Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Firing Rate Homeostasis for Dynamic Neural Field Formation

Claudius Gläser, Frank Joublin

2011

Preprint:

This is an accepted article published in IEEE Transactions on Autonomous Mental Development. The final authenticated version is available online at: https://doi.org/[DOI not available]

Firing Rate Homeostasis for Dynamic Neural Field Formation

Claudius Gläser and Frank Joublin

***Abstract*—Dynamic neural fields are recurrent neural networks which aim at modeling cortical activity evolution both in space and time. A self-organized formation of these fields has been rarely explored previously. The main reason for this is that learning-induced changes in effective connectivity constitute a severe problem with respect to network stability. In this paper we present a novel network model which is able to self-organize even in face of experience-driven changes in the synaptic strengths of all connections. Key to the model is the incorporation of homeostatic mechanisms which explicitly address network stability. These mechanisms regulate activity of individual neurons in a similar manner as cortical activity is controlled. Namely, our model implements the homeostatic principles of synaptic scaling and intrinsic plasticity. By using fully plastic within-field connections our model further decouples learning from topological constraints. For this reason, we propose to incorporate an additional process which facilitates the development of topology preserving mappings. This process minimizes the wiring length between neurons. We thoroughly evaluated the model using artificial data as well as continuous speech. Our results demonstrate that the network is able to self-organize, maintains stable activity levels, and remains adaptive to variations in input strength and input distribution.**

***Index Terms*—Development, Dynamic Neural Field, Hebbian Plasticity, Homeostasis, Topology Preservation**

I. Introduction

Dynamic neural field theory provides a mathematical framework by which cortical processing can be modeled at a mesoscopic level. Thereby, neural fields constitute maps in which activity is propagated via extensive lateral connections between neuron populations [1]. Due to the variety in exhibited dynamic behavior [2], neural fields have become a popular technique for modeling spatio-temporal activity flow. Originally developed as a model of pattern formation on the neural tissue [3], neural fields later have been successfully applied in various domains [4], [5], [6]. Even though dynamic neural field theory describes a general network model, a lack in understanding how neural fields can self-organize limits their applicability. Particularly when considering autonomously developing systems, a self-organization of representations is advantageous. This is because representations internal to a developing system are not fixed. They rather are adaptive to allow for a continuous incorporation of new knowledge. The ability to learn and adapt, thus, plays a pivotal role in any developing system.

Unfortunately, learning and adaptation have only rarely been investigated in the context of dynamic neural fields. Learning most often focuses on the synaptic weights of input projections to the neural field, thereby adapting the input-driven dynamics, but leaving the self-driven dynamics unchanged. This is due to the fact that even small learning-induced changes in the connectivity of the field can result in a significantly altered dynamic behavior of the network [7], [8]. Hence, the incorporation of synaptic plasticity is challenging with regard to maintaining the network in stable operation modes.

The LISSOM model [9] constitutes a significant advance in this respect. In contrast to conventional models it additionally features lateral connections that undergo Hebbian plasticity. More precisely, the LISSOM model uses an interaction kernel that is initially wide and roughly Mexican Hat shaped. Subsequently, however, activity-driven learning results in a die off of synapses which sharpens and fine-tunes the kernel. The development of the interaction kernel hence can be compared to the gradually decreasing Gaussian neighborhood that is applied when training the popular Kohonen maps [10]. Due to a change in the within-field connectivity, the LISSOM model is sensitive to the used parameter settings. This is why it additionally comprises an adaptation of the neuron transfer functions that follows a predefined regime.

In the recent years, significant progress has been made in the understanding of homeostatic principles like synaptic scaling [11] or intrinsic plasticity [12]. These principles describe local adaptation mechanisms which regulate neuronal activity following input variations. Homeostasis consequently gained attention as self-regulation may ultimately lead a way to overcome stability issues. The Adaptive LISSOM model (ALISSOM) [13] constitutes an extension of LISSOM by means of two homeostatic mechanisms. Firstly, it uses Triesch's intrinsic plasticity model [14] to adapt neuronal transfer functions and thus replaces the previously predefined regime. Secondly, ALISSOM applies activity-dependent synaptic scaling on the afferent input connections. However, since ALISSOM does not use homeostatic principles to alter the within-field connections, it exhibits a similar parameter sensitivity as the LISSOM model. This is due to the fact that activity propagation within a neural field is largely affected by the balance between excitation and inhibition within the field. The question how a balanced lateral interaction can emerge from self-regulation is consequently an open issue.

Similar to LISSOM our network model differs from conventional neural fields insofar as we do not make any assumption on the connectivity of the field. In other words, all synaptic weights – afferent projections to the field as well as lateral connections within the field – are plastic and change via experience-driven learning. As a direct consequence, we have to apply additional mechanisms to circumvent unfavorable network behavior. We previously proposed homeostatic pro-

The authors are with the Honda Research Institute Europe, Carl-Legien-Strasse 30, 63073 Offenbach, Germany, phone: +49-69-89011-702, fax: +49-69-89011-749, e-mail: {firstname.lastname}@honda-ri.de.

cesses [15] which address this stability issue. These processes operate purely locally with the objective of regulating the average activity of individual neurons towards some target level. The specific way in which self-regulation is implemented within our model is based on recent findings on homeostatic principles applied in the central nervous system. Similar to ALISSOM we use an intrinsic plasticity mechanism as well as synaptic scaling to adjust the network behavior. In contrast, however, we not only scale afferent projections to the field, but also lateral connections by which the excitation-inhibition ratio is altered. The network hence dynamically balances cooperation and competition between the model neurons in an activity-dependent manner.

In this paper we recapitulate our model, justify it by evidence from neurobiological findings, and further thoroughly evaluate it in a series of experiments. We will show that the combination of learning and homeostasis enables our network to self-organize a dynamic neural field while keeping the network in a proper dynamical regime. This increase in flexibility not only allows the network to develop new representations, but also to adapt existing ones to changed environmental conditions. Since the developed mappings do not necessarily have to be topology preserving, we further present an extension of the model which facilitates topology preservation [16]. More precisely, this is achieved via an additional process which aims at minimizing the wiring length between model neurons. This process is independent of learning and runs in parallel to it. We will experimentally show, that the incorporation of this principle significantly decreases the number of topological defects within the developed mappings.

The rest of the paper is organized as follows. In section II we present the structure of our network model and show how learning is incorporated. In section III we then review homeostatic principles applied in the brain and present a specific implementation of them. Afterwards, we discuss our approach for developing topology preserving mappings in section IV. Finally, we present results of various experiments in section V before we give a summary in section VI.

