USING NENGO.AI FOR DEVELOPING MODELS OF SPEECH PRODUCTION AND SPEECH PERCEPTION

Bernd J. Kröger

Neurophonetics Group, Phoniatrics Department, Medical Faculty

RWTH-Aachen University, Aachen, Germany

www.speechtrainer.eu

Introduction

- Motivation: Why neural models?
- The NENGO.ai framework
- Our speech processing model (production and perception)
  - architecture of the modeled brain (modeling patients)
  - Scenarios: simulated tasks -> simulated behavior (modeling screenings)
  - speech movement generation
- Concluding remarks

Motivation: Why neural models?

Neural models

- can simulate macroscopic behavior if they are large-scale models
  - whole brain plus motor out and sensory in components:
  - CNS (whole brain) plus PNS towards vocal tract and ears
- can give clear definitions of neural dysfunctions
  - we can “destroy” neurons in specific parts of model
  - we study the resulting behavior of model: -> symptoms of speech disorders

Motivation: Why neural models?

Models are simplifications of reality and thus help to understand

- basic functional principles of neural processes, (here: for speech production and perception)
- relationship between neural processes and behavior and thus: e.g., how a neural dysfunction may change behavior (-> how a … leads to specific symptoms)?

The NENGO.ai framework

- "Neural ENGineering Objects" framework:
- www.nengo.ai
- Nengo is open source
- Python based: -> runs on: Windows, Mac OS X and Linux
- Profound scientific background:
  - Eliasmith, C. (2013). How to build a brain: A neural architecture for biological cognition. *Oxford University Press*.

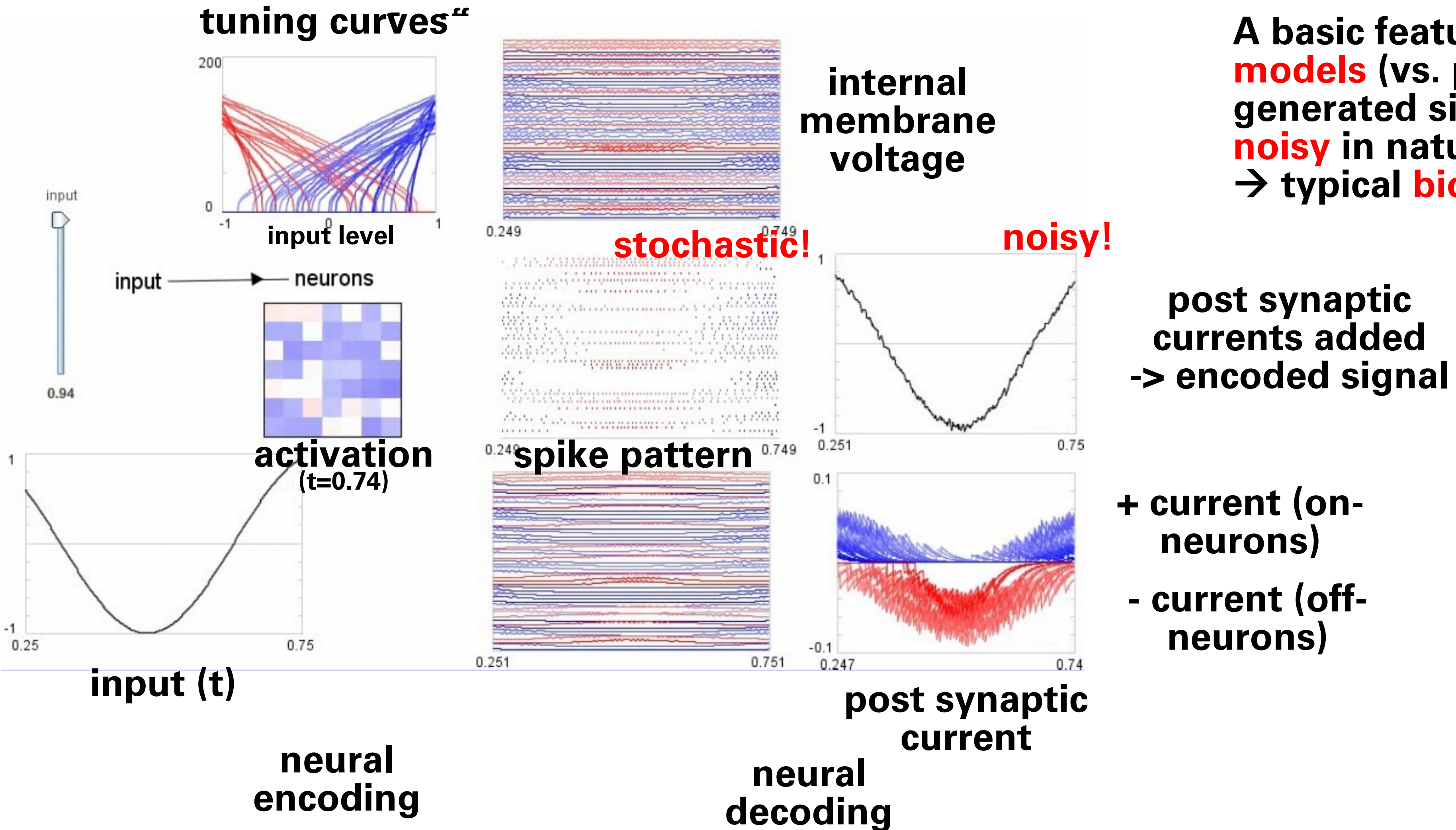
NENGO comprises

- basic neuron models (e.g., LIF-model) -> generation of spikes
- concepts or strategies for connecting neurons:
  - neuron ensembles, neuron buffers, short-term memories, associative memories, binding buffers, …
- strategies for coding functional **items** which need to be processed by the model (concepts, phonol. forms, motor plans)
- fully developed model for for neural process control over time (BG-Thalamus-complex: cortico-cortical control loop)
- strategy for connecting the periphery: -> large-scale model
  - sensory input: eyes and/or ears
  - motor output: arms and/or the speech articulators

Basis of our model: LIF neurons

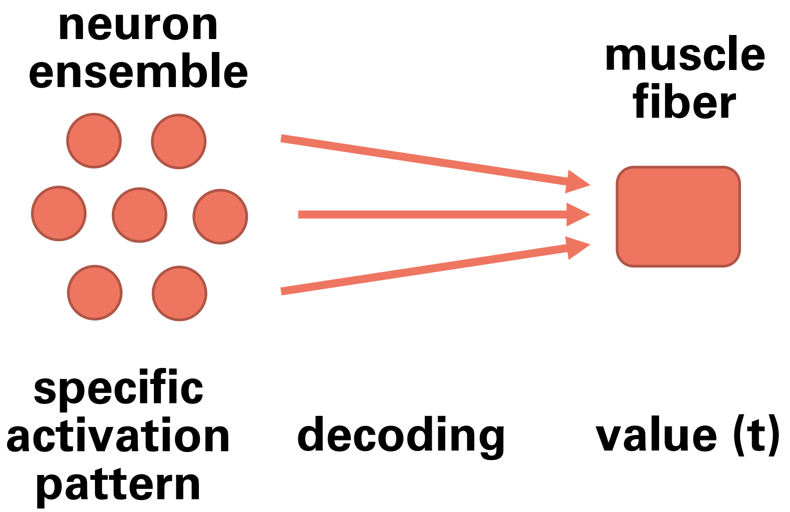
- 50 neurons -> lead to a sufficient coding of input values

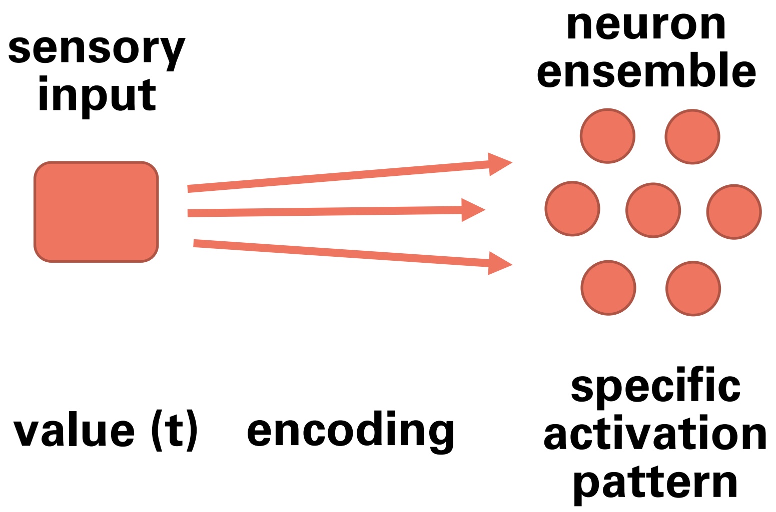
A basic feature of spiking neuron models (vs. parametric models): generated signals are stochastic and noisy in nature
→ typical biological signals



50-100 neurons -> neuron ensemble

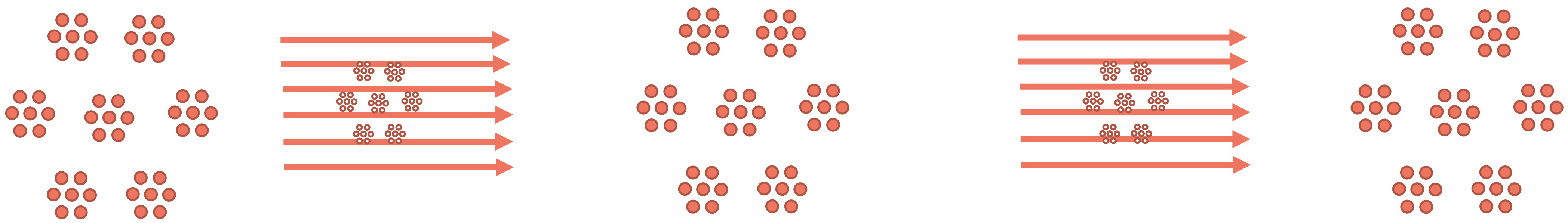
- represent **values**
  - e.g., muscular activation strength
  - or: sensory input value





50-100 ensembles -> neuron buffer

- Represent **<u>items</u>**:
  - motor plan of a syllable (complex!)
  - phonetic or phonological forms (sounds, syllables words)
  - lexemes (language-specific)
  - concepts (meanings) ((-> "thoughts"))



<u>neuron buffer</u> =
number of ensembles

neural connection =
<u>associative memory</u>

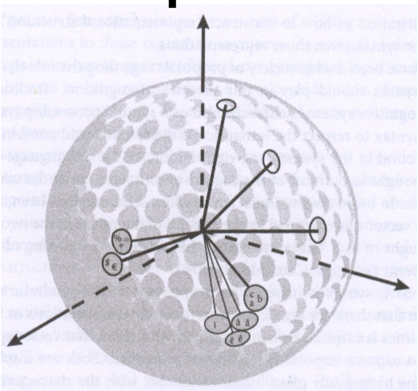
Semantic Pointer Architecture

- buffer represents complex information (**<u>items</u>** to process)
- mathematically: items are represented by vectors (called: S-pointers)
- SPA (Semantic Pointer Architecture is an important conceptual part of NENGO)

NENGO: S-pointer-networks

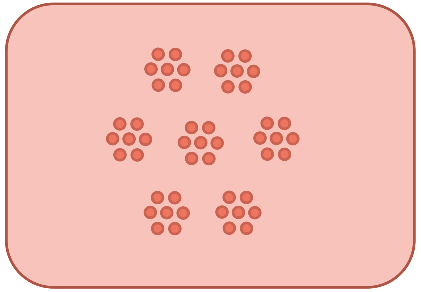
- … define an ordering of items:
- define the degree of similarity of items: e.g.,:
  - at concept level: <dog> - <cat> -> animals; <chair> - <table> -> furniture
  - at phonological level: /car/ - /cat/ -> begin with /k/; /far/ - /fat/ -> with /f/
- (mathematically: similar S-pointers point in a similar direction)

