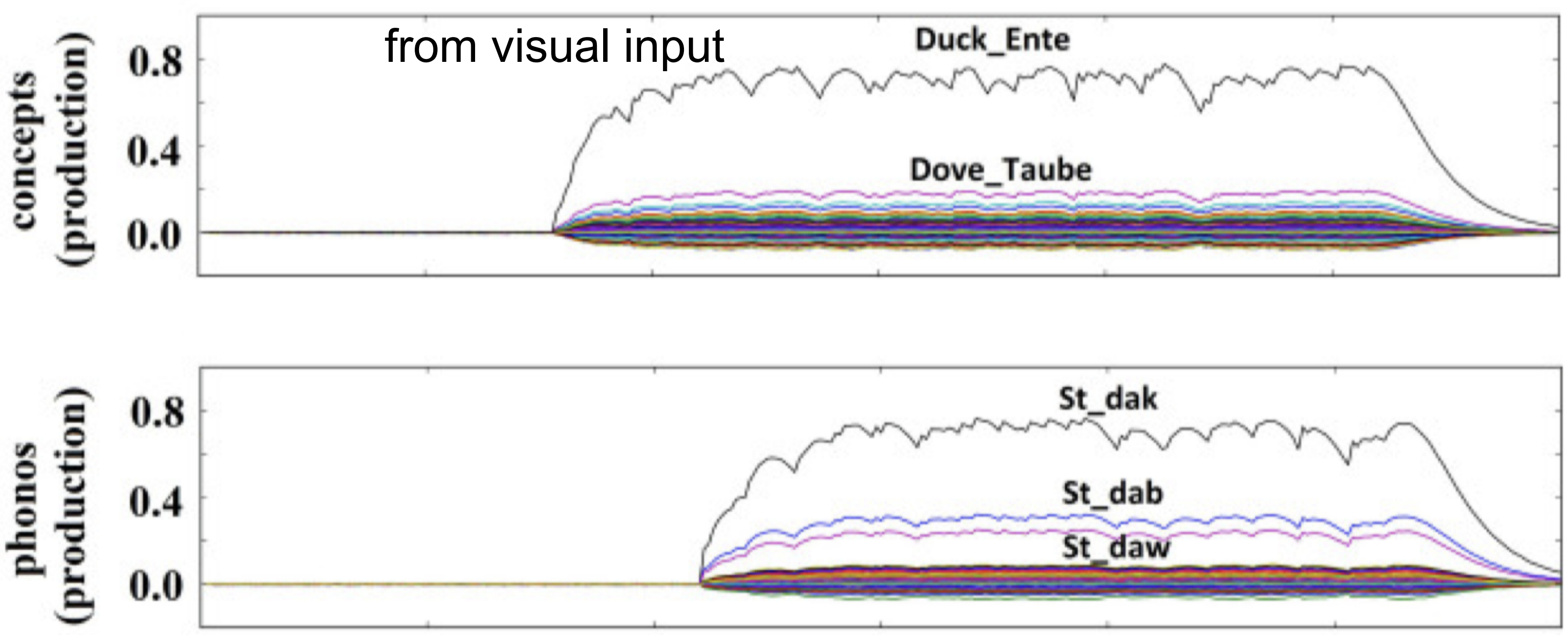


St_dak appears mainly because of S-pointer network similarity relations!

the wrong item (distractor item) wins

Simulation example: production

Picture naming, **no** distractor word



semantic level coactivation

the correct item wins

phono level coactivation

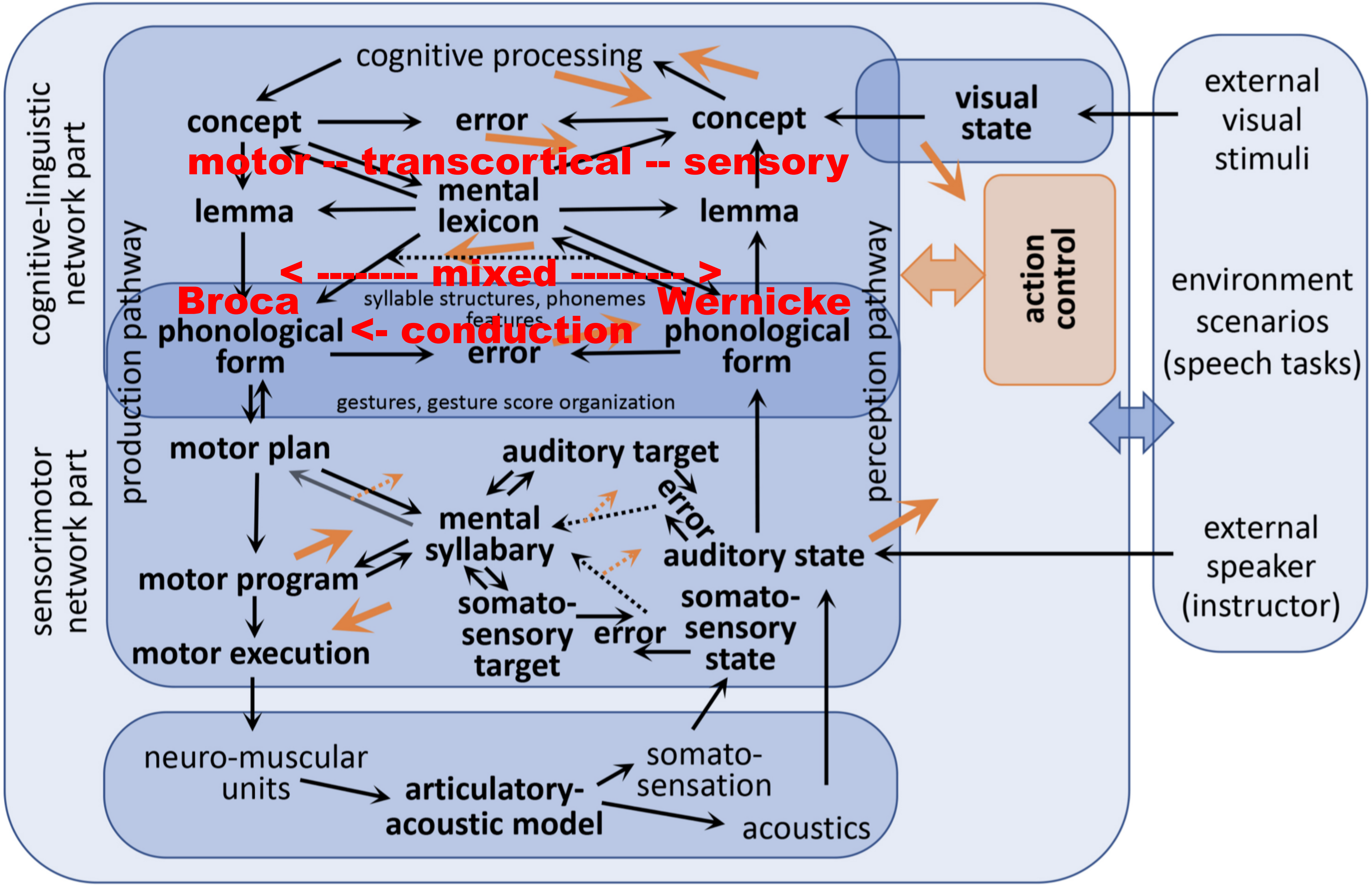
Goals of modeling
* beside: neurobiological plausibility
* simulation of psycholinguistic testings / experiments
* further goals: simulation of diagnostic testings (screenings)
* simulation of therapeutic treatment scenarios (if the model is capable of learning)

Neural model of mental lexicon

including word processing pathways

types of aphasia

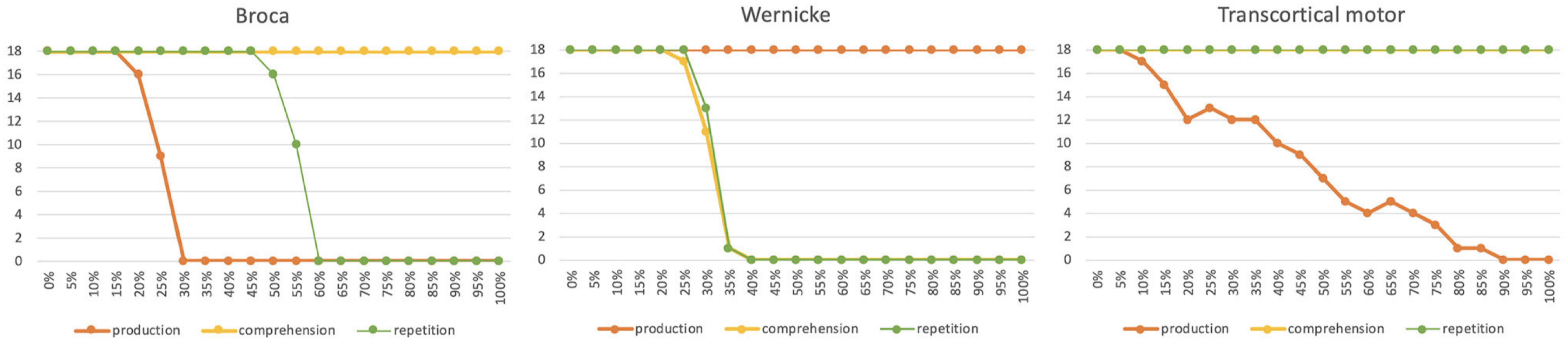
state buffers or associative memories: percentage of ablated neurons (inserting "neural dysfunctions")



