How could a self-organizing associative speech action repository (SAR) be represented in the brain?

Bernd J. Kröger[1,2], **Stefan Heim**[3,4,5,6]

[1]Neurophonetics Group, Department of Phoniatrics, Pedaudiology, and Communication Disorders, Medical School, RWTH Aachen University, Aachen, Germany
[2]Cognitive Computation and Applications Laboratory, School of Computer Science and Technology, Tianjin University, Tianjin, P.R.China
[3]Section Structural Functional Brain Mapping, Department of Psychiatry, Psychotherapy, and Psychosomatics; Medical School, RWTH Aachen University, Aachen, Germany
[4]Research Centre Jülich, Institute of Neuroscience and Medicine (INM-1 and INM-3), Jülich, Germany
[5]Section Clinical and Cognitive Neurosciences, Department of Neurology, Medical School, RWTH Aachen University, Aachen, Germany
[6]JARA – Translational Brain Medicine, Jülich and Aachen, Germany

1 Introduction

The concept of a mental lexicon as a repository of semantic and phonological representations of words is well established (Levelt 1999, Elman 2004). In parallel it can be assumed that a repository exists for motor representations of the most frequent syllables of a speakers language in order not to burden the brain with “compiling” or “online assembling” a motor plan in real time for each syllable during speech production (concept of mental syllabary, see Levelt & Wheeldon 1994, Cholin et al. 2006).

Over the last decade we have computer-simulated the knowledge or skill acquisition needed for the built up of such a mental syllabary. This knowledge or skill acquisition occurs during the first years of speech and language learning (Kröger et al. 2009 and 2011). A computer simulation of speaking skill acquisition first of all needs a quantitative definition of the neural maps and neural mappings establishing the mental syllabary, called “speech action repository” or SAR in our quantitative approach.

In this paper we give a brief description of the SAR and we try to underline hypotheses concerning potential brain locations of the neural maps and mappings constituting the SAR based on results of functional brain imaging studies.

2 The quantitative model of a speech action repository (SAR)

In a computer-implemented model for simulating human behavior like acquisition, production, or perception of speech, each component needs to be defined quantitatively. The speech action repository in our approach comprises four neural maps and three mappings (see also Kröger et al. 2009 and 2011). The model neurons within the phonetic map represent phonetic states of frequent syllables and are capable to activate a motor plan (also called higher level motor state), a higher level auditory and a higher level somatosensory state. Thus if a phonetic state is activated at the level of the phonemic map (also called P-Map), the model directly activates skill knowledge, i.e. the model "knows" how to articulate that syllable as well as how the production of that syllable should "feel" (activated somatosensory state) and how that syllable should "sound" (activated auditory state). The neural mappings between phonetic map and sensorimotor maps are "dense", which means that each neuron of the phonetic map is connected with each neuron of each sensorimotor map (green arrows in Fig. 1).