D-dimensional space



vector ordering in D-space

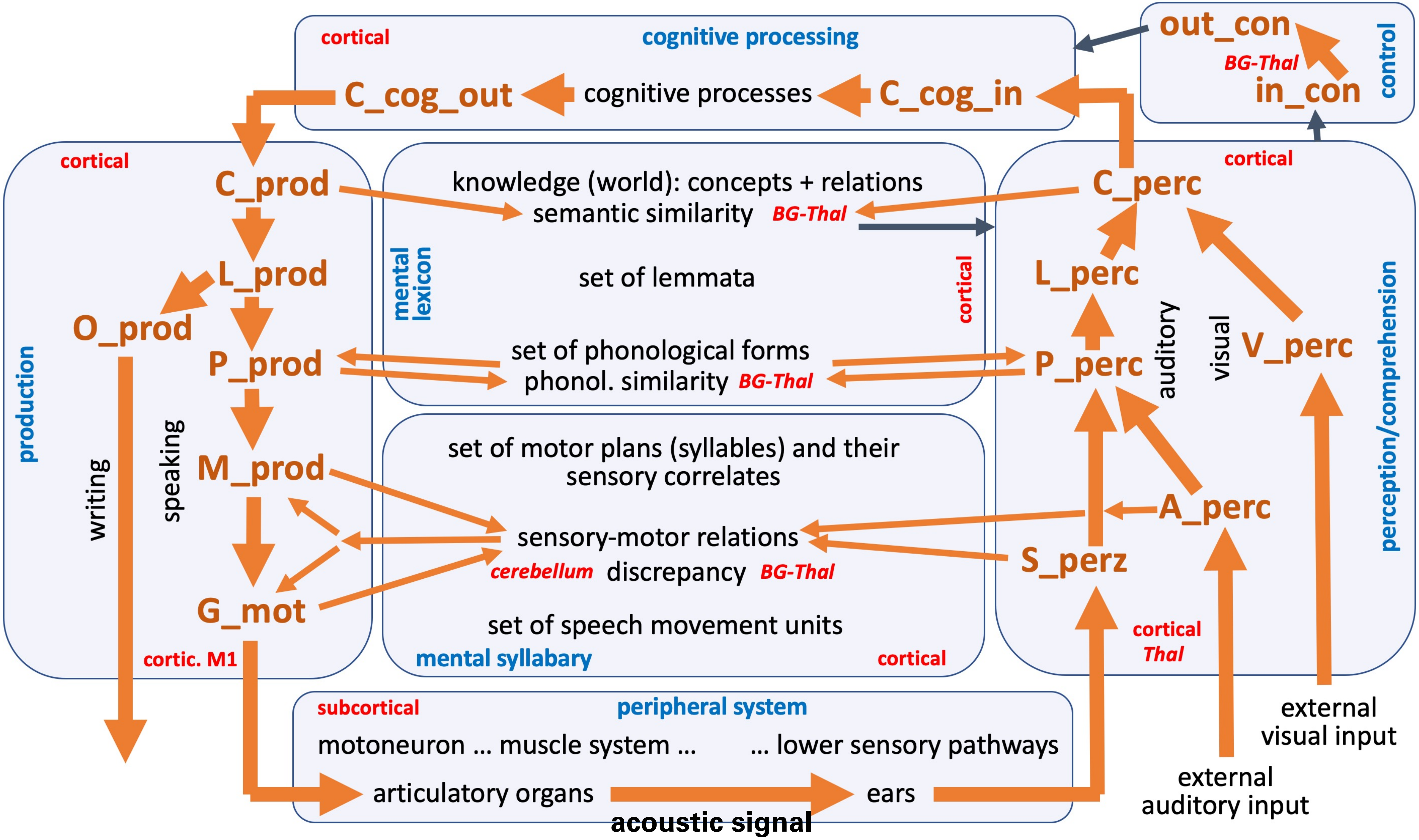
dog
cat
…
man
woman
…
table
chair



buffer A

Speech processing model: Architecture

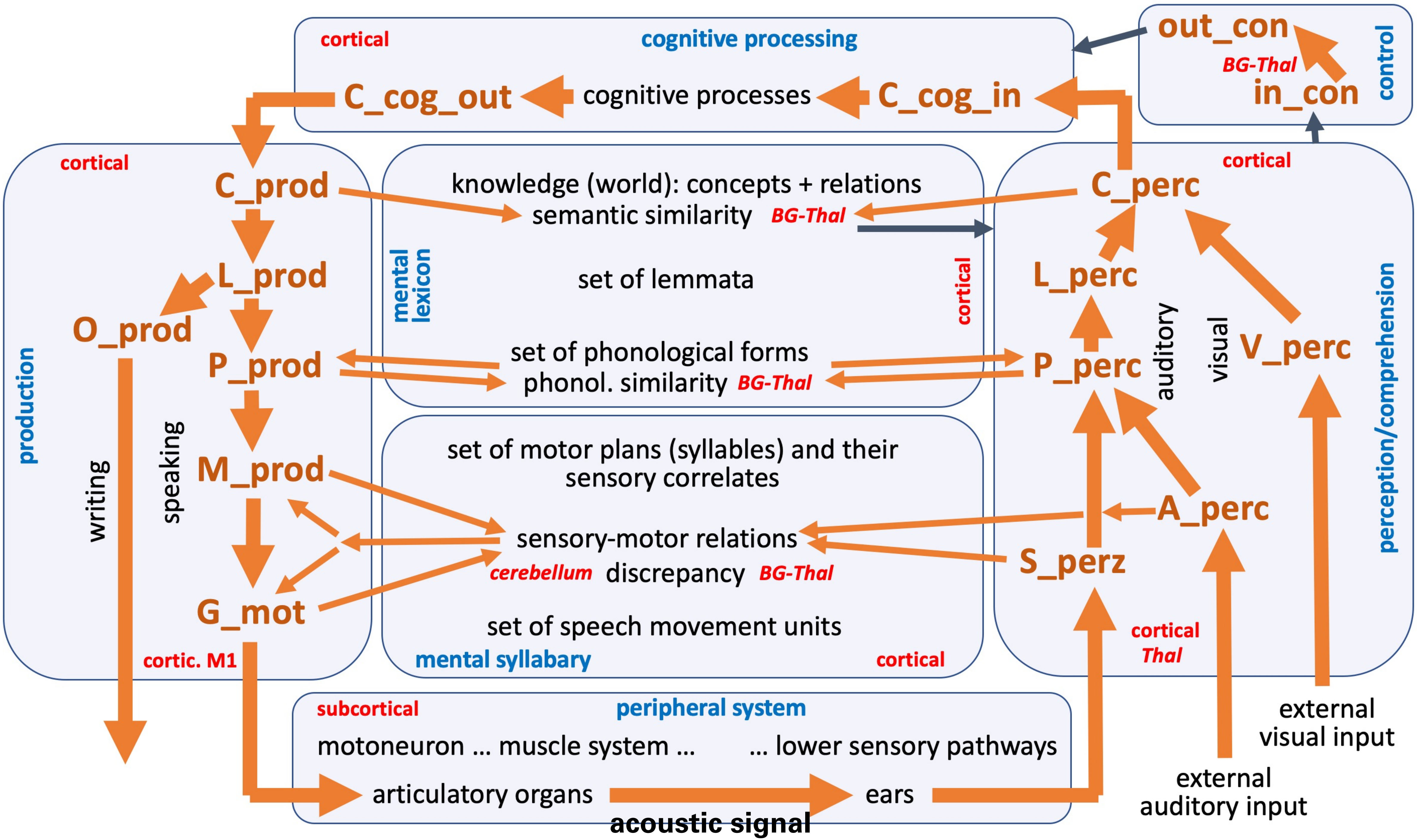
- … is a large-scale neural model
- Perception / production pathways: a chain of buffers, connected by associative memories (arrows)
- Knowledge and skill repositories: mental lexicon, mental syllabary implemented as S-pointer networks (concepts, phonological forms, motor plans)



cortical
cognitive processing
C_cog_out
cognitive processes
C_cog_in
out_con
BG-Thal
in_con
control
cortical
C_prod
L_prod
O_prod
P_prod
M_prod
G_mot
production
writing
speaking
cortic. M1
knowledge (world): concepts + relations
semantic similarity
BG-Thal
mental lexicon
set of lemmata
cortical
set of phonological forms
phonol. similarity
BG-Thal
set of motor plans (syllables) and their sensory correlates
sensory-motor relations
cerebellum
discrepancy
BG-Thal
set of speech movement units
mental syllabary
cortical
cortical
C_perc
L_perc
P_perc
auditory
visual
V_perc
A_perc
S_perz
perception/comprehension
cortical
Thal
subcortical
peripheral system
motoneuron ... muscle system ...
... lower sensory pathways
articulatory organs
ears
acoustic signal
external visual input
external auditory input

Speech processing model: Architecture

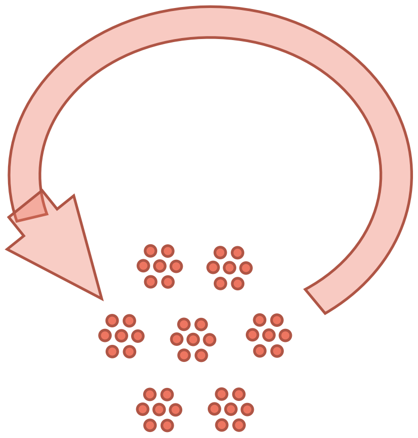
- in addition:
- action control by BG and Thalamus model
- internal feedback loops: semantic level, phonological level, sensorimotor level
- external feedback loop: motor-articulatory-acoustic-auditory



cortical
cognitive processing
out_con
BG-Thal
in_con
control
C_cog_out
cognitive processes
C_cog_in
cortical
C_prod
knowledge (world): concepts + relations
semantic similarity
BG-Thal
cortical
C_perc
L_prod
mental lexicon
set of lemmata
cortical
L_perc
O_prod
production
P_prod
set of phonological forms
phonol. similarity
BG-Thal
P_perc
auditory
visual
V_perc
perception/comprehension
writing
speaking
M_prod
set of motor plans (syllables) and their sensory correlates
sensory-motor relations
cerebellum
discrepancy
BG-Thal
A_perc
S_perz
G_mot
set of speech movement units
mental syllabary
cortical
cortical
Thal
cortic. M1
subcortical
peripheral system
motoneuron ... muscle system ...
... lower sensory pathways
articulatory organs
ears
acoustic signal
external visual input
external auditory input

The cognitive processing level

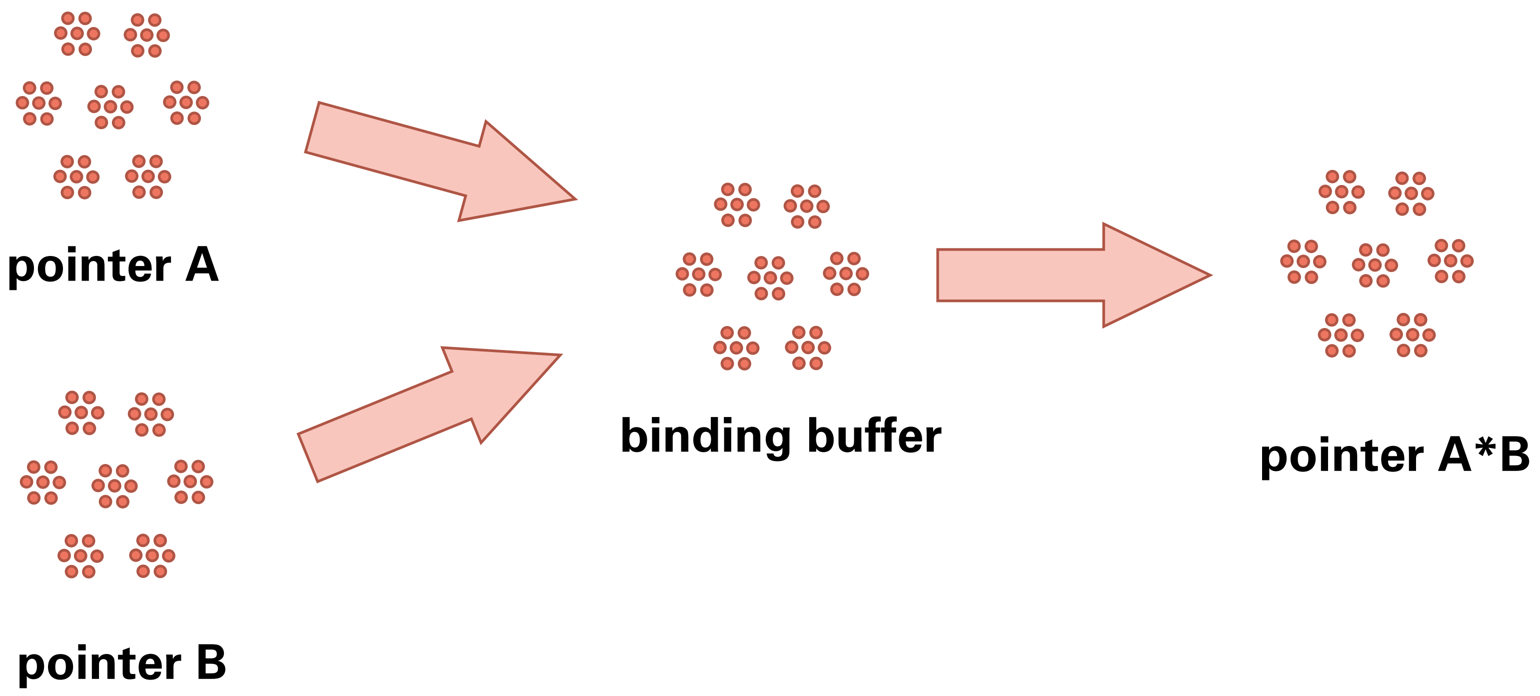
- Short-term memories, e.g., for memorizing word lists (-> serial recall task, perception + production tasks)
- realized as recursive buffers (neural within-associations)



recursive buffer

The cognitive processing level

- Binding / unbinding is realized by specific neural structures -> binding buffers vs. normal S-pointer buffers
  - at concept level: <dog> - <cat> -> animals; <chair> - <table> -> furniture
  - <dog>*<is_a> -> <animal>

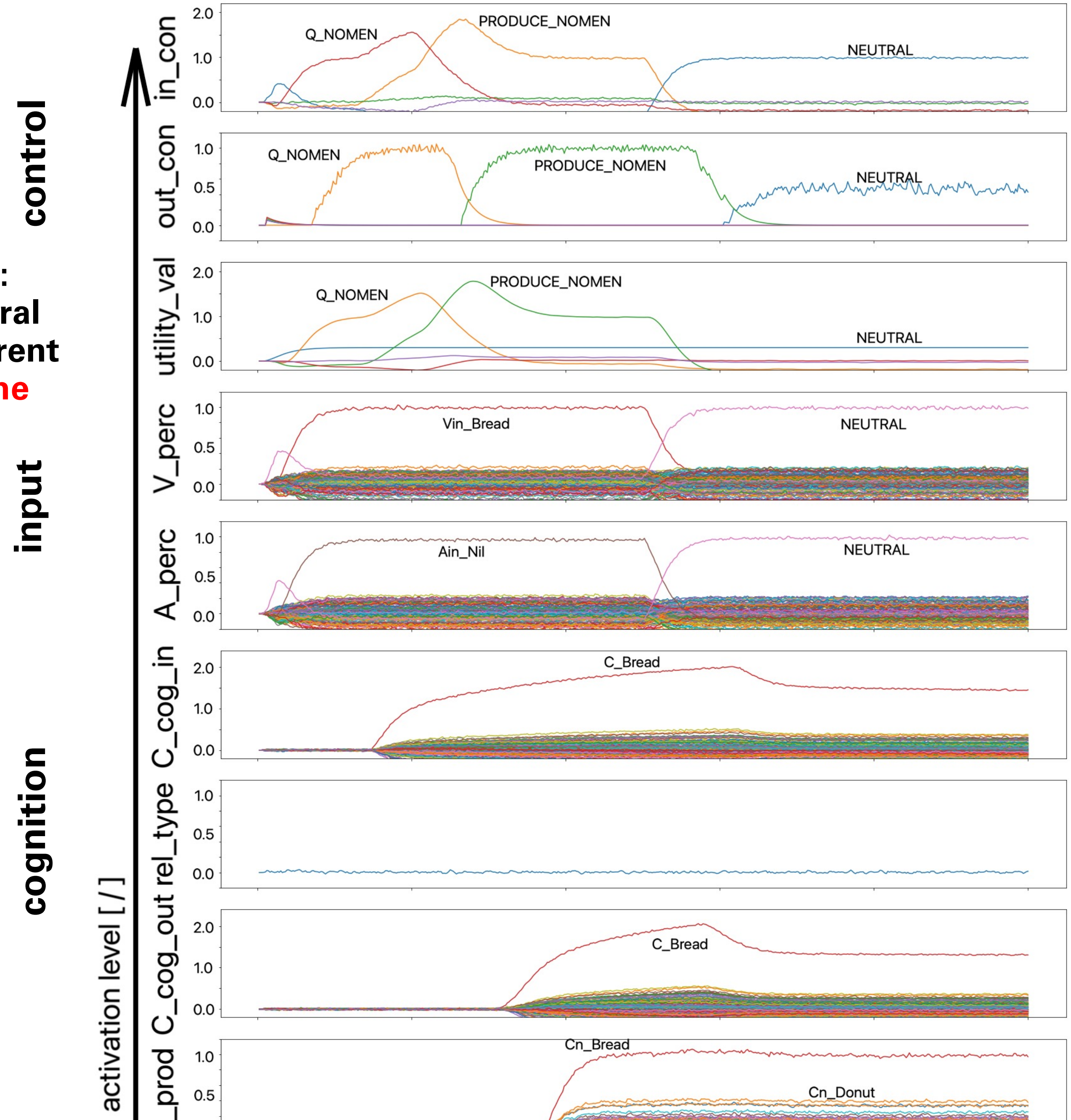


The definition of scenarios

- So far: the neural architecture of the model:
- In addition: we need to define the task which needs to be executed
- -> definition of scenarios
- Scenarios: speech tasks: (medical screenings)
  - Picture naming (visual input -> production)
  - word comprehension (auditory input -> naming of superordinate term)
  - word (logatome / syllable) repetition