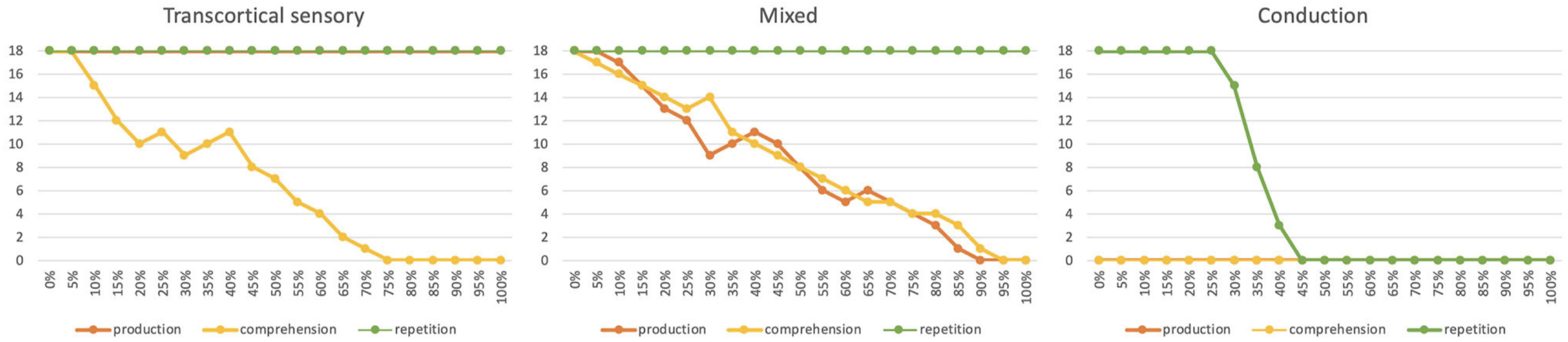
Kröger 2020 frontiers in

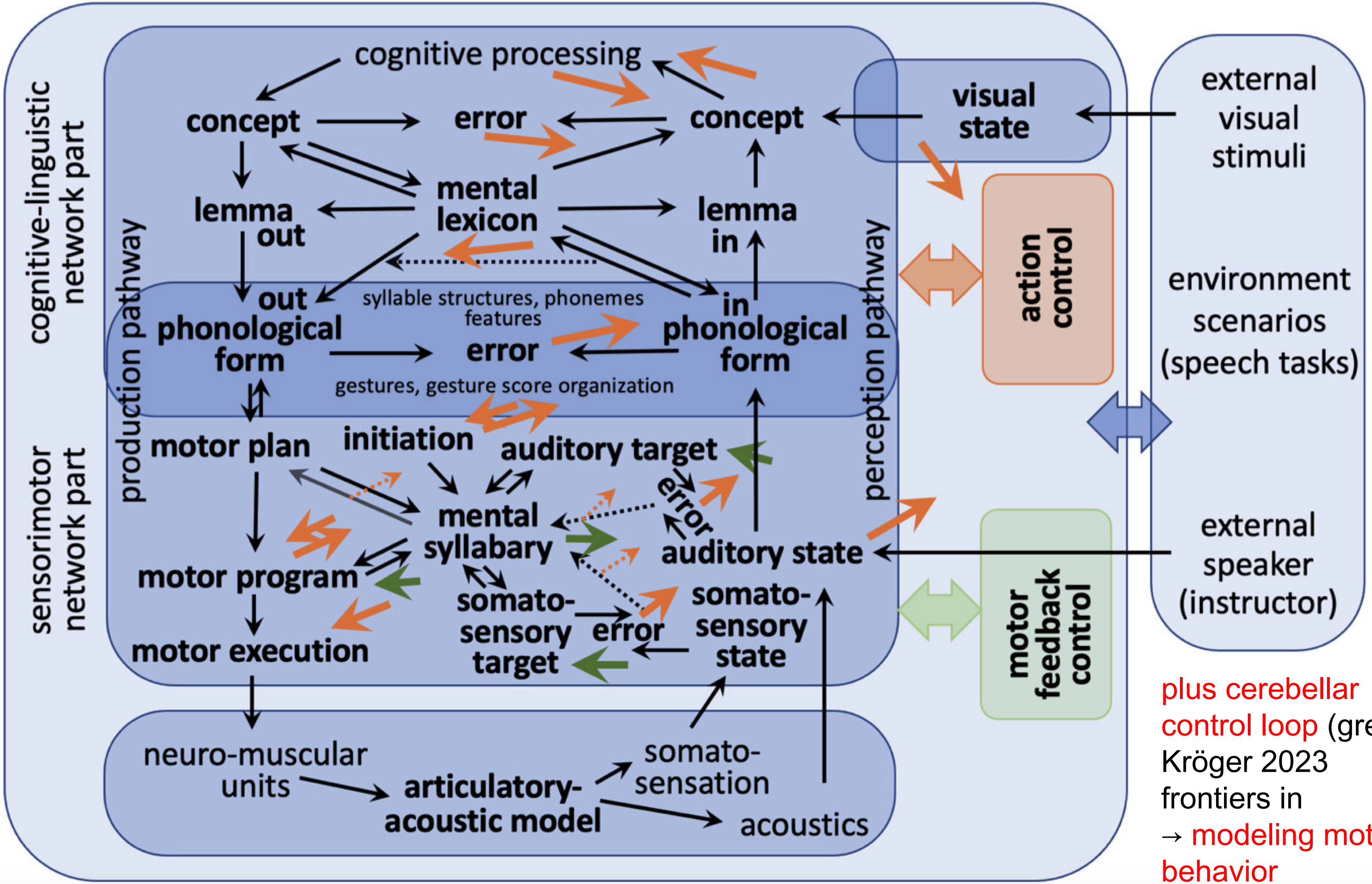
simulation results: modeling aphasia

Kröger et al. 2020



Picture naming / auditory presentation of words and pointing / auditory presentation and repetition