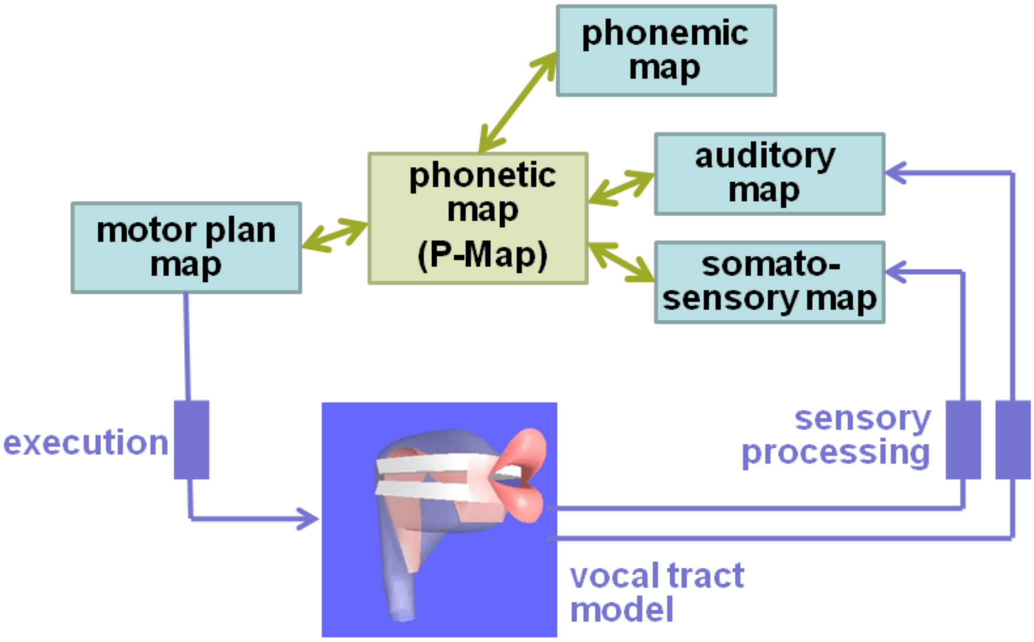

Figure 1 - Structure of the computer-implemented model of speech production, perception, and acquisition. Motor plan map is equivalent to higher level motor map, see Fig. 2 and Fig. 3. Motor plan map, higher level auditory map, higher level somatosensory map, phonetic map (also called P-Map), and the mappings in between these maps build up the speech action repository (SAR). Phonemic map is part of the mental lexicon.

The motor and sensory maps within SAR are higher level neural representations. They comprise the motor plan as well as the sensory impression (or state) of a syllable and thus, these maps are part of short term memory. The activation patterns within these maps (motor plan and sensory maps) change from syllable to syllable.

The motor plan state rather describes the temporal coordination of speech action units (Kröger et al. 2010) than detailed muscle activation patterns which occur at the level of the primary motor map (lower level motor map). Thus the motor plan activation pattern co-activates lower level motor patterns for each syllable under production and subsequently these patterns generate muscle forces and result in model articulator movements (Fig. 1: forward pathway including execution).

The model articulator displacements or model articulator movements are calculated at the peripheral level by the vocal tract model (Kröger & Birkholz 2007). This module is capable of generating 3D-geometries of vocal organs, vocal tract tube geometries, as well as acoustic speech signals. This geometrical and acoustic information is further used for generating higher level somatosensory (proprioceptive and tactile) and auditory higher level states for each syllable (see Fig. 1: feedback pathway, including sensory processing). The resulting higher level sensory states produced by the feedback pathway can be compared with the initially activated higher level sensory states already stored in the SAR in order to detect production errors.

The assumption of storing not just motor plans but higher level sensory information as well is based on the theory of error signal generation and feedback correction of motor commands as developed by Guenther et al. (2006). Thus our SAR not just stores motor plans for each frequent syllable of a target language but as well stores the sensory (auditory, proprioceptive, and tactile) targets for each frequent syllable. This storage of sensory targets is definitely needed, if modeling of adaptation effects due to auditory or mechanical articulatory perturbations is desired (Guenther 2006).

Last but not least it should be mentioned that the P-Map can be interpreted as a “connector hub” in brain network theory (Bullmore & Sporns 2009). This map connects motor and sensory states and thus has the function of a sensorimotor interface (see below).

3 Hypotheses concerning the representation of an SAR in the brain

The computational functioning of our model of speech production, speech perception, and speech acquisition is well documented (Kröger et al. 2009 and 2011). But especially the association of our postulated phonetic map (P-Map, Fig. 1) with specific brain regions is not a trivial task.

The location of lower level (or primary) motor and sensory maps is obvious. The lower level motor map for the speech organs is located in the inferior part of the primary motor cortex (BA 4), the lower level somatosensory map is located in the inferior part of the primary somatosensory cortex (BA 3) and the lower level auditory map is located in the primary auditory cortex (BA 41, BA 42, see e.g. Kandel et al. 2000 and see light blue ovals in Fig. 2).