Similarity plots: display the neural activity of different buffers **over time**

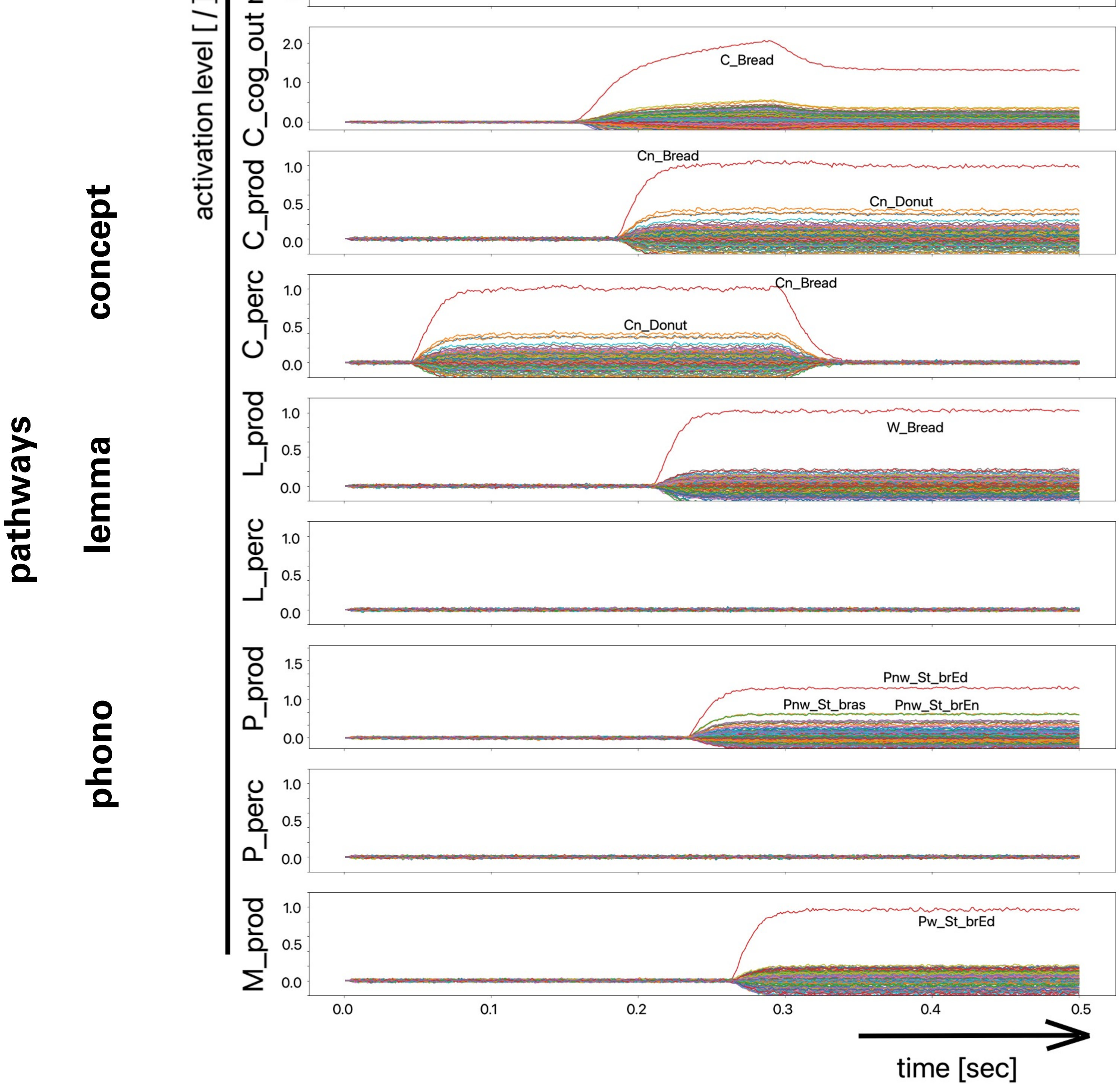
gives information about the function-ning of the model: we can see what is going on in each part of the model at each point in time



← priming: 100 msec: "look@pic and activate noun" +150 msec : "utter noun" c

Picture naming

Kröger et al. (2020)



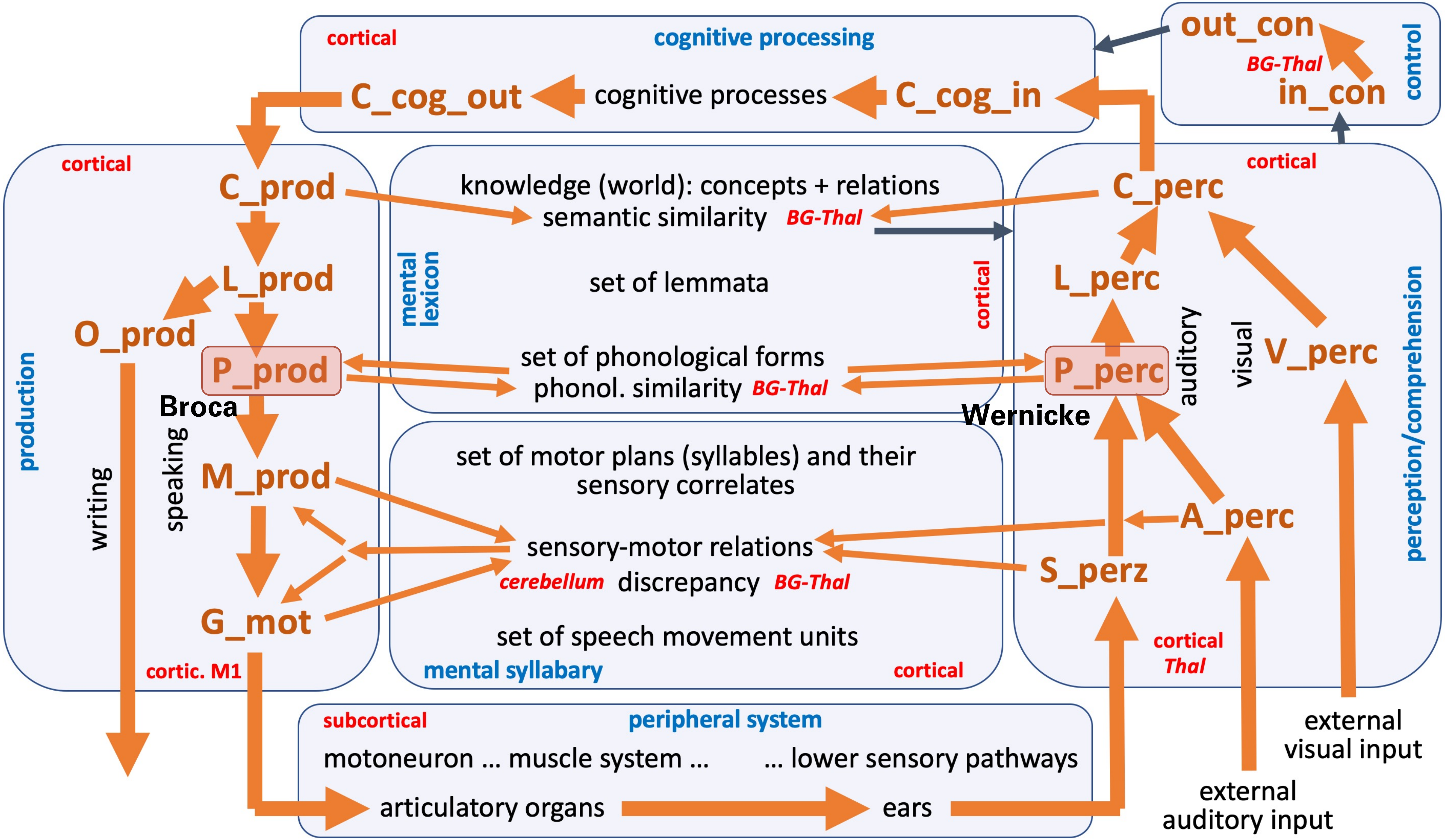
Kröger et al. (2020)

Example:
Simulation of disordered speech

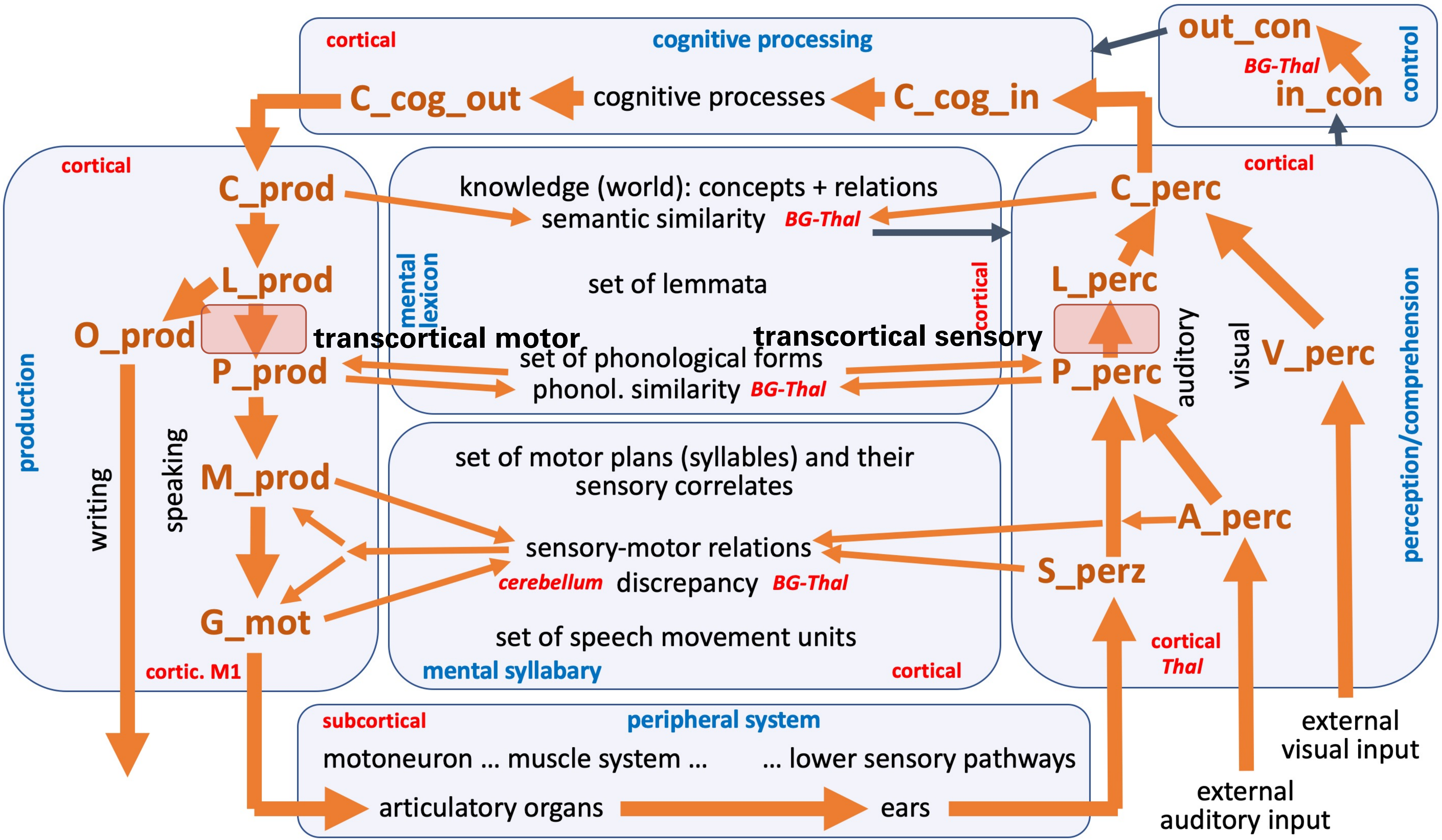
- Goal -> modelling of different types of aphasia
- Want to see typical symptoms:
- Simulation of three tasks:
  - picture naming
  - word comprehension
  - logatome repetition
- So: we need neural damage at different parts of the model

Implementation of neural dysfunctions

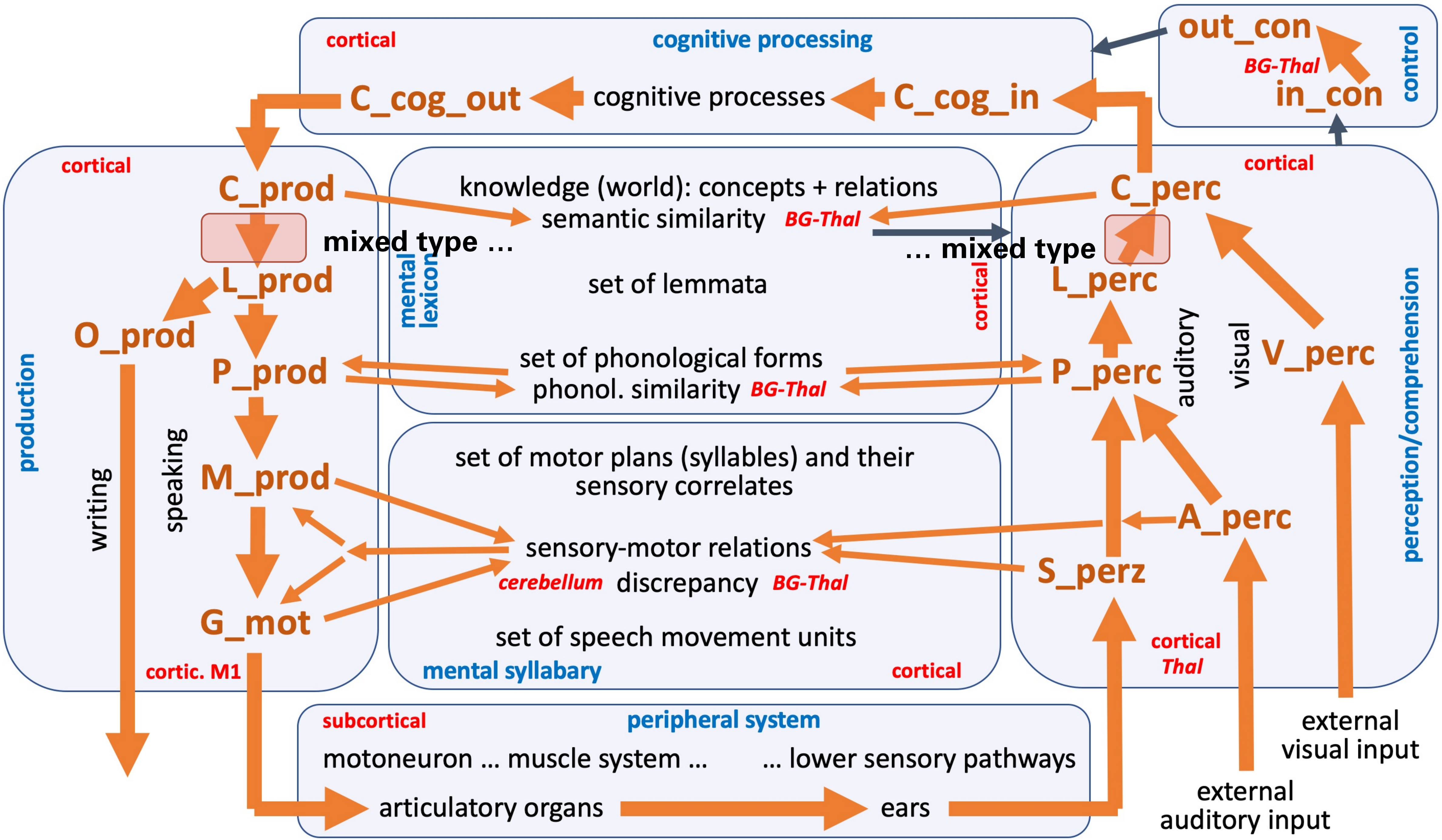
- Lesions at different levels of the model:
  - Phonological state buffer -> Broca / Wernicke aphasia
  - Associative memories between phono-lemma buffer -> transcortical motor / sensory aphasia
  - Associative memories between lemma-concept buffers -> mixed aphasia
  - Neural association between phono-phono buffers -> conduction aphasia



