Self-Organization and Associative Learning as a Basis for Cognitive and Sensorimotor Modeling of Speech Production, Perception, and Acquisition

Bernd J. Kröger, Cornelia Eckers, Christiaine Neuschaefer-Rube
({bkroeger, ceckers, cneuschaefer}@ukaachen.de)
Department of Phoniatics, Pedaudiology, and Communication Disorders, RWTH Aachen University,
Pauwelsstr. 30, 52074 Aachen, Germany

Keywords: sensorimotor modeling; cognitive modeling; speech production; speech perception; speech acquisition.

The Structure of the Model

A computational model has been proposed which is capable of simulating early phases of speech acquisition, speech production, and speech perception. The model comprises two main modules, i.e. mental lexicon and action repository (Fig. 1). The mental lexicon activates semantic and phonological representations of words (cognitive level, Li et al. 2004) while the action repository activates sensory and motor representations of syllables (cf. Levelt & Wheeldon 1994, Guenther et al. 2006, Kröger et al. 2009, Kröger et al. 2011a).

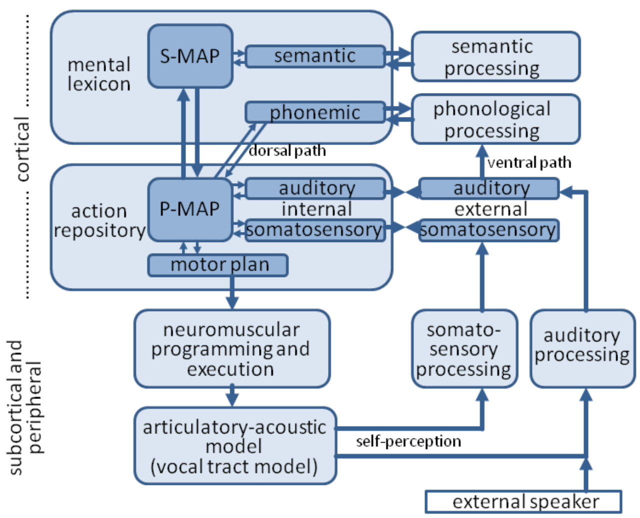

Figure 1: Structure of our cognitive-sensorimotor model of speech processing. Light blue boxes indicate processing modules; dark blue boxes indicate self-organizing maps (i.e. semantic map S-MAP and phonetic map P-MAP) or neural state maps, i.e. the semantic, phonemic, auditory, somatosensory, and motor plan state map (for a detailed description of this model see Kröger et al. 2011a).

The sensorimotor part of our model has been implemented and tested by simulating early phases of speech acquisition (i.e. babbling phase and imitation phase) and performing production and perception tests after learning (Kröger et al. 2009 and 2011b). The detailed structure of the sensorimotor part of our model is given in Fig. 2 (top). Cortical regions associated with specific neural maps within our approach, are displayed in Fig. 2 (bottom).

A characteristic feature of our approach is that we assume two self-organizing maps – i.e. a semantic map (S-MAP, Fig. 1) and a phonetic map (P-MAP, Fig. 1) – which on the one hand are associated with each other and which on the other hand are associated with state maps, representing current semantic, motor plan or sensory activation patterns of speech items (cf. Li et al. 2004, Zhao et al. 2011).

Acquisition of Skills and Knowledge

Acquisition is simulated in our approach by applying a huge amount of training items to the model. These training items represent stimuli, which are exposed to a newborn and later on to a toddler (i.e. to the model) during the first two years of lifetime.

Acquisition starts with "babbling", i.e. a training phase which is mainly language independent. Here the model generates random motor patterns (motor plan states) and produces appropriate auditory and somatosensory patterns (auditory and somatosensory states). Motor plan and sensory states are exposed to the model nearly simultaneously and thus allow associative learning, i.e. an association of specific motor plan states with corresponding sensory states (Kröger et al. 2009). This learning leads to an adjustment of neural connections between state maps and the self-organizing phonetic map (P-MAP). Neurons within the phonetic map represent specific sensorimotor states and these states are ordered with respect to phonetic features within this map (i.e. self-organization). Thus, after learning co-activation of a motor plan state is possible, if a specific sensory (e.g. auditory) state is activated. In this way, initial sensorimotor knowledge is acquired for "proto-vowels" and "proto-CV-syllables" (the term "proto" refers to the fact that these phonetic states are not necessarily language specific; cf. Kröger et al. 2009).

This initial sensorimotor knowledge later on allows "imitation training" since after initial babbling the model is able to imitate external auditory stimuli. Imitation training leads to a further adjustment of neural connections between P-MAP and state maps and leads to a further ordering of states within the P-MAP, which now leads to language-specific speaking skills (Kröger et al. 2011b). Beside further development of the action repository, imitation training is also the starting point for building up the mental lexicon (ibid.). Training items for imitation training comprise sensorimotor states, which result from imitation trials performed by the model itself, but in addition comprise a semantic representtation of the word which is currently imitated. This allows a parallel self-organization of the S-MAP and adjusts the neural connections between both self-organizing maps (P-MAP and S-Map) as well as the neural connections to all state maps. Thus, imitation training in addition leads to the formation of first language specific knowledge (i.e. phonological representation of words, see Kröger et al. 2011a).

