SOM'S AND GSOM'S IN BIOLOGICALLY INSPIRED MODELS OF SPEECH PROCESSING

Bernd J. Kröger[1,2] & Mengxue Cao[3]

[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]Laboratory of Phonetics and Speech Science, Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China

bernd.kroeger@rwth-aachen.de, mengxuecao@outlook.com

Abstract: Modeling speech processing in a biologically inspired way can be done by using growing self-organizing maps. But this approach is highly abstract because each "node" here represents an ensemble of "real" neurons, in our interpretation a cortical column. Moreover neural spikes trains are not modeled in this approach. Rather a mean rate of neural activation is taken as basic processing variable for each node. In this paper the concept of growing self-organized maps is reviewed and its neuroscientific relevance is discussed (i) from the viewpoint of spatial and temporal integration (cortical columns and activity rates) and (ii) from the viewpoint of basic neural principles like self-organization and associative learning in speech processing.

1 Introduction: SOMs and GSOMs

The synaptic associations between neurons are organized and modified during learning. Thus, only a part of cortical and subcortical organization is pre-determined by our DNA, while a main part of synaptic links especially at the level of the neocortex result from learning, i.e. from our daily interaction with our environment. For example during babbling the toddler learns sensorimotor associations from exploring random motor states. Later on, during imitation of speech items produced by a caretaker (teacher), the toddler may be awarded for correct or at least understandable word productions (reinforcement learning) but still in this case the learning stimuli occur in a more or less random fashion.

An interesting concept, how the neocortex may organize these learning results (this knowledge) is the concept of self-organization, resulting from non-supervised learning, where the adjustment of synaptic link weights mainly follows a "winner takes all" principle and where the adaptation of link weights is always possible even during later stages of learning (neural plasticity). A quantitative concept called "self-organizing maps" (SOMs), also known as self-organizing feature maps was introduced by Kohonen [1-4]. Here, the self-organizing map in most cases is realized as a 2D map, which could be interpreted as a highly abstract representation of a cortical neural map. These maps highlight features, inherent to the training items by a specific spatial ordering of states; e.g. vocalic states are ordered with respect to phonetic dimensions like high-low or front-back; e.g. consonantal states are ordered with respect to phonetic features like place and manner of articulation [5]. Self-organizing maps can be seen as a nonlinear generalization of principal component analysis. Competitive learning ("winner takes all" principle) is a further important characteristic for the adjustment of link weights between a self-organizing map and its input/output maps [5].

A first main shortcoming of the concept of self-organizing maps may be its simplicity. In these days, most computational or theoretical neuroscientists use spiking neuron approaches

(e.g. [6]), where the temporal resolution is high (i.e. at the level of 1 ms) and where each model neuron is assumed to represent a biological one (i.e. high spatial resolution). Moreover the detailed spiking behavior of single neurons as well as the spike-train behavior of neuron groups is modeled. In contrast, the concept of self-organizing maps, which already has been developed during the 70th and 80th of the last century, is different. Here "model neurons" or "nodes" represent a whole ensemble of cortical neurons. Nodes are used as central processing units and even neural activity is accumulated over time intervals (one processing step here may represent 20 to 50 ms intervals). Thus, self-organizing maps belong to the group of *activation rate models*. But it will be shown below that activation rate models and the concept of "activation rates" as well as of "model neurons", representing an ensemble of neighboring biological neurons, can be interpreted quite concrete in biological terms [7].

A further shortcoming of the concept of self-organizing maps may be that the number of model neurons for the self-organizing map needs to be predefined, which is biologically implausible, because due to neural plasticity the number of neurons which are involved in the formation of a self-organizing map should increase during learning in relation to the increasing number of words or syllables which are learned. Thus, algorithms for training of *growing* self-organizing maps (GSOMs) have been developed and it has been shown that GSOMs have advantages in performance with respect to learning effort and feature representation [8].

2 Biological Inspiration: Spatial Aspects

We hypothesize that each computational "node" within a 2D-self-organizing map represents one or more (neighboring) cortical columns, because the ordering of features occurring in SOMs and GSOMs are found as well in cortical maps (see also [25-29] in section 4 of this paper). The concept of cortical columns introduced in 1957 by Mountcastle [9] has become a very attractive concept for combining functional and anatomical aspects of the neocortex (for a review see [10, 11]). Cortical columns here are defined as "basic information processing elements of the cortex, with each column being responsible for analyzing a small range of stimuli" ([11], p.7). Beyond this definition, which is focusing on sensory cortex (i.e. parts of the cortex analyzing somatosensory cortex [9], visual cortex [12], or auditory cortex [13, 14]), cortical columns have been found within motor cortex [15, 16] as well as within higher level supra- or hypermodal and cognitive cortex regions [17].

In cortical columns, cortical layer 4 processes the input signal and the signal is forwarded to output layers 2 and 3 by pyramidal cells [18]. Furthermore, columnar neural activity is processed vertically (i.e. within a cortical column) by pyramidal cells in deep layers 5 and 6. Inhibitory lateral (i.e. horizontal) cortical connections with other cortical columns within a cortical map are established by interneurons, occurring within all cortical layers. In addition some short range excitatory horizontal connections are formed by the neurons of layer 4 (see [11], p. 7).

But the concept of cortical columns has been critically reviewed by other researchers with respect to anatomical aspects [19] as well as with respect to functional aspects [20]. One argument is the failure of finding discrete boundaries between columns in many cases, but the fact of strong vertical connectivity and the fact of existence of input as well as output function, to our opinion are sufficient as biological arguments for the concept of 2D-cortical maps composed of "nodes" or "model neurons". In addition, a horizontal connectivity within one as well as from cortical column to other cortical columns has been reported [11]. Within a cortical column, excitatory synaptic links can be assumed as well from the viewpoint of "model neurons" in order to strengthen their activity, while synaptic connections towards neighboring columns (neighboring model neurons) are mainly inhibitory form the viewpoint

of self-organizing maps in order to underpin the winner takes all principle (lateral inhibition [3], pp. 177ff).

3 Biological Inspiration: Temporal Aspects

While spiking neuron models are capable of generating complex spike trains at the level of a millisecond time scale, time representation is raw in neurocomputational *rate models* like self-organizing and growing self-organizing maps. In the case of rate models the network and its synaptic link weight values develop and/or change one time for each training steps. Therefore, each training step represents one time step. Here an input stimulus is modeled by a specific activation pattern of input neurons, which may result in a sum of spikes representing the brain activity related to that input stimulus. The resulting change in link weights between input layer and self-organizing map is influenced mainly by the fact whether a stimulus is awarded or not, and thus the time window for stimulus activation and for the adaptation of synaptic link weights with respect to a stimulus may last one or more seconds [21]. Thus time is not represented *directly* within this network. Only a temporal succession of training items (training stimuli) and a co-occurring temporal succession of changes of synaptic link weights is modeled.

After learning, if the model is producing or perceiving a speech item (production or perception mode of the network), a rate network can be used for calculating activation patterns at the input level from winner neuron activation at the level of the self-organizing map. But here as well no specific modeling of temporal aspects occurs. Here, each time step is used to activate a SOM or GSOM neuron and subsequently a specific neural activation pattern of a whole syllable or word [5]. If stimuli have an intrinsic temporal representation, this intrinsic time is an inherent part of the neural input or output representation. A sequence of time intervals is coded by a (spatial) sequence of model neurons as it may occur in working memory (see our definition of neural representation of auditory states or spectrograms and our neural representation of motor plans [22]).

Here the *multilayer Hebbian-learning model* of Garagnani et al. [7], leading to "Hebbian neuronal circuits" as highly specific functional units, offers a neurobiologically more realistic perspective for modeling sensorimotor aspects of syllables and words. This multilayer Hebbian-learning model especially takes into account important neural functional principles (e.g. neural excitation and inhibition and thus long-term potentiation (LTP) and long-term depression (LTD)) in a less abstract, i.e. in a neurobiologically more realistic way as it is done in SOM or GSOM theory.

Last but not least *recurrent neural networks* (rate based or spiking based) could bring more neurobiological realism as well, because temporal features of sensory or motor stimuli can be coded here intrinsically (intrinsic time representation [23, 24]).

4 Neural Self-Organization, Neural Plasticity, and Associative Learning

A major argument for biological realism of topology preserving cortical maps in speech processing comes from studies which prove spatial ordering e.g. of phonetic features for vowels as well as for consonants [25-29]. These topological aspects of speech sound ordering with respect to phonetic sound features can be modelled using self-organizing maps [5, 30]. Moreover it has been shown that words can be organized very effectively with respect to their semantic features by using self-organizing maps [31, 32]. The resulting self-organizing phonetic as well as semantic maps vary locally during training due to the influence of new learning items and thus show plasticity [33]. Even in the case of lower-level sensor maps – reflecting tonotopy, retinotopy as well as somatotopy – neural plasticity can be modelled by

using self-organizing maps [34]. This feature of cortical topology preservation at high supra- or hypermodal levels cannot be modelled currently by spiking neuron models.