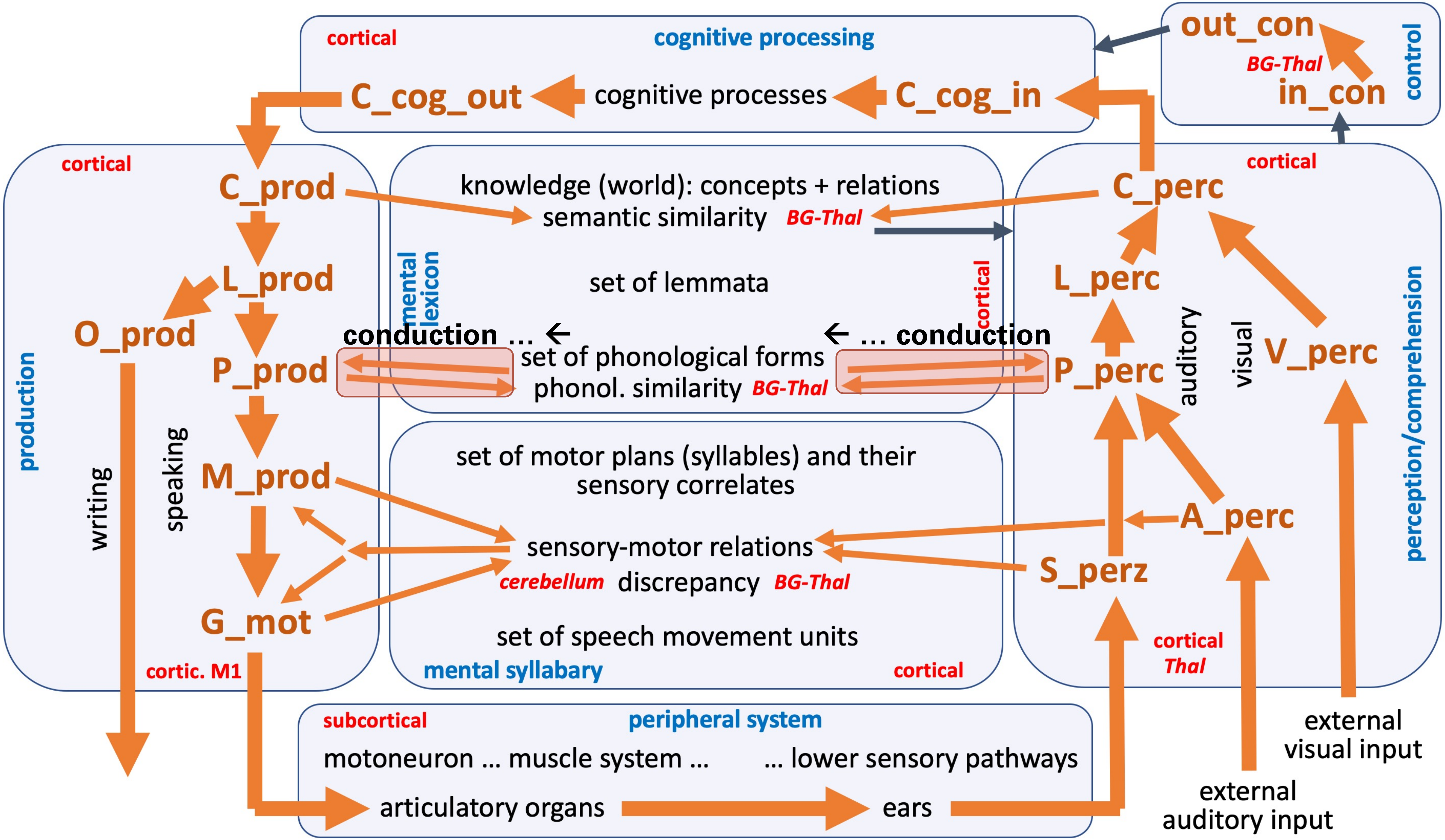
cortical
cognitive processing
C_cog_out
cognitive processes
C_cog_in
out_con
BG-Thal
in_con
control
cortical
production
C_prod
L_prod
O_prod
P_prod
Broca
writing
speaking
M_prod
G_mot
cortic. M1
knowledge (world): concepts + relations
semantic similarity
BG-Thal
mental lexicon
set of lemmata
cortical
set of phonological forms
phonol. similarity
BG-Thal
set of motor plans (syllables) and their
sensory correlates
sensory-motor relations
cerebellum
discrepancy
BG-Thal
set of speech movement units
mental syllabary
cortical
cortical
C_perc
L_perc
P_perc
Wernicke
auditory
visual
V_perc
perception/comprehension
A_perc
S_perz
cortical
Thal
subcortical
peripheral system
motoneuron … muscle system …
… lower sensory pathways
articulatory organs
ears
external
visual input
external
auditory input



cortical
cognitive processing
C_cog_out
cognitive processes
C_cog_in
out_con
BG-Thal
in_con
control
cortical
C_prod
L_prod
O_prod
P_prod
M_prod
G_mot
production
writing
speaking
cortic. M1
knowledge (world): concepts + relations
semantic similarity
BG-Thal
mental lexicon
set of lemmata
cortical
transcortical motor
transcortical sensory
set of phonological forms
phonol. similarity
BG-Thal
set of motor plans (syllables) and their sensory correlates
sensory-motor relations
cerebellum
discrepancy
BG-Thal
set of speech movement units
mental syllabary
cortical
cortical
C_perc
L_perc
P_perc
auditory
visual
V_perc
A_perc
S_perz
perception/comprehension
cortical
Thal
subcortical
peripheral system
motoneuron ... muscle system ...
... lower sensory pathways
articulatory organs
ears
external visual input
external auditory input



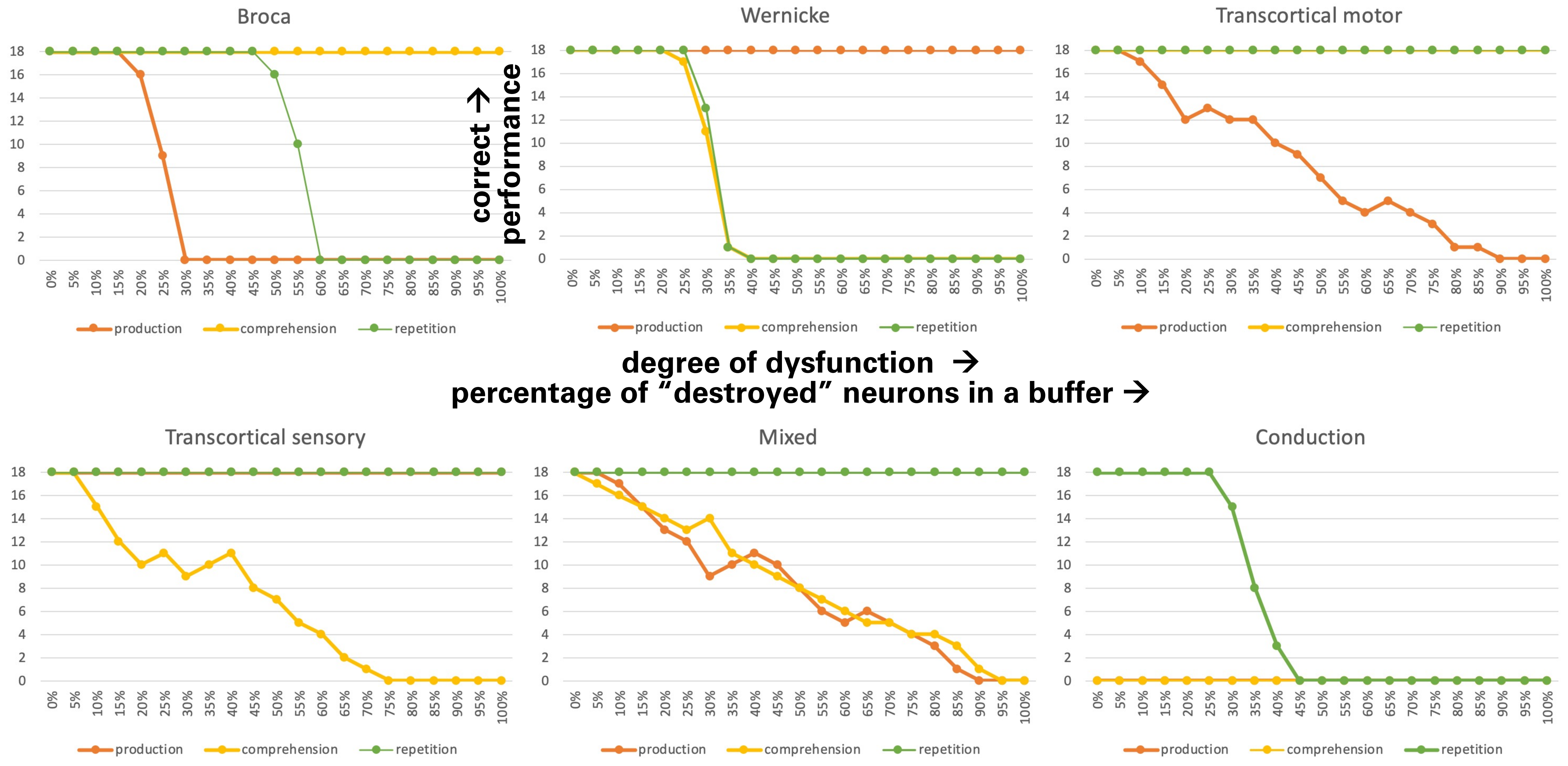
cortical
cognitive processing
C_cog_out
cognitive processes
C_cog_in
out_con
BG-Thal
in_con
control
cortical
C_prod
mixed type …
L_prod
O_prod
P_prod
M_prod
G_mot
production
writing
speaking
cortic. M1
knowledge (world): concepts + relations
semantic similarity
BG-Thal
set of lemmata
mental lexicon
cortical
set of phonological forms
phonol. similarity
BG-Thal
set of motor plans (syllables) and their sensory correlates
sensory-motor relations
cerebellum
discrepancy
BG-Thal
set of speech movement units
mental syllabary
cortical
cortical
C_perc
… mixed type
L_perc
P_perc
auditory
visual
V_perc
A_perc
S_perz
cortical
Thal
perception/comprehension
subcortical
peripheral system
motoneuron … muscle system …
… lower sensory pathways
articulatory organs
ears
external visual input
external auditory input



cortical
cognitive processing
C_cog_out
cognitive processes
C_cog_in
out_con
BG-Thal
in_con
control
cortical
C_prod
L_prod
O_prod
P_prod
M_prod
G_mot
production
writing
speaking
cortic. M1
knowledge (world): concepts + relations
semantic similarity
BG-Thal
mental lexicon
set of lemmata
cortical
conduction ... ←
← ... conduction
set of phonological forms
phonol. similarity
BG-Thal
set of motor plans (syllables) and their sensory correlates
sensory-motor relations
cerebellum
discrepancy
BG-Thal
set of speech movement units
mental syllabary
cortical
cortical
C_perc
L_perc
P_perc
auditory
visual
V_perc
A_perc
S_perz
perception/comprehension
cortical
Thal
subcortical
peripheral system
motoneuron ... muscle system ...
... lower sensory pathways
articulatory organs
ears
external visual input
external auditory input

Neural dysfunctions

- Lesions are realized by ablation (“die off” / “die back” process) for a percentage of neurons of a buffer
- -> a percentage of neurons is “destroyed” / “switched off”
- -> decrease in task performance



Kröger et al. (2020)

The definition of scenarios

- So far: simple tasks and implementation of neural dysfunctions
- we want to demonstrate the power of the neural model:
- We tried to simulate more complex tasks:
  - to simulate overburdening -> may lead to speech errors
- Further speech screenings:

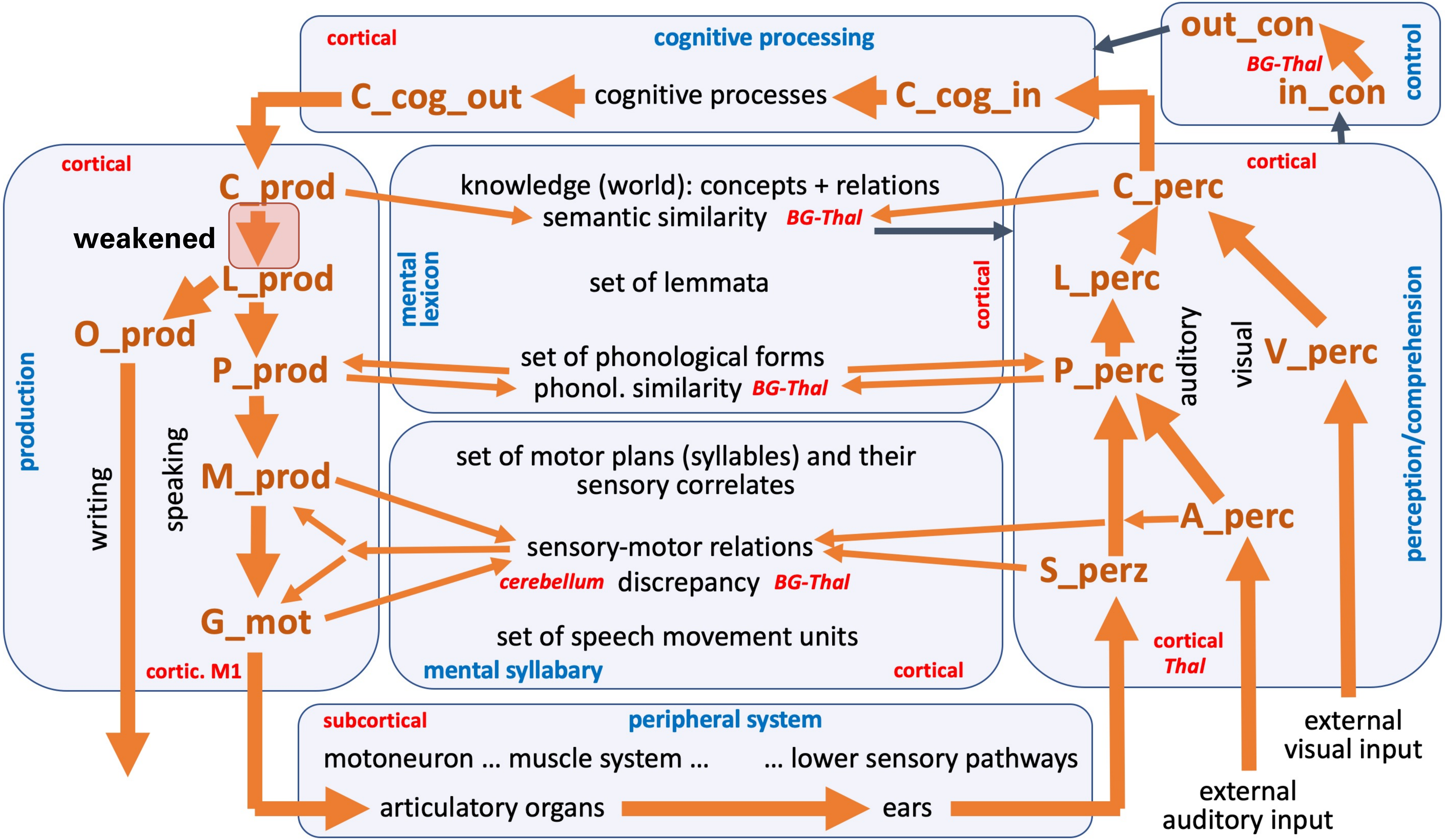
Result:

- Typical symptoms occurring in different subtypes of aphasia can be modelled correctly
- E.g.;
  - decrease performance of speech production -> Broca
  - ... perception -> Wernicke
  - ...