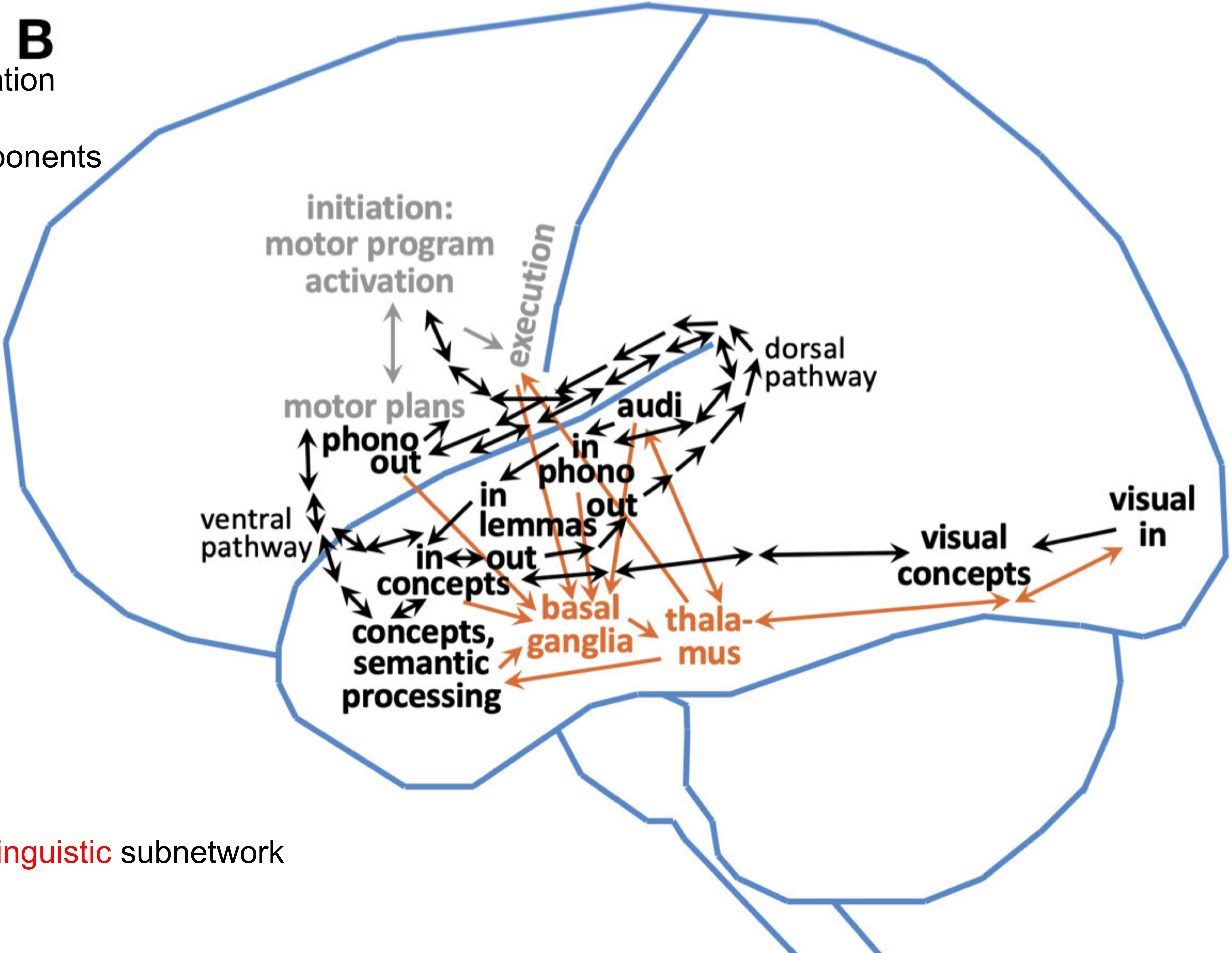




plus cerebellar control loop (green)
Kröger 2023
frontiers in
→ modeling motor behavior

B

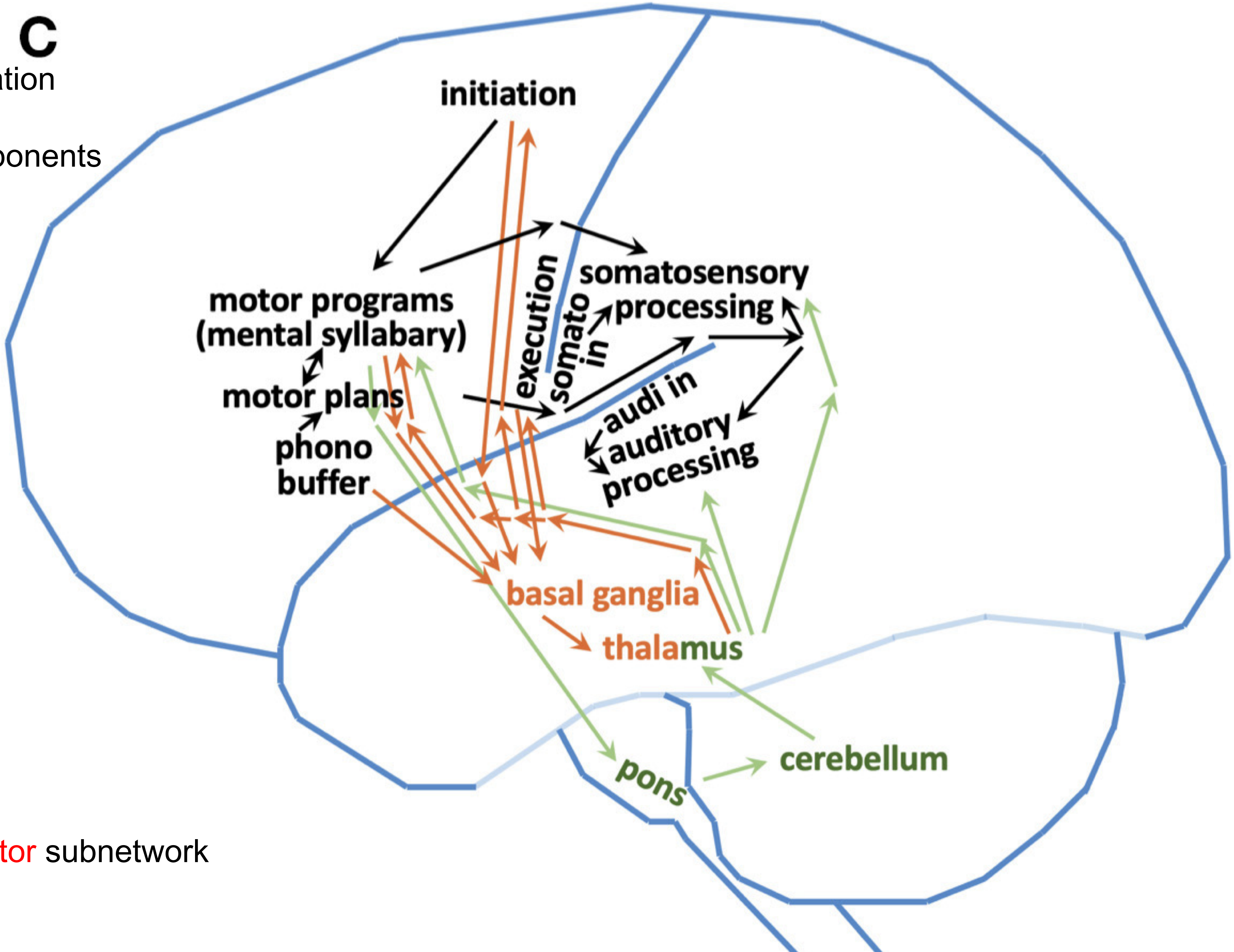
anatomical location
of
functional components