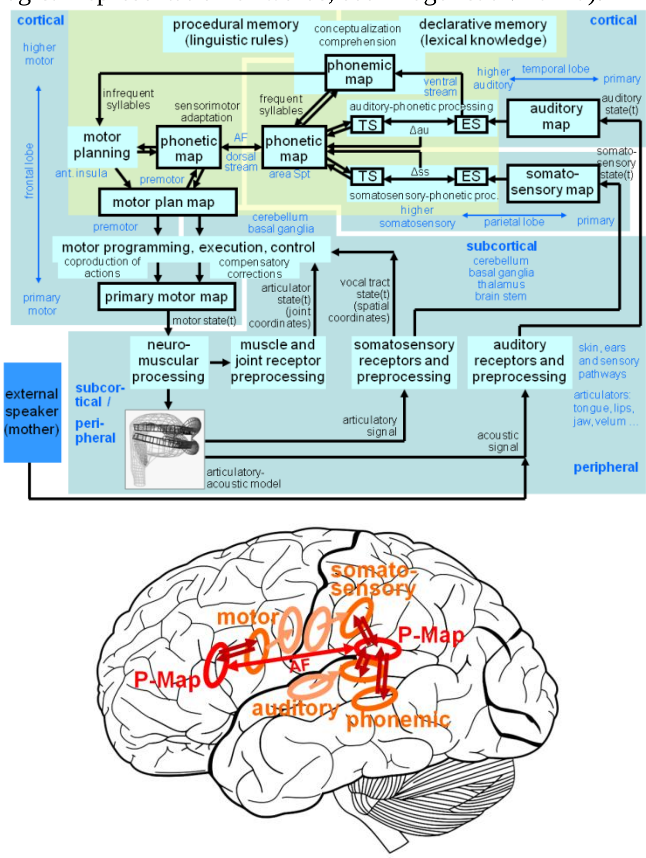

Figure 2: Structure of the sensorimotor part of the model (top) and cortical regions related to specific neural maps occurring within the model (bottom). Top: The model comprises a feedforward pathway (motor) and three feedback pathways (lower and higher level somatosensory and auditory). Outlined boxes indicate neural maps; other boxes indicate neural processing modules, which are not specified in detail. Single arrows indicate neural pathways for forwarding information; double arrows indicate neural mappings which are involved in information processing. The light green area indicates higher processing levels which activate syllables as entire units; lower levels (primary cortical maps and subcortical sturcutres) are capable of processsing smaller temporal units of production and perception. TS: map for trained sensory states (already acquired); ES: map for external sensory states (currently produced); Δau: auditory error signal; Δss: somatosensory error signal. Bottom: Cortical locations of implementations of the primary neural maps (pink: primary motor map, primary somatosensory map, and primary auditory map), of the neural state maps (orange: phonemic map, motor plan map, somatosensory and auditory state maps within the somatosensory-phonetic and auditory-phonetic processing modules), and of two mirrored representations of the phonetic map (red). Neural mappings between state maps and phonetic map are indicated by dark red arrows; The neural pathway between the two mirrored representations of the phonetic map is indicated by a red arrow; AF: arcuate fascilicus.

Future Work

The phonetic as well as the semantic map are the central maps for self-organization and enable associative learning in our approach. While the gross structure of our model is in accordance with the well-known models introduced by Guenther et al. (2006) and by Li et al. (2004), the reality of the phonetic map (e.g. in its mirrored location, joined via the AF, as postulated in Fig. 2) needs to be proved by brain imaging experiments.

Acknowledgments

This work was supported in part by German Research Council (DFG) grants Kr 1439/13-1 and Kr 1439/15-1.

References

Guenther, F.H., Ghosh, S.S. & Tourville, J.A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96, 280-301.

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

Kröger, B.J., Birkholz, P. & Neuschaefer-Rube, C. (2011a). Towards an articulation-based developmental robotics approach for word processing in face-to-face communication. *PALADYN Journal of Behavioral Robotics*, 2, 82-93.

Kröger, B.J., Birkholz, P., Kannampuzha, J., Kaufmann, E. & Neuschaefer-Rube, C. (2011b). 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. LNCS 6800*. Berlin: Springer, pp. 287-293.

Levelt, W.J.M. & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239-269.

Li, P., Farkas, I. & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, 17, 1345-1362.

Zhao, X., Li, P. & Kohonen, T. (2011). Contextual self-organizing map: software for constructing semantic representations. Behavior Research Methods 43, 77-88.