Further goal: demonstrate the power of our model

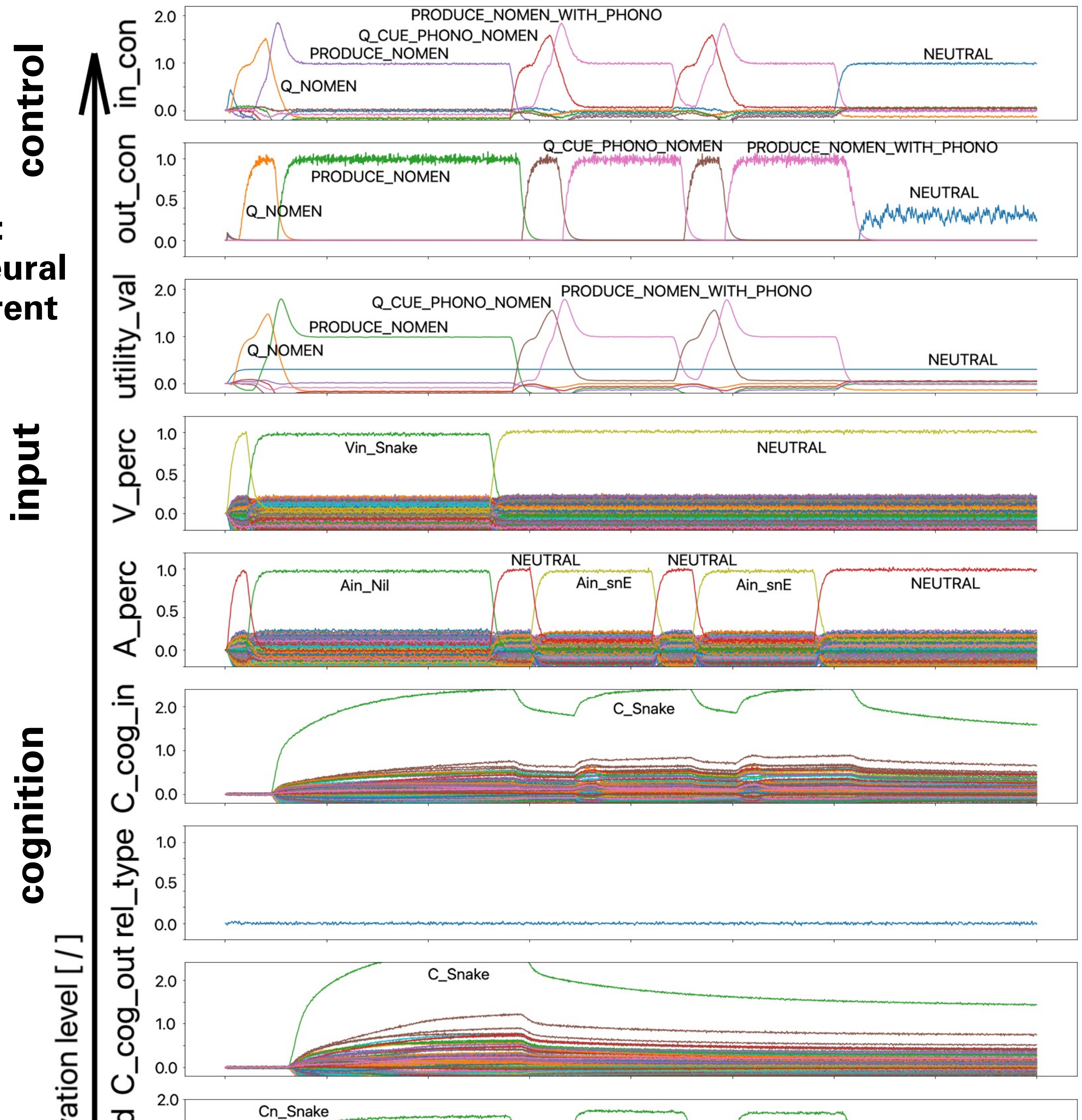
- Can we model more complex tasks?
- Can we model overburdening situations?
- Further speech screenings: (-> demonstrating realism of model)
  - E.g.: Picture naming in case of lexical access problems ("tip of the tongue" case):
    - Problem to find a correct lexeme (in case of full concept activation!)
    - facilitation (help) by introducing phonological and semantic cues -> increase in performance

Kroeger BJ, Stille C, Blouw P, Bekolay T, Stewart TC (2020) Frontiers in Computational Neuroscience, 14:99. www.speechtrainer.eu **-> publications**



cortical
cognitive processing
C_cog_out
cognitive processes
C_cog_in
out_con
BG-Thal
in_con
control
cortical
C_prod
weakened
L_prod
O_prod
P_prod
M_prod
G_mot
production
writing
speaking
cortic. M1
knowledge (world): concepts + relations
semantic similarity
BG-Thal
mental lexicon
set of lemmata
cortical
set of phonological forms
phonol. similarity
BG-Thal
set of motor plans (syllables) and their sensory correlates
sensory-motor relations
cerebellum
discrepancy
BG-Thal
set of speech movement units
mental syllabary
cortical
cortical
C_perc
L_perc
P_perc
auditory
visual
V_perc
A_perc
S_perz
cortical
Thal
perception/comprehension
subcortical
peripheral system
motoneuron … muscle system …
… lower sensory pathways
articulatory organs
ears
external visual input
external auditory input

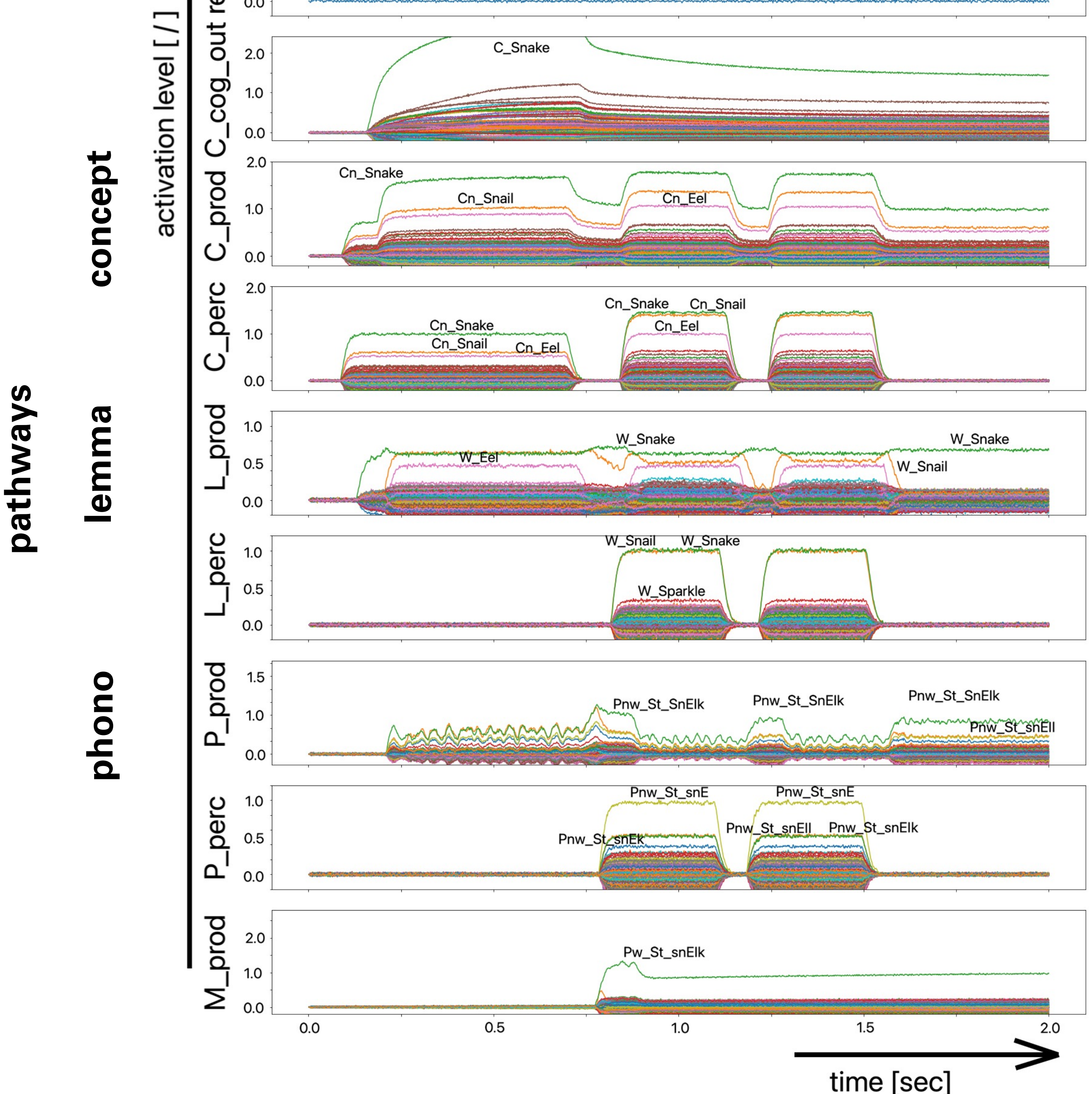
Similarity plots: representing neural activity of different buffers



← priming: 100 msec: "look and think" + 600 msec : "utter noun"
+ later: "listen to cues" + "utter noun"

Picture naming plus phonological cue

Kröger et al. (2020)



phonological cue in!

Kröger et al. (2020)

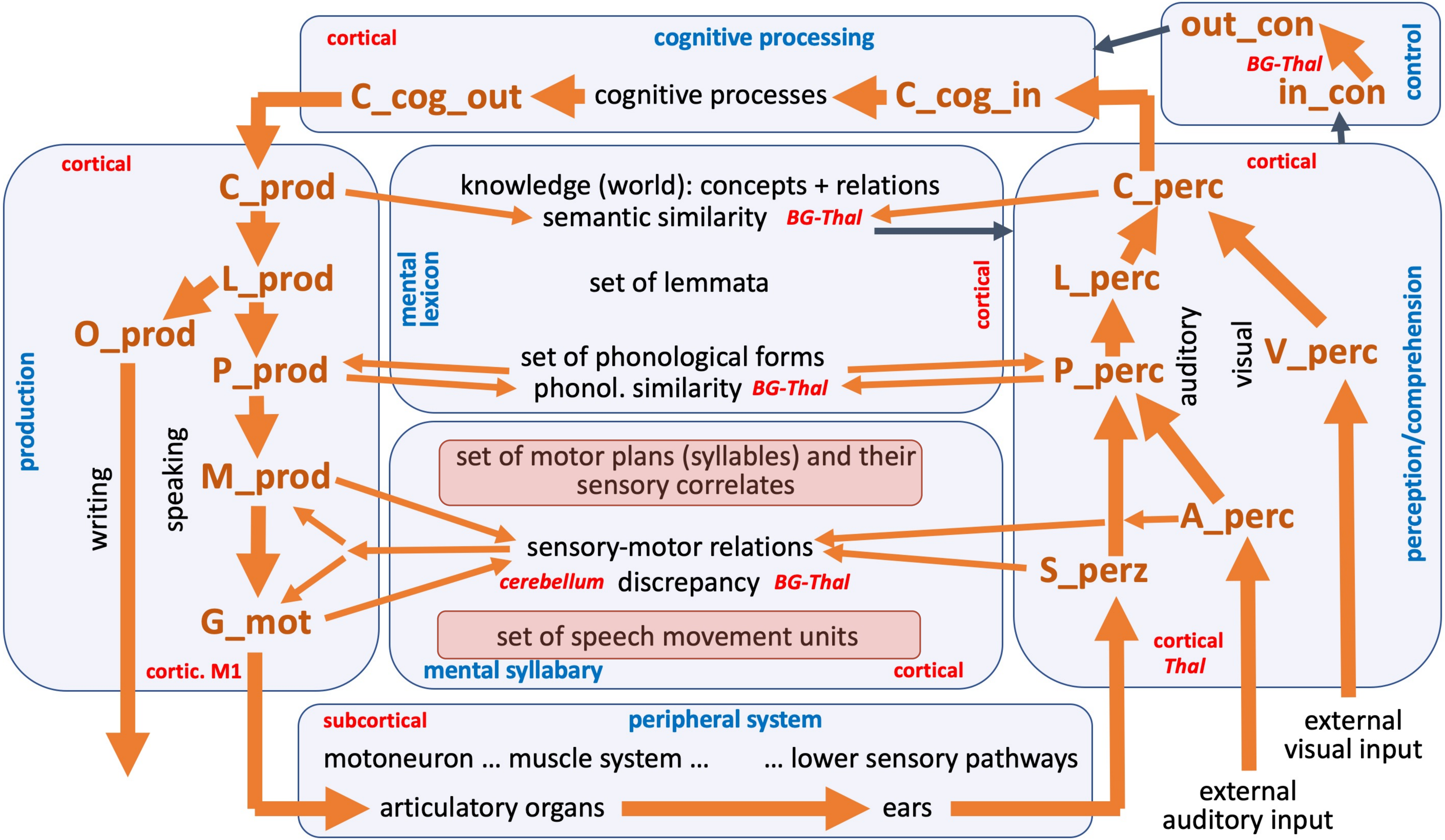
Conclusion: cues are helpful

- 2 different types of cues are helpful: semantic / phonological
- Starting with 20-45% of correct word naming (because of neural dysfunction)
- increase to about 55-70% as result of cues
- (this type of scenario sometimes is used in lexical access screenings)

Kröger et al. (2020)

The sensorimotor part of the model

- Results shown so far -> cortical / lexical part of the model
- Now: how to model the sensorimotor part?
- How to generate articulatory speech movement patterns using a spiking neuron model?
- How to generate speech gesture scores (speech movement unit / speech action scores)?



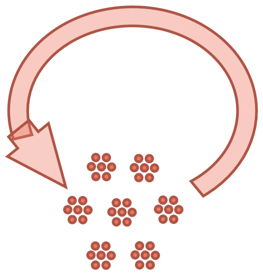
cortical
cognitive processing
C_cog_out
cognitive processes
C_cog_in
out_con
BG-Thal
in_con
control
cortical
C_prod
L_prod
O_prod
P_prod
M_prod
G_mot
production
writing
speaking
cortic. M1
knowledge (world): concepts + relations
semantic similarity
BG-Thal
mental lexicon
set of lemmata
cortical
set of phonological forms
phonol. similarity
BG-Thal
set of motor plans (syllables) and their sensory correlates
sensory-motor relations
cerebellum
discrepancy
BG-Thal
set of speech movement units
mental syllabary
cortical
cortical
C_perc
L_perc
P_perc
auditory
visual
V_perc
perception/comprehension
A_perc
S_perz
cortical
Thal
subcortical
peripheral system
motoneuron … muscle system …
… lower sensory pathways
articulatory organs
ears
external visual input
external auditory input

The sensorimotor part of the model

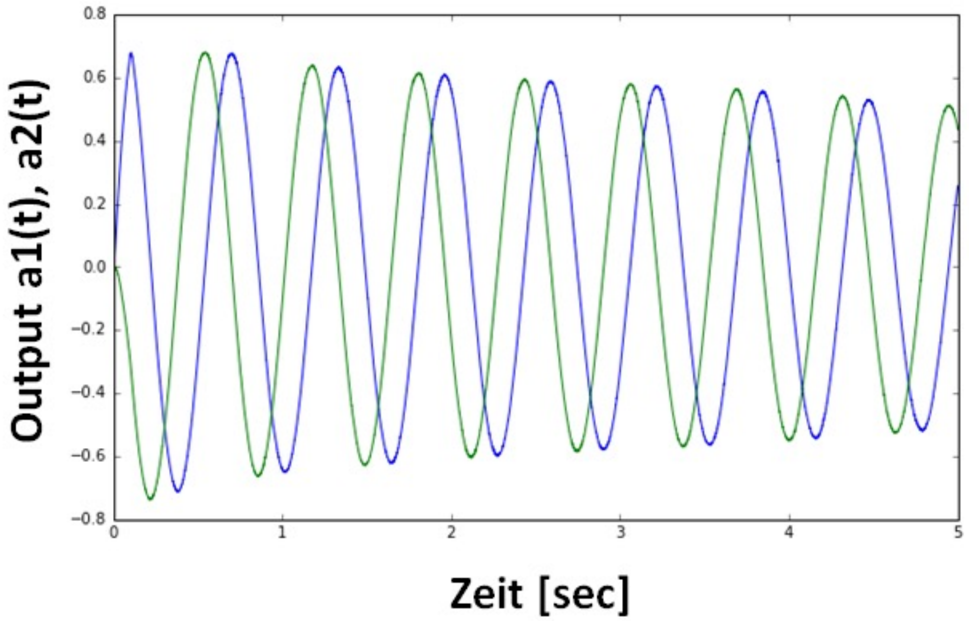
Idea: Basic concept for the production side: realized using

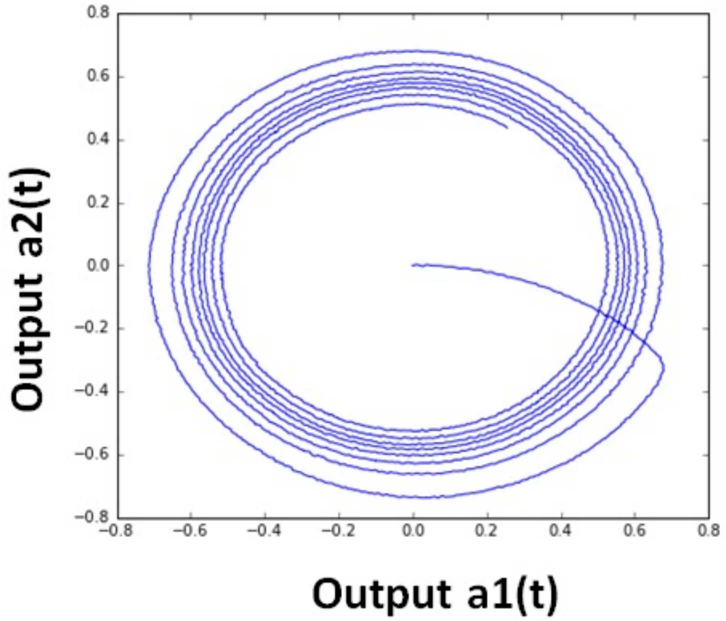
- motor plan (= gesture score) realized as neural oscillator
- gesture (= speech movement unit SMU: vocalic, consonantal, velopharyngeal, glottal) -> neural oscillator