II. Recurrent Neural Network Model

A. Dynamic Neural Field Theory

Dynamic neural fields model the spatio-temporal evolution of activity in the brain. Thereby, the neural tissue is considered to be a two-dimensional plane on which neurons are distributed. The neurons are stimulated by externally applied inputs which evoke an activity within the field. Spatio-temporal response patterns are obtained by propagating activity through extensive lateral interactions between the model neurons. This dynamic spread of activity can be formally described by Amari's field equation [1]:

$$\tau \frac{\partial u(x,t)}{\partial t} = -u(x,t) \quad + \quad \int w(x,x') \cdot f(u(x',t))\, dx' \quad + \quad S(x,t) + h. \tag{1}$$

Here t denotes time, $u(x,t)$ the local membrane potential of a population of neurons at position x of the cortical plane, and $S(x,t)$ the stimulus applied to this neuron population.

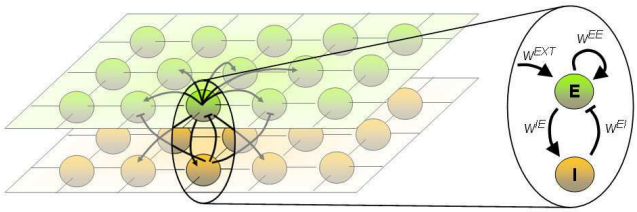


Fig. 1. The structure of the recurrent neural network.

Furthermore, neurons feature a rest potential h which is approached in absence of any other input. The monotonically increasing non-linear function relating the potential of neurons to their activities is termed f. Finally, the lateral connectivity of neurons located at position x' to neurons located at position x of the neural tissue is defined by $w(x,x')$. This interaction kernel is typically fixed and distance-dependent, i.e. $w(x,x') = w(|x-x'|)$. In most previous models a Mexican Hat connectivity is chosen. It implements an excitation between nearby neurons and an inhibition between distal ones. Hence, activity propagation within the field can result in spatially focused regions of activity – also known as activity bubbles.

By using a distance-dependent interaction kernel, models inherently rely on the assumption that any input can be adequately mapped onto the two-dimensional plane. Since this is obviously not true, we propose not to use a fixed interaction kernel. We rather suggest to use plastic lateral connections which can undergo experience-driven changes in synaptic strength. In other words, we believe that it is advantageous not only to learn afferent projections to the field, but also to develop interaction kernels which are appropriate for the inputs that should be represented within the field. The network model we present in the following consequently does not make any assumption on the lateral connection strengths.

B. The Network Structure

Fig. 1 shows the structure of our recurrent neural network model. Similar to the Wilson-Cowan model [3] it is composed of interconnected excitatory units E and inhibitory units I. The different types of units are distributed on a two-dimensional grid mimicking the cortical plane. The difference to the Wilson-Cowan model lies in the connectivity between the units. Where Wilson and Cowan applied an all-to-all connectivity, our model consists of the following connection patterns: External input to the network is provided by afferent projections (w^{EXT}) to the excitatory units. Activity within the field can be propagated via connections between the units. Thereby, the lateral connectivity consist of excitatory connections from E-cells to other E-cells (w^{EE}) as well as I-cells (w^{IE}). Additionally, E-cells receive inhibitory projections (w^{EI}) originating from I-cells.

By discretization of Amari's field equation the spatio-temporal evolution of activity within our network can be described by two differential equations. We use the variables u and v to describe the membrane potentials of the excitatory and inhibitory units, respectively. We further subset an index i

to refer to the unit located at position x_i of the cortical plane:

$$\tau_E \frac{du_i}{dt} = -u_i + \sum_j g(d_{ij}) \cdot w_{ij}^{EE} \cdot f(u_j) - \sum_j w_{ij}^{EI} \cdot f(v_j) + \sum_j w_{ij}^{EXT} \cdot s_j + h^E \quad (2)$$

$$\tau_I \frac{dv_i}{dt} = -v_i + \sum_j g(d_{ij}) \cdot w_{ij}^{IE} \cdot f(u_j) + h^I. \quad (3)$$

Here, the membrane potentials are updated according to the time constants τ_E and τ_I. In absence of any input the potentials u_i and v_i approach the rest potentials h^E and h^I, respectively. Furthermore, the synaptic weight of a connection from unit j to unit i is denoted w_{ij}^* where $* \in \{EE, EI, IE, EXT\}$ describes the type of connection. The relation between the membrane potentials and the activities of units is described by the sigmoidal transfer function f which is of the form

$$f(z) = \frac{1}{1 + \exp\left(-\gamma(z - \theta)\right)}. \quad (4)$$

Thereby, θ is the threshold value at which a neuron exhibits an activity of 0.5, whereas γ is a gain factor which specifies the dynamic range of membrane potentials a neuron is most sensitive to.

In the update equations for the membrane potentials we additionally incorporated a modulation factor g. This factor affects the efficiency of excitatory lateral connections as a function of the distance $d_{ij} = \| x_i - x_j \|_2$ between the pre- and postsynaptic units. More precisely, we use the following function to implement an exponential decrease in connection efficiency when distance increases:

$$g(d) = \exp\left(-\frac{d^2}{2\sigma^2}\right). \quad (5)$$

We consequently define that excitatory lateral connections between nearby units are more efficient than those between distant units. One possible interpretation of g is that of a connection probability between neurons that becomes smaller as their distance increases. Since inhibitory cells are supposed to have a broader connectivity range, they are not modulated by g in this model.

It is, however, worth noticing that the incorporation of a distance-dependent modulation factor fundamentally differs from using distance-dependent interaction kernels. The latter implies that the synaptic weight values of lateral connections are chosen as a function of the distance between the pre- and postsynaptic units. This is not the case for our model, since we do not make any assumption on the synaptic weight values themselves. Large synaptic weight values can consequently compensate for a decrease in connection efficiency.

Interestingly this also means that we withdraw the topological constraints that drive other models to develop topology preserving mappings. Hence, we hypothesize that our model produces mappings which show more topological defects than mappings developed by other approaches. For this reason, our computational model incorporates an additional and independently running process which explicitly addresses the issue of how the development of topology preserving mappings can be facilitated. We will later introduce this process in section IV. A recent study of van Hooser et al. [17] provides evidence in favor of our model. They found orientation-sensitive cells in the primary visual cortex (V1) of a highly visual rodent, the gray squirrel. These cells are similar to those found in V1 of primates, but in contrast to primate V1 the orientation-selectivity did not smoothly vary across the cortical surface. This and other findings [18] suggest that a topology preserving self-organization is not a fundamental principle of mammalian cortical development; it rather seems to depend on other mechanisms missing in rodents.

C. Hebbian Learning

As previously mentioned, all connections within our network model undergo experience-driven changes in synaptic strength. Thereby, the used learning regime is twofold: Firstly, it incorporates a learning rule that adapts connection weights according to the input patterns presented to the network. Secondly, it also comprises self-regulatory processes that keep the neural field in a stable state. The latter will be the focus of the following section. Here, we describe how model neurons develop appropriate representations of the input patterns via Hebbian plasticity. The learning principle stated by Hebb's rule can be shortly summarized as *cells that fire together, wire together*. In other words, if the postsynaptic cell repeatedly fires following a stimulation by the presynaptic cell, the synapse linking both cells is strengthened. To circumvent unconstrained weight growth we apply Oja's rule [19] which incorporates an activity-dependent leakage term:

$$\Delta w_{ij}^* \propto \eta_i \cdot \xi_j - w_{ij} \cdot \eta_i^2. \quad (6)$$

Here, w_{ij}^* is the synaptic weight; η_i and ξ_j are the pre- and postsynaptic activities, respectively. Since it can be shown that Oja's rule extracts the principal component from its inputs [19], it constitutes a suitable learning technique for adapting the synaptic weights of a neural field.

III. Homeostatic Plasticity

Homeostasis refers to the property of a system to regulate its internal environment to compensate for fluctuations in the external environment. It thus ensures system stability via self-maintenance of a proper operation mode. For neural fields, a stable operation strongly depends on balanced levels of excitation and inhibition in the network. Too much inhibition will obviously lead to vanishing activity whereas a high level of excitation may result in runaway activity. This problem is even more severe for developing systems as learning continuously changes network connectivity [20], [21]. Computational models of network development consequently have to incorporate homeostatic mechanisms to cope with these changes. In the following we highlight recent advances in the understanding of the processes regulating neuronal activity and show how similar principles can be used within our network model.

A. Homeostasis in the Central Nervous System

Self-regulation in the central nervous system is achieved by numerous mechanisms which act at different network scales [22]. At a macroscopic level the large diversity in neuron types – particularly the heterogeneity of interneurons – plays a vital role in activity control [23]. Similarly, activity can be controlled at a microscopic level, e.g. by altering the strengths of synapses via homeostatic synaptic plasticity [11] or by varying internal neuron parameters which have an influence on the excitability of the neuron [12]. In the following we focus on these locally operating processes.

A stable network operation constitutes itself in proper levels of network activity. Stability hence could be a consequence of activity control. In fact, studies in neuroscience provide compelling evidence for activity regulation at the level of individual neurons. For example it has been shown that neurons compensate for ongoing changes in input strength [24]. In these experiments, neuron cultures are typically placed in pharmacological substances like tetrodotoxin (TTX) which deprives the activity of the respective neurons. When this blockade is released, neurons exhibit significantly increased firing rates compared to control values. Reverse activity patterns are obtained when inhibitive inputs are blocked. However, even if an input blockade persists, cell activity gradually returns to the control level again. It could be shown that this activity regulation depends on synaptic scaling [25] as well as on altering the function relating current to firing rate [26].

Synaptic scaling, i.e. the scaling of synapses of afferent projections to a neuron, is mediated by the activity-dependent release of the neurotrophin BDNF (brain-derived neurotrophic factor) [27]. This has two important implications: Firstly, the activity of (postsynaptic) inhibitory interneurons is regulated based on the activity of the (presynaptic) excitatory cells which release the BDNF [28]. Secondly, synaptic scaling changes the ratio between excitation and inhibition within the network. This is due to the opposite effects that BDNF has on the scaling of excitatory synapses on pyramidal neurons and interneurons, respectively [29]. In other words, a high BDNF level weakens synapses on excitatory neurons, but strengthens those on inhibitory neurons and vice versa.

The transfer function of a neuron describes a dynamic range of input strengths to which a neuron is sensitive to. Learning-induced connectivity changes can easily results in inputs that do not match this dynamic range. For example inputs that are too weak, such that a neuron will not fire, or inputs that are too strong, such that firing saturates. That is why an adjustment of the transfer function – so called intrinsic plasticity – is reasonable as it shifts the sensitive region such that it matches the average input level [21]. It is further known that this kind of self-regulation is also effected by the release of BDNF [30].

B. Modeling Homeostatic Plasticity

In the following we describe how we implemented a biologically inspired dynamic self-regulation in detail. Due to the activity-dependent nature of homeostasis, we first estimate the average activity level $\bar{A}_i$ of a neuron i via an integration of instantaneous activities:

$$\bar{A}_i(k) = (1 - \frac{1}{\tau_H}) \cdot \bar{A}_i(k-1) + \frac{1}{\tau_H} \cdot A_i(k) \qquad (7)$$

Here, k is a discrete time index, $A_i(k) = f(u_i(k))$ the instantaneous activity, and τ_H defines the time scale on which integration takes place. $\bar{A}_i$ consequently can be related to intracellular calcium concentrations as they provide a correlate of a neuron's firing statistic [31].

Next, we model the BDNF release of an excitatory unit i (E–cell) given its mean activity $\bar{A}_i^E$ and a target rate $\widehat{A}$ as

$$BDNF_i^E(k) = 1 + \beta_H \left(\frac{\bar{A}_i^E(k) - \widehat{A}}{\widehat{A}} \right), \qquad (8)$$

where β_H is a homeostatic learning rate. If an E-cell's mean activity exceeds its target level, the cell's release of BDNF will be greater than 1. Conversely, the BDNF value is smaller than 1, when the cell is less active than the target level.

For the case of synaptic scaling in Kohonen-type SOMs, DeSieno previously suggested an additive scaling factor based on a neurons mean firing rate [32]. It is, however, known that multiplicative synaptic scaling is performed in the central nervous system [25]. This has the computationally attractive feature of leaving the relative difference in synaptic weights unchanged. A multiplicative scaling factor for SOMs, which is similar to our modeled BDNF level, has been suggested in [33]. However, even though our model uses the same factor for the scaling of connection weights, the way in which synaptic weights are adjusted differs fundamentally. Firstly, our learning regime combines Hebbian plasticity in form of Oja's rule with a BDNF-mediated scaling. Secondly, we further take the opposite effects of BDNF on the connections to excitatory and inhibitory cells into account. In summary, our model uses the following weight update equations:

$$w_{ij}^{EXT}(k) = \frac{w_{ij}^{EXT}(k-1) + \alpha \cdot \Delta\widetilde{w}_{ij}^{EXT}(k)}{BDNF_i^E(k) \cdot BDNF_j^{EXT}(k)} \qquad (9)$$

$$w_{ij}^{EE}(k) = \frac{w_{ij}^{EE}(k-1) + \alpha \cdot \Delta\widetilde{w}_{ij}^{EE}(k)}{BDNF_i^E(k) \cdot BDNF_j^E(k)} \qquad (10)$$

$$w_{ij}^{EI}(k) = \left[w_{ij}^{EI}(k-1) + \alpha \cdot \Delta\widetilde{w}_{ij}^{EI}(k)\right] \cdot BDNF_i^E(k) \qquad (11)$$

$$w_{ij}^{IE}(k) = \left[w_{ij}^{IE}(k-1) + \alpha \cdot \Delta\widetilde{w}_{ij}^{IE}(k)\right] \cdot BDNF_j^E(k) \qquad (12)$$

Here, α denotes a learning rate and $\Delta\widetilde{w}_{ij}^{*}(k)$ the weight change according to Eq. (6).

In addition to synaptic scaling we model homeostatic intrinsic plasticity by altering the transfer functions of individual excitatory units. Given a sigmoidal transfer function f according to Eq. (4), a neuron's intrinsic excitability can be changed by dynamically adjusting the gain and threshold parameter γ and θ, respectively. In a recent work, Triesch [14] derived an update formula for both parameters based on information theory. The difference to the mechanism applied by our model is twofold: Firstly, we restrict adaptation to the threshold parameter θ and, secondly, we express the rate of

adaptation in terms of the released BDNF level:

$$\begin{aligned}\theta_i^E(k) &= \theta_i^E(k-1) + \left(BDNF_i^E(k) - 1\right) \\ &= \theta_i^E(k-1) + \beta_H \cdot \left(\frac{\bar{A}_i^E(k) - \widehat{A}}{\widehat{A}}\right). \end{aligned} \quad (13)$$

Homeostatic plasticity, as it is incorporated within our network model, consequently can be summarized as follows. If an excitatory neuron's average activity level exceeds its control level, the neuron releases a lot of BDNF. In turn, BDNF mediates a downscaling of synaptic weights of excitatory connections to the neuron, whereas those of inhibitory ones are upscaled. The high level of BDNF additionally triggers a decrease in the intrinsic excitability of the neuron by increasing the threshold value of its transfer function. The reverse is true when a neuron's activity level lies below its target level.

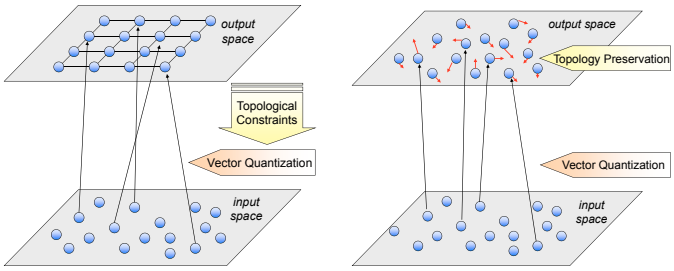
IV. Developing Topology Preserving Mappings

Dynamic neural fields, in contrast to self-organizing maps (SOMs) like the popular Kohonen maps [10], are recurrent networks in which activity is dynamically propagated. Even though we consider dynamic neural fields advantageous over Kohonen maps, we will discuss the issue of topology preservation also with respect to Kohonen maps. This is because our method for topology preservation is not limited to our network model; it rather can be applied to Kohonen maps, too.

When training SOMs two goals are pursued simultaneously. Firstly, SOMs perform vector quantization of the input space. They consequently strive for a minimization of the quantization error. Secondly, they incorporate topological constraints into the vector quantization process in order to develop topology preserving mappings (see Fig. 2 (a)). These constraints are defined in terms of fixed neighborhood relations between the map units. Unfortunately, when mapping higher-dimensional data onto the two-dimensional output space the two objectives most often can not be simultaneously satisfied. In such cases, SOMs privilege the minimization of the quantization error at the cost of an increase in the number of topological defects within the developed mappings.

A. Existing Approaches

Over the past, various techniques for enhancing topology preservation during map formation have been proposed. These methods most often rely on fixed neighborhood relations between map units, but adjust the width of an active neighborhood over the course of training. This dynamic adjustment is often based on global heuristics such as a gradual decrease in the size of the active neighborhood. Alternatively, more sophisticated local measurements like input novelty [34], topology defects [35], or the degree of local folding [36] can be used. Only a few approaches do not rely on fixed neighborhood relation. These methods rather apply a two-stage process in which vector quantization is performed first. The result of vector quantization is subsequently used to construct neighborhood relations. One example is the building of tree-like neighborhoods via a hierarchical clustering of the codebook vectors [37].



(a) (b)

Fig. 2. An illustrative comparison between (a) the conventional SOM learning algorithm and (b) the proposed system for enhancing topology preservation in SOMs.

B. Topology Preservation via Wiring Length Minimization

The technique we propose is related to this two-stage method in different respects. The first one is the release (or at least a relaxation) of the topological constraints from the process of vector quantization. As already discussed in section II this is due to the fact that our network does not rely on a fixed lateral connectivity, but rather features plastic within-field connections. Similarly to the two-stage model we thus propose to incorporate an additional process which is specifically concerned with enhancing topology preservation. As it is illustrated in Fig. 2 (b) this means that the objective function of minimizing topological defects is no longer implicitly defined via topological constraints, but rather explicitly by a process running in parallel to vector quantization.

The key difference to the two-stage model is how this process enhances topology preservation. Here, we suggest that it changes the positions of the units in the output space. We consequently consider model neurons not to be distributed on a fixed two-dimensional grid, but rather allow them to move on the cortical plane such that they get close to other neurons with similar receptive fields.

In the following we assume neurons to be fully laterally connected, i.e. each neuron features connections to all other neurons of the SOM. Furthermore, we define the connection weight w_{ij} between two units i and j to be proportional to the similarity between the receptive fields (codebook vectors) RF_i and RF_j of the units, e.g. by

$$w_{ij} \quad \propto \quad \frac{1}{\parallel RF_i - RF_j \parallel_2}. \quad (14)$$

Let $d_{ij} = \parallel x_i - x_j \parallel_2$ denote the distance between two units i and j. Then we suggest to adjust the position of a unit i based on the local objective of minimizing the unit's weighted wiring length WL_i to other units of the SOM.

$$WL_i = \sum_j w_{ij} \cdot d_{ij}^2 - \lambda \cdot \sum_j \ln(d_{ij}) \quad \longrightarrow \quad \min \quad (15)$$

Here, we include an additional penalization term (weighted by a factor λ) which prevents units to coincide at similar locations. Since the minimization of the distance between units with large connection weights produces a map layout where nearby units have similar receptive fields, wiring length minimization enhances topology preservation.

The movement of neurons on the cortical plane does not seem to be biologically plausible at a first glance. Even

though neurogenesis [38] – the process of continuous neuron creation from brain stem cells and their subsequent migration to target areas – describes neuron movements in the cortex, it is questionable whether neurogenesis can alter the layout of whole maps. Wiring length minimization, however, seems to be a biological principle. More precisely, it has been shown that functional brain areas as well as neuron populations within functional areas are positioned in an optimal way with respect to the achieved wiring length [39], [40]. Furthermore, a link between neuronal morphology and wiring length has been established [41]. Whether neurogenesis or structural plasticity arising from the outgrowth of axons and dendrites constitute the biological underpinning of an optimal wiring length remains an open question. Therefore, we consider our framework as a reasonable abstraction of the real biological mechanisms.

C. Implementation of Wiring Length Minimization

To minimize Eq. (15) the unit positions can be adapted using multiple optimization techniques, e.g. gradient descent or evolutionary algorithms [42]. Here we apply the gradient descent method insofar as the position of unit i is updated according to $\Delta x_i = -\gamma \cdot \partial WL_i / \partial x_i$ with

$$-\frac{\partial WL_i}{\partial x_i} = \sum_j 2 w_{ij} d_{ij} \cdot \frac{x_j - x_i}{d_{ij}} - \sum_j \frac{2\lambda}{d_{ij}} \cdot \frac{x_j - x_i}{d_{ij}}. \quad (16)$$

This formula illustrates that the map can be interpreted as an elastic network in which units exert forces on each other (see Fig. 3). Firstly, the lateral connections act like springs with spring constants chosen proportional to the connection weights w_{ij}. A connection between two units consequently exerts an attraction force F^+ on the units. Thereby, F^+ gets stronger when the connection weight w_{ij} or the distance d_{ij} between the units increases. Secondly, repulsion forces F^- act between the units. These forces are independent of the connection strengths. They rather solely depend on the distance d_{ij} between the units, i.e. F^- gets stronger when the distance decreases.

When wiring length minimization is applied to our network model described in section II, the weight values of the learned lateral connections can directly be used as connection weights w_{ij}. This is possible, because lateral connections learned via Hebbian plasticity constitute a measure for the similarity between the receptive fields of different neurons. We consequently obtain the following local objectives, where (17) and (18) hold for an excitatory unit i and an inhibitory unit i, respectively.

$$WL_i^E = \sum_{j \in E} w_{ij}^{EE} d_{ij}^2 + \sum_{j \in I} (w_{ij}^{EI} + w_{ji}^{IE}) d_{ij}^2 - \lambda \cdot \sum_{j \in E} \ln(d_{ij}) \quad (17)$$

$$WL_i^I = \sum_{j \in E} (w_{ij}^{IE} + w_{ji}^{EI}) d_{ij}^2 - \lambda \cdot \sum_{j \in I} \ln(d_{ij}) \quad (18)$$

Here, it is important to note that our model does not incorporate repulsion forces between excitatory and inhibitory neurons. This is because both neuron types are considered to be placed on separate layers of the neural map (cf. Fig. 1).

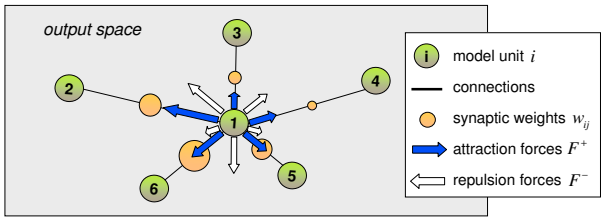


Fig. 3. The attraction and repulsion forces exerted on model units depend on the strengths of the connections as well as the distances between the units.

V. Results

We performed a series of experiments to evaluate our recurrent neural network model. In the following we first present results for a simulation where we applied the network to learn an associative mapping between artificially created multi-modal inputs. We further used the same kind of experiment to investigate how changes in stimuli strength or stimuli distribution affect neural field formation. Next, we estimated the influence of different parameter settings, i.e. different target firing rates, with respect to the developed mappings and, finally, we assessed the use of wiring length minimization for developing topology preserving mappings. The latter will be done both in the context of multi-modal association learning as well as for developing phoneme representations from continuous speech. In the following sections we provide details on the obtained results.

A. Learning Associations between Multi-modal Inputs

Due to their competitive processing regime, dynamic neural fields are particularly suited to learn multi-modal associations. Here we applied our computational model to associative learning in the domain of reference frame transformation, which is a particularly important issue for robotic applications involving eye-hand coordination. Artificial agents as well as animals have to be able to flexibly transform between different frames of reference, such as body-, head-, or eye-centered coordinates [43]. Agents consequently have to be equipped with an intermodal body calibration scheme which can either be innately given to the system or, more importantly, be autonomously acquired in the early stages of development [44]. For the latter the key aspect is that simultaneously present stimuli become associated in unified representations which can later be used for the transformation from one modality into another [45].

For modeling the body calibration process we restricted ourselves to a one-dimensional eye-hand coordination task. Thereby, a simulated agent performs random hand and eye movements, i.e. target gaze and hand positions are randomly chosen whereas a linear dynamic model produces smooth transitions between subsequent target positions. This kind of behavior emulates the self-exploratory actions that can be observed in early infancy. The agent further perceives

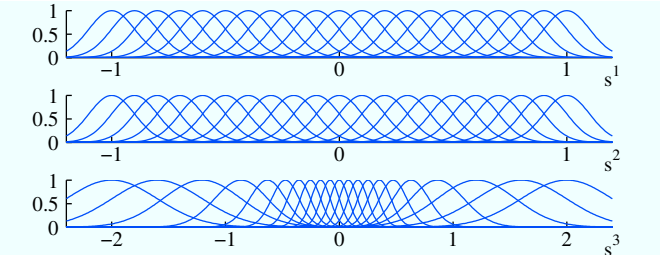


Fig. 4. The receptive fields of the neurons used for coding the gaze position in body-centered coordinates (s^1), the hand position in body-centered coordinates (s^2), and the hand-position in an eye-centered reference frame (s^3).

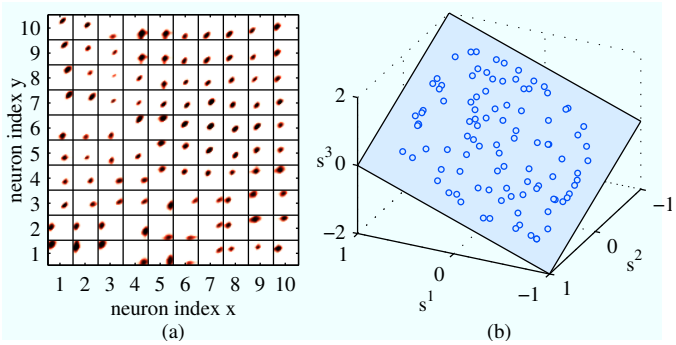


Fig. 5. (a) The receptive fields of all excitatory units are shown. Here, each subimage corresponds to the response pattern of a particular neuron to different combinations of s^1 (x-coordinate) and s^2 (y-coordinate). Dark colors represent strong neuron activity whereas light colors correspond to weak responses. (b) The positions of the excitatory neurons in the feature space as obtained by calculating the center of masses of their receptive fields.

its resulting gaze and hand positions in different reference frames. In the experiment we use three stimuli s^1,s^2,s^3 with $s^1, s^2 \in [-1, 1]$ and s^3= s^1– s^2, where s^1 and s^2 mimic the gaze and hand position in a body-centered reference frame, respectively, as well as s^3 representing the hand position in eye-centered coordinates. A specific body state consequently yields stimuli that produce individually ambiguous activities in each input modality. Their combination, however, provides a unique description of the body state. In our setup each of the stimuli is represented by a population of 21 neurons with partly overlapping Gaussian-shaped receptive fields (see Fig. 4). For s^1 and s^2 the receptive fields have a standard deviation of 0.1 and their centers were uniformly placed in the interval $[-1, 1]$. The centers of the receptive fields for s^3 have been uniformly sampled from the inverse of the cumulative density function of the normal distribution with standard deviation 0.4. Thereby, the receptive fields have a standard deviation that is half the distance to their nearest neighbor.

To learn associations between the different modalities we use the following system setup: Our recurrent neural network is composed of 100 excitatory units and 100 inhibitory units, both arranged on a 10x10 grid. The lateral connections within the field are initialized with uniform weight values, whereas the weights of afferent projections to the field are initialized with small random values. The time constants of the model are set to $\tau_E = \tau_I = 5$ and $\tau_H = 10^4$. This large homeostatic time constant τ_H ensures that average firing rates are based on a long time interval and not affected by moment-to-moment fluctuations in activity. We further use learning rates of $\alpha = 10^{-3}$ as well as $\beta_H = 10^{-4}$, i.e. Hebbian plasticity is faster than homeostatic plasticity. Finally, we apply a target activity level of $\widehat{A} = 0.05$. In the present experiment we do not perform wiring length minimization between the model units and learning is carried out at each time step, i.e. we do not make any assumption on when learning takes place. This is in contrast to the LISSOM models [9], [13], which also feature plastic lateral connections. There the synaptic weights were changed only after the network settled into a stable state following stimuli presentation.

When applying the network to the sequentially arising stimuli, different phases can be observed over the course of development. Initially the model units cooperate via lateral excitation such that the whole field grossly adapts to the input pattern distribution. However, afterwards an increased lateral inhibition implements a competition between the model units. More precisely, the excitatory units compete for the representation of different input patterns. This competitive learning facilitates a diversification of the units and results in a specialization of the units to distinct input patterns. After several input patterns have been presented, we fixed the network weights and calculated the receptive fields of the excitatory units. Therefore, we applied different stimuli combinations and recorded the units' activities after the field activity settled into a stable state. The resulting receptive fields are depicted in Fig. 5 (a) where we plot the response pattern of each excitatory unit to different combinations of s^1 and s^2. As can be seen, each neuron specializes to a particular combination of the stimuli. We further calculated the center of masses of the receptive fields. By doing so we obtain the positions of the neurons in the feature spaces. Fig. 5 (b) illustrates that the neurons cover the whole feature space (the s^1-s^2-s^3-plane), i.e. each input pattern is adequately represented by the neural field.

Lastly, we investigated whether the incorporated homeostatic mechanisms drive the individual neurons towards some target firing rate. Therefore, we recorded how the average activity levels of all excitatory neurons develop over time. This has been done for two simulations using target firing rates of $\widehat{A} = 0.05$ and $\widehat{A} = 0.1$, respectively. Fig. 6 plots the medians of the resulting activity levels. The regions around the medians depict the upper and lower quartiles of the activity level distributions. The plot illustrates that the neurons' average activities quickly raise towards the specified target firing rates. Additionally, we could previously show that the overall activity within the field approaches a level which is proportional to the target firing rate [15]. In summary this shows that the applied locally operating mechanisms are suited not only to regulate an individual neurons activity, but also to regulate the activity within a population of neurons.

B. *Effect of Changes in Stimuli Strength*

Once neurons are equipped with the ability to regulate their activity, the question arises whether the same homeostatic

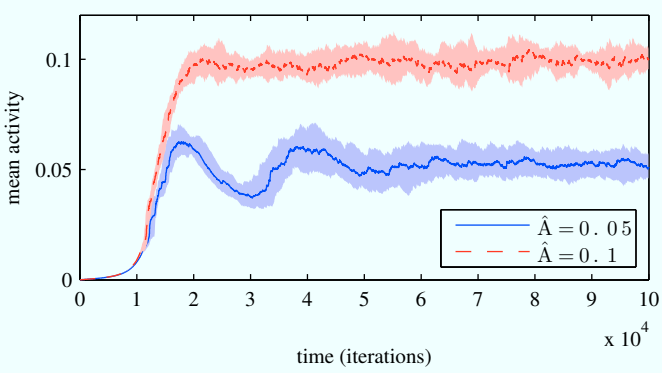


Fig. 6. The median of the average activity levels of all excitatory neurons is plotted for two simulations using a target firing rate of $\widehat{A} = 0.05$ and $\widehat{A} = 0.1$, respectively. Regions around the medians depict the upper and lower quartiles of the respective activity level distributions.

processes are suited to adapt the dynamic neural field to long-lasting changes in the input stimuli. This includes changes in stimuli strength as well as changes in the input pattern distribution. Here, we first concentrate on the former aspect. The latter will be investigated in the next section.

To test the ability of our network to adapt to changing input strengths we simulated a biological experiment in which the input to a neuron is blocked [24] (see section III-A for a description of the experiment). More precisely, we modeled the blockade of excitatory inputs by an attenuation of the input amplitude to 20% compared to normal operation. After a while this blockade is released again. We recorded the evolution of the BDNF level $BDNF_1^E$ as well as the transfer function threshold θ_1^E of the respective neuron. The corresponding plots are shown in Fig. 7. Here, time $t = 0$ denotes the onset of stimulus depression, whereas at time $t = 100$ the blockade is released. At both times, we further recorded the responses of the neuron to a specific input pattern, once using normal inputs strength and once using an input strength depressed to 20% of normal operation. The inset plots at time $t = 0$ show that the presentation of the input pattern without blockade produces a stable and large response of the neuron. In contrast, the depressed input pattern is too weak to produce a significant increase in the neuron's membrane potential such that the neuron remains inactive. Due to this behavior, the neuron's mean activity level will decrease after the onset of blockade at $t = 0$ (not shown). As a result, the neuron compensates for this change in a similar way as biological neurons do: It decreases its BDNF level, which results in an upscaling of excitatory, but a downscaling of inhibitory synapses. It further decreases its threshold and thus changes its transfer function towards higher excitability. The result of this regulation is depicted by the inset plots at time $t = 100$. The depressed input pattern now induces the same response of the neuron as the normal pattern did at time $t = 0$. However, if we release the blockade, i.e. present the undepressed input pattern, then the neuron shows a much larger sensitivity to the pattern. This reflects itself in the significantly increased membrane potential evoked by the input and consequently a faster and prolonged response of the

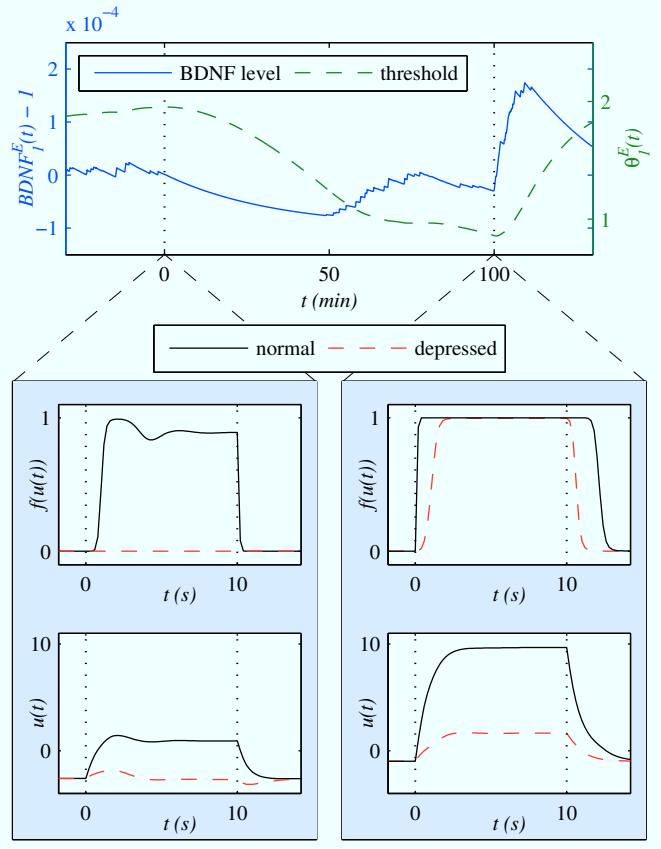


Fig. 7. The evolution of a neuron's BDNF level and transfer function threshold following an input blockade at $t = 0$ as well as a release of the blockade at $t = 100$. The blockade has been modeled by an attenuation of input strengths to 20% of normal operation. The insets shows the response of the neuron to the normal as well as the depressed input pattern when they are presented at times $t = 0$ and $t = 100$, respectively.

neuron.

In summary these results show that the homeostatic mechanisms enable individual neurons to compensate for changes in the strengths of their inputs. From a computational point of view this is advantageous as it allows the network to cope with inputs which may continuously change their amplitude over a long time scale.

C. Effect of Changes in the Input Pattern Distributions

Next, we investigated whether our network model is able to adapt an already developed mapping to a changed input pattern distribution. Biological neurons can cope with such changes [46]. They are even able to adapt to sudden and significant changes such as those following the amputation of a limb [47]. For example it has been shown that digit amputation in raccoon forces affected neurons in primary somato-sensory cortex to reorganize their receptive fields [48]. More precisely, neurons that become silent after amputation (due to missing input) subsequently expand their receptive field to large regions of adjacent digits or the palm and finally shrink them again such that the neuron becomes selectively responsive to inputs stemming from the new receptive field. To test our computational model we perform an experiment of

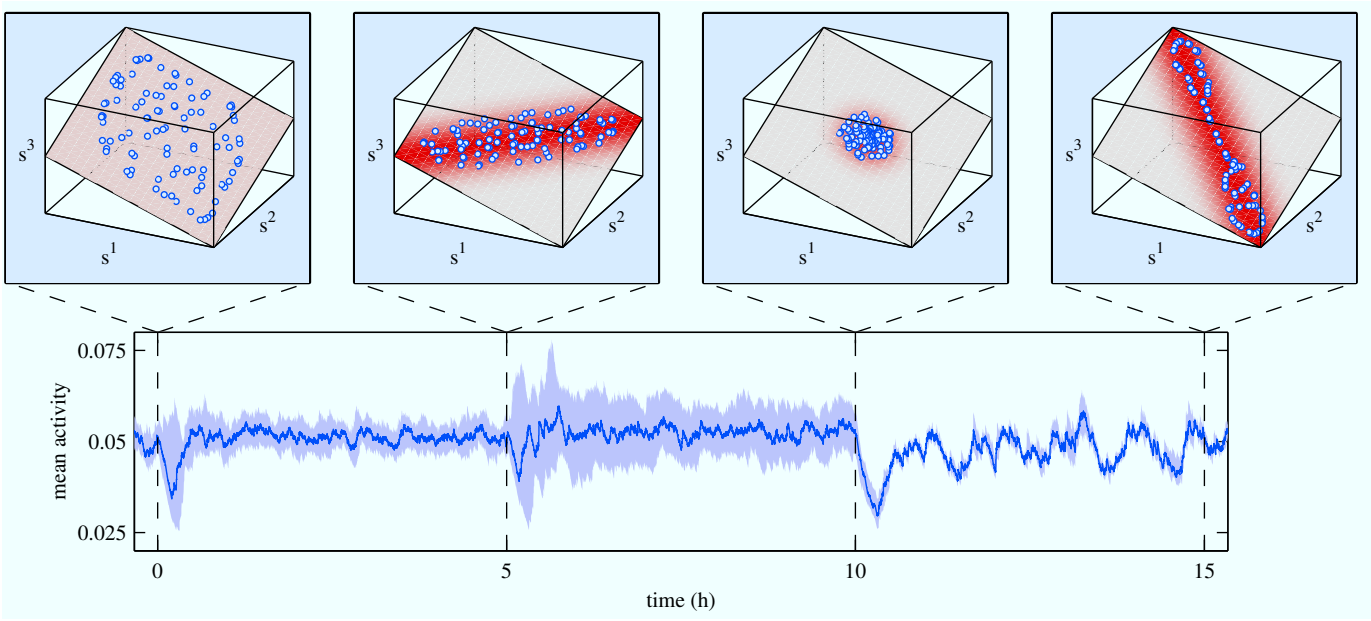


Fig. 8. The initial uniform distribution from which stimuli are sampled is repeatedly changed at time steps $t = 0$, $t = 5$, and $t = 10$. The insets depict the distribution of the developed receptive fields overlaid to the previously applied sampling distribution. The bottom panel shows how the average activity levels of the neurons develop over time. Therefore, the median as well as the upper and lower quartiles of the activity level distribution are plotted.

a similar kind. Therefore, we first let the network develop a mapping, but subsequently we significantly change the input pattern distribution at regular time intervals.

The results of this simulation are depicted in Fig. 8. Here, the network initially has been trained with uniformly sampled input stimuli, i.e. $s^1 \sim \mathcal{U}(-1,+1)$, $s^2 \sim \mathcal{U}(-1,+1)$, and $s^3 = s^1 - s^2$. As shown in the inset at time $t = 0$, the network developed a mapping where the receptive fields of the excitatory neurons are nicely distributed in the input space. At time $t = 0$ we applied the first change in the input pattern distribution. At subsequent time steps, stimuli with $s^3 \sim \mathcal{N}(0, 0.09)$ have been sampled and presented to the network. Neurons, which previously developed a receptive field corresponding to large absolute values of s^3, consequently do not receive inputs anymore. In contrast, those neurons, with receptive fields already lying close to $s^3 = 0$, now become activated by more input patterns than before the change. This is reflected in the neurons' average activity level distribution which is plotted at the bottom panel of Fig. 8. Since most neurons do not become activated anymore, the median as well as the lower quartile of the distribution decrease during the time steps following $t = 0$. However, the change in the input pattern distribution also increases the average activity level of some neurons, such that the upper quartile of the distribution rises. The homeostatic processes consequently try to compensate for these changes in order to maintain stable activity patterns. More precisely, the less active neurons increase their sensitivity such that they become active for other stimuli as well. In other words, they expand their receptive fields and step in competition for responsibility in representing those other stimuli. The subsequent competition between neurons lets the receptive fields shrink again such that the neurons become selectively responsive to the new stimuli. As a result of this adaptation process, the average activity level of all neurons approaches the target level again. The inset at time $t = 5$ shows the distribution of the new receptive fields which nicely resembles the applied input pattern distribution.

In the following, we apply two more changes to the input pattern distribution: The first change occurs at time $t = 5$, where we start to sample input patterns according to $s^1 \sim \mathcal{N}(0, 0.04)$ and $s^2 \sim \mathcal{N}(0, 0.04)$. The second change is applied at time $t = 10$, from where on input patterns get sampled according to $(s^2 - s^1) \sim \mathcal{N}(0, 0.09)$. Both disturbances are visible as changed activity levels of the neurons. The homeostatic processes consequently force the neurons to reorganize their receptive fields. The insets at time $t = 10$ and $t = 15$ show that these reorganizations develop receptive fields whose distributions resemble the applied input pattern distributions.

These results illustrate the ability of the network not only to cope with changes in stimuli strength, but also to adapt to changes in the stimuli distribution. The latter is particularly interesting for technical applications where stimuli may suddenly change, e.g. due to failure of individual system components. In such cases, neurons, which specialized for inputs stemming from failed components, can develop new receptive fields and consequently become responsive to outputs of intact system devices.

D. Influence of the Target Activity Level Parameter

In the following, we evaluate the influence of different parameter settings. Due to the incorporation of homeostatic processes, the number of free parameters that have to be controlled reduces to the time constants as well as the target activity of individual neurons. We already discussed the time scale at which homeostatic processes have to operate. More

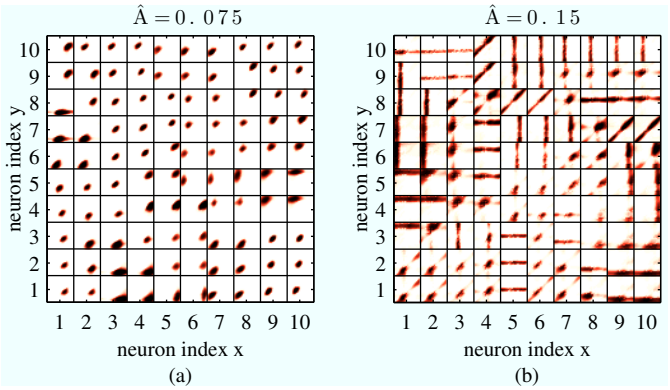


Fig. 9. The receptive fields of all excitatory units are shown for two simulation runs: using a target activity of (a) $\widehat{A} = 0.075$ and (b) $\widehat{A} = 0.15$. Here, each subimage corresponds to the response pattern of a particular neuron to different combinations of s^1 (x-coordinate) and s^2 (y-coordinate). Dark colors represent strong neuron activity whereas light colors correspond to weak responses.

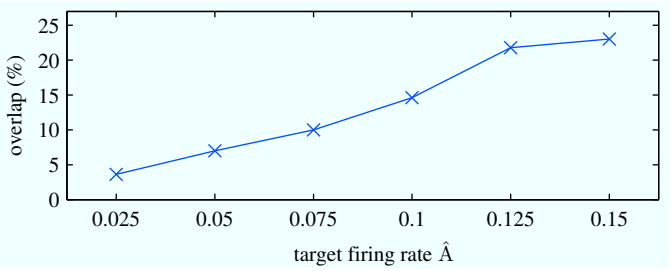


Fig. 10. The influence of the target firing rate parameter on the overlap between the developed receptive fields is depicted.

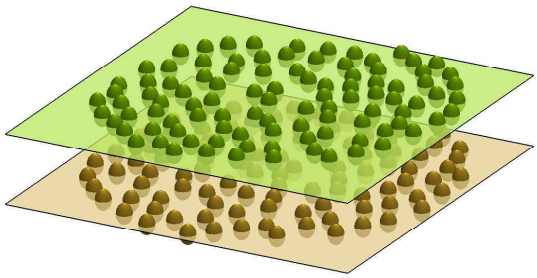


Fig. 11. The final layout of the model units as obtained when using the wiring length minimization process during neural field formation.

precisely, these processes have to be fast enough to compensate for long-lasting activity changes induced by Hebbian learning. However, they additionally have to be slow enough in order not to destroy the moment to moment fluctuations which carry the input signal information. What remains is an evaluation on the influence of the target activity parameter.

As shown in section V-A, the target firing rate determines the average activity levels of individual neurons as well as the overall activity within the field. We consequently hypothesized that the target firing rate has an influence on the size of the developed receptive fields. In this case the parameter would further have an effect on the overlap between the receptive fields of individual neurons and therewith the sparsity of the developed representation. To validate this hypothesis we performed multiple simulations using different target activity levels. The developed receptive fields for two of these simulations are exemplarily shown in Fig. 9. In (a) we see that at a target activity level of $\widehat{A} = 0.075$ neurons specialize to specific combinations of the three stimuli. In (b) we see that an increase in the target firing rate to $\widehat{A} = 0.15$ yields significantly larger receptive fields. To reach the target activity level individual neurons specialize to single input modalities. This is shown by the horizontal (s^1), vertical (s^2), or diagonal (s^3) response patterns of the neurons. We further calculated the overlap between the receptive fields for each parameter setting, respectively. The corresponding result is plotted in Fig. 10. Here, we observe a steady increase in the overlap when the target activity is increased. These results validate our initial hypothesis on the influence of the target firing rate parameter.

E. Topology Preservation

We next discuss the additional incorporation of the wiring length minimization (WLM) process. As described in section IV the process adapts the neuron positions such that neurons with strong lateral connections become adjacent to each other. Since a lateral connection between model units only features a large synaptic weight when the neurons' receptive fields are similar to each other, WLM should facilitate the development of topology preserving mappings. When using WLM in the multi-modal association learning experiment, we finally obtain the spatial neuron layout depicted in Fig. 11.

To estimate the effect of WLM on the topology preserving properties of the developed mapping, we compared the results of two simulation runs: one using WLM and one not using WLM. Here, we first investigate whether our interpretation of a neural field to be an elastic net (with lateral connections exerting attraction forces on units) is a suitable choice for minimizing wiring length. To do so, we calculated the total weighted wiring length (Eq. (15)) for both simulation runs. To compensate for different spatial scales we further normalized the distances between units by their mean. The evolution of the resulting measure is depicted in Fig. 12. The total wiring length of the simulation without WLM increases at the beginning. This can be attributed to an initial rough adjustment of the lateral connection weights. A competition between the model units subsequently induces a "die off" of many synapses, which let the total weighted wiring length decrease over time. When incorporating WLM we observe a similar trend, except that the initial increase in total wiring length vanishes. Most importantly, however, our implementation of WLM results in a smaller total weighted wiring length. