Neurocomputational Model of Speech Production, Speech Perception, and Speech Acquisition

Summarizing Our Work

Bernd J. Kröger

Department of Phoniatrics, Pedaudiology, and Communication Disorders
RWTH Aachen University, Germany

Outline

- Part 1: The neurocomputational model: production and acquisition: structure of the model and gaining knowledge
- Part 2: The neurocomputational model: perception: auditory perception (CP) and audio-visual perception: McGurk-Effect
- Part 3: The concept of speech actions and its relation to manual and facial actions in face-to-face communication

Outline: Part 2

- Introduction
- Simulation of Speech Perception
  - Categorical Perception
  - McGurk Effect
- Conclusions

Outline: Part 2

- Introduction
- Simulation of Speech Perception
  - Categorical Perception
  - McGurk Effect
- Conclusions

Introduction

- Perception is an integral part of each production model, because speech acquisition (needed for acquiring knowledge for the production model) needs
  - self-perception (for babbling and imitation training) as well as
  - perception of external speakers (perception of communication partners: care taker, mother; for imitation training)
- Is the production model capable of showing typical effects of speech perception, i.e. Categorical Perception, McGurk-Effect, … ?

Outline: Part 2

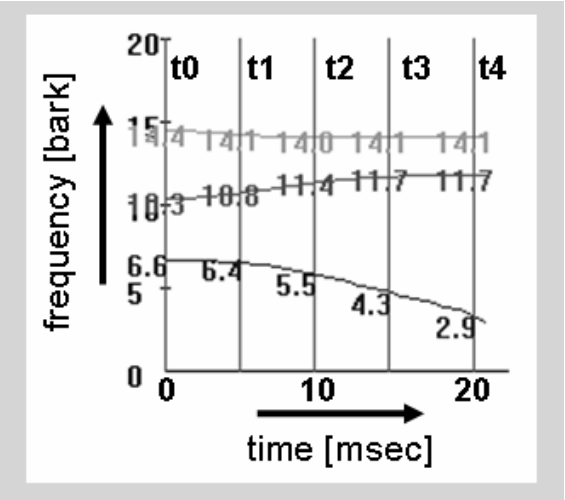
- Introduction
- Simulation of Speech Perception
  - Categorical Perception
  - McGurk Effect
- Conclusions

Categorical Perception

- Whether perception is continuous or categorical can be measured by performing identification and discrimination experiments.
- Basis:
  - an acoustically equidistant stimulus continuum
  - a pool of around 20 listeners for performing the experiments
- Further question is: are consonants (place of articulation in CV) perceived more categorical than vowels?

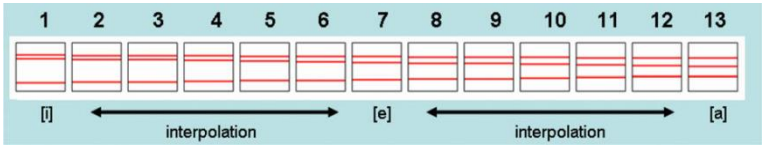
Categorical Perception

- Two stimulus continua for V: from /i/ … to /a/ and for CV from /ba/ … to /ga/
- Identification experiment: do you hear /i/, /e/, or /a/? …. /ba/, /da/, or /ga/?
- Discrimination experiment: you get ABX with A and B of constant distance (1-3, 2-4, 3-5, …); Question: X equals A or B?



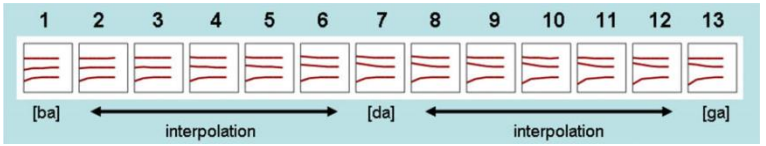
acoustically equidistant

13 V- and CV-Stimuli



[i] interpolation [e] interpolation [a]

acoustically equidistant



[ba] interpolation [da] interpolation [ga]

Categorical Perception

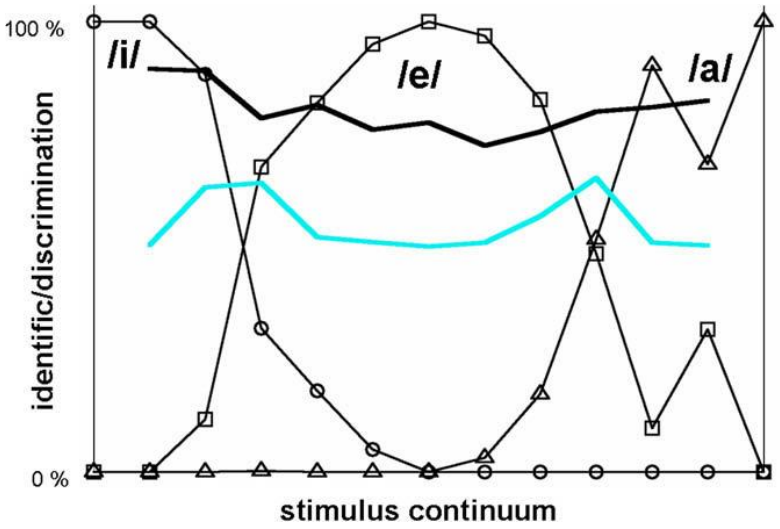
- Typical results:
  - Identification: phoneme regions in acoustic space
  - Discrimination: stronger categorical perception for CV than for V (see phoneme boundaries from measured discrimination!)

V = /i e a/

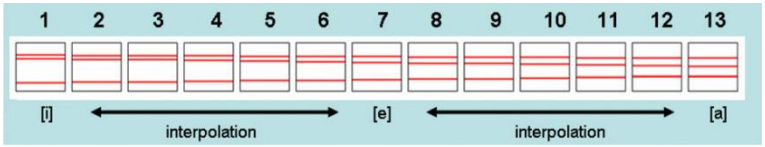
identi-fication

discrimi-nation

calcu-lated discrimi-nation

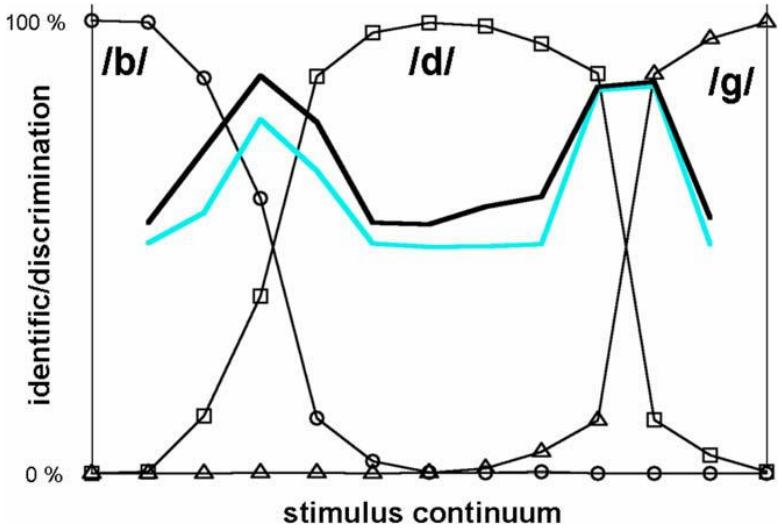


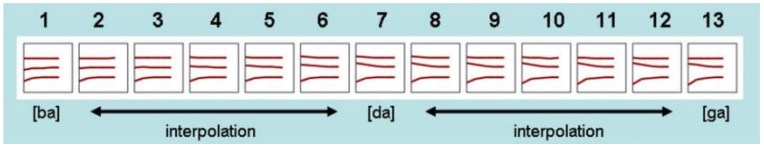
13 V- and CV-Stimuli



[i] interpolation [e] interpolation [a]

CV = /ba da ga/





[ba] interpolation [da] interpolation [ga]

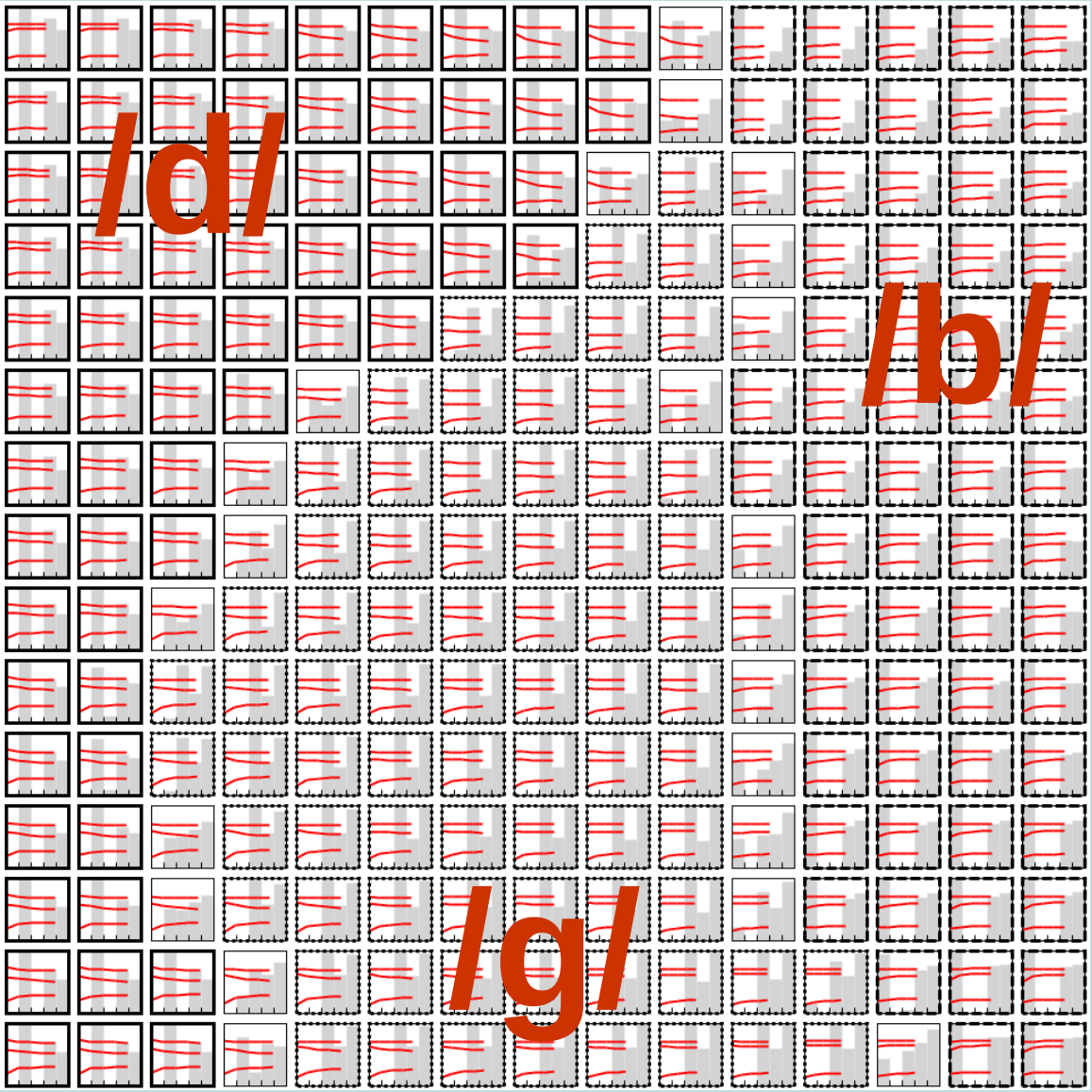
Question:

Can we explain this typical effect of stronger categorical perception for consonants than for vowels using the (dorsal) perception pathway as described above in our neurocomputational model?

To answer this question

- we have to clarify in detail, how to measure identification and discrimination in the model?
- we have to train 20 different instances of the model as "virtual listeners"

Reminder: Phonetic Map and Model Neurons



Display of:

15x15 model neurons (nodes)

Each: representing a phonetic CV state

Neural links towards auditory, motor plan and phonemic map

Auditory link weights: formant transitions

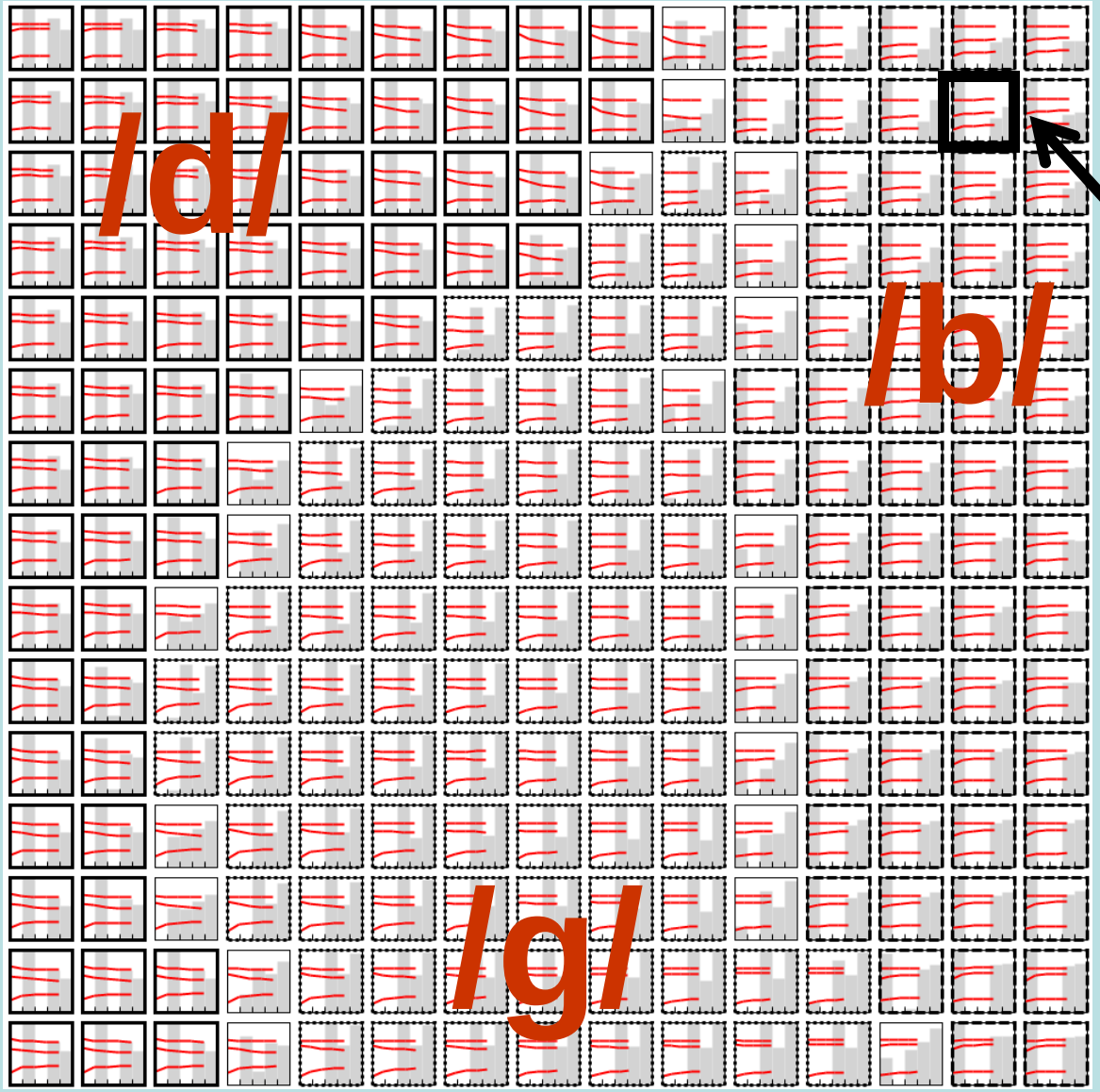
Motor plan link weights: 5 bars (grey)

- first three: closure: lab/api/dors

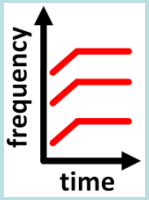
- last two: vocalic: back-front, low-high

Phonemic link weights > 0.8: edges of the neuron boxes: solid line, dashed line, dotted lines)

Measuring Identification and Discrimination



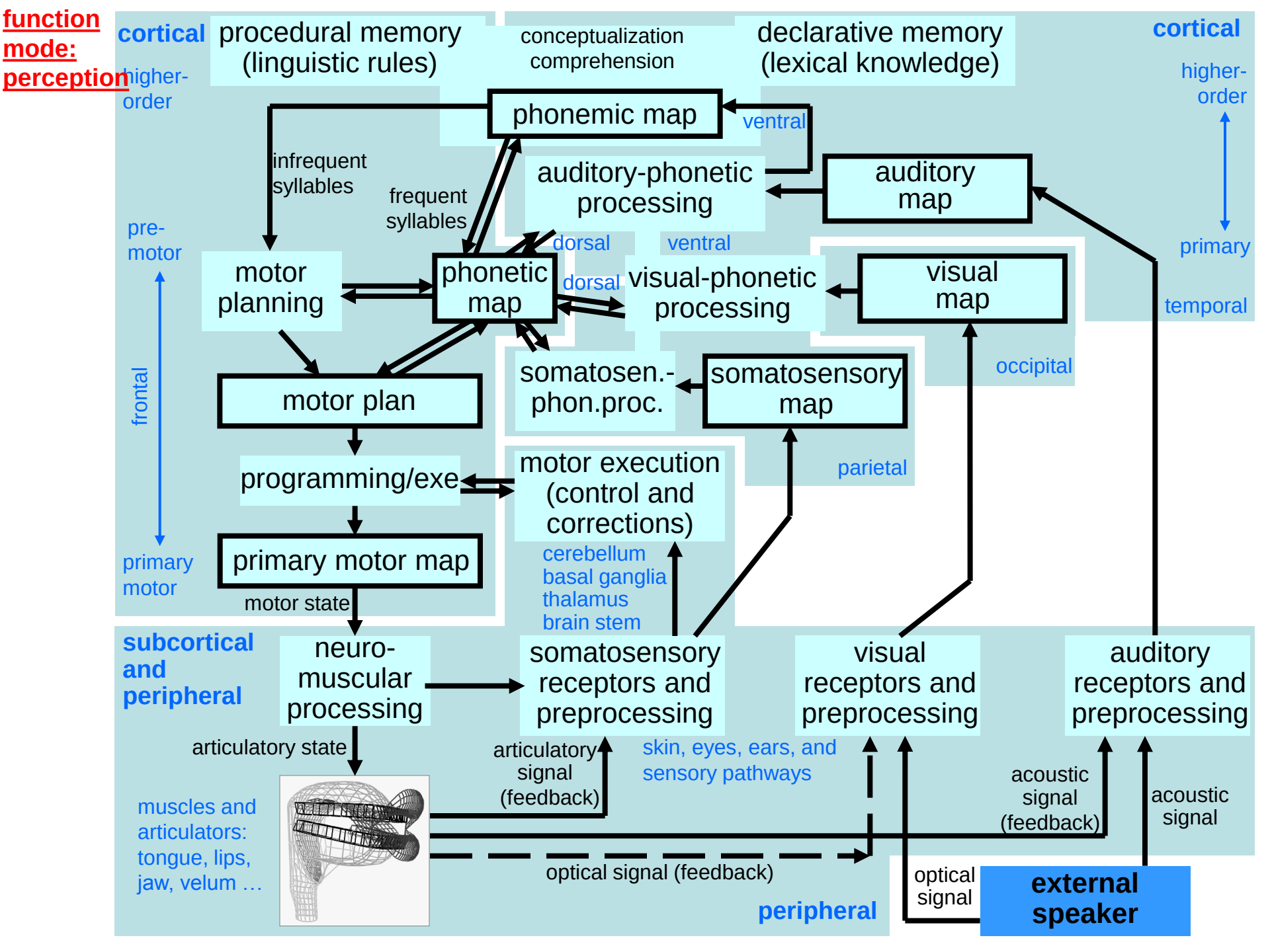
Identification: neuron with highest degree of activation for a stimulus