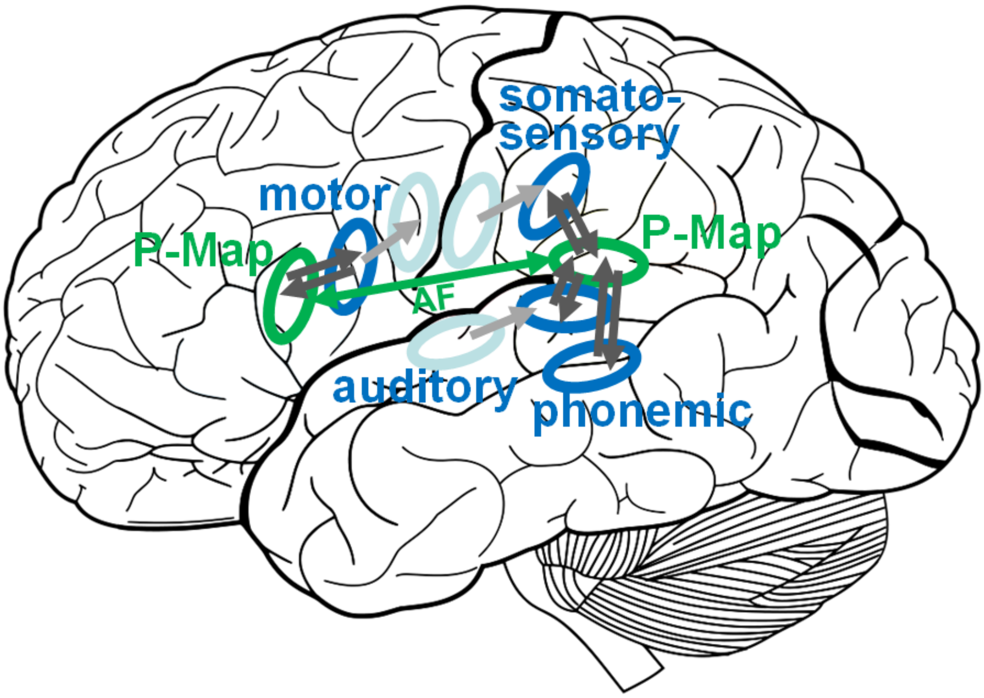

Figure 2 – Hypothetical locations of (lower and higher level) neural maps and mappings of our model. Green: two mirrored locations of the phonetic map (P-Map) possibly interconnected by the arcuate fasciculus (AF); dark blue: higher level motor maps, sensory maps, and phonemic map; light blue: lower level (or primary) motor and sensory maps; higher level motor map is equivalent with motor plan map in Fig. 1. Arrows indicated mappings between neural maps.

Higher level sensory maps are assumed to be located in brain regions found for the sensory error maps in the Guenther model (also called DIVA model, Guenther 2006, Golfinopoulos et al. 2010), i.e. posterior superior temporal lobe / planum temporale for higher level auditory map (BA 22 and Fig. 2) and inferior parietal lobe / ventral somatosensory cortex for higher level somatosensory map (BA 5 and Fig. 2). The phonemic map is assumed to be located in the posterior mid temporal lobe (Hickok & Poeppel 2007 and Fig. 2). The location of the higher level motor map (motor plan map) can be assumed to be within the medial and dorsolateral premotor cortex, supplementary motor area, anterior insula, as well as superior cerebellum (Riecker et al. 2005; the location of motor plan map as part of premotor cortex is shown in Fig. 2).

A first hypothetical location for a hypermodal sensorimotor interface – as is one of the main functions of the phonetic map in our model (see Fig. 1) – is postulated by Hickok & Poeppel (2007) as “area Spt”, which is located at the posterior end of the Sylvian fissure, i.e. the boundary between parietal and temporal lobe. But this location – as any other single location for the phonetic map (P-Map in Figure 2) – causes a severe problem from the viewpoint of neural functioning.