Associative learning is one of our major learning principles and can be modelled by SOMs and GSOMs if we do not take input training items just from one input domain – as is usually done in the case of Kohonen's SOMs and in the case of most GSOM applications – but from *two or more input domains*, e.g. auditory, somatosensory and motor domain in parallel in the case of speech processing [5]. In early phases of speech acquisition like babbling, this leads to an association of motor with sensory states. Later on during imitation training this leads to an association of sensorimotor states with phonemic states [5]. Thus, SOMs as well as GSOMs not just lead to an ordering of states with respect to phonetic or semantic features (in case of speech processing) but it occurs as well as an association of auditory, somatosensory, motor, and phonemic representations at the level of the phonetic map. This main feature, i.e. the capability of associating states from sensory and motor domains is a further argument for the biological plausibility of SOMs and GSOMs, because no spiking neuron model exists, which clearly indicates this feature.

In the case of spiking neuron models, a first approach is available to model learning of associations between sensory or motor stimuli and their reward, if these stimuli are rewarded in a communication situation. Reinforcement learning can be modelled here by associating a sensory or motor stimulus with a reward stimulus [35, 36]. This method is not directly comparable to the situation of a direct association of auditory and motor states as described above for SOMs and GSOMs, but gives a deep insight in behaviour of long-term potentiation (LTP) as well as long-term depression (LTD) resulting from spike-timing-dependent plasticity (STDP) of synaptic connections as well as resulting from dopamine modulations, where the later modulation is mainly responsible for long-term effects with respect to associating a motor or sensory state with a reward stimulus.

5 Discussion and Conclusions

If the goal is the development of a large-scale model for speech acquisition, speech production, and speech perception, rate models like SOMs or GSOMs seem to be a reasonable solution. But its neurobiological plausibility is questionable because of its high degree of abstraction. These models are in a way abstract because (i) they use abstract "model neurons" (or "nodes"), because (ii) no specific inherent time representation is given, and because (iii) no explicit modelling of neural inhibition is done. But these models are capable to represent sensory input and motor output in a reasonable way and are capable to learn and store sensorimotor skills (speech action repository [5, 30]) as well as lexical knowledge (e.g. semantic knowledge [31-33]). They are abstract but could be interpreted carefully as biologically motivated to a specific degree, because they incorporate important neurofunctional principles like self-organization, associative learning, Hebbian learning, adaptation, and neural plasticity.

Currently there still seems to be a huge gap for replacing these self-organizing map models by large-scale spiking neuron models. First approaches are still available [24, 36] but these approaches are still far away from modelling the complex interactions between speech learning (acquisition), speech production, and speech perception as it is already possible by using neural rate models, especially SOM- and GSOM-based neural models.

Literature

[1] Kohonen T (1982) Self-organized formation of topologgically corect feature maps. Biological Cybernetics 43, 59-69

[2] Kohonen T (1990) The self-organizing map. Proceedings of the IEEE 78, 1464-1480

[3] Kohonen T (2001) Self-Organizing Maps (Springer, 3rd edition)

[4] Kohonen T, 2013. Essentials of the self-organizing map. Neural Networks 37, 52-65

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

[6] Kasabov N (2010) To spike or not to spike: A probabilistic spiking neuron model. Neural Networks 23, 16-19

[7] Garagnani M, Wennekers T, Pulvermüller F (2008) A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain. European Journal of Neuroscience 27, 492-513

[8] Alahakoon D, Halgamuge SK, Sirinivasan B (2000) Dynamic self-organizing maps with controlled growth for knowledge discovery. IEEE Transactions on Neural Networks 11, 601-614

[9] Mountcastle VB (1957) Modality and topographic properties of single neurons of cat's somatic sensory cortex. Journal of Neurophysiology 20, 408-434

[10] Mountcastle VB (1997) The columnar organization of the neocortex. Brain 120, 701-722

[11] Goodhill GJ, Carreira-Perpinan MA (2002) Cortical columns. In: Encyclodedia of Cognitive Science (John Wiley & Sons Ltd.) URL: http://onlinelibrary.wiley.com/doi/10.1002/0470018860.s00356/abstract

[12] Hubel DH, Wiesel TN (1977) Functional architecture of the macaque monkey visual cortex. Proceedings of the Royal Society of London B 198, 1-59