cognitive-linguistic subnetwork

C

anatomical location
of
functional components

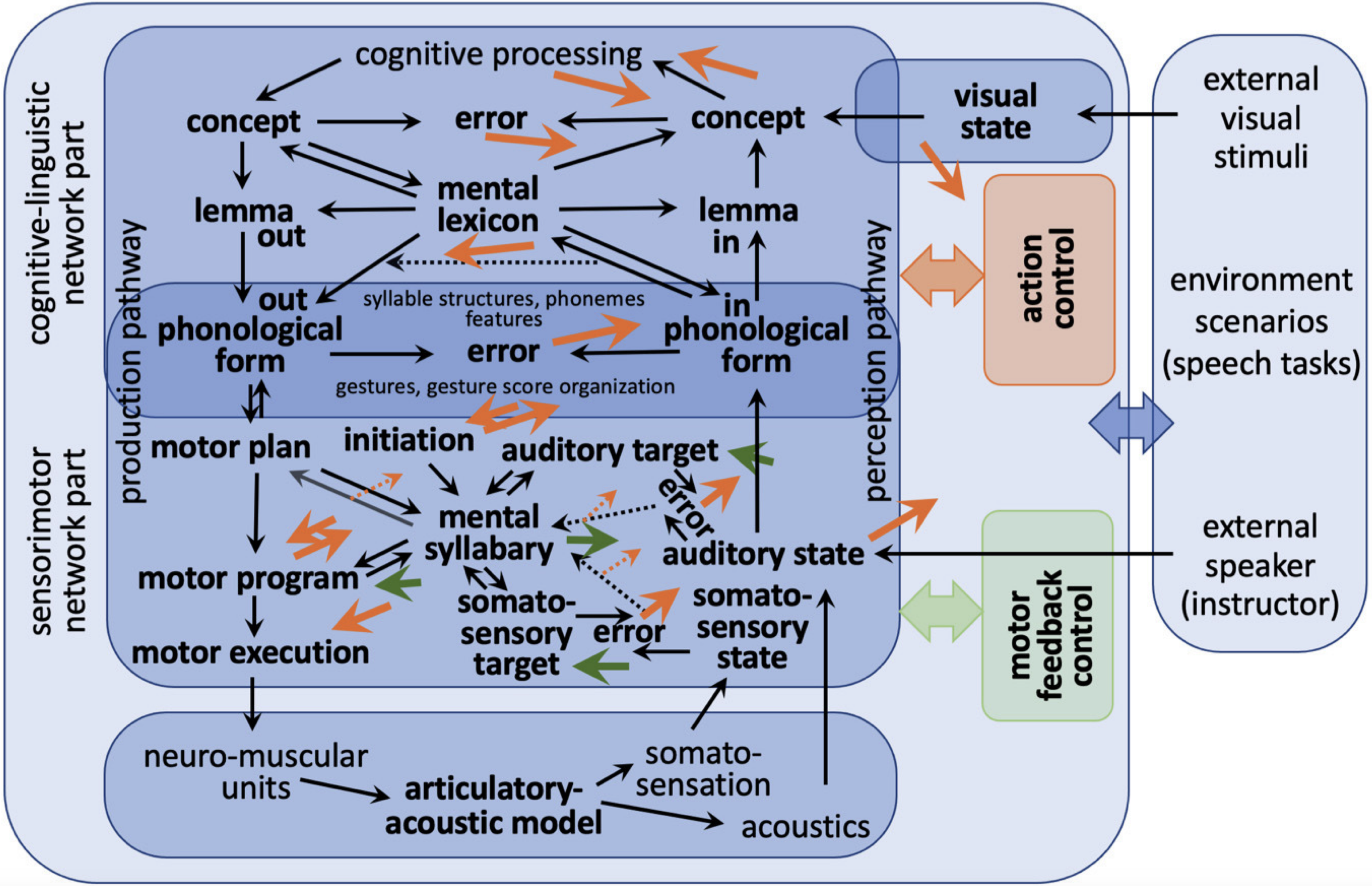


sensorimotor subnetwork

The syllable level and articulation

- three levels on production side:
  - (i) phonological (= **raw gesture score** already: gestures as lexical units) vs.
  - (ii) **motor plan** (= **fully specified gestures** but not fully specified muscle activity patterns = not the articulator positions level) <-> premotor, and:

    sensory representations (higher sensory processing level as well)
- Beyond: (iii) motor realization = **motor program**
  - fully specified **muscle activation pattern** (fully specified movements)
  - is that stored in mental syllabary for frequent syllables?
  - because: fast adaptation in case of injury at the level of articulatory system (glossectomy, bite block experiments, …)
- we (Aachen/Geneva) separate planning and programming in the same way: Jouen, Fougeron & Laganaro (2024), Kröger (2022) frontiers in, Kröger (2023) JIN
- shortcoming here (Kröger & Bekolay 2022): todo: coupling of a neuro-muscular acoustic-articulatory model (there is one: Sanguineti et al. 1998)