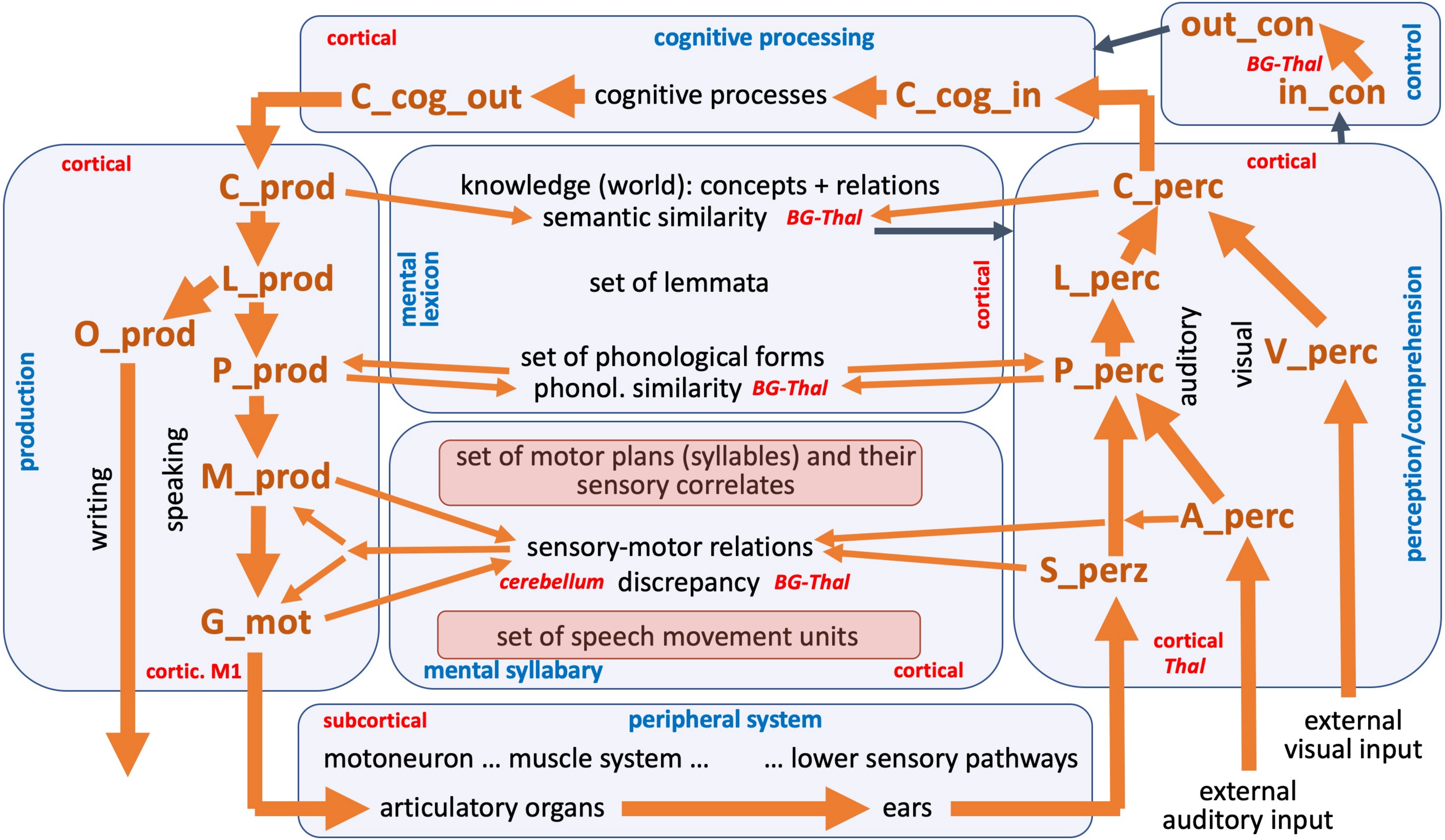
Neural oscillators are buffers:



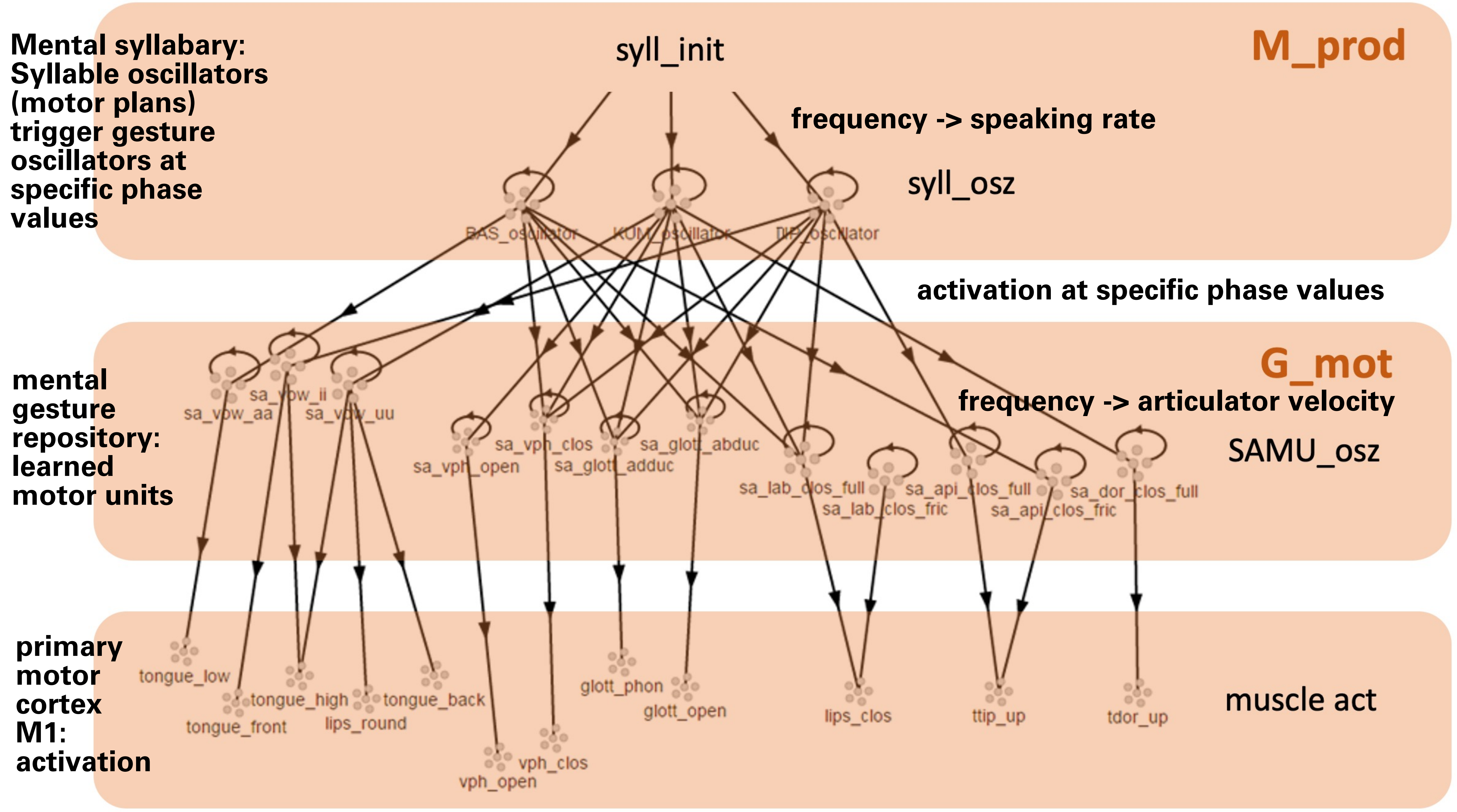
- Buffers with recursive neural connections -> short-term memories or oscillators
- activation transfer from 50% of the neurons to the other 50% of the neurons of the buffer and back ...





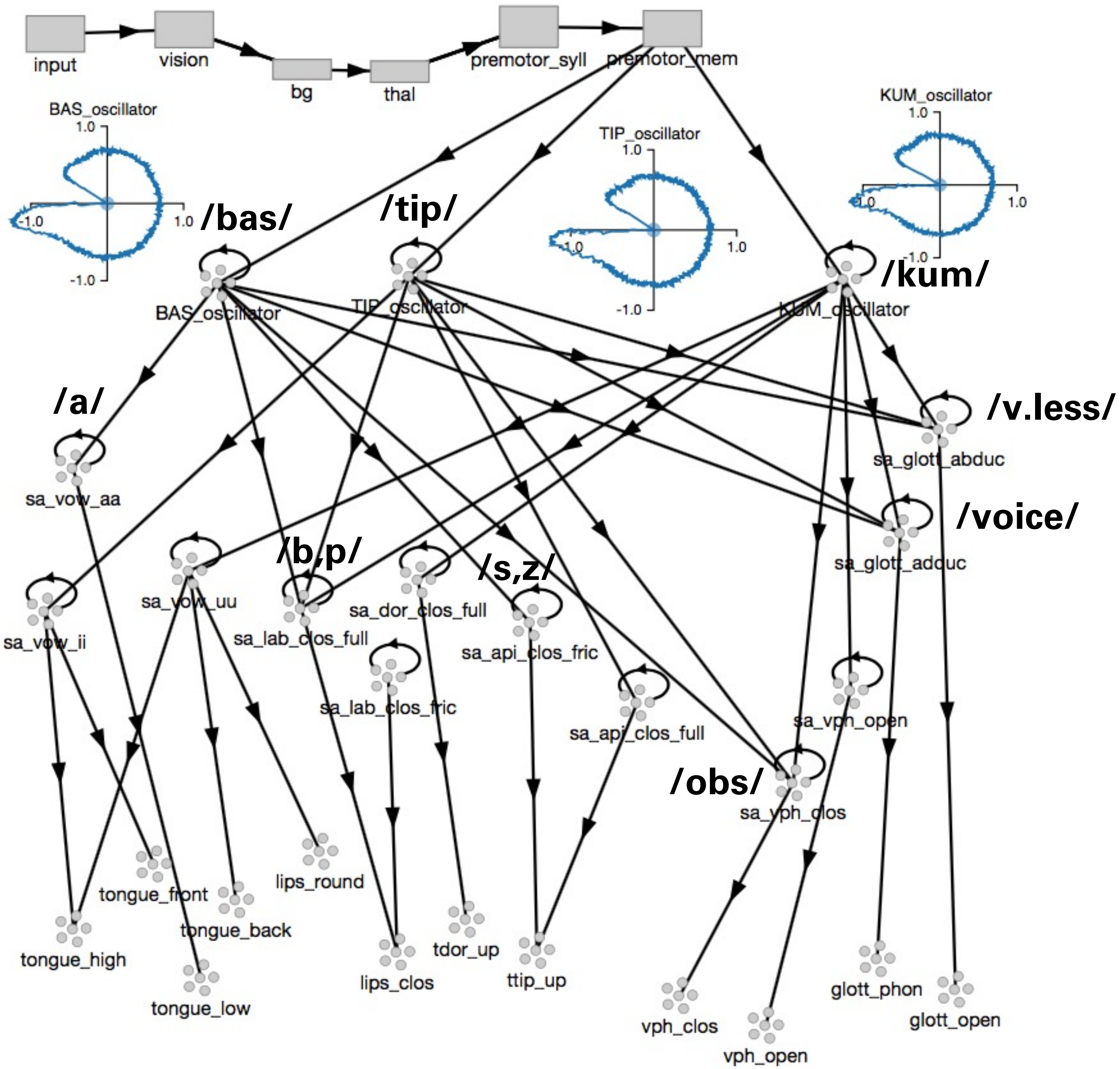


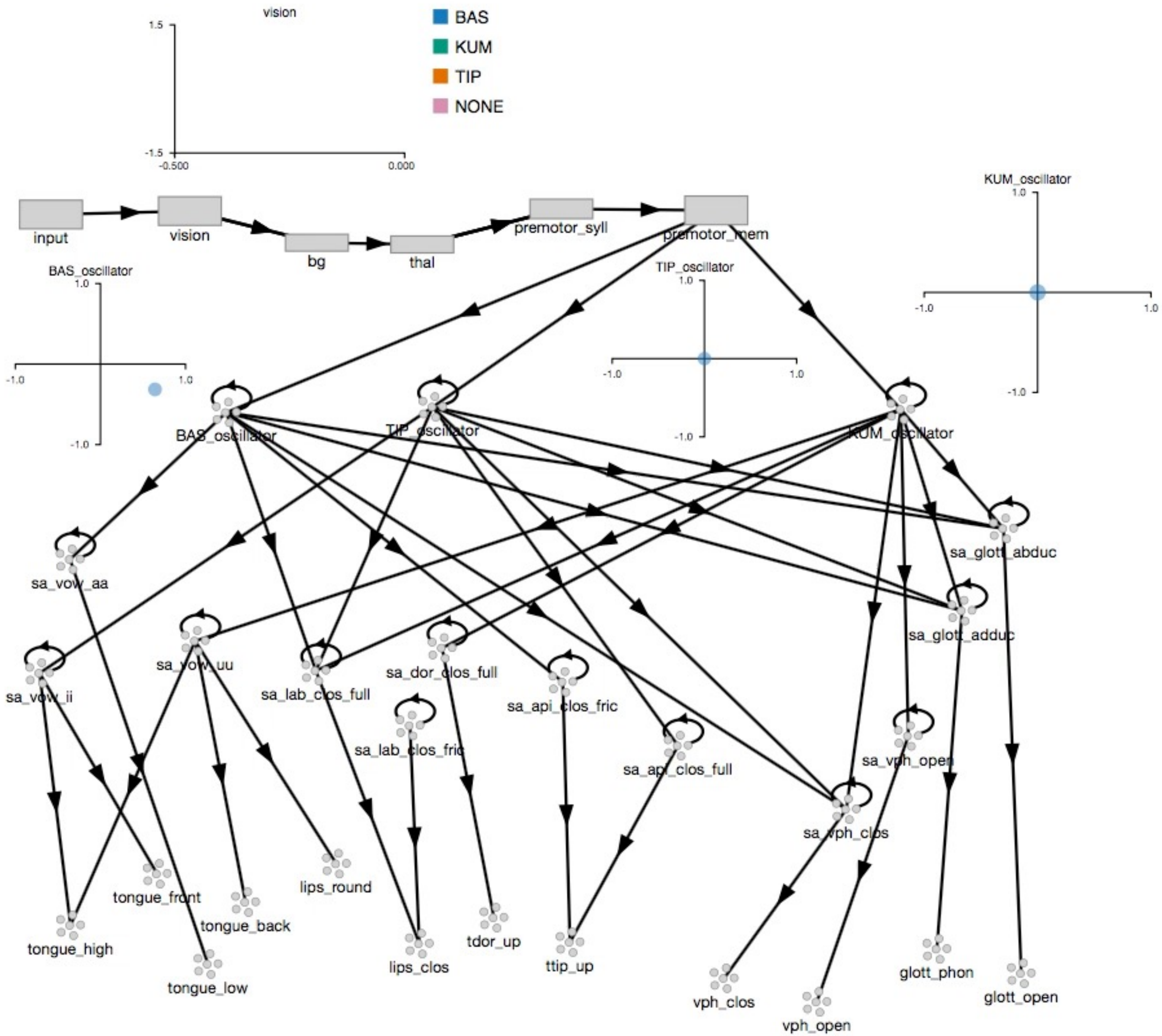
cortical
cognitive processing
C_cog_out
cognitive processes
C_cog_in
out_con
BG-Thal
in_con
control
cortical
production
C_prod
L_prod
O_prod
P_prod
M_prod
G_mot
writing
speaking
cortic. M1
mental lexicon
knowledge (world): concepts + relations
semantic similarity
BG-Thal
set of lemmata
set of phonological forms
phonol. similarity
BG-Thal
cortical
set of motor plans (syllables) and their sensory correlates
sensory-motor relations
cerebellum
discrepancy
BG-Thal
set of speech movement units
mental syllabary
cortical
cortical
C_perc
L_perc
P_perc
auditory
visual
V_perc
A_perc
S_perz
perception/comprehension
cortical
Thal
subcortical
peripheral system
motoneuron ... muscle system ...
... lower sensory pathways
articulatory organs
ears
external visual input
external auditory input


Mental syllabary: Syllable oscillators (motor plans) trigger gesture oscillators at specific phase values
syll_init
M_prod
frequency -> speaking rate
syll_osz
BAS_oscillator
KUM_oscillator
TIP_oscillator
activation at specific phase values
G_mot
frequency -> articulator velocity
mental gesture repository: learned motor units
sa_vow_aa
sa_vow_ii
sa_vow_uu
sa_vph_open
sa_vph_clos
sa_glott_adduc
sa_glott_abduc
sa_lab_clos_full
sa_lab_clos_fric
sa_api_clos_full
sa_api_clos_fric
sa_dor_clos_full
SAMU_osz
primary motor cortex M1: activation
tongue_low
tongue_front
tongue_high
lips_round
tongue_back
vph_open
vph_clos
glott_phon
glott_open
lips_clos
ttip_up
tdor_up
muscle act

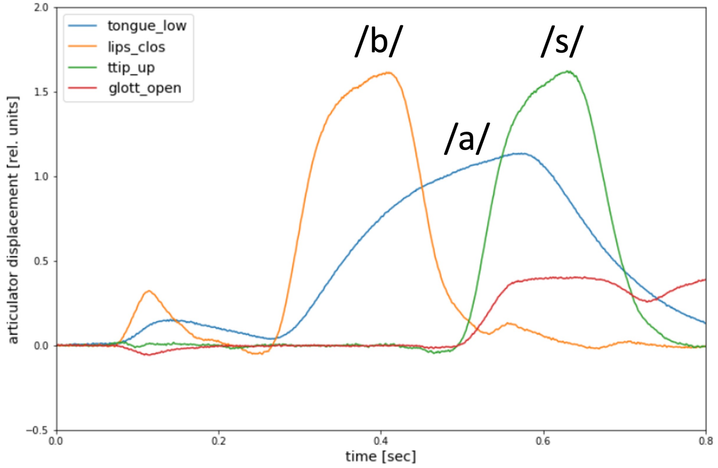
Example word: /bas kum tip/

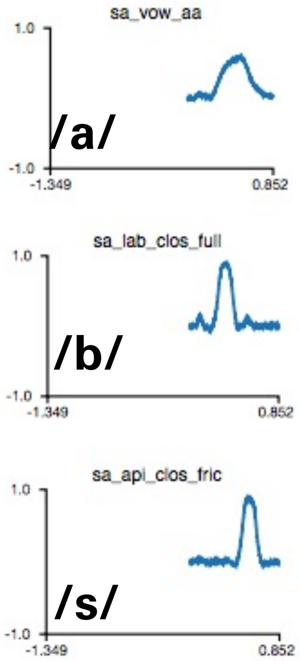
picture is generated by Nengo_gui →

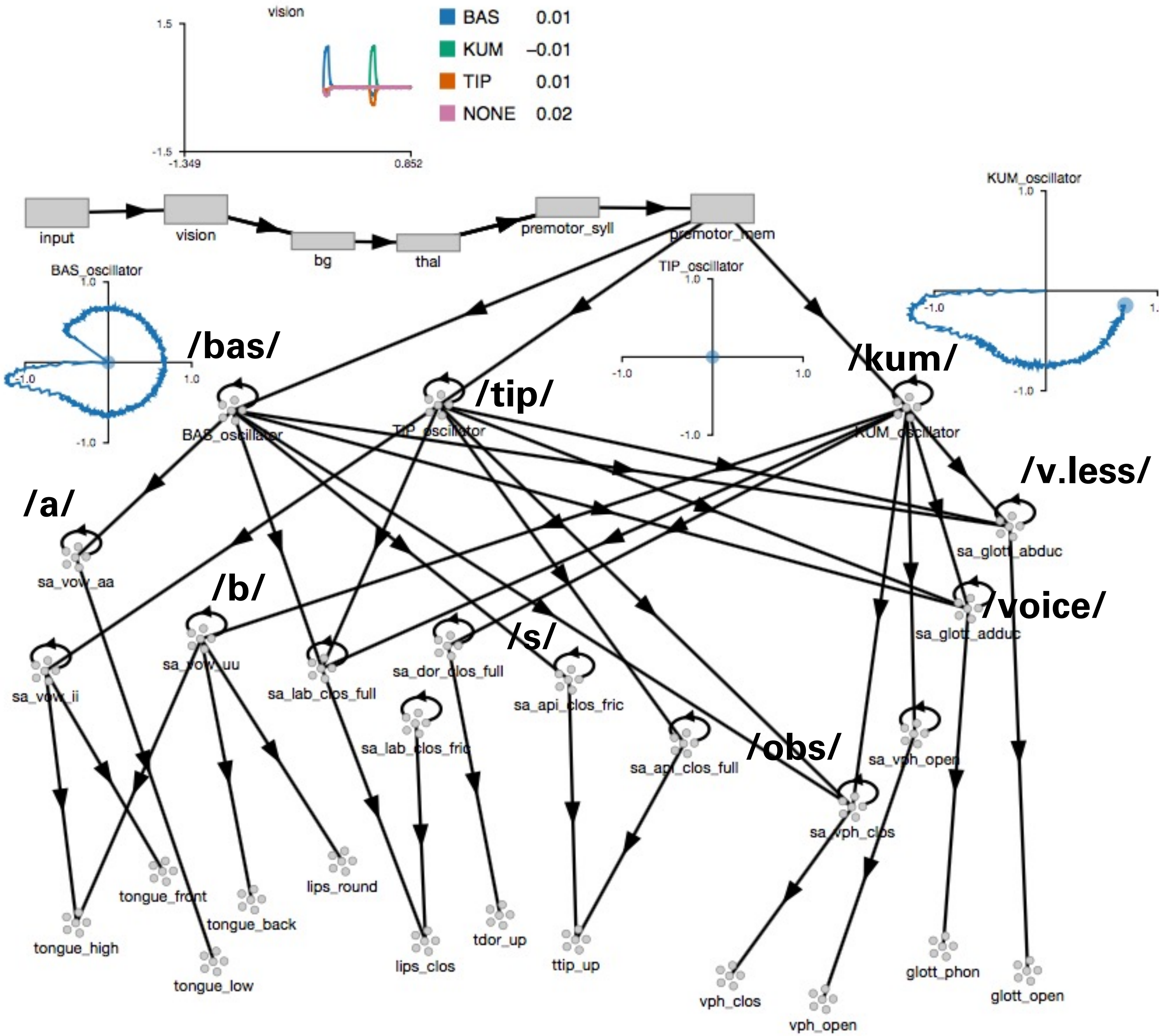



vision
1.5
-1.5
-0.500
0.000
BAS
KUM
TIP
NONE
input
vision
bg
thal
premotor_syll
premotor_mem
BAS_oscillator
1.0
-1.0
1.0
-1.0
TIP_oscillator
1.0
-1.0
1.0
-1.0
KUM_oscillator
1.0
-1.0
1.0
-1.0
BAS_oscillator
TIP_oscillator
KUM_oscillator
sa_vow_aa
sa_vow_uu
sa_vow_ii
sa_lab_clos_full
sa_dor_clos_full
sa_lab_clos_fric
sa_api_clos_fric
sa_api_clos_full
sa_glott_abduc
sa_glott_adduc
sa_vph_open
sa_vph_clos
tongue_high
tongue_front
tongue_back
tongue_low
lips_round
lips_clos
tdor_up
ttip_up
vph_clos
vph_open
glott_phon
glott_open

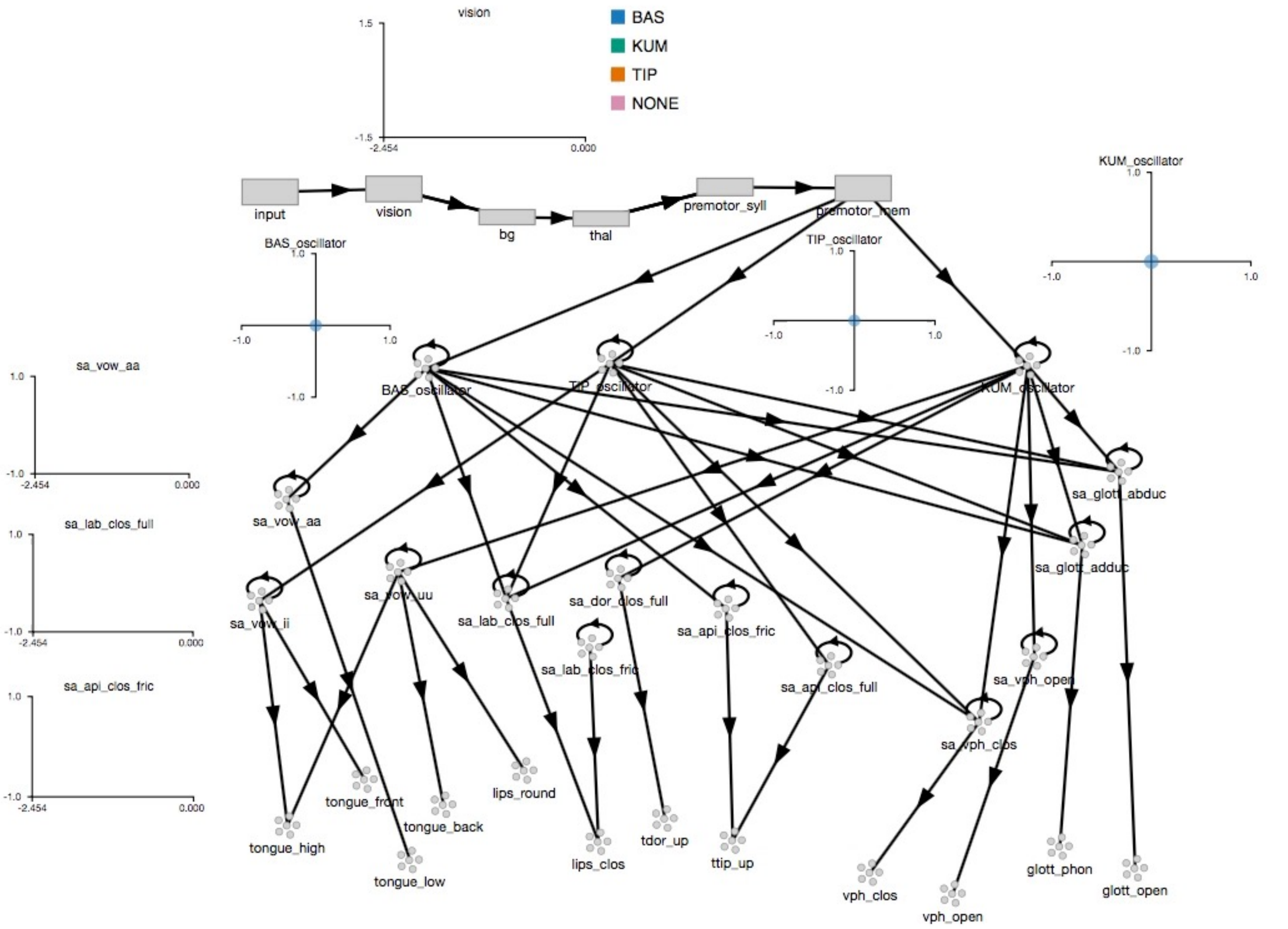
Syllable /bas/





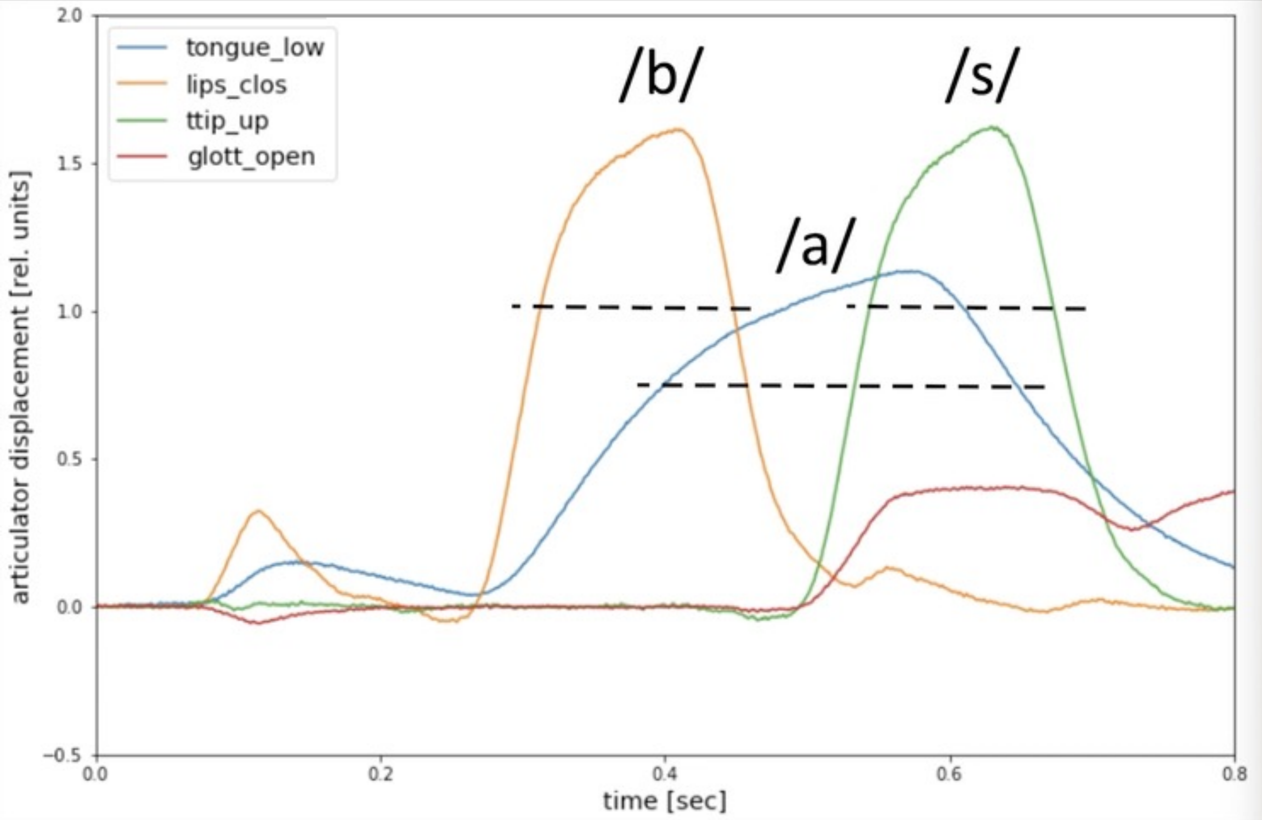


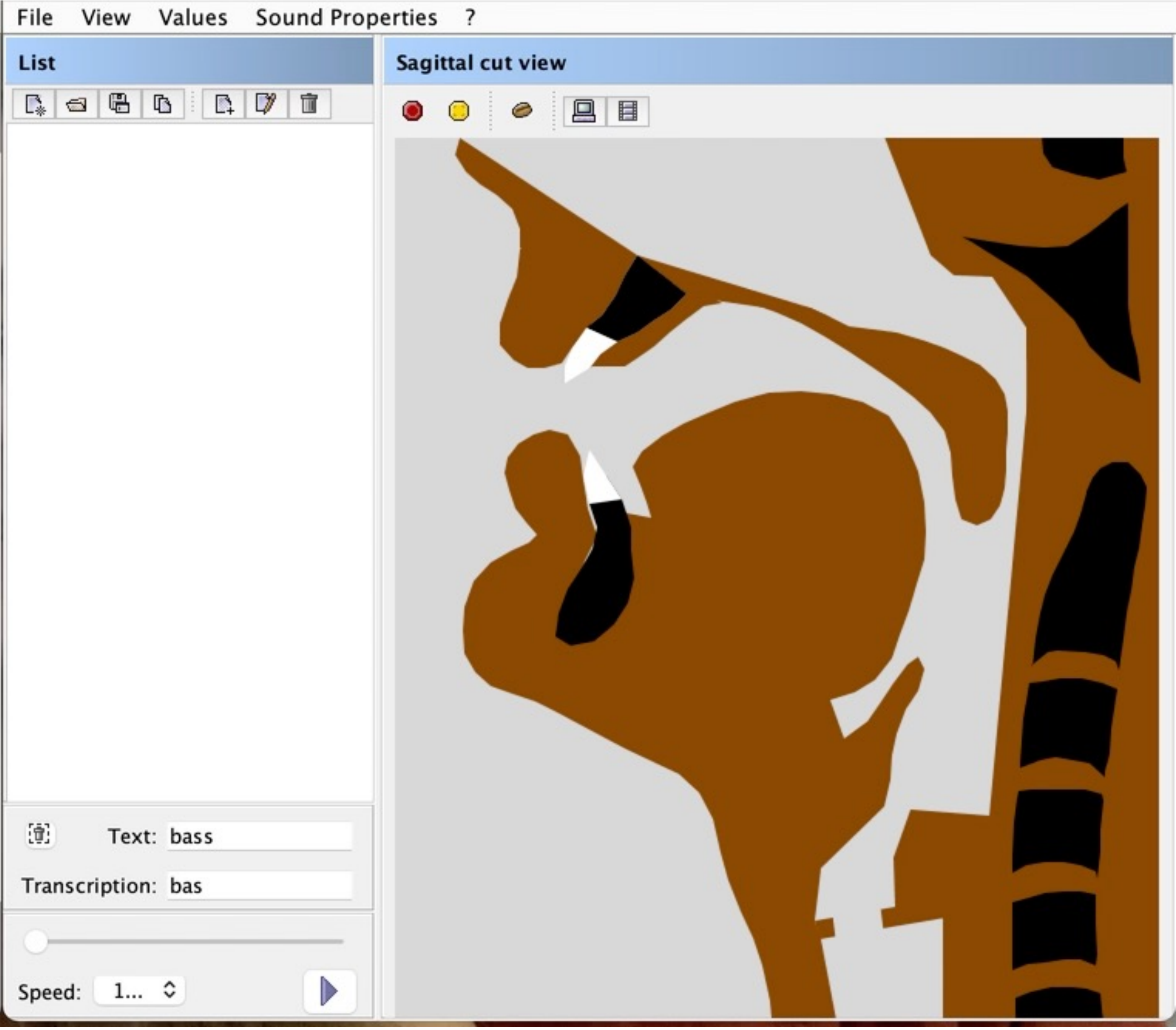
Gesture oscillators: degree of neural activation
-> degree of realization of a gesture
-> "abstract" movement trajectory



Gesture oscillators: degree of neural activation -> degree of realization of a gesture -> “abstract” movement trajectory

2D-articulatory model





Why oscillators?

- Oscillators define time intervals (<-> duration of one oscillation cycle)
- Oscillators are able to define timing of actions:
  - actions may start at specific points in time within an oscillation cycle -> phase values

Application of our approach: Modeling different speaking rates

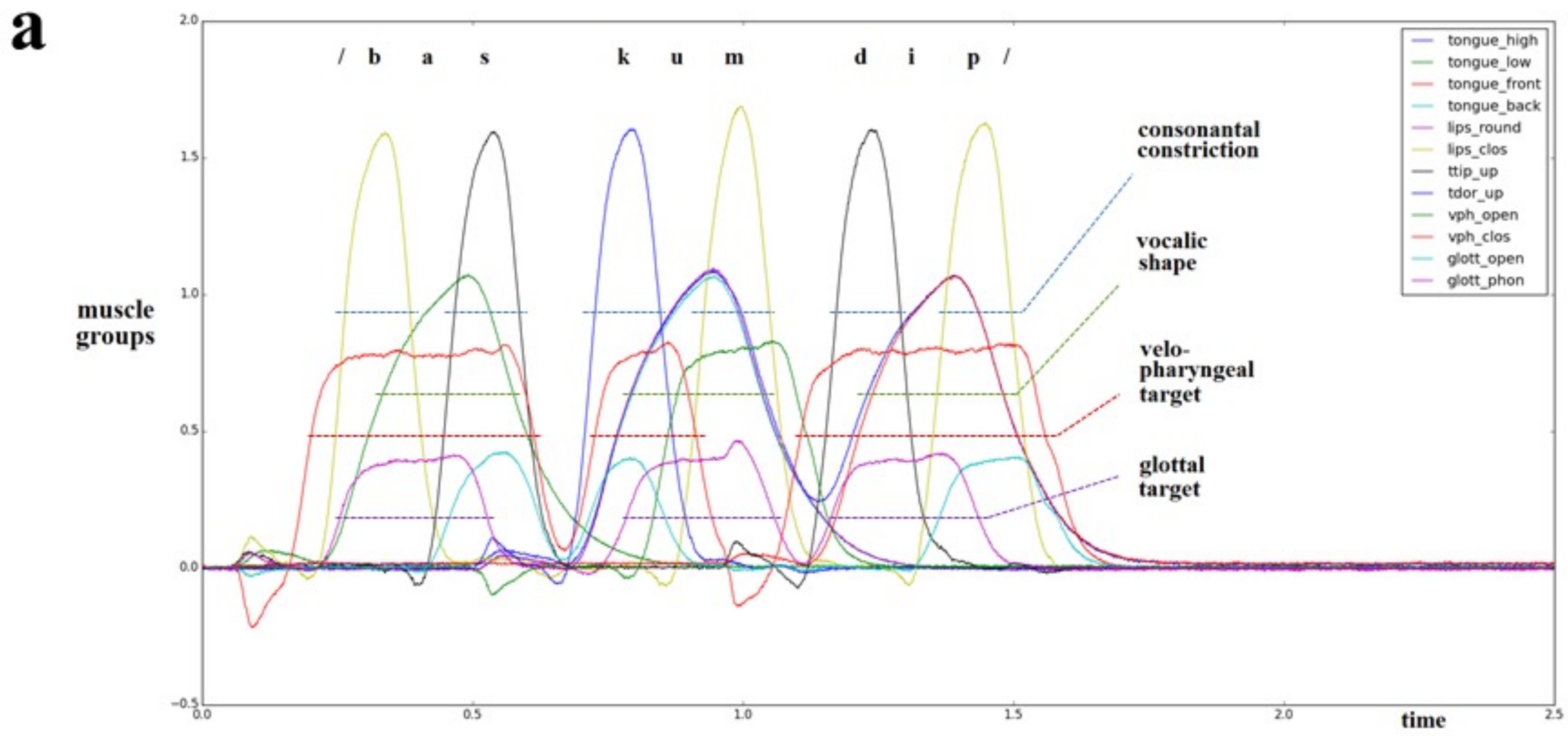
- phase values (coded in the motor plan / syllable level) stay constant
- oscillator frequency for gestures stays constant (coding vocalic and consonantal articulator velocities)
- oscillator frequency of syllable oscillators (motor plan level) is allowed to vary!! -> (increases with increasing speaking rate)

Kroeger, BJ, Bekolay T, Blouw P, Stewart TC (2020) Proceedings of the International Seminar on Speech Production ISSP2020. Yale University, New Haven, CT.
www.spechtrainer.eu -> publications

Results

- constant phasing of actions leads to correct production of speech sounds at each speaking rate:
- slow speaking rate:

no increase in articulatory effort per speech action

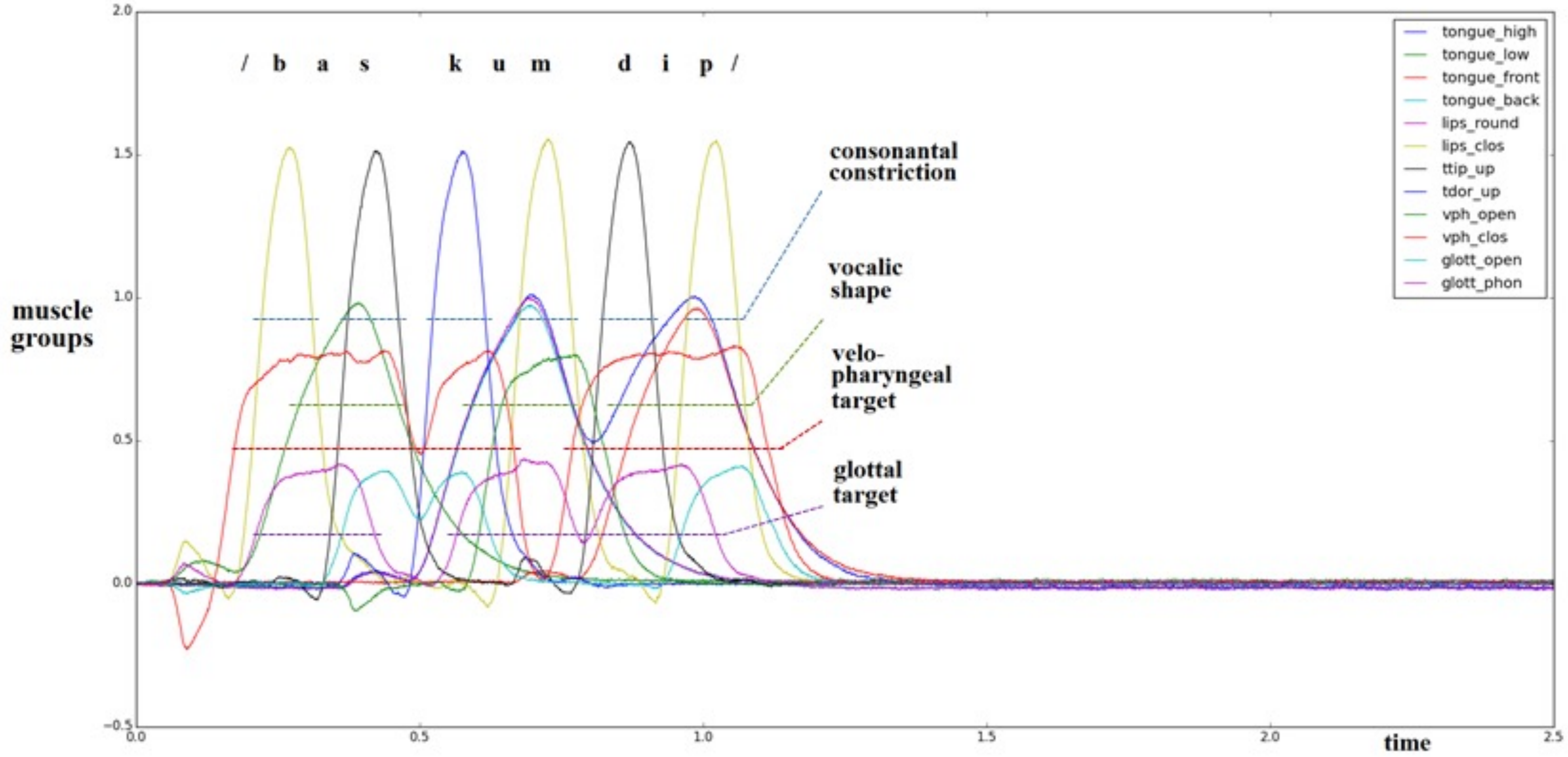


Results

- constant phasing of actions leads to correct production of speech sounds at each speaking rate:
- normal speaking rate:

no increase in articulatory effort per speech action

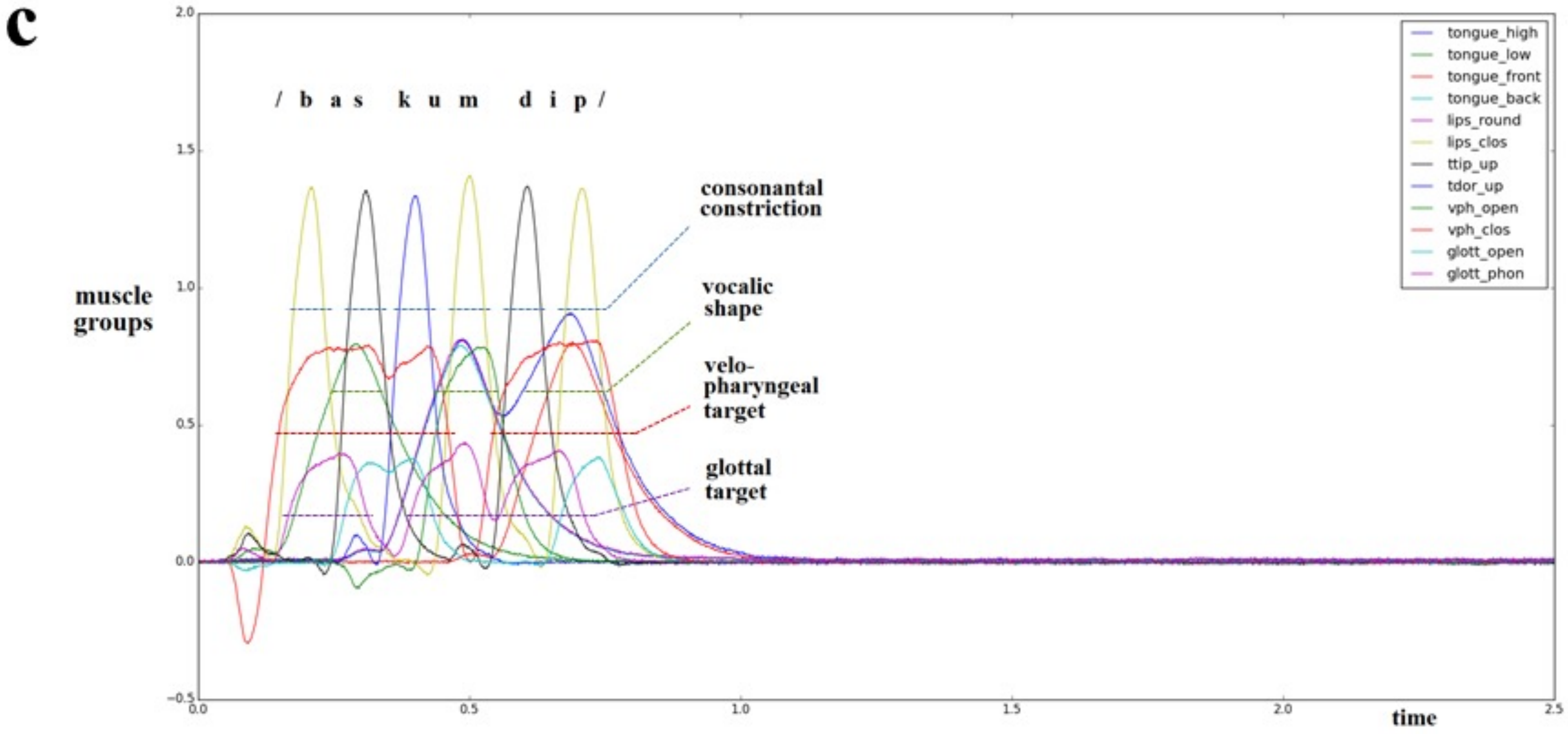
b



Results

- constant phasing of actions leads to correct production of speech sounds at each speaking rate:
- fast speaking rate:

no increase in articulatory effort per speech action



Overall conclusions

- the large-scale model approximates human behavior surprisingly well in many cases: e.g.,
  - Modeling different types of aphasia -> decrease of performance for production or perception or repetition tasks, depending on the type of dysfunction
  - cues are helpful in case of “tip of the tongue”-situations: -> increase of correct word naming from 20-45% (no cues) to 55-70% -> cues are helpful in specific situations of overburdening (borderline situations)
- a basic concept for sensorimotor part of the model exists

Benefits for medical research from modeling:

Medical / speech therapy research:

- creation of specific “model patients” suffering from specific neural dysfunctions with specific degrees of severity
- multiple simulations can be performed: with each specific model patient
- that allows to optimize medical screenings (evaluate its sensitivity for a specific disorder)

THANK YOU VERY MUCH

Download this talk as pdf from:
www.speechtrainer.eu -> publications