Incoming auditory stimulus

Best auditory match?
→ winner neuron

Is the winner neuron linked to a phonemic state?
Yes: /b/


function mode: perception
cortical
procedural memory (linguistic rules)
conceptualization comprehension
declarative memory (lexical knowledge)
cortical
higher-order
higher-order
phonemic map
ventral
infrequent syllables
frequent syllables
auditory-phonetic processing
auditory map
pre-motor
dorsal
ventral
primary
motor planning
phonetic map
dorsal
visual-phonetic processing
visual map
temporal
occipital
frontal
somatosen.-phon.proc.
somatosensory map
motor plan
parietal
programming/exe
motor execution (control and corrections)
primary motor map
cerebellum
basal ganglia
thalamus
brain stem
primary motor
motor state
subcortical and peripheral
neuro-muscular processing
somatosensory receptors and preprocessing
visual receptors and preprocessing
auditory receptors and preprocessing
articulatory state
articulatory signal (feedback)
skin, eyes, ears, and sensory pathways
acoustic signal (feedback)
acoustic signal
muscles and articulators: tongue, lips, jaw, velum …
optical signal (feedback)
optical signal
external speaker
peripheral

function mode: perception

auditory state of a syllable produced by external speaker

learned motor plan state, auditory state, and phonemic state

nearest neighbor!

cortical

higher-order

procedural memory (linguistic rules)

conceptualization comprehension

declarative memory (lexical knowledge)

cortical

higher-order

primary

temporal

occipital

parietal

frontal

pre-motor

primary motor

phonemic map

ventral

infrequent syllables

frequent syllables

self

auditory-ph processing

ext

auditory map

dorsal

ventral

dorsal

motor planning

phonetic map

visual-phonetic processing

visual map

somatosen.-phon.proc.

somatosensory map

motor plan

programming/exe

motor execution (control and corrections)

cerebellum

basal ganglia

thalamus

brain stem

primary motor map

motor state

subcortical and peripheral

neuro-muscular processing

somatosensory receptors and preprocessing

visual receptors and preprocessing

auditory receptors and preprocessing

articulatory state

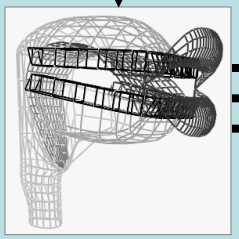
articulatory signal (feedback)

skin, eyes, ears, and sensory pathways

acoustic signal (feedback)

acoustic signal

muscles and articulators: tongue, lips, jaw, velum …



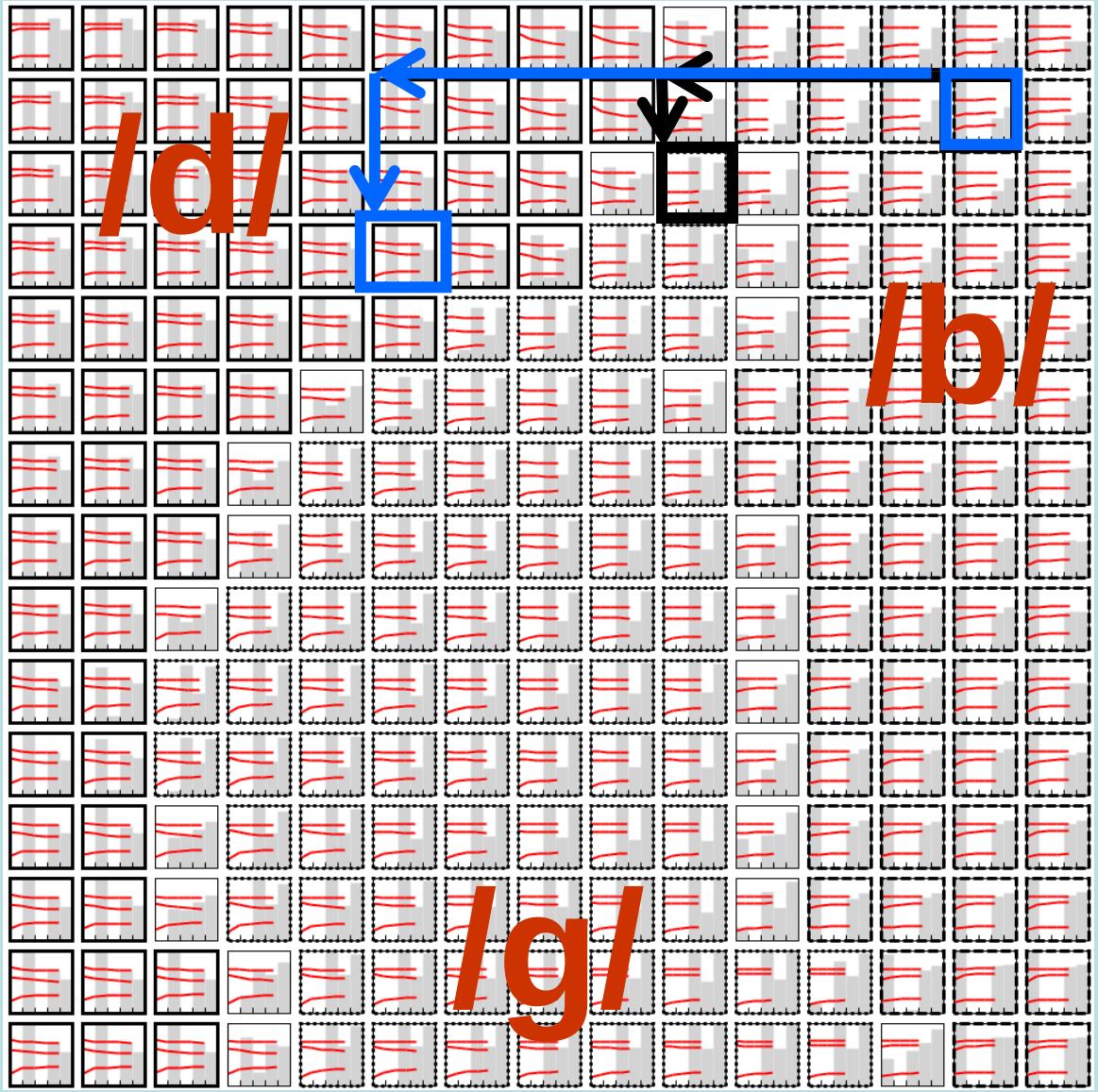
optical signal (feedback)

optical signal

external speaker

peripheral

Measuring Identification and Discrimination



Identification: neuron with highest degree of activation for a stimulus

Discrimination: city-block-distance of neurons activated for stimuli A and B

Assumption: Discrimination is better the higher the distance of both stimuli at the level of the phonetic map

Training of Different Instances of the Model:

Starting with a different link weight initialization of SOM

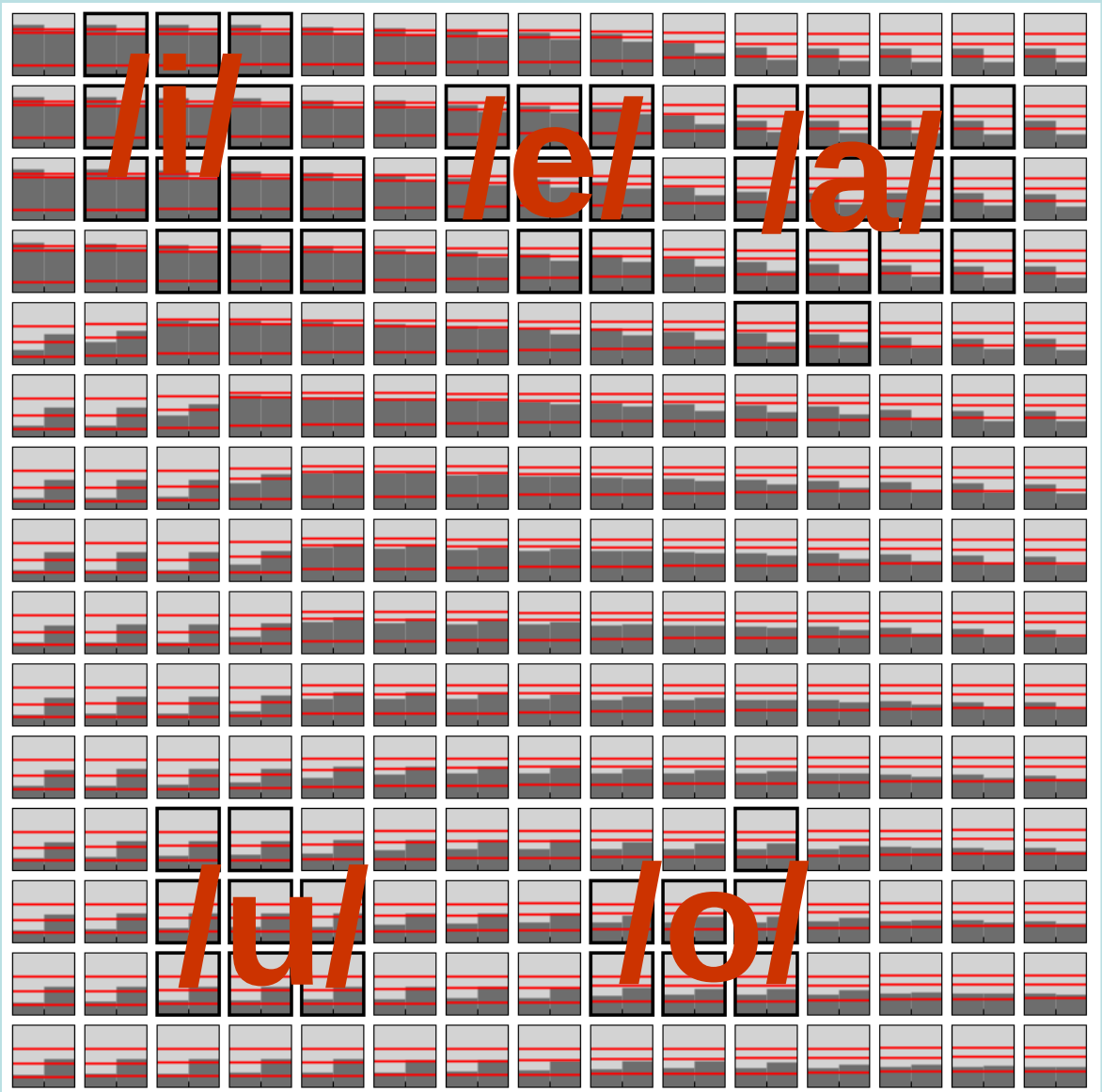
Applying different training items (different items from a pool of V- or CV-representations)

Applying a different ordering / randomization of training items

Leads to: different SOMs / “different brains” / different listeners

Some examples:

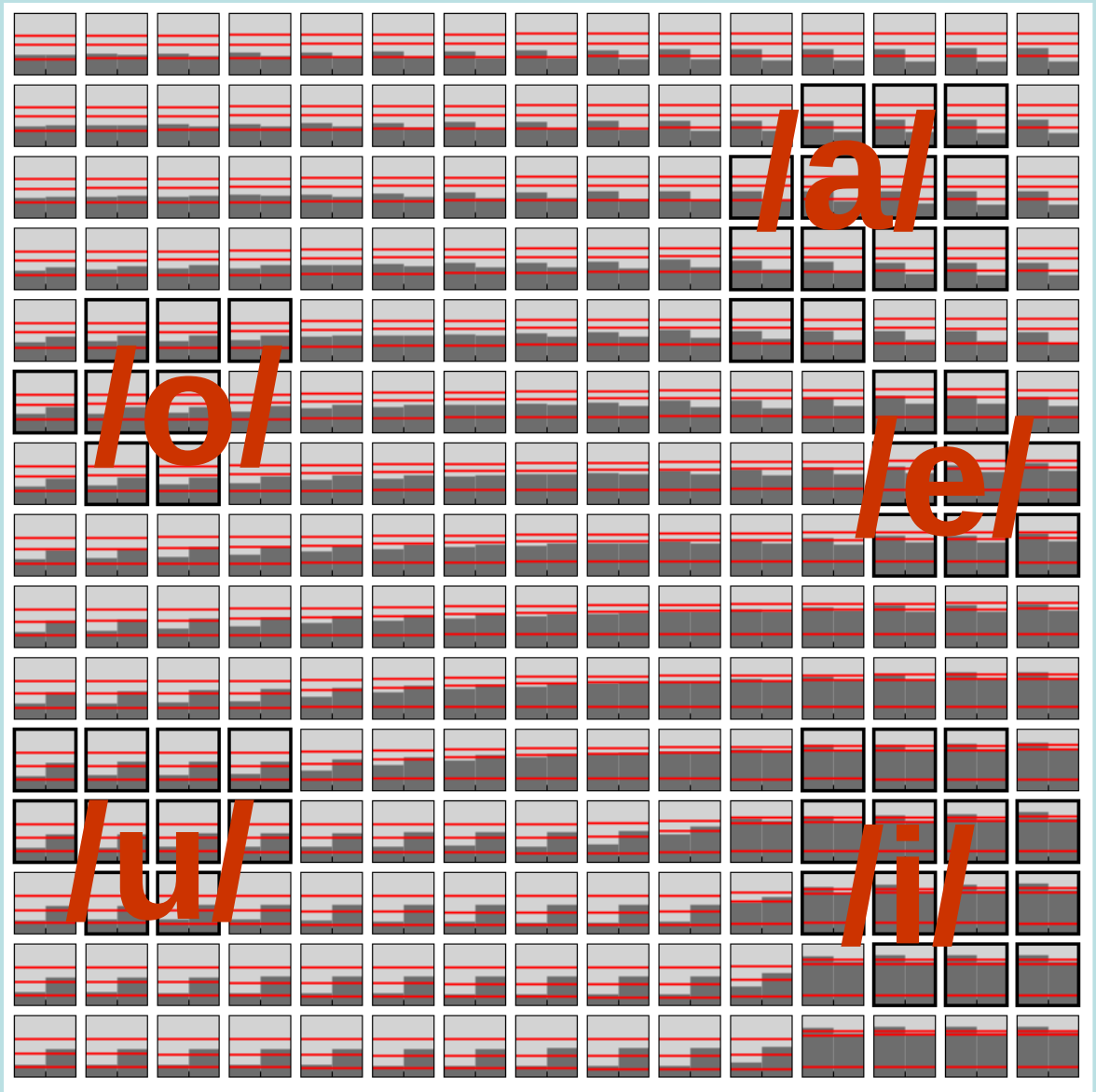
V-SOM (Brain 1 from 20)



Association of auditory and motor states: ok

Phonetic ordering (high-low; front-back): ok

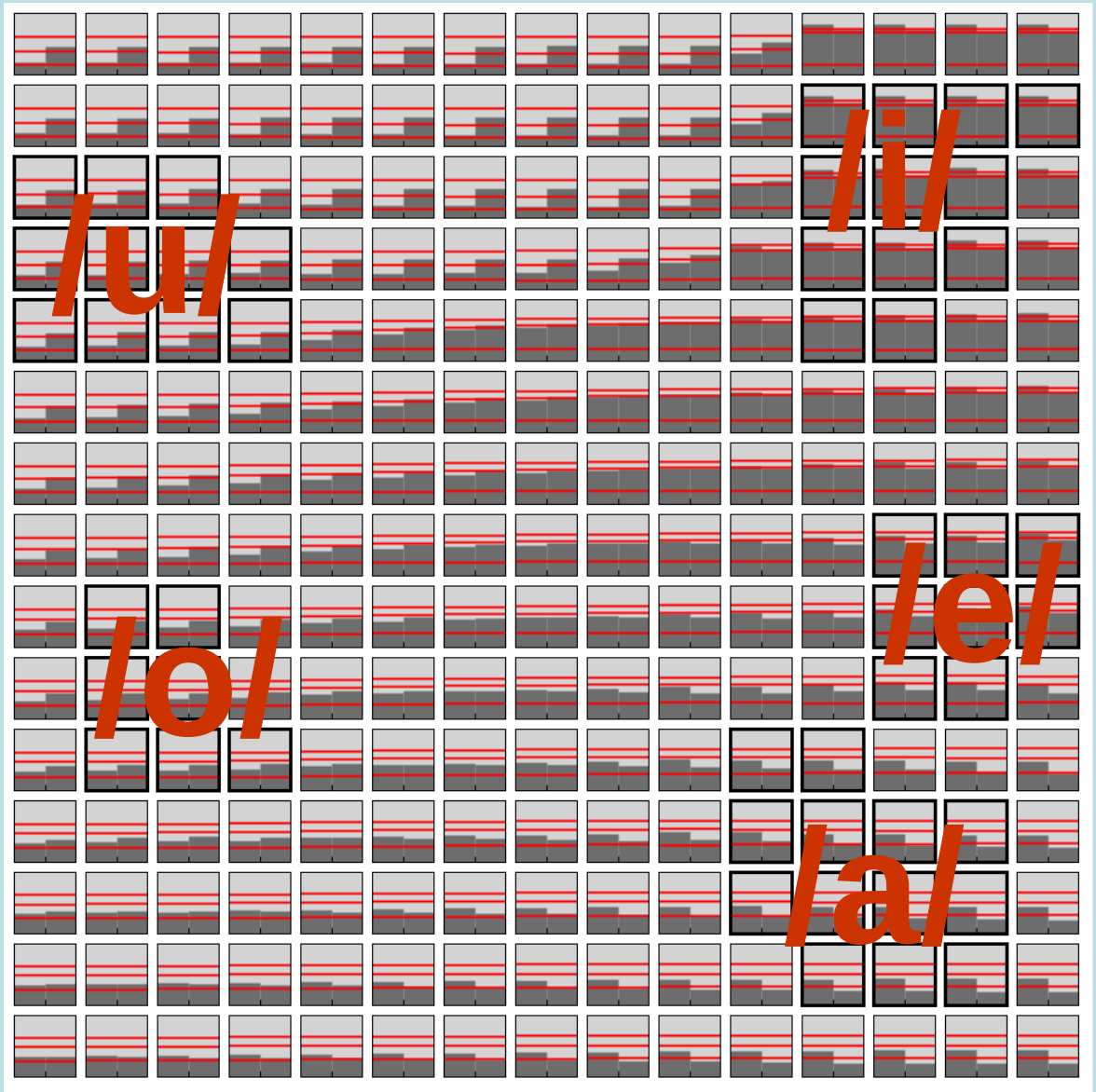
V-SOM (Brain 2 from 20)