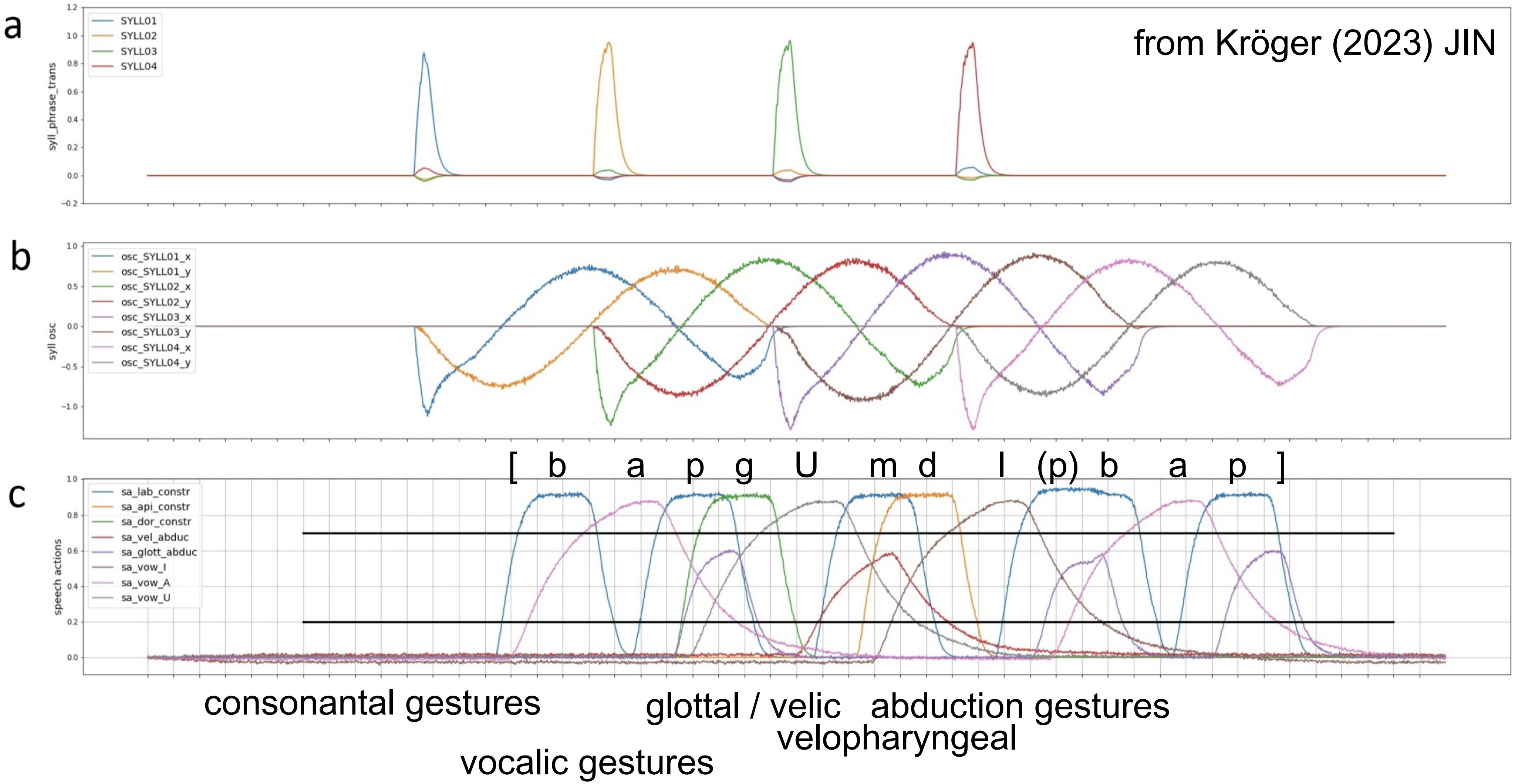
premotor level

primary motor level

Kröger 2023

Simulation of planning

planning results will be stored in mental syllabary



syllable pulses (initialization)

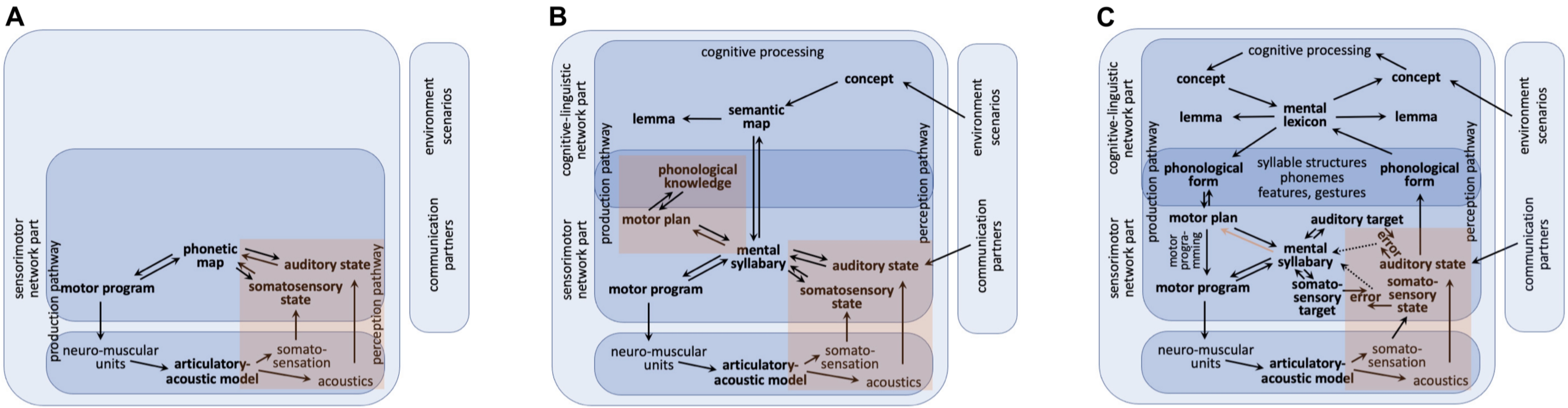
syllable oscillators: Defining intrinsic timing (phase values) -> allows phasing

gestures: potential degree of realization (premotor level)

The complete speech processing model

- In agreement with Common Model of Cognition (CMC)
- But:
  - recurrent neural networks (Google OpenAI) allow training but stay “unstructured”
  - Training (modeling speech acquisiton):
    - still complex for biologically inspired models (spiking neuron approaches like NEF and SPA);
    - and: there is no approach for network growth

Developmental model of word processing



babbling phase:
playing around with own speech organs

imitation phase:
triangulation processes

adult (matured):
word production-comprehension

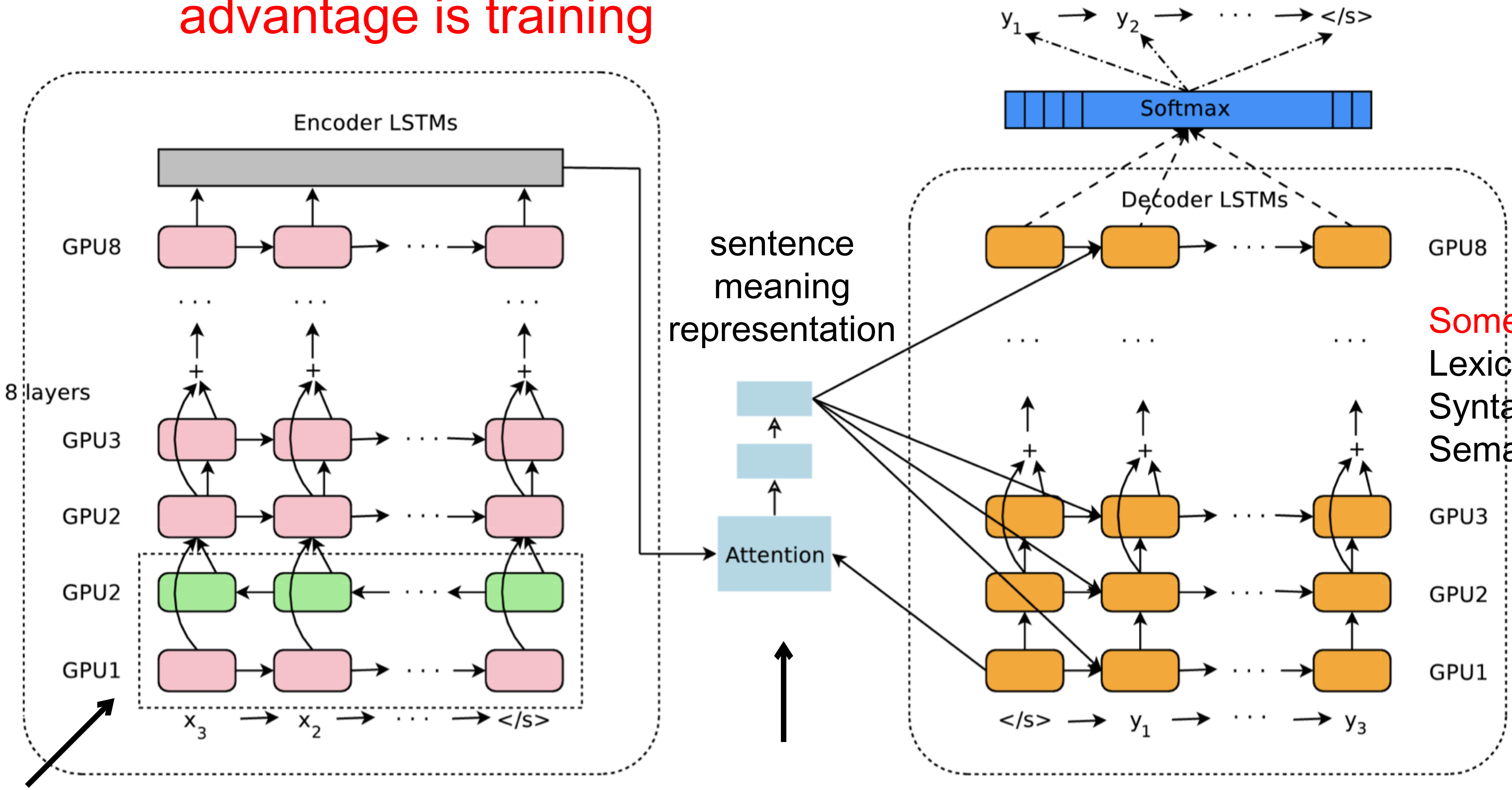
In red: sensory feedback regions for enabling sensory-motor integration

from: Kröger et al. 2022 frontiers in

Google's machine sentence translation network model (Wu et al. 2016)

advantage is training

language 2
word sequence output



sentence meaning representation

Somewhere inside:
Lexical processing
Syntactic processing
Semantic processing

Relatively equally structred / unstructured ensemble of modules

word sequence input
language 1

Interpretation of the neural activation patterns appearing here is difficult!!!