Because in our neurocomputational approach the model neurons within the phonetic map represent single phonetic states (i.e. specific phonetic realizations of a syllable), and because these local representations of phonetic states within the P-Map lead to distributed neural co-activations within motor plan and higher level sensory maps, the neural mappings between phonetic map and motor plan map as well as between phonetic map and sensory maps (auditory and somatosensory map) need to be dense. In other words, each neuron within the phonetic map must be connected with each neuron within the motor plan, within the auditory, and within the somatosensory map. And each synaptic link weight within these mappings needs to be adjusted during speech acquisition (Kröger et al. 2009 and 2011). Thus, the dense neural mappings between phonetic map and motor map as well as between phonetic map and the higher level sensory maps comprises the whole knowledge concerning the facts how a frequent syllable is executed, how its production feels and how that syllable sounds.

We assume that neural maps which need a dense neural interconnectivity are not allowed to be far apart in space for economical as well as anatomical reasons. Thus, we assume that it could be allowed to connect auditory and somatosensory map via a phonetic map located in between these two sensory maps as is indicated in Fig. 2 by the posterior location of the P-Map (area Spt). But it is not likely, that a dense and information-bearing neural mapping is allowed to bridge such a long distance from area Spt to the motor plan map, located mainly in the premotor cortex. Therefore we hypothesize that a second or “mirrored” representation of the P-Map exists in the frontal lobe (anterior representation of P-Map), which at any time indicates the same neural activation pattern as is activated within the posterior representation of the P-Map. In this case only a sparse “long distance” connection between the posterior part of Sylvian fissure and an area within or near the premotor cortex needs to be established. From our neurocomputational viewpoint this connection just needs to be simple parallel connection of axons with a maximal exhibitory synaptic link weight between the neurons of the two representations of the P-Map (Fig. 3), because only one identical neural activation pattern needs to be forwarded from P-Map to P-Map.