Association of auditory and motor states: ok

Phonetic ordering (high-low; front-back): ok

V-SOM (Brain 4 from 20)



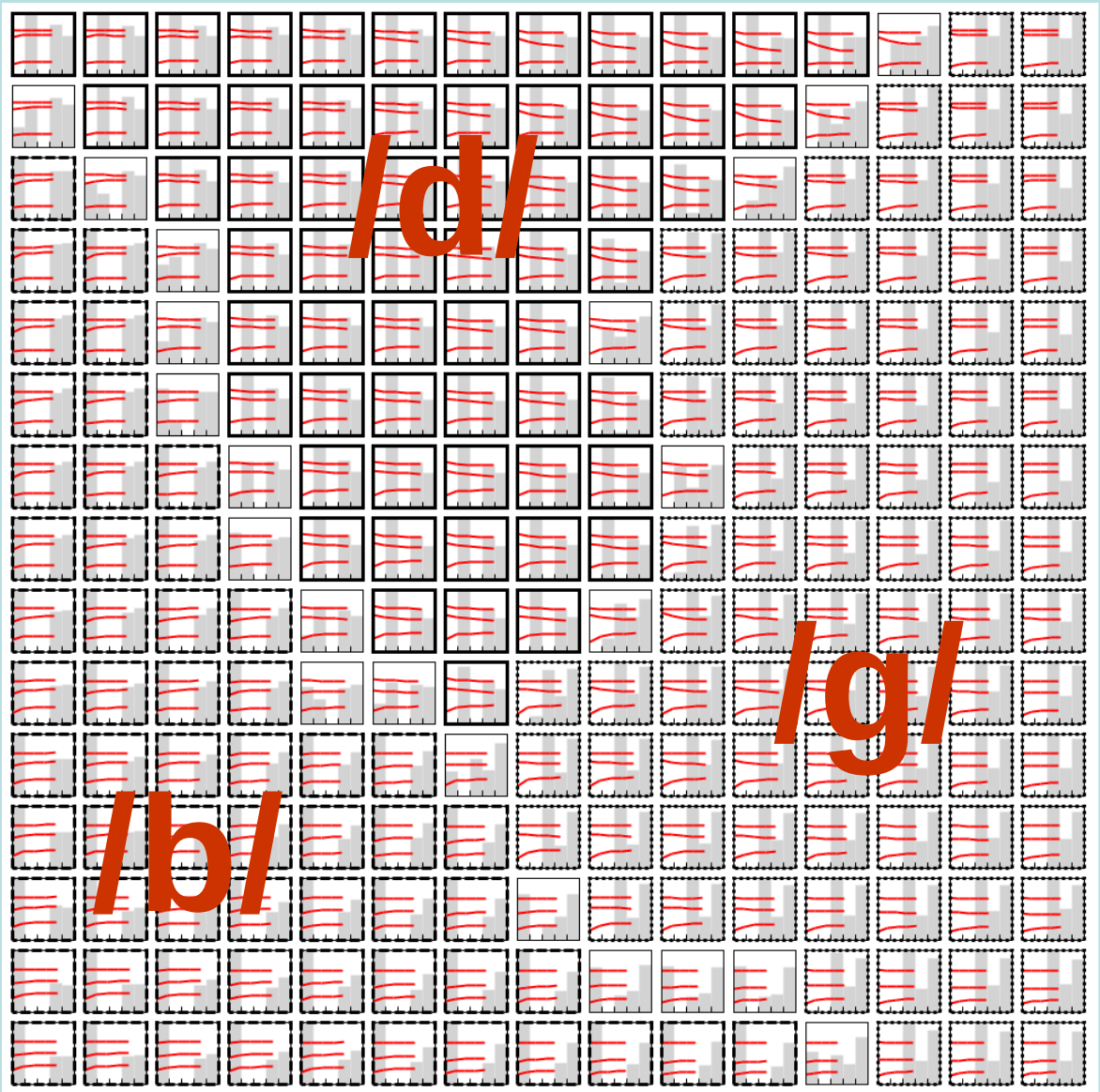
Association of auditory and motor states: ok

Phonetic ordering (high-low; front-back): ok

… the examples indicate: high-low front-back - dimensions occur in different directions for different instances of the model (different brains)

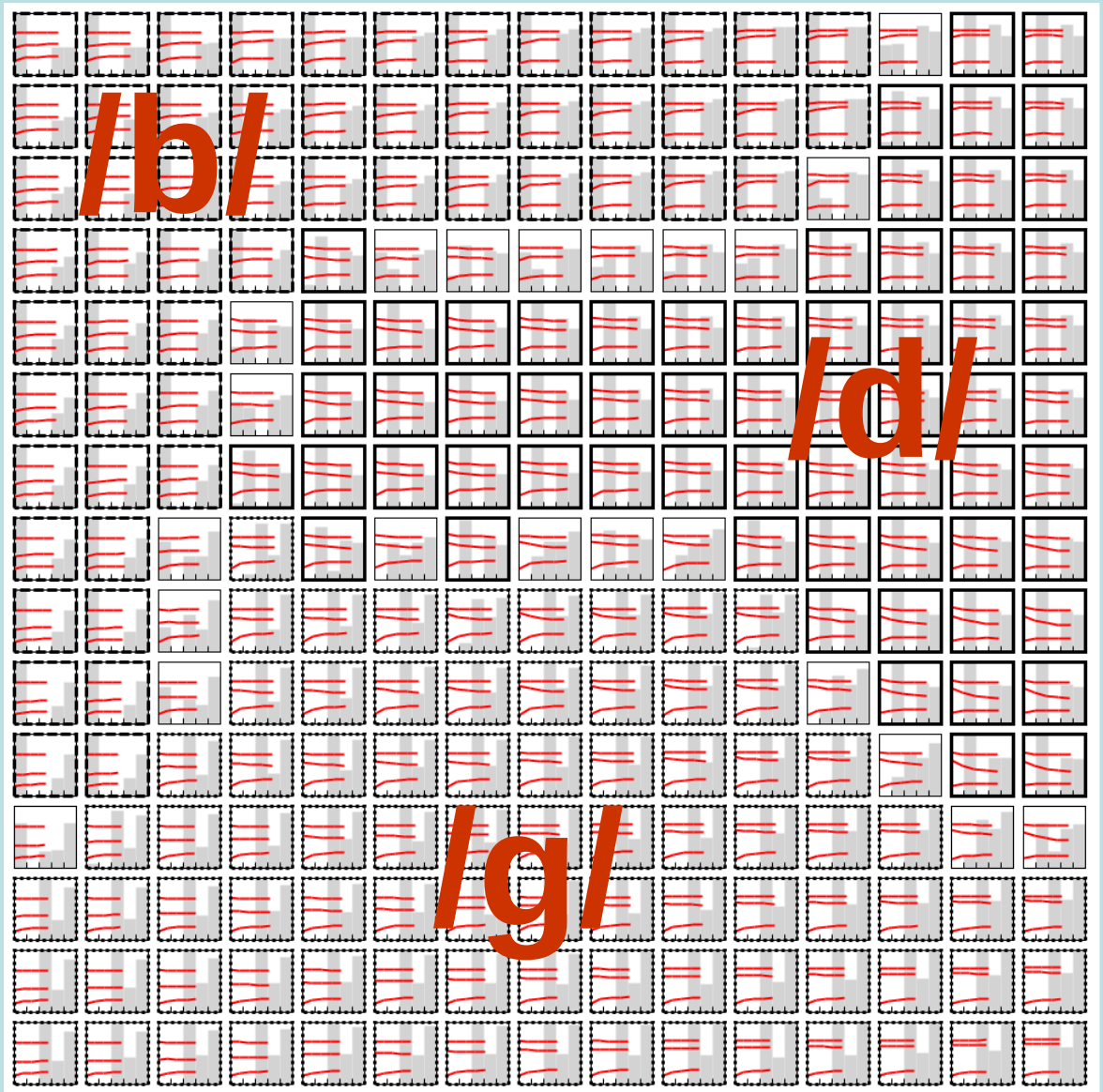
→ could be a problem for imaging experiments! Only use single subjects

CV-SOM (Brain 2 from 20)



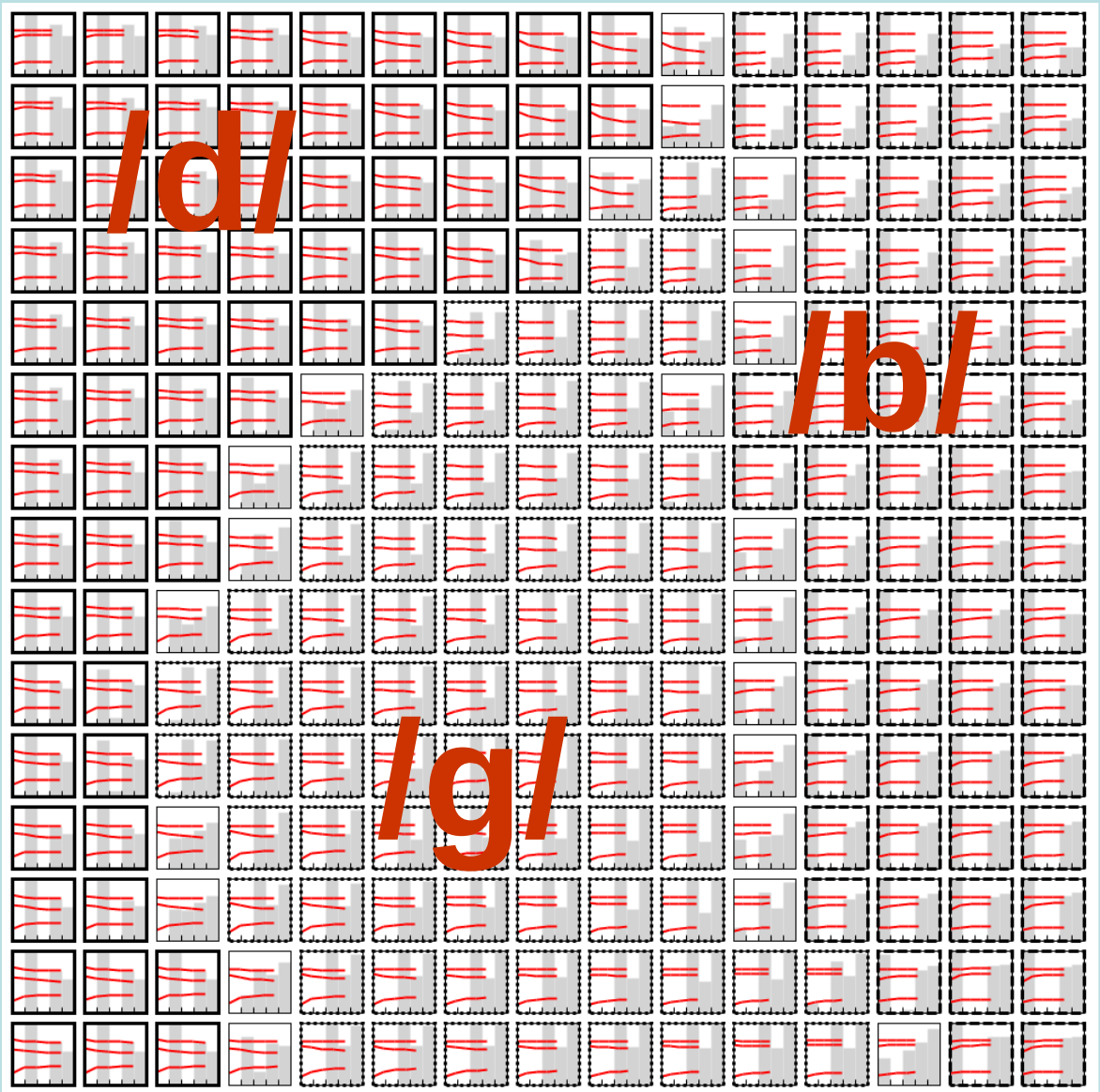
brain 2

VC-SOM (Brain 3 from 20)



brain 3

VC-SOM (Brain 4 from 20)

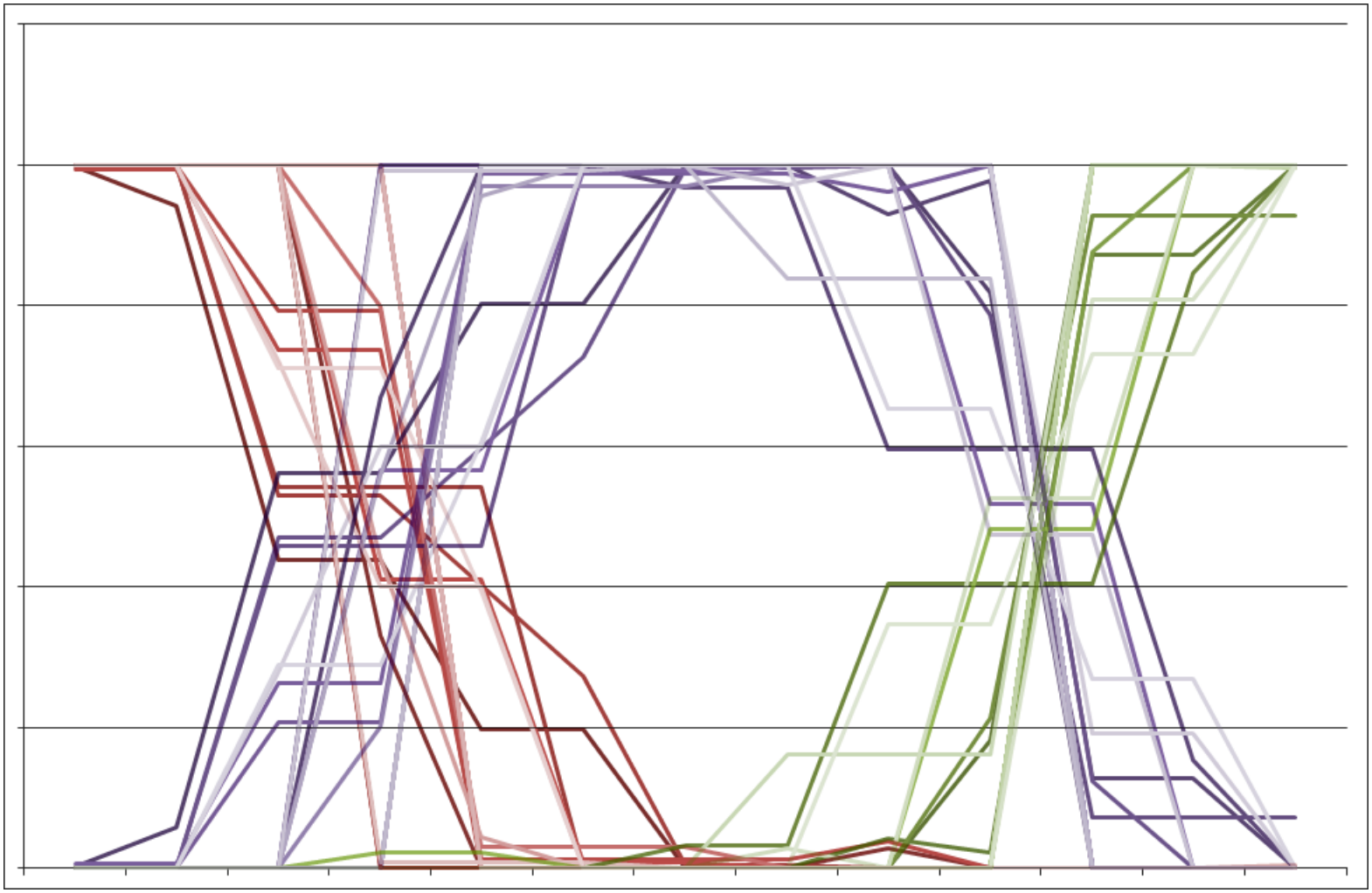


In all cases:
/b/-, /d/-, /g/-regions are compact or cohesive and not splitted in parts

… but in different locations within the CV-map

Identification Scores for CV: 20 brains

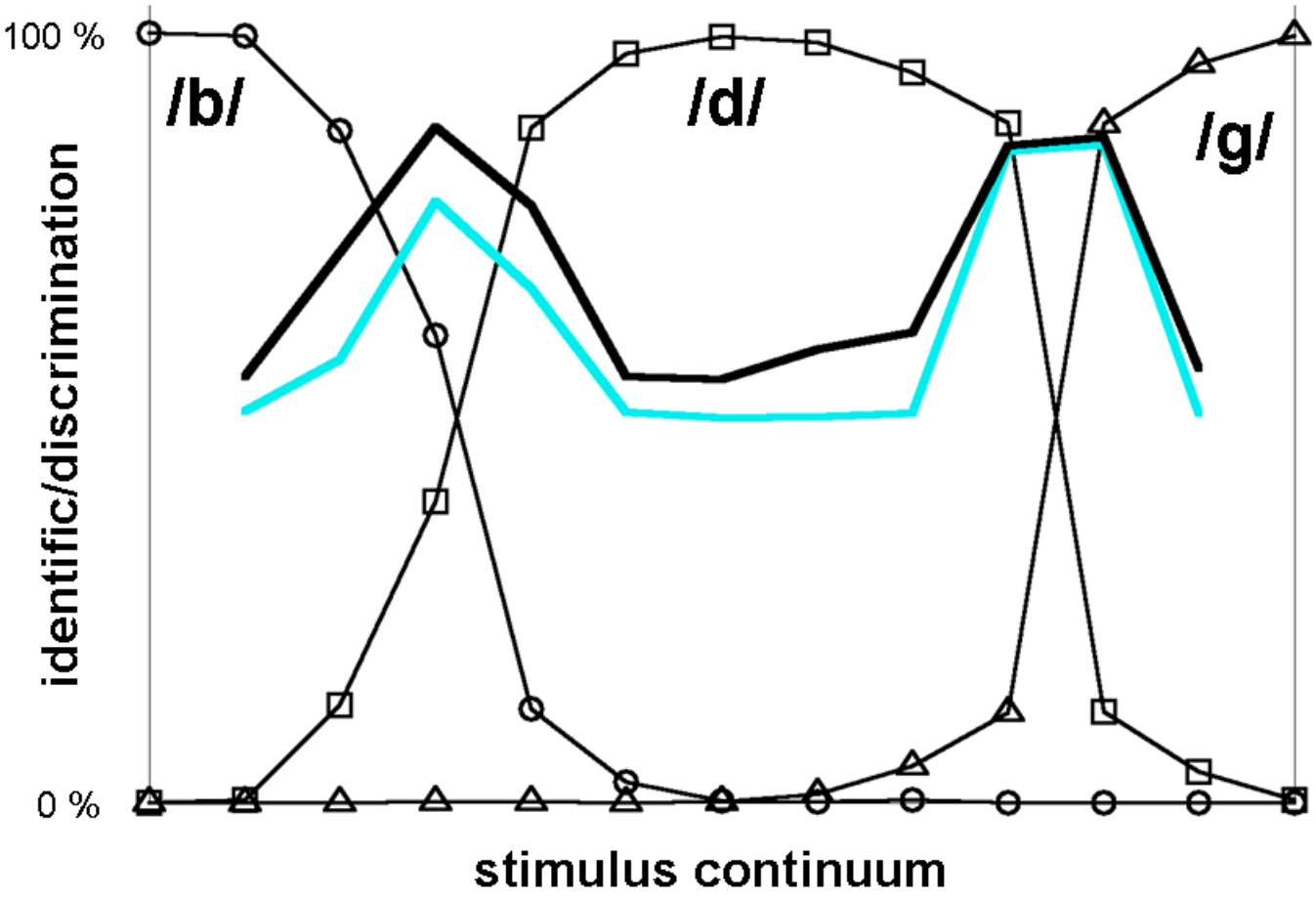
... slightly different identification scores per instance as occur naturally



Brain 01 to 20: form light to dark color

Identification and Discrimination Scores: Consonant Perception

Typical categorical perception: lower discrimination scores within phoneme regions in comparison to phoneme boundaries



identification

discrimination (measured)

discrimination (calculated)

20 listeners

Calculated Discrimination

Discrimination calculated on the basis of measured identification (Liberman et al. 1957)

$$p_{\text{discr}} = 0.5 + 0.5 \cdot \sum_{i=1}^{3} (p_{id}(a,i) - p_{id}(b,i))^2$$

p_{discr} = calculated discrimination of two stimuli a and b, which are identified as i = 1, 2, 3 (/b/, /d/, /g/) with probabalilty p_{id}.

That is: discrimination which is based exclusively on linguistic information, i.e. on differences in identification: $p_{id}(a)$-$p_{id}(b)$

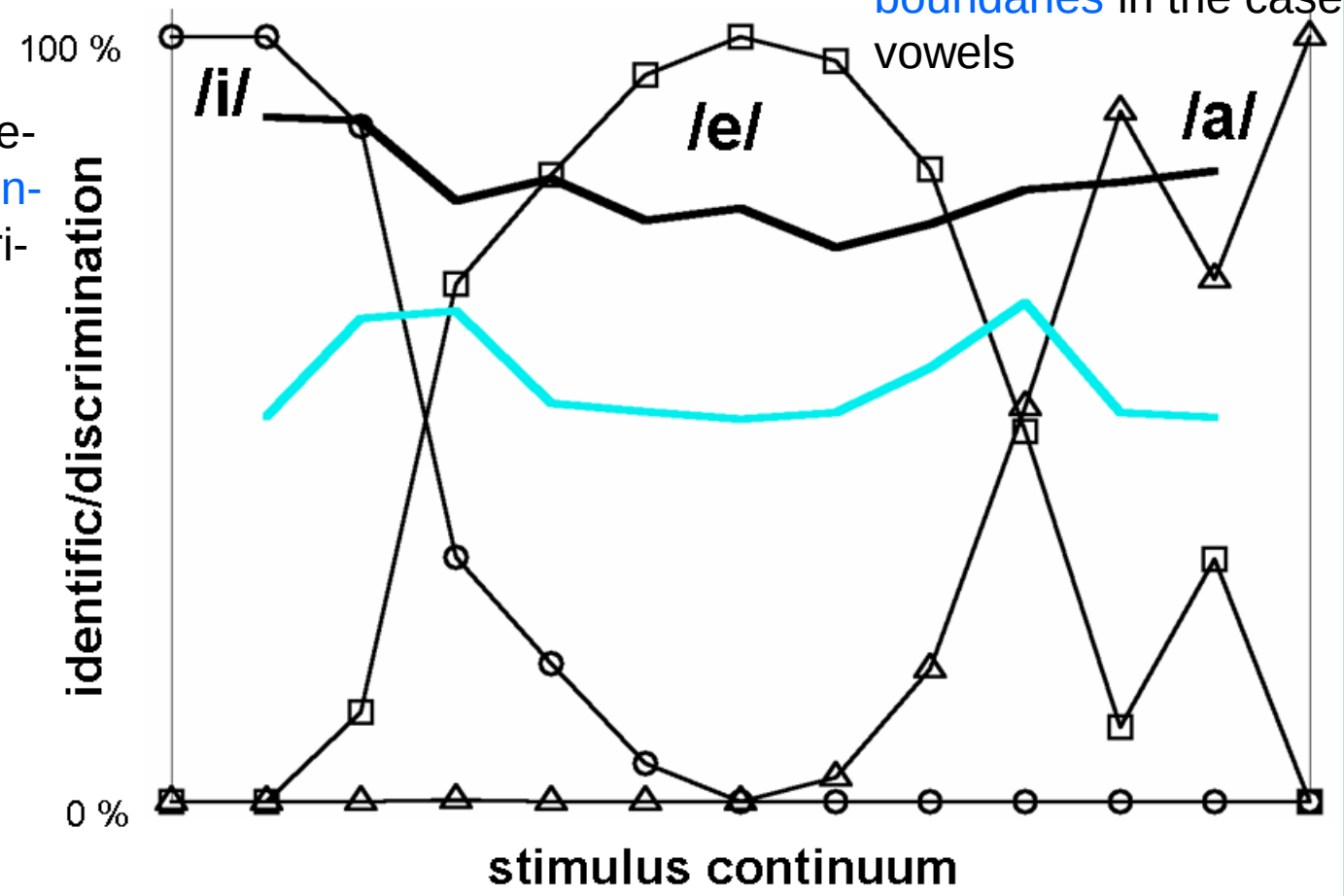
Calculated discrimination is comparable to measured discrimination for CV, but:

Identification and Discrimination Scores: Vowel Perception

Much higher percentage of measured than calculated discrimination in comparison to consonants

Interpretation: this difference represents a non-linguistic (non-categorical) component in vocalic sound perception

And very important: Measured discrimination indicates: no phoneme boundaries in the case of vowels



identification

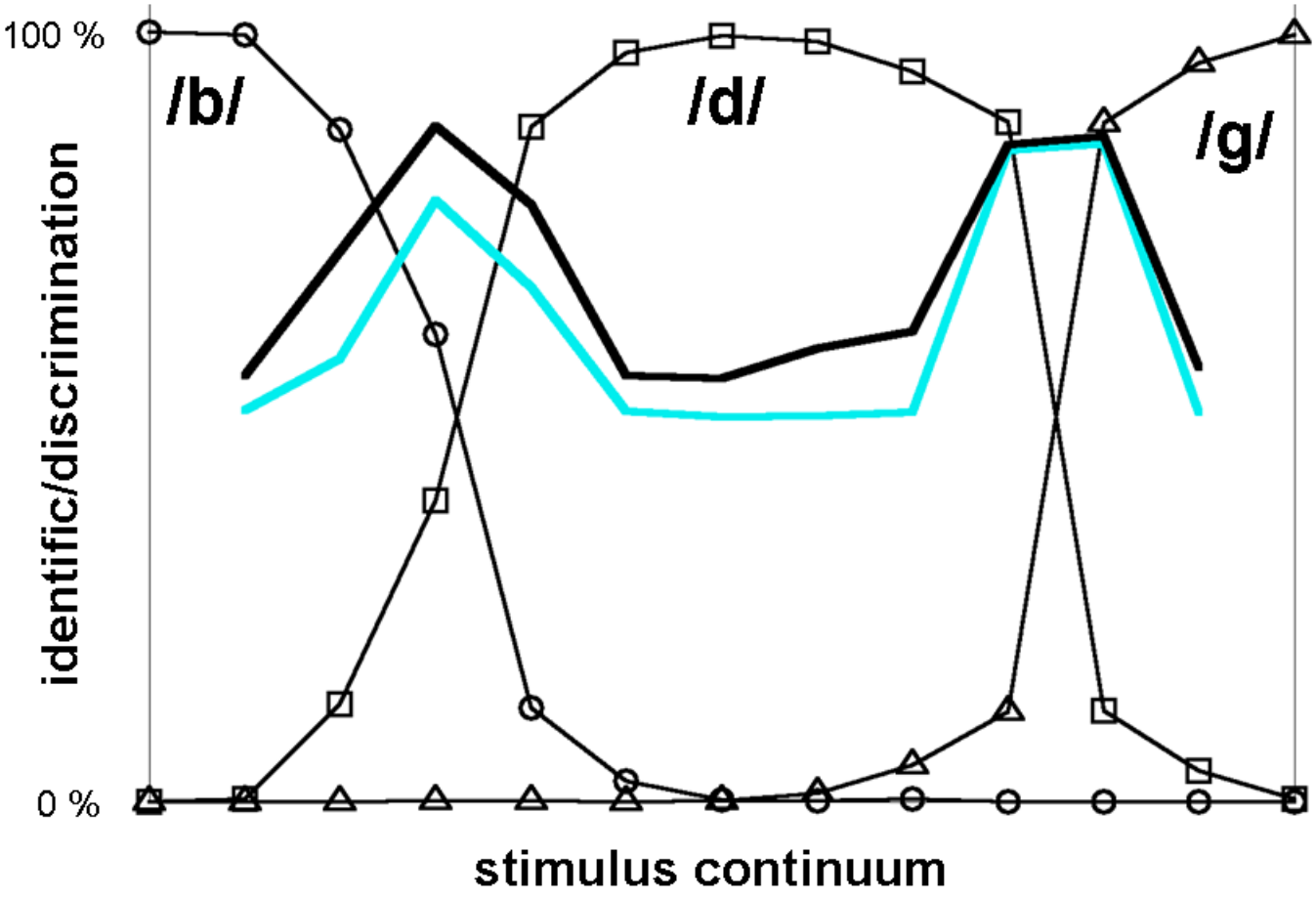
discrimination (measured)

discrimination (calculated)

20 listeners

Identification and Discrimination Scores: Consonant Perception

But: strong phoneme boundaries in the case of consonants



identification

discrimination (measured)

discrimination (calculated)

20 listeners

Can we find Neurophonetic Reasons for this Difference in Categorical Perception?

Yes! These differences in categorical perception of consonants and vowels result from topological differences concerning the ordering of phonetic states within V- and CV-SOMs:

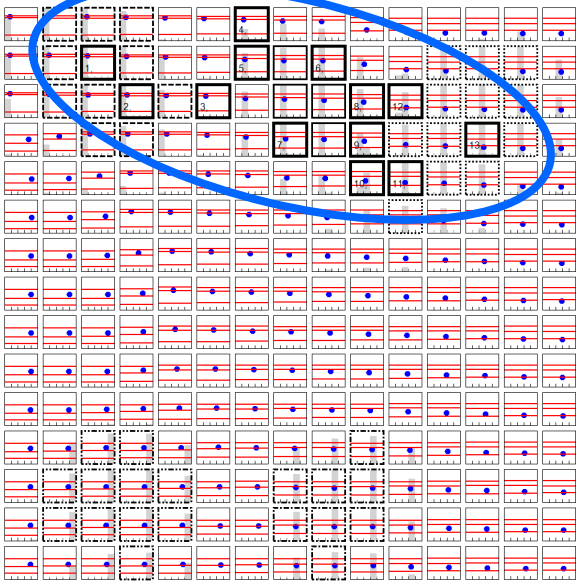
Neurophonetic (microscopic) Reasons

- Marked boxes: Neurons, representing the stimuli of the stimulus continua
- The V-stimuli are „continuously distributed“ within the V-SOM space
- The CV-stimuli are more “clusterd” within the CV-SOM space

V = /i e a o u/

one big cluster

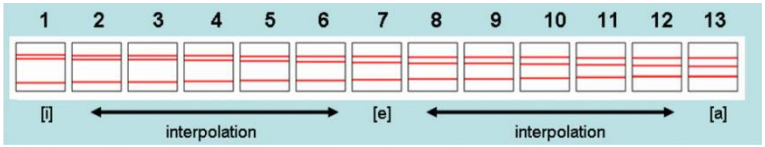
supramodal phonetic map including phoneme regions



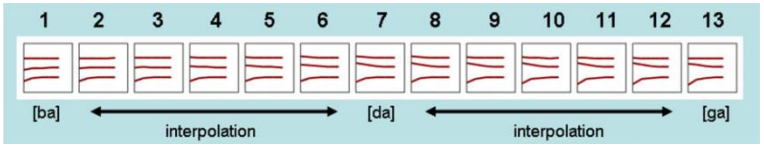
CV = /ba da ga/

three small clusters

13 V- and CV-stimuli



[i] [e] [a]

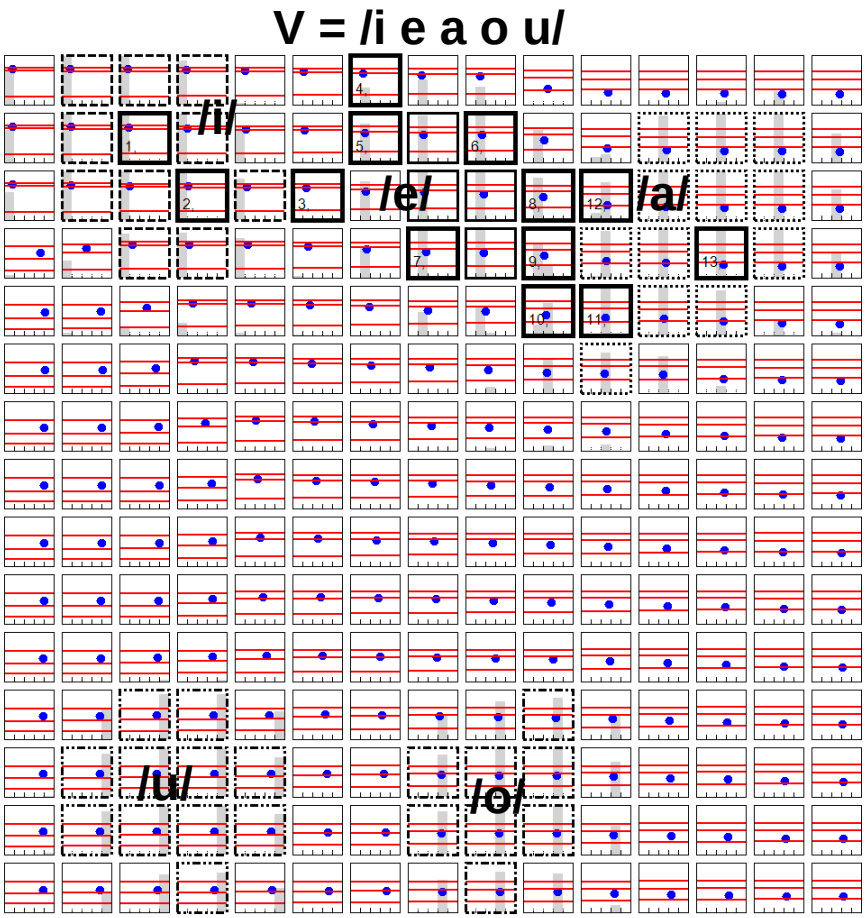


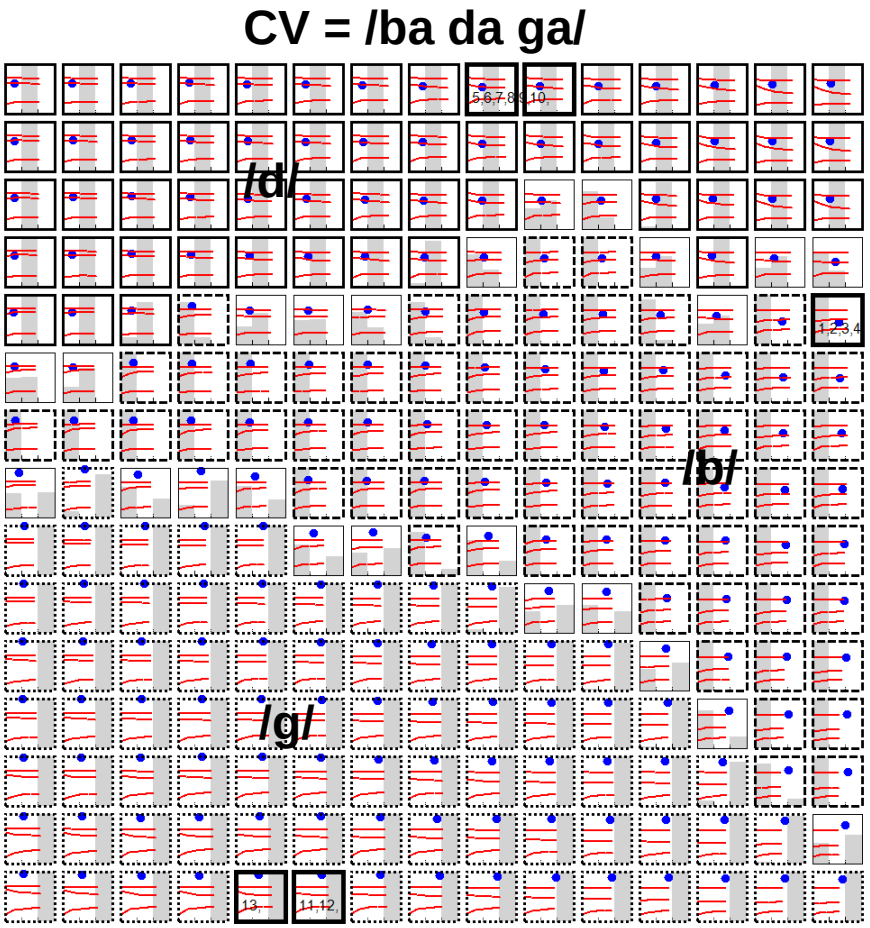
[ba] [da] [ga]

Neurophonetic Reasons

Thus:

- Case V-SOM: one (big) “stimulus cluster” (CL) covers three “phoneme regions” (PRs)
- Case CV-SOM: three (small) “stimulus clusters” (CLs) occur; and each “stimulus cluster” occurs within just one phoneme region





Neurophonetic Reasons

If we look at all 20 instances (20 brains):

- This situation holds for 19 instances in the case of V-SOMs (95%)
- But this holds only for 11 instances in the case of CV-SOMs (55%) (unfortunately: 8 instances show one cloud for two consonants)
- But: the number of "convenient" stimulus clusters (i.e. fulfilling our demand: 1CL covers 1 PR) over all 20 brains in the case of CV-SOMs is 41 from 60 (68.3%)

Type of instance	Number of Instances		Number of Clusters over all Instances	
	/CV/	/V/	/CV/	/V/
1 CL covers 3 PR	1	**19**	1	**19**
1 CL covers 2 PR	8	0	8	1
1 CL covers 1 PR	**11**	1	**41**	1
Total / Maximum	20	20	60	20

One More Question:
Does our neurophonetic approach account for perception of phonetic features in infancy?

- Newborns can differentiate phonetic features (distinctive categories) of many languages (Kuhl 2004);
- Within our model: that may result from the initially non-ordered phonetic map at the starting point for babbling and imitation
- If the phonetic maps becomes ordered during babbling and imitation, the toddler is no longer sensitive to distinctive features of other languages
- Phoneme regions = regions of low discrimination!

Outline: Part 2

- Introduction
- Simulation of Speech Perception
  - Categorical Perception
  - McGurk Effect
- Conclusions

The McGurk-Effect

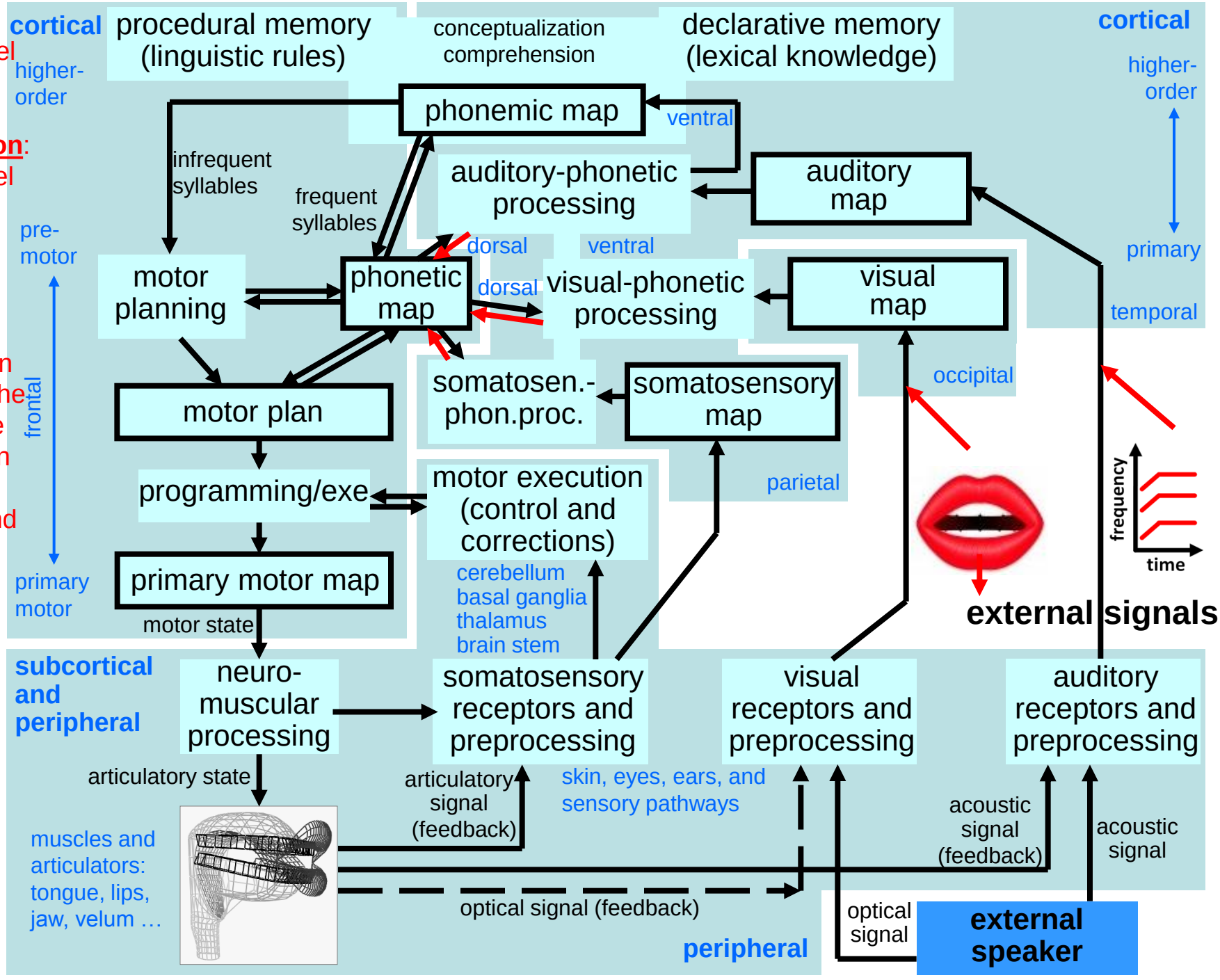
- The McGurk-Stimulus is not “natural”; i.e. can not be produced by a human;
- it is a “dubbing” (wrong synchronization) of a visual stimulus with an auditory stimulus:

  $[ba]_{acoustic} + [ga]_{optical} = /da/_{percept}$

  we “hear” (perceive) /da/ (not produced physically)
- Conclusion drawn by many researchers: “perception” is not isolated auditory or visual etc.; perception (better: “identification” of sounds, to “comprehend” s.th.) is hypermodal or multimodal
- At first glance: that fits well with our model: identification of sounds is done by a processing at the level of the hypermodal phonetic map (i.e.: is not an isolated auditory or visual process);

This model includes **sensory integration**: at the level of the phonetic map

dorsal pathway in terms of the dual route perception model of Hickok and Poeppel (2007) Nature Reviews Neuro-science



The McGurk-Effect

- So: Does the McGurk-Effect occur in our model?
- If yes: Can we find a microscopic (neurophonetic) explanation for the McGurk-Effect based on the structure and functioning of our model?

The McGurk-Effect

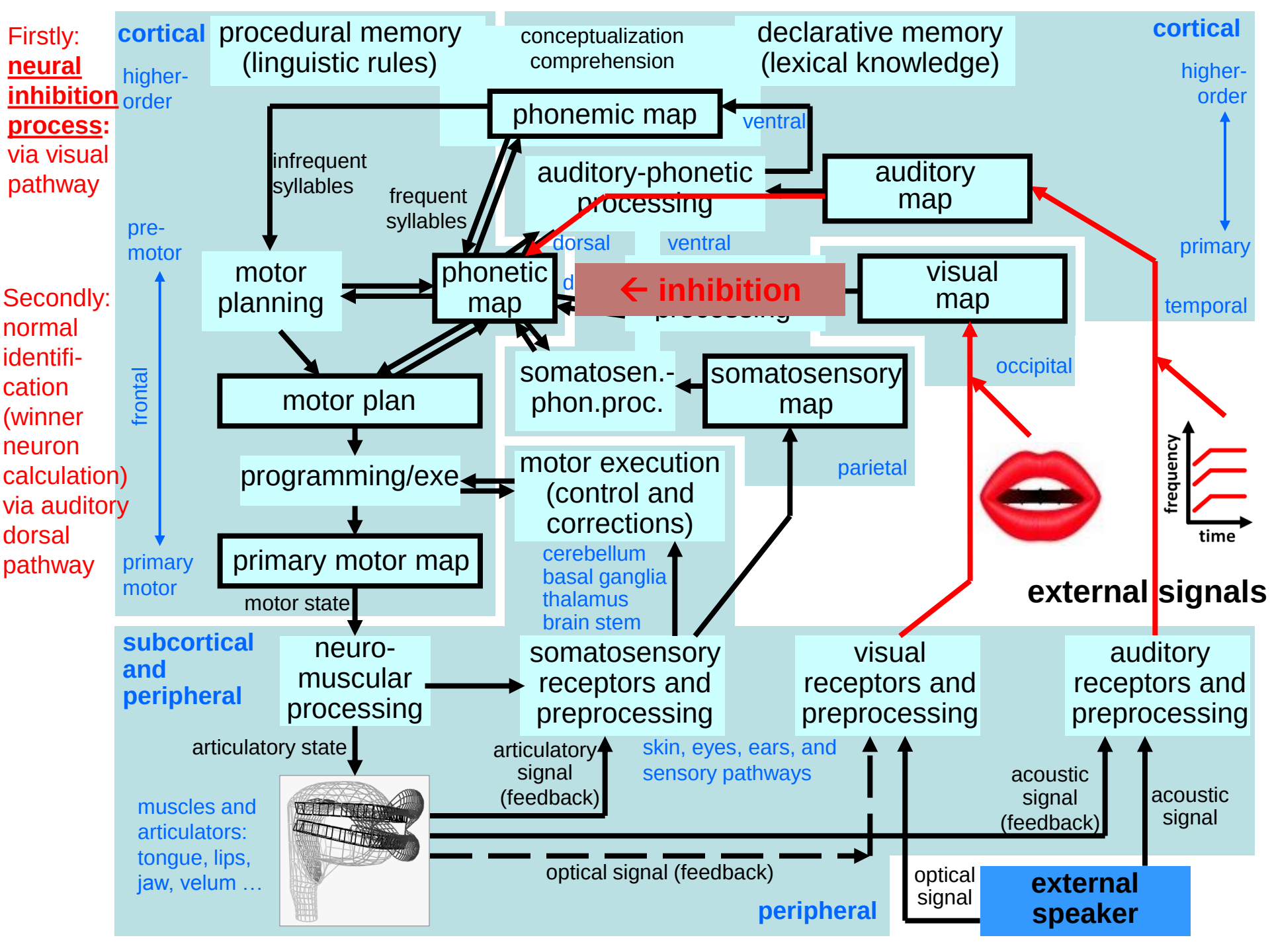
Some further requirements in order to observe the “McGurk-Effect” are:

- We need an excellent timing of the acoustic and optic stimulus with respect to the release of the closure
- Round about 20 listeners are needed for the experiment; we will see: not all listeners come to the same result, but most “hear” /da/! → we have to train 20 different instances of the model!

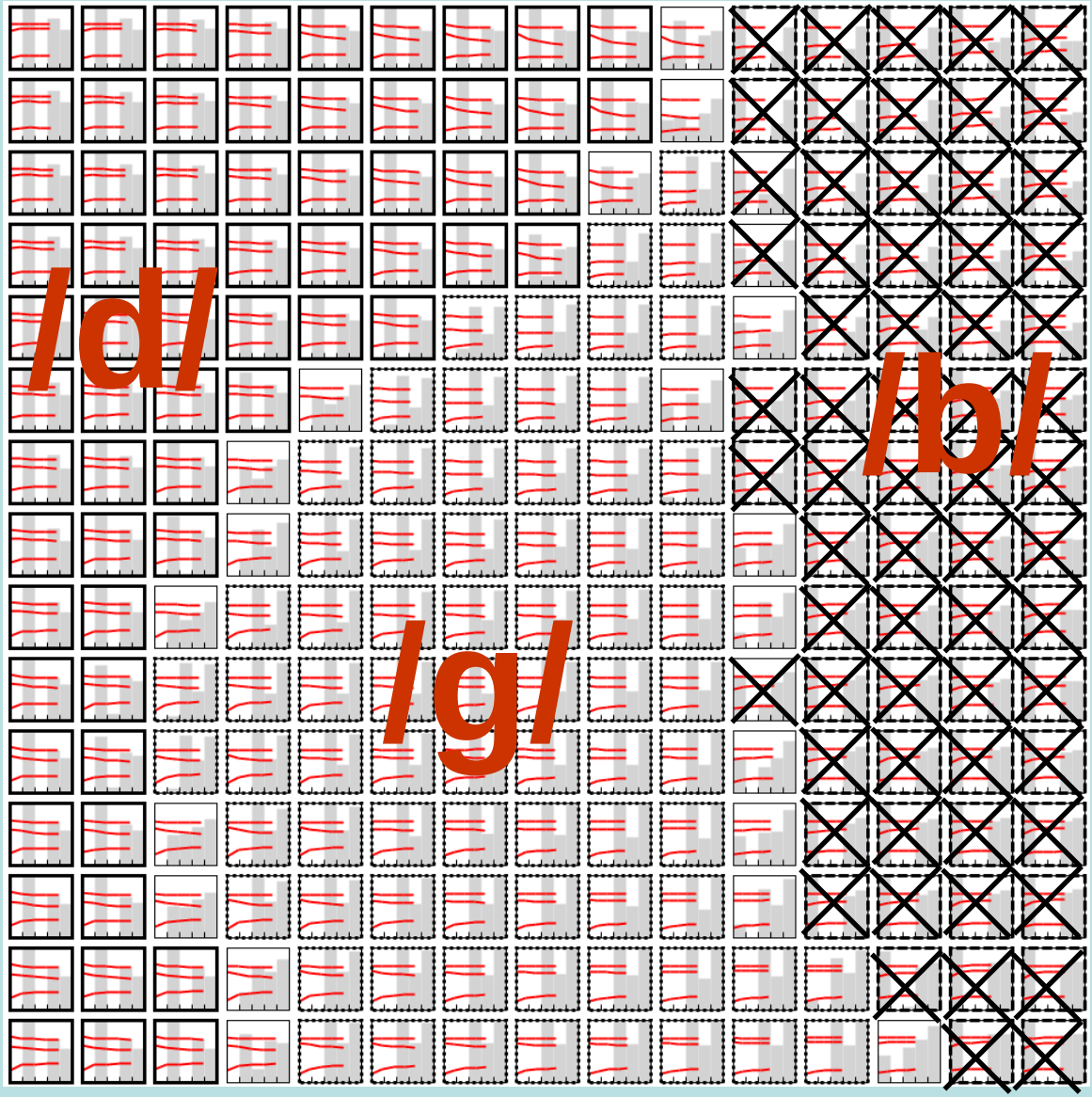
Identification procedure, assumed:

- Start with: Babbling and imitation training for 20 instances of the model using "natural" CV-stimuli for a language comprising /ba/, /da/, /ga/

  → CV-phonetic map is already established for 20 instances of the model on the basis of natural stimuli
- Now: Apply the McGurk-Stimulus;
  Assumption for its identification processing: two-step process:
  - First step: Postulate an inhibition process, initiated by the visual percep-tion pathway: If no visual labial closure occurs in the visual stream, all neurons, representing a visual labial closure at the level of the phonetic map are inhibited (can not be activated)
  - Second step: Now, a normal identification process occurs via the audi-tory perception pathway by identifying the winner neuron at the level of the phonetic map (on the basis of the remaining non-inhibited neurons)


Firstly:
neural inhibition process:
via visual pathway
Secondly:
normal identifi-cation (winner neuron calculation)
via auditory dorsal pathway
cortical
procedural memory (linguistic rules)
conceptualization comprehension
declarative memory (lexical knowledge)
cortical
higher-order
higher-order
phonemic map
ventral
infrequent syllables
frequent syllables
auditory-phonetic processing
auditory map
pre-motor
dorsal
ventral
primary
motor planning
phonetic map
← inhibition
visual map
temporal
frontal
occipital
somatosen.-phon.proc.
somatosensory map
motor plan
programming/exe
motor execution (control and corrections)
parietal
primary motor map
primary motor
cerebellum
basal ganglia
thalamus
brain stem
motor state
external signals
frequency
time
subcortical and peripheral
neuro-muscular processing
somatosensory receptors and preprocessing
visual receptors and preprocessing
auditory receptors and preprocessing
articulatory state
articulatory signal (feedback)
skin, eyes, ears, and sensory pathways
acoustic signal (feedback)
acoustic signal
muscles and articulators: tongue, lips, jaw, velum …
optical signal (feedback)
optical signal
external speaker
peripheral

Identification of McGurk-Stimulus: [ba]$_{aud}$+[ga]$_{vis}$



Display of phonetic map:

Auditory link weights: formant transitions

Motor plan link weights: 5 bars (grey)

- first three: closure: lab/api/dors (reflects: visual link weights because of perfect association: motor - visual)

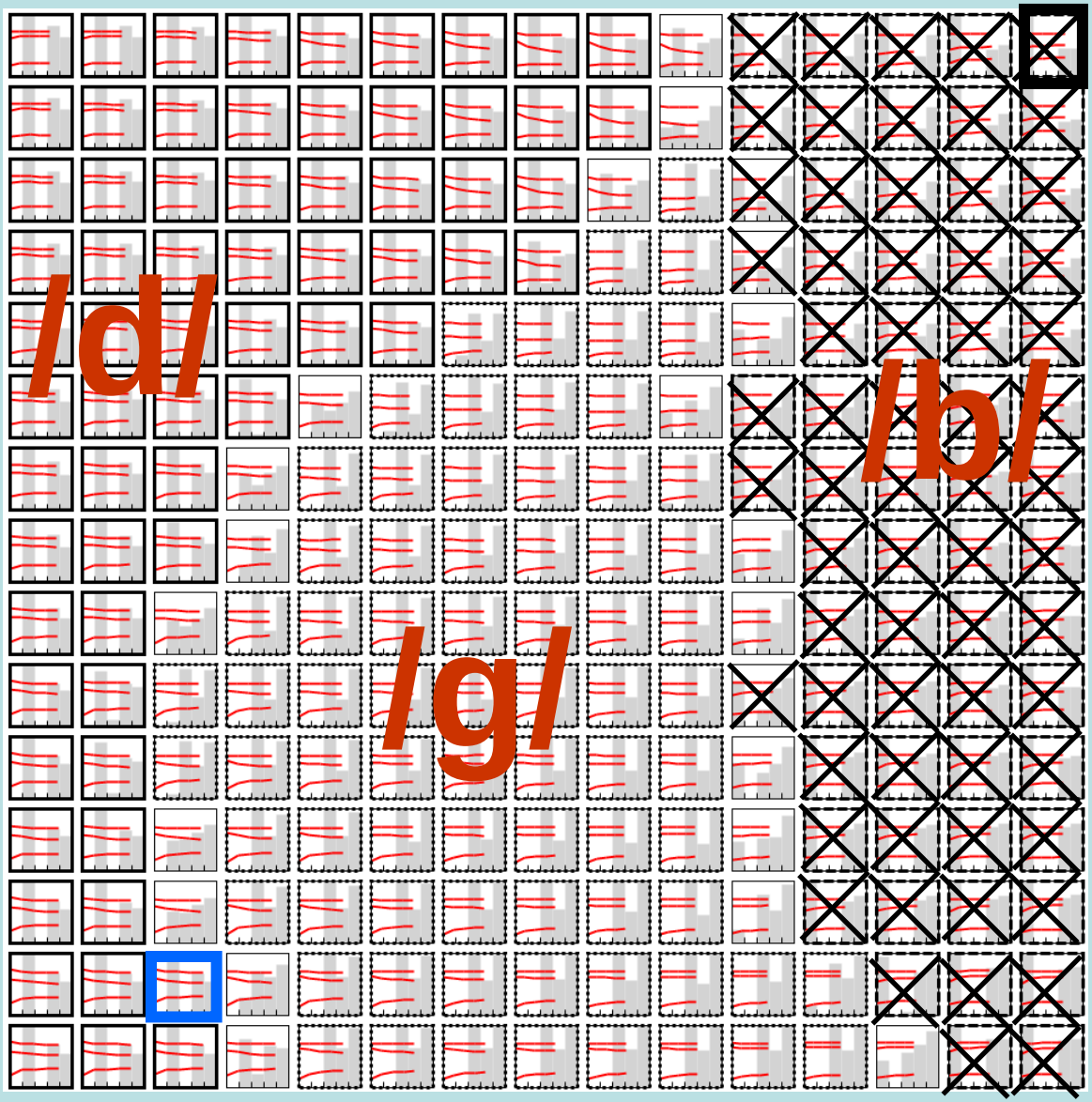
- last two: vocalic: back-front, low-high

Phonemic link weights > 0.8: edges of the neuron boxes: solid line, dashed line, dotted lines)

From visual pathway:

Inhibited neurons, if the visual stimulus does show a labial closure -> visual inhibition region

Identification of McGurk-Stimulus: [ba]$_{aud}$+[ga]$_{vis}$



Now: the auditory pathway:

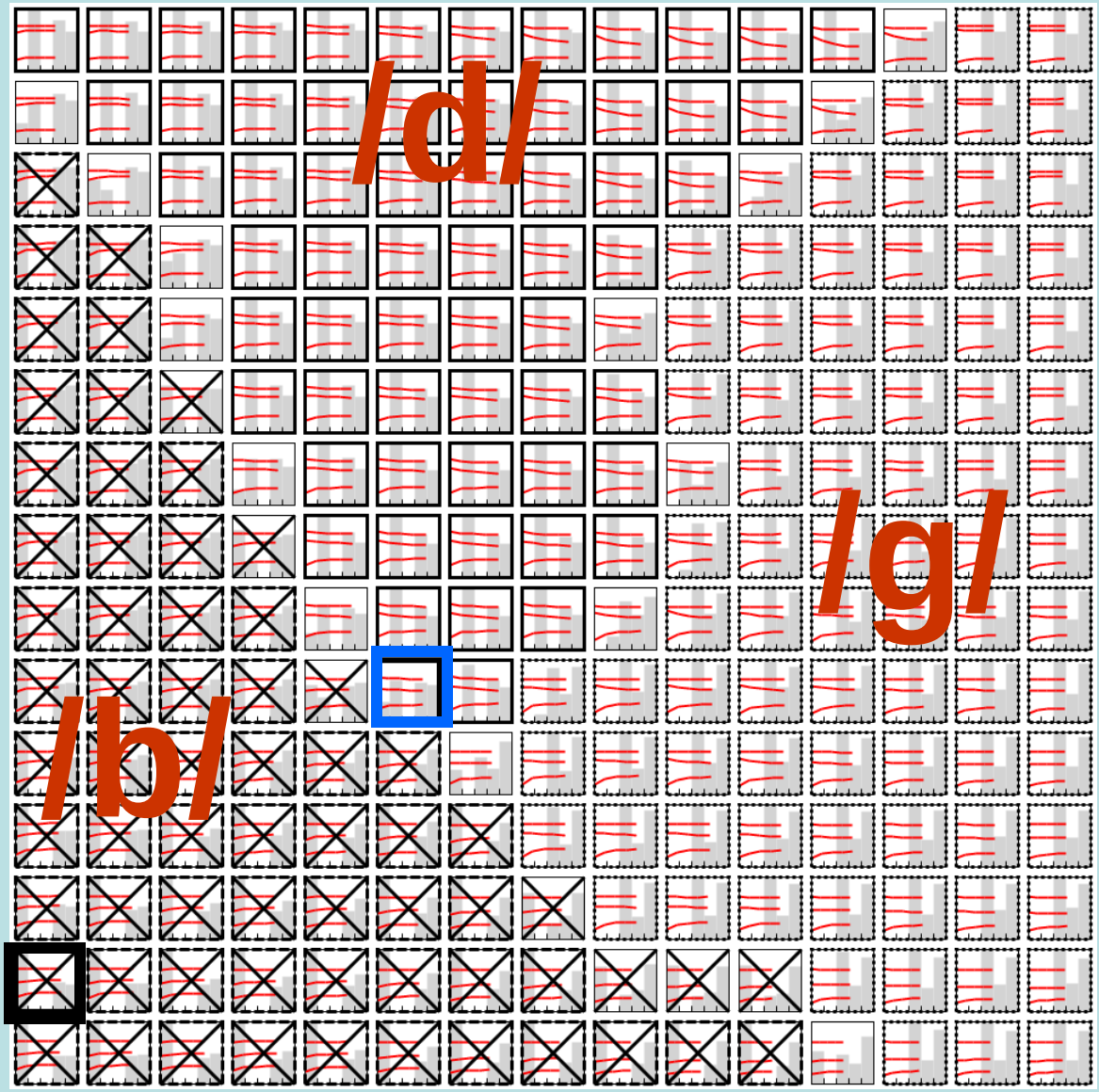
Winner neuron of auditory excitation for auditory [ba] (without inhibition) is within the visual inhibition region for the McGurk-Stimulus → can not be the final winner!

The new next best match: (surprisingly here; but not for all brains) is far away from the inhibition region!

Resulting identification (i.e. phonemic activation) is 100% /da/, but on the edge to /ga/.

Is this brain a typical “McGurk”-listener? Other brains:

Virtual Listener 2: [ba]$_{aud}$+[ga]$_{vis}$

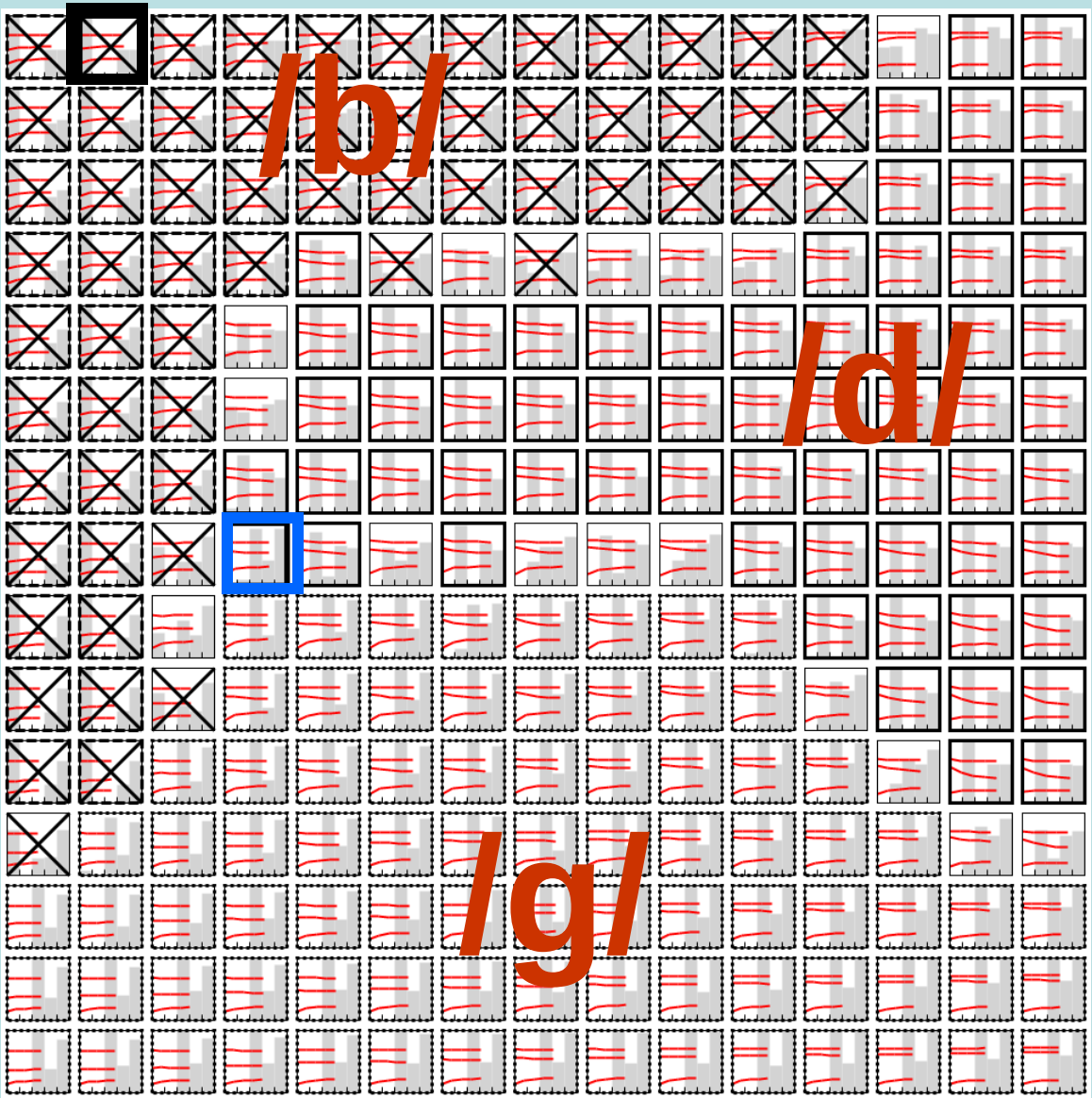


The next best match is near the inhibited region: is between the /b/ and /d/ region

Resulting identification (i.e. phonemic activation) is between /ba/ (33%) and /da/ (67%), see gray bars.
(assuming: motor plan link weight value for lab/api/dor equals phonemic identification rate)

Assuming a perfect association: motor – phonemic (which occurs during imitation training)

Virtual Listener 3: [ba]$_{aud}$+[ga]$_{vis}$



The next best match is near the inhibited region, and: between the /b/, /d/, and /g/ region

Resulting identification (i.e. phonemic activation) is between /da/ (9%) and /ga/ (91%), see gray bars.

Results: McGurk Effect

Listener	/b/	/d/	/g/
1	0.20	**0.80**	0.00
2	0.33	**0.67**	0.00
[illegible]	[illegible]	[illegible]	[illegible]
4	0.00	**1.00**	0.00
5	0.34	**0.66**	0.00
6	0.27	**0.73**	0.00
7	0.00	**1.00**	0.00
8	0.00	**1.00**	0.00
9	0.04	**0.96**	0.00
10	0.34	**0.66**	0.00
11	0.00	**1.00**	0.00
12	0.15	**0.85**	0.00
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
15	0.34	**0.66**	0.00
[illegible]	[illegible]	[illegible]	[illegible]
17	0.00	**1.00**	0.00
[illegible]	[illegible]	[illegible]	[illegible]
19	0.26	**0.74**	0.00
20	0.39	**0.61**	0.00

Three different types of listeners:

Listener type 1 ★ always perceives /d/ (→ 5 virtual listeners); spatial separation of the winner neuron from the inhibited region within the phonetic map

Listener type 2 ★ mainly perceives /d/ but /b/ in some cases as well (→ 10 virtual listeners); no spatial separation from inhibited region within the phonetic map

Exclude listener type 3 → remain 15 brains

Listener type 3 mainly perceives /g/ but /b/ in some cases as well (→ 5 virtual listeners); no spatial separation as well how realistic ?

No /g/-identification in the experiment (McGurk and MacDonald 1976, Nature)!

Tab 1: Phonemic link weight values for the McGurk winner neuron within the phonetic map for each of 20 virtual listeners
→ Probability of /b-d-g/-identification

Results: McGurk Effect

Listener	/b/	/d/	/g/
1	0.20	**0.80**	0.00
2	0.33	**0.67**	0.00
3	0.00	0.09	**0.92**
4	0.00	**1.00**	0.00
5	0.34	**0.66**	0.00
6	0.27	**0.73**	0.00
7	0.00	**1.00**	0.00
8	0.00	**1.00**	0.00
9	0.04	**0.96**	0.00
10	0.34	**0.66**	0.00
11	0.00	**1.00**	0.00
12	0.15	**0.85**	0.00
13	0.19	0.00	**0.81**
14	0.08	0.00	**0.92**
15	0.34	**0.66**	0.00
16	0.21	0.00	**0.79**
17	0.00	**1.00**	0.00
18	0.29	0.00	**0.71**
19	0.26	**0.74**	0.00
20	0.39	**0.61**	0.00

Case	/b/	/d/	/g/
20 brains	0.17	0.62	0.21
15 selected	**0.28**	**0.72**	**0.00**

Tab 2: Probability of /b-d-g/-perception resulting from our model
All 20 brains: /d/ is still perceived most often!
Just 15 brains: → Close to the results of McGurk and MacDonald 1976, Nature

Tab 1: Phonemic link weight values for the McGurk winner neuron within the phonetic map for each of 20 virtual listeners
→ Probability of /b-d-g/-perception in the

A Neurofunctional Model of Speech Production Including Aspects of Auditory and Audio-Visual Speech Perception

Kröger and Kannampuzha (2008) Proceedings of Int. Conf. of AVSP, Moreton Island, Queensland, Australia

A Neurofunctional Model of Speech Production Including Aspects of Auditory and Audio-Visual Speech Perception

Bernd J. Kröger & Jim Kannampuzha

Department of Phoniatrics, Pedaudiology and Communication Disorders, University Hospital Aachen and Aachen University, Germany

bkroeger@ukaachen.de, jkannampuzha@ukaachen.de

Outline: Part 2

- Introduction
- Simulation of Speech Perception
  - Categorical Perception
  - McGurk Effect
- Conclusions

Conclusions

- Starting point: A model of speech production, which gained its knowledge during babbling and imitation training:
- Categorical Perception and McGurk-Effect result in a straight forward way
- CP and McGurk-Effect can be interpreted at the neural (microscopic) level / at the level of the phonetic self-organizing supramodal map

Many Thanks for Your Attention !!!

References:

www search: Bernd Kroeger homepage

www.speechtrainer.eu

Outline

- Part 1: The neurocomputational model: production and acquisition: structure of the model and gaining knowledge
- Part 2: The neurocomputational model: perception: auditory perception (CP) and audio-visual perception: McGurk-Effect;
- Part 3: The concept of speech actions and its relation to manual and facial actions in face-to-face communication

Outline: Part 3

- Motivation
- The Organization of Motor Plans
  - Speech
  - Facial Expressions
  - Manual Gesturing
- Action Hierarchy
- Conclusions

Outline: Part 3

- Motivation
- The Organization of Motor Plans
  - Speech
  - Facial Expressions
  - Manual Gesturing
- Action Hierarchy
- Conclusions

Motivation: Face-to-Face Communication

- Passive listening (e.g. looking TV) / i.e. passive learning is not effective if a toddler starts learning a language! Moreover:
- Speech acquisition needs face-to-face communication between toddler and language expert (caretaker)
- Toddler uses triadic communication scenarios in order to learn words: point and look at an object, look at caretaker, thus: the toddler forces the caretaker to produce the word
- Then: active imitation of the word in order to learn its motor plan, its auditory state, … for storing it in the action repository (phonetic map)
- And: we need motivation (open stance; positive emotional state) in order to be capable to learn (hippocampus, limbic system) → sociable robot!

MIT: Kismet-Project, Breazeal 2004

Towards an articulation-based developmental robotics approach for word processing in face-to-face communication

Kröger et al. (in press)
PALADYN Journal of Behavioral Robotics

PALADYN. JOURNAL OF BEHAVIORAL ROBOTICS
DOI: 10.2478/s13230-011-0016-6 Online First

REVIEW ARTICLE

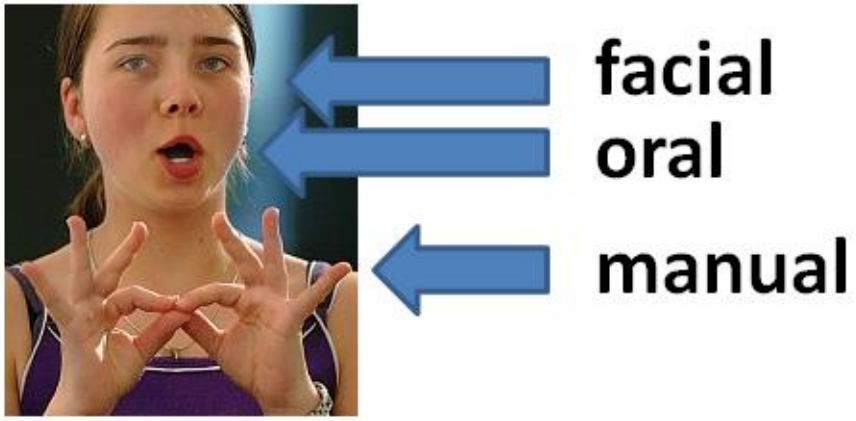
Towards an articulation-based developmental robotics approach for word processing in face-to-face communication

Bernd J. Kröger, Peter Birkholz und Christiane Neuschaefer-Rube

Motivation: Movement Actions

Movement actions are the basic units of production in all domains of face-to-face communication:

- communication via speech; conveys the meaning of an utterance (as well as emotional states by voice quality etc.)
- communication via facial expressions; e.g. indicating an emotional state (e.g. happiness, fear, …)
- communication via manual gesturing; e.g. indicating a direction by pointing (helps to express meanings; to underline important parts …)

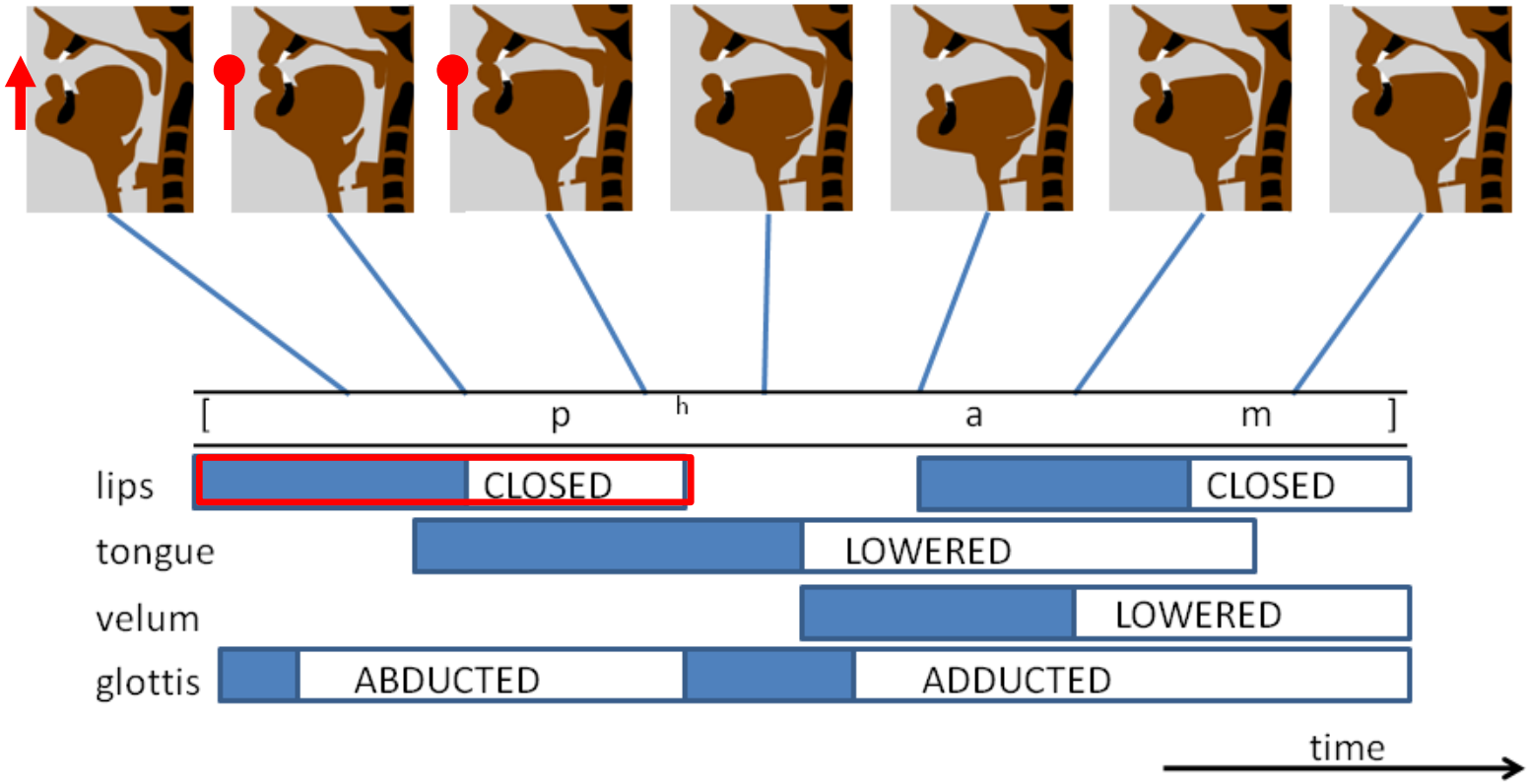


Outline: Part 3

- Motivation
- The Organization of Motor Plans
  - Speech
  - Facial Expressions
  - Manual Gesturing
- Action Hierarchy
- Conclusions

Speech: Motor Plan for the Word “palm”

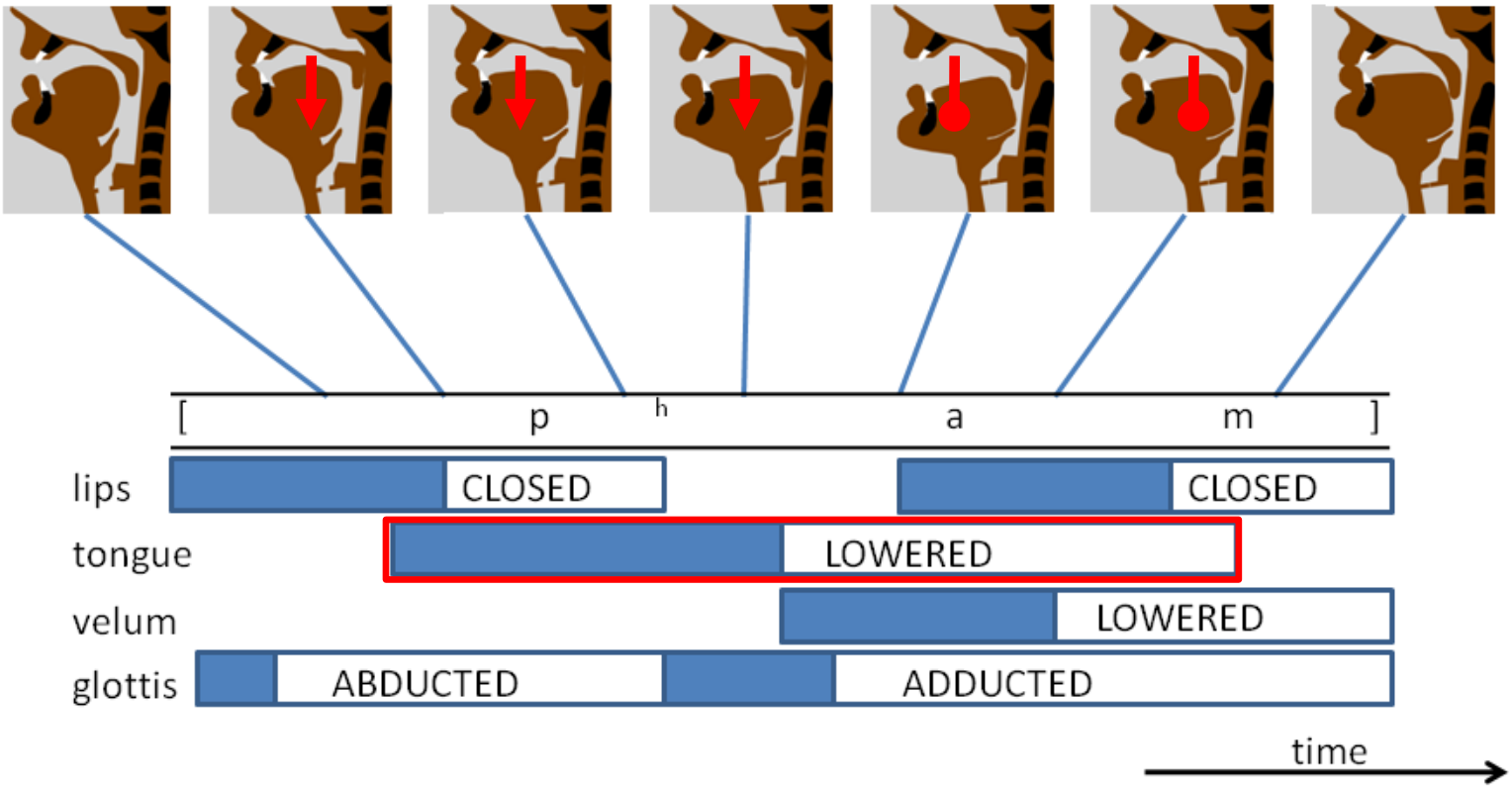
score of: vocal tract movement actions (Browman and Goldstein 1989, 1992, Goldstein 2006, Kröger et al. 1992, 1995, 2008, 2010 and 2011):



- movement actions comprises movement phase (blue) and target phase (white)

Speech: Motor Plan for the Word “palm”

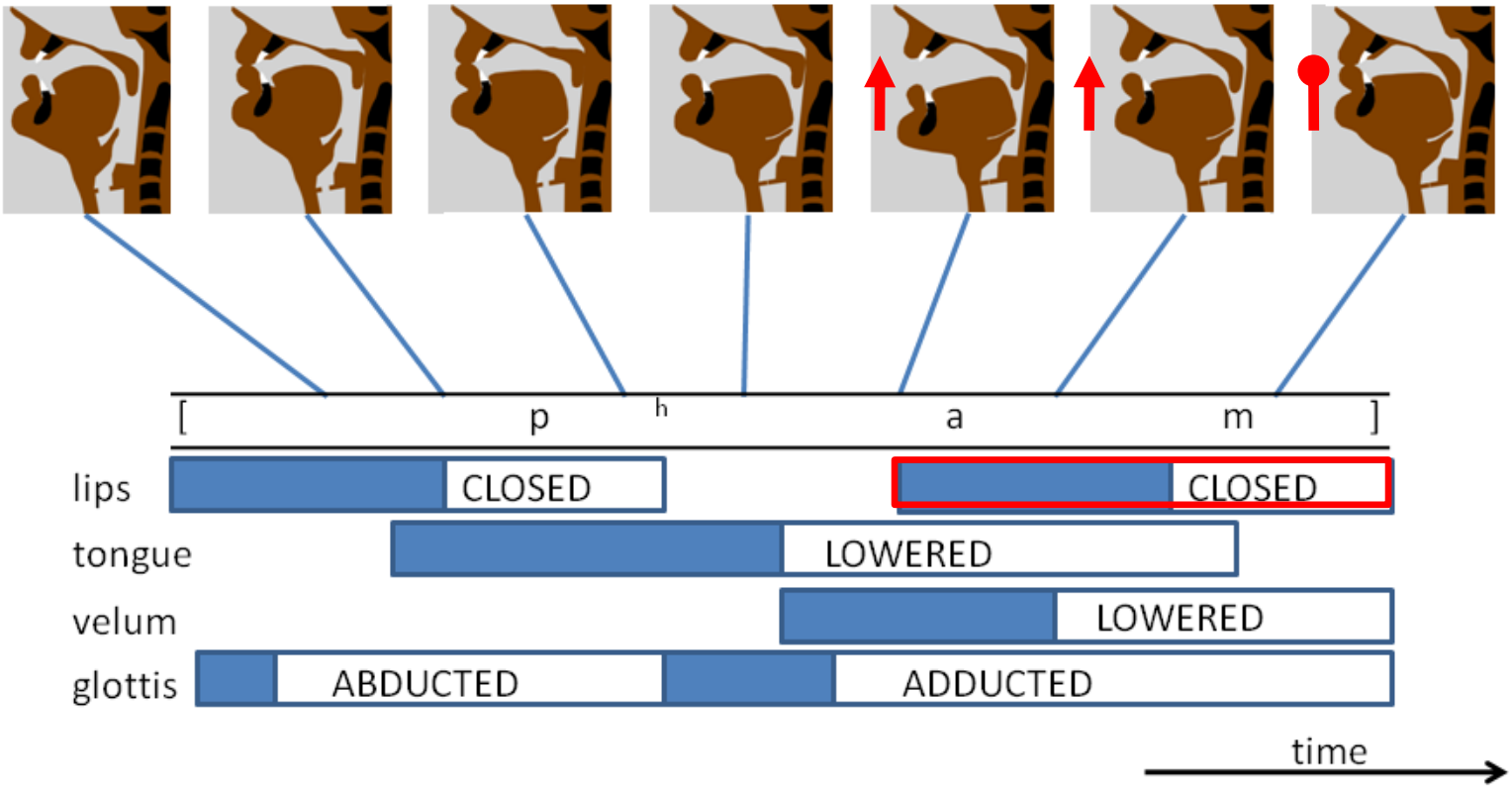
score of: vocal tract movement actions :



- movement actions comprises movement phase (blue) and target phase (white)

Speech: Motor Plan for the Word “palm”

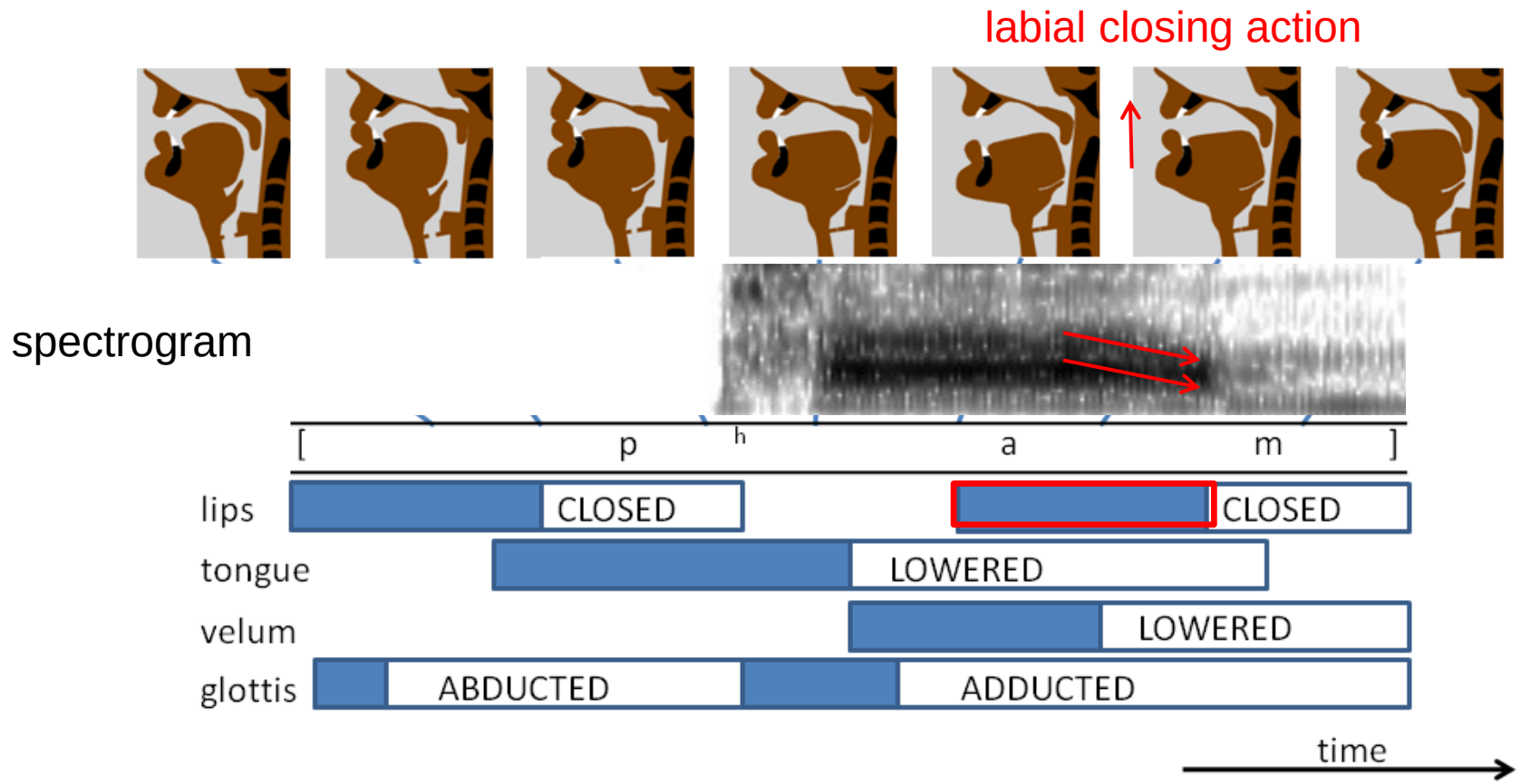
score of: vocal tract movement actions:



- movement actions comprises movement phase (blue) and target phase (white)

Speech: Motor Plan for the Word “palm”

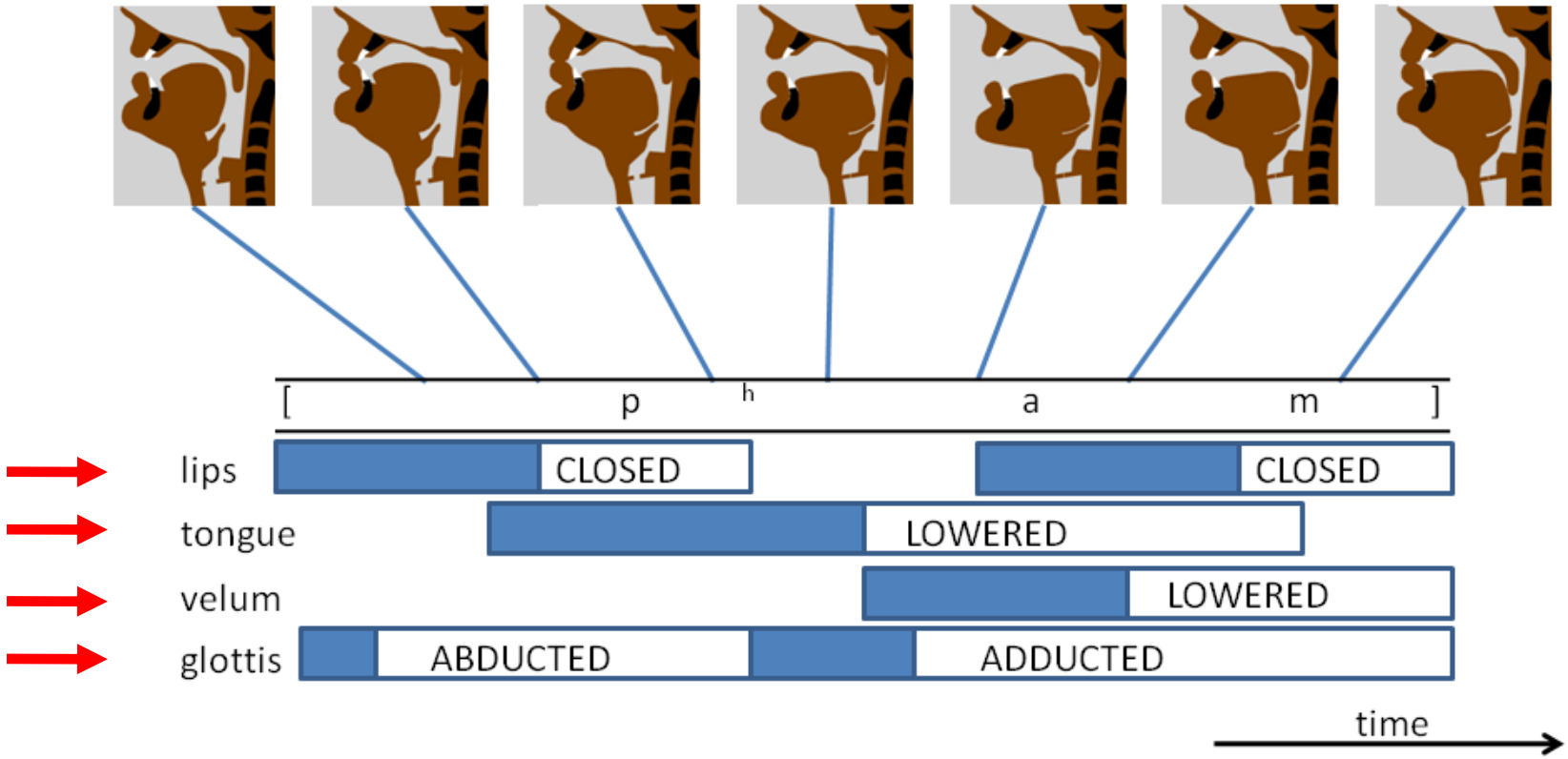
score of: vocal tract movement actions:



- movement actions comprises movement phase (blue) and target phase (white)
- Importance of movement phase: e.g.: the second consonantal movement
- results in a formant transition, which codes the place of articulation: labial!

Speech: Motor Plan for the Word “palm”

score of: vocal tract movement actions:



Keep in mind: the organization of the motor plan:
The existence of parallel tiers for temporal coordination of movement actions.

This occurrs in a parallel way for facial expressions and for manual gesturing!

Outline: Part 3

- Motivation
- The Organization of Motor Plans
  - Speech
  - Facial Expressions
  - Manual Gesturing
- Action Hierarchy
- Conclusions

Facial Expressions: Motor Plan for “happy”

basic units: facial action units (AUs):


neutral


happy

AU2: outer brow raising
AU4: brow lowering
AU5: upper lid raising
AU6: cheek raising + lid compressor
AU7: lid tightening
AU12: lips corner pulling
AU15: lip corner depressing
AU26: jaw dropping

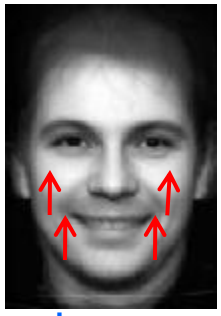
The concept of actions is already well established for describing facial expressions since 1976:
→ Facial Action Coding System FACS (Ekman and Friesen 1976, Cohn 2007)

Facial Expressions: Motor Plan for “happy”

basic units: facial action units (AUs):

mainly two AUs constitute the facial expression “happy”:

neutral



happy

AU2: outer brow raising
AU4: brow lowering
AU5: upper lid raising
AU6: cheek raising + lid compressor
AU7: lid tightening
AU12: lips corner pulling
AU15: lip corner depressing
AU26: jaw dropping

The concept of actions is already well established for describing facial expressions since 1976:

→ Facial Action Coding System FACS (Ekman and Friesen 1976, Cohn 2007)

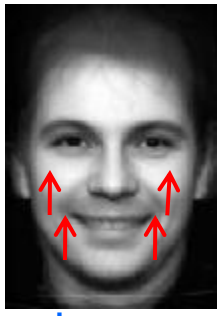
Facial Expressions: Motor Plan for “happy”

basic units: facial action units (AUs):

mainly two AUs constitute the facial expression “happy”:

the motor plan is relatively simple (compared to speech): temporally co-occurring (synchronous) movement actions:

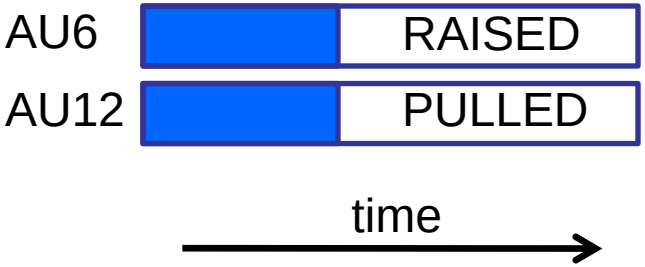
neutral



happy

AU2: outer brow raising
AU4: brow lowering
AU5: upper lid raising
AU6: cheek raising + lid compressor
AU7: lid tightening
AU12: lips corner pulling
AU15: lip corner depressing
AU26: jaw dropping

Importance of movement phase:



Two behaviors exist in order to separate a spontaneous smile from an acted smile:

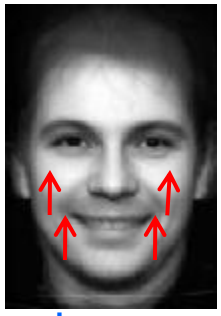
Facial Expressions: Motor Plan for “happy”

basic units: facial action units (AUs):

mainly two AUs constitute the facial expression “happy”:

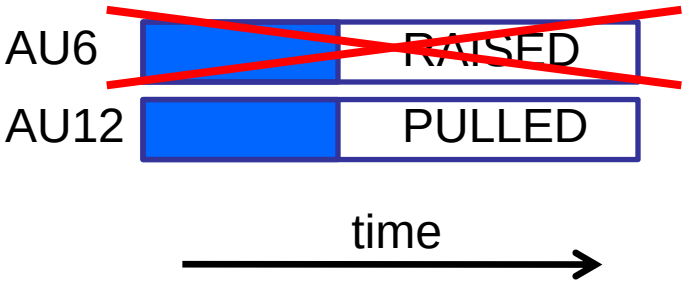
the motor plan is relatively simple (compared to speech): temporally co-occurring (synchronous) movement actions:

neutral



happy

AU2: outer brow raising
AU4: brow lowering
AU5: upper lid raising
AU6: cheek raising + lid compressor
AU7: lid tightening
AU12: lips corner pulling
AU15: lip corner depressing
AU26: jaw dropping



Two behaviors exist in order to separate a spontaneous smile from an acted smile:
- no eye activity (lid compression) → acted

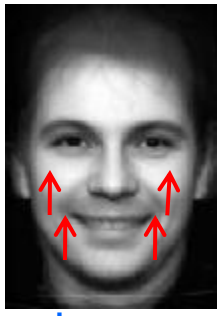
Facial Expressions: Motor Plan for “happy”

basic units: facial action units (AUs):

mainly two AUs constitute the facial expression “happy”:

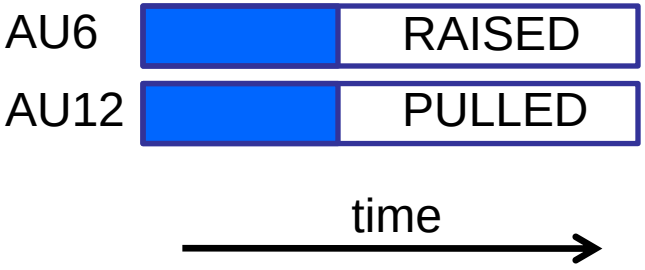
the motor plan is relatively simple (compared to speech): temporally co-occurring (synchronous) movement actions:

neutral



happy

AU2: outer brow raising
AU4: brow lowering
AU5: upper lid raising
AU6: cheek raising + lid compressor
AU7: lid tightening
AU12: lips corner pulling
AU15: lip corner depressing
AU26: jaw dropping



Two behaviors exist in order to separate a spontaneous smile from an acted smile:
- no eye activity (lid compression) → acted
- faster movement phase → spontaneous

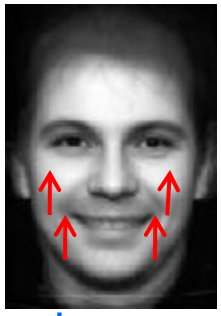
Facial Expressions: Motor Plan for “happy”

basic units: facial action units (AUs):

mainly two AUs constitute the facial expression “happy”:

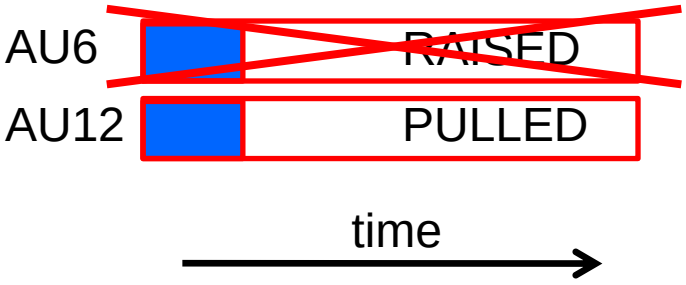
the motor plan is relatively simple (compared to speech): temporally co-occurring (synchronous) movement actions:

neutral



happy

AU2: outer brow raising
AU4: brow lowering
AU5: upper lid raising
AU6: cheek raising + lid compressor
AU7: lid tightening
AU12: lips corner pulling
AU15: lip corner depressing
AU26: jaw dropping



Two behaviors exist in order to separate a spontaneous smile from an acted smile:
- no eye activity (lid compression) → acted
- faster movement phase → spontaneous

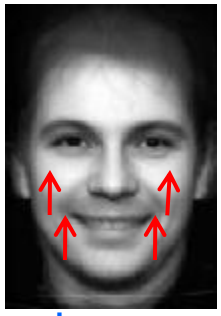
Facial Expressions: Motor Plan for “happy”

basic units: facial action units (AUs):

mainly two AUs constitute the facial expression “happy”:

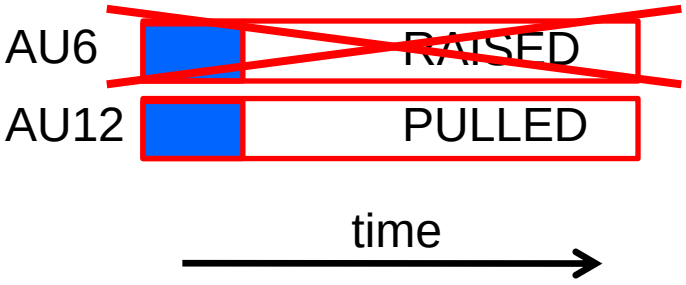
the motor plan is relatively simple (compared to speech): temporally co-occurring (synchronous) movement actions:

neutral



happy

AU2: outer brow raising
AU4: brow lowering
AU5: upper lid raising
AU6: cheek raising + lid compressor
AU7: lid tightening
AU12: lips corner pulling
AU15: lip corner depressing
AU26: jaw dropping



Two behaviors exist in order to separate a spontaneous smile from an acted smile:
- no eye activity (lid compression) → acted
- faster movement phase → spontaneous

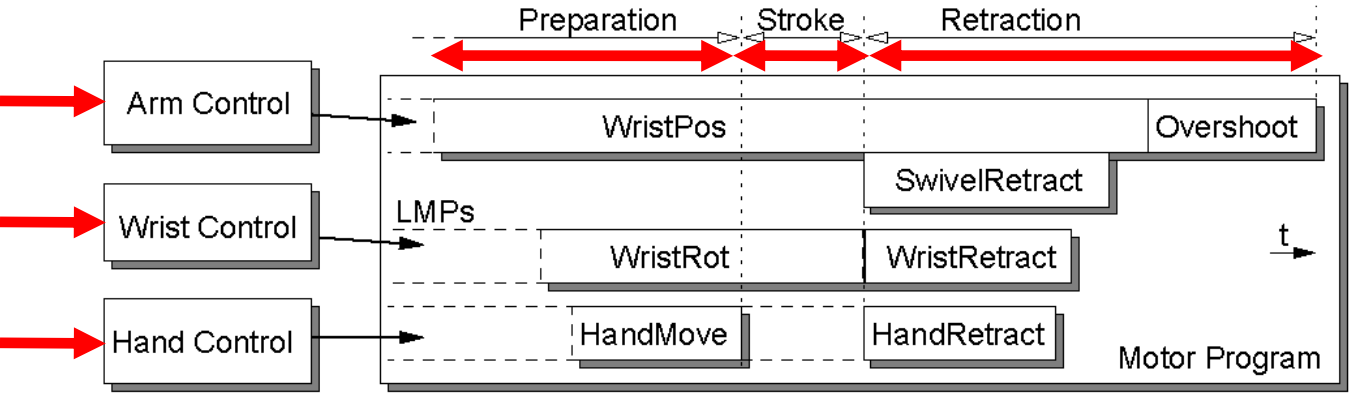
Outline: Part 3

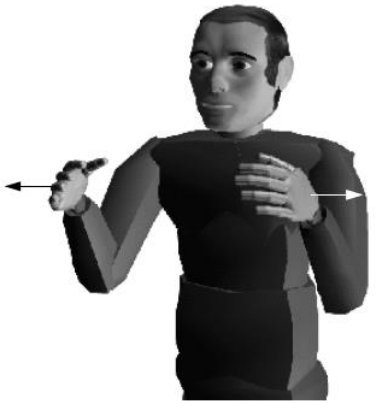
- Motivation
- The Organization of Motor Plans
  - Speech
  - Facial Expressions
  - Manual Gesturing
- Action Hierarchy
- Conclusions

Manual Gesturing: Motor Plan for “this size”

- basic units: manual action units:
- organized on three different articulator tiers
- three major “temporal phases” constitute one (meaningful) manual gesture (above the level of movement actions)

  → this kind of complex organization of movement actions also occurs for speech, where the syllable is the major temporal organization unit





see: Kopp & Wachsmuth 2002

Outline: Part 3

- Motivation
- The Organization of Motor Plans
  - Speech
  - Facial Expressions
  - Manual Gesturing
- Action Hierarchy
- Conclusions

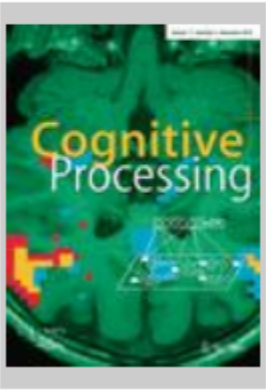
Comparable Action Hierarchy in all 3 Domains:

		Oral (speech)	Facial	Manual
Function (FML)	meaningful unit (intention)	word, utterance	facial expression, e.g. happiness,fear, surprise …	gesture, e.g. pointing in a specific direction
Behavior (BML)	unit of temporal coordination of actions	syllable	one unit per expression: synchronous movement actions	preparation, stroke, retrac-tion phase, …
	movement actions (distinctive features)	vocal tract action	facial action unit e.g. AU12	manual action unit
	articulators coordi-nated by a move-ment action	lips, tongue, velum, …	one per action: e.g. mouth corners, eye lids, …	arm, wrist, handform (fin-gers)
	muscle groups controlling each articulator	e.g. tongue: hyoglossus, genioglossus, styloglossus,…	e.g. mouth corners: zygomaticus major, risorius, …	e.g. arm: mus-cles of: shoulder to upper arm to forearm

A model for production, perception, and acquisition of actions in face-to-face communication

Kröger et al. (2010)
Cognitive Processing 11: 187-205

COGNITIVE PROCESSING
Volume 11, Number 3, 187-205, DOI: 10.1007/s10339-009-0351-2



REVIEW

A model for production, perception, and acquisition of actions in face-to-face communication

Bernd J. Kröger, Stefan Kopp und Anja Lowit

Outline: Part 3

- Motivation
- The Organization of Motor Plans
  - Speech
  - Facial Expressions
  - Manual Gesturing
- Action Hierarchy
- Conclusions

Conclusions

Starting with a neurocomputational model for production, perception and acquisition of speech (part 1 and 2 of this talk), it is a goal of our current work to extend the model in order to take into account the whole process of face-to-face communication, including speech, facial and manual actions

This is for example important in order to model speech acquisition in a more realistic way!

We try to replace our “connectionist rate neuron (node)” approach (SOMs and GSOMs) by a spiking neuron approach -> replace nodes by spiking neuron ensembles

Many Thanks for Your Attention !!!

References:

www search: Bernd Kroeger homepage

www.speechtrainer.eu