[13] Schreiner CE (1995) Order and disorder in auditory cortical maps. Current Opinion in Neurobiology 5, 489-496

[14] Schreiner CE, Winer JA (2007) Auditory cortex mapmaking: principles, projections, and plasticity. Neuron 56, 356-365

[15] Asanuma H (1975) Recent developments in the study of the columnar arrangement of neurons within the motor cortex. Physiological Reviews 55, 143-156

[16] Hatsopoulos NG (2010) Columnar organization in the motor cortex. Cortex 46, 270-271

[17] Silver MA, Kastner S (2009) Topographic maps in human frontal and parietal cortex. Trends in Cognitive Sciences 13, 488-495

[18] Mumford D (1992) On the computational architecture of the neocortex. II: the role of cortico-cortical loops. Biological Cybernetics 66, 241-251

[19] da Costa NM, Kevan KAC (2010) Whose cortical column would that be? Frontiers in Neuroanatomy DOI: 10.3389/fnana.2010.00016

[20] Horton JC, Adams DL (2005) The cortical column: a structure without a function. Philosophical Transactions of the Royal Society B 360, 837-862

[21] Izhikevich EM (2007) Solving the distal reward problem through linkage of STDP and dopamine signaling. Cerebral cortex 17, 2443-2452

[22] Kannampuzha J, Eckers C, Kröger BJ (2011) Training einer sich selbst organisierenden Karte im neurobiologischen Sprachverarbeitungsmodell MSYL. In: Kröger BJ, Birkholz P (eds.) Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2011 (TUDpress, Dresden, Germany), pp. 154-163

[23] Lazar A, Pipa G, Triesch J (2009) SORN: a self-organizing recurrent neural network. Frontiers in Computational Neuroscience 3(23) doi:10.3389/neuro.10.023.2009

[24] Kiebel S, Yildiz B (2012) How does the brain recognize speech? Modelling using hierarchical recurrent neural networks. In: Wolff M (ed.) Studientexte zur Sprach-

kommunikation: Elektronische Sprachsignalverarbeitung 2012 (TUDpress, Dresden, Germany), pp. 96-103
[25] Obleser J, Lahiri A, Eulitz C (2004) Magnetic brain response mirrors extraction of phonological features from spoken vowels. Journal of Cognitive Neuroscience 11, 31-39
[26] Shestakova A, Brattico E, Soloviev A, Klucharev V, Huotilainen M (2004) Orderly cortical representation of vowel categories presented by multiple exemplars. Cognitive Brain Research 21, 342-350
[27] Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP (2006). Vowel sound extraction in anterior superior temporal cortex. Human Brain Mapping 27, 562-571
[28] Obleser J, Leaver A, Van Meter J, Rauschecker JP (2010) Segregation of vowels and consonants in human auditory cortex: Evidence for distributed hierarchical organization. Frontiers in Psychology 1. doi:10.3389/fpsyg.2010.00232.
[29] Scharinger M, Isardi WJ, Poe S (2011) A comprehensive three-dimensional cortical map of vowel space. Journal of Cognitive Neuroscience 23, 3972-3982
[30] Kröger BJ, Birkholz P, Kannampuzha J, Kaufmann E, Neuschaefer-Rube C (2011) Towards the acquisition of a sensorimotor vocal tract action repository within a neural model of speech processing. In: Esposito A, Vinciarelli A, Vicsi K, Pelachaud C, Nijholt A (eds.) Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues. LNCS 6800 (Springer, Berlin), pp. 287-293
[31] Ritter H, Kohonen T (1989) Self-organizing maps. Biological Cybernetics 61, 241-254
[32] Li P. Farkas I, MacWhinney B (2004) Early lexical development in a self-organizing neural network. Neural Networks 17, 1345-1362
[33] Cao M, Li A, Fang Q, Kröger BJ (2013) Growing self-organizing map approach for semantic acquisition modeling. Proceedings of of 4th IEEE Conference on Cognitive Infocommunications.
[34] Ritter H, Schulten K (1986) On the stationary state of Kohonen's self-organizing sensory mapping. Biological Cybernetics 54, 99-106
[35] Izhikevich EM (2007) Solving the distal reward problem through linkage of STDP and dopamine signaling. Cerebral Cortex 17, 2443-2452
[36] Warlaumont AS (2012) A spiking neural network model of canonical babbling development. Proceedings of the 2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL). DOI: 10.1109/DevLrn.2012.6400842