Comparison our model vs. NLP-approaches:

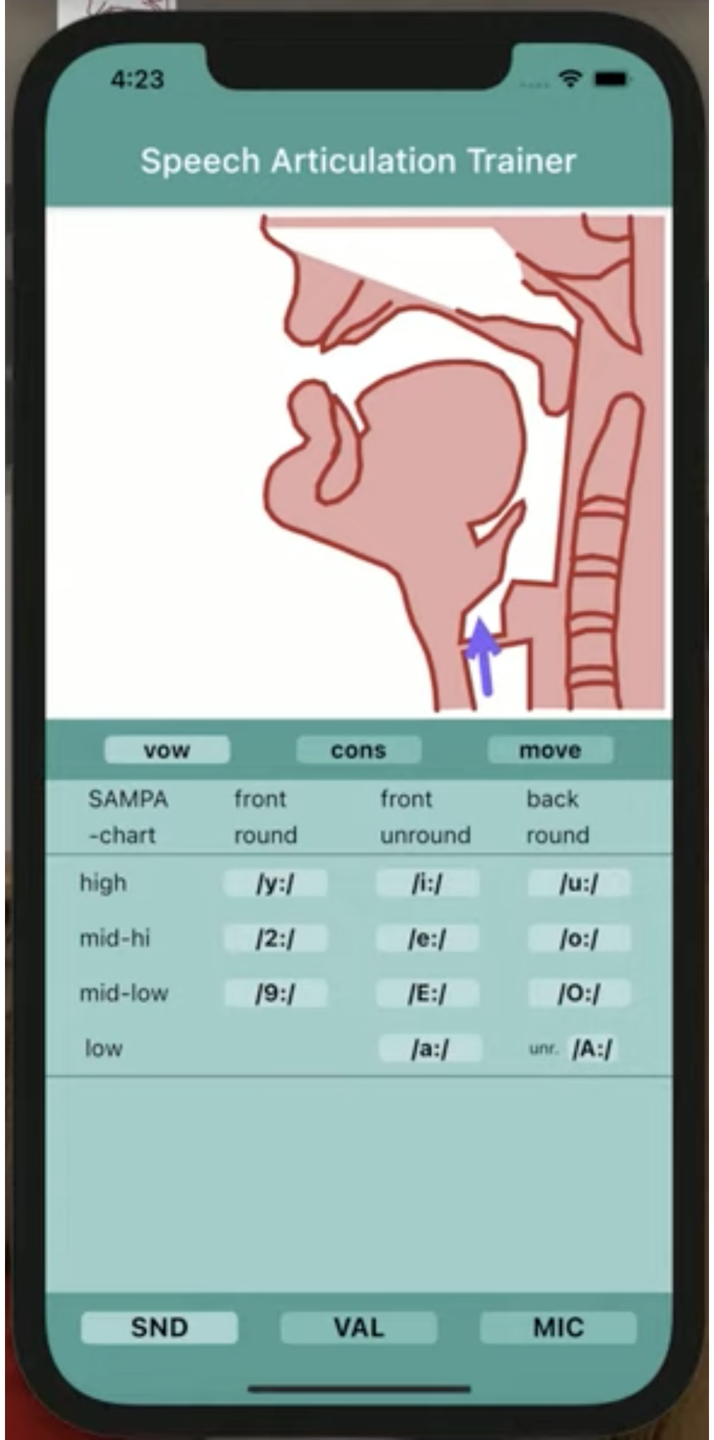
- Simple nearly unstructured recurrent NLP network models can profit from powerful training algorithm (free **ontogenetic** development of the model; individual development of a model following birth)
  - Example: Google translation system (see slide above)
  - Leads to **free formation** of lexical, syntactic, semantic **representations**
  - Leads to **free formation** of the **inner structure** of the network layers which represent the *modules* (lexical, syntactic, semantic processing) in a **largely interwoven** manner (parallel and hierarchical architecture)
  - These models outperform neurobiologically based networks, outperform our models as well as humans

Discussion of our approach:

- Our neurobiologically plausible approach:
  - **predefinition** of **module architecture** and
  - **predefinition** of lexical, syntactic, semantic **representations**
  - Advantage: the model is able
    - to simulate network growth during speech acquisition, and
    - to simulate different stages of speech acquisition -> next slide (remark: no growth modeling in NLP networks!)
  - The model is able
    - to model neural damage by insertion of neural dysfunctions in specific functional or anatomical parts of the model (stroke, traumatic brain injury) and
    - to simulate screenings and thus: symptoms of speech disorders / speech errors
    - To give high quality results concerning the association of neurofunctional damage and resulting speech symptoms -> will increase our knowledge about / our definitions available for different types of speech disorders

(provocative) conclusion:

- Having in mind that google's translation neural network outperforms humans :
- The **phylogenetic** development of the brain (evolutionary history of brain development of humans during last 50.000 years) → the structure of the brain limits performance of training
- **Evolution** limits performance?
- Hypothesis: training of an unstructured brain network would take too long, longer than parents can care for the child by a specific **ontogenetic** development
- ((like: some animal must be able to get up, stand upright, and move immediately after birth))



Thank you for your attention

Bernd J. Kröger
(neuroscientist, physicist, phonetician)

www.speechtrainer.eu → publications
www.speechtrainer.eu → the app

google scholar → Bernd J. Kröger

ORCID: https://orcid.org/0000-0002-4727-2957