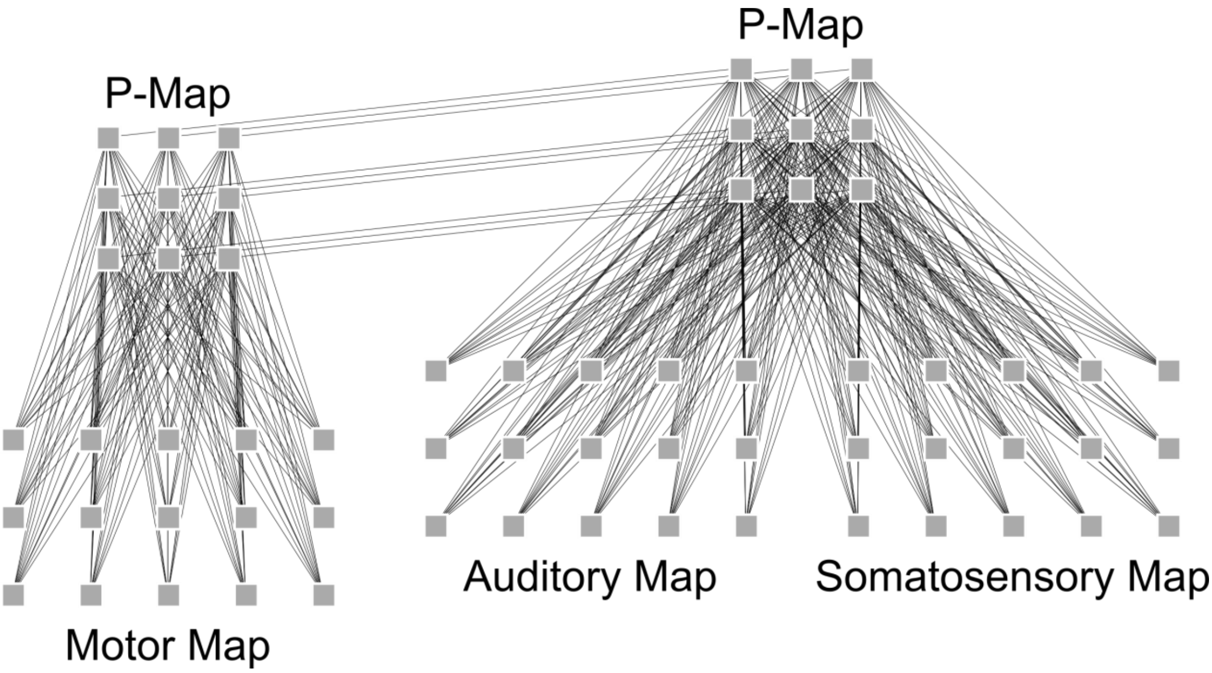

Figure 3 – Detailed neural structure of the computer-implemented model of the speech action repository (SAR), i.e. higher level motor and sensory maps and mappings in between via two representations of the P-Map. Boxes represent model neurons within each map; lines between boxes represent (bidirectional) synaptic connections between model neurons. For hypothetical locations of two P-Map representations, see text.

4 Conclusions and further work

The assumption concerning a „mirrored" representation of a P-Map is at least anatomically underlined by the existence of a neural fiber bundle, called “arcuate fasciculus” connecting Broca and Wernicke area (Kandel et al. 2000). This idea of connecting sensory and motor regions of the brain is a very traditional one in neurolinguistics. But in terms of our model this assumption mainly results from economy principles for neural connections based on the topology of our neurocomputational model. Furthermore it should be stated that the interconnection of Broca and Wernicke area originally is not a purely phonetic concept – i.e. a concept describing the level of mental syllabary – but a concept for connecting the mental lexicon with motor execution, which in comparison to our modeling approach is located at higher levels of the speech production and speech perception mechanism.

Thus, it is one of our next goals to do brain imaging experiments which are based on “mental syllabary tasks” in order to underpin our assumption, i.e. to find brain activation patterns which underline the existence of two representations of the phonetic map as part of a speech action repository (SAR).

Acknowledgements

The first author of this paper wants to thank Prof. Dr. Georg Heike, who drew my attention to phonetic and articulatory modeling already 25 years ago and who conducted my first simulation experiments concerning vocal tract modeling. Moreover we would like to thank Cornelia Eckers for carrying out Figure 3.

This work is partially supported by the National Natural Science Foundation of China (Grant No. 61233009)

References

Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience 10, 186-198

Cholin J, Levelt WJM, Schiller NO (2006) Effects of syllable frequency in speech production. Cognition 99, 205-235

Elman JL (2004) An alternative view of the mental lexicon. Trends in Cognitive Sciences 8, 301-306

Golfinopoulos E, Tourville JA, Guenther FH (2010) The integration of large-scale neural network modeling and functional brain imaging in speech motor control. NeuroImage 52, 862-874

Guenther FH (2006) Cortical interaction underlying the production of speech sounds. Journal of Communication Disorders 39, 350-365

Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. Brain and Language 96, 280-301

Hickok G, Poeppel D (2007) Cortical neuroanatomy of speech processing. Nature Reviews Neuroscience 8, 393-402

Kandel ER, Schwartz JH, Jessell TM (2000) Principles of Neural Science (4th ed., McGraw-Hill, New York)

Kröger BJ, Birkholz P (2007) A gesture-based concept for speech movement control in articulatory speech synthesis. In: A Esposito, M Faundez-Zanuy, E Keller, M Marinaro (eds.) Verbal and Nonverbal Communication Behaviours, LNAI 4775 (Springer, Berlin, Heidelberg), pp. 174-189

Kröger BJ, Kannampuzha J, Neuschaefer-Rube C (2009) Towards a neurocomputational model of speech production and perception. Speech Communication 51, 793-809

Kröger BJ, Kopp S, Lowit A (2010) A model for production, perception, and acquisition of actions in face-to-face communication. Cognitive Processing 11, 187-205

Kröger BJ, Birkholz P, Kannampuzha J, Neuschaefer-Rube C (2011) Towards the acquisition of a sensorimotor vocal tract action repository within a neural model of speech processing. In A Esposito, A Vinciarelli, K Vicsi, C Pelachaud, A Nijholt (eds.) Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues (Springer, Berlin), pp. 287-293

Levelt WJM (1999) Models of word production. Trends in Cognitive Sciences 3, 223-232

Levelt WJM, Wheeldon L (1994) Do speakers have access to a mental syllabary? Cognition 50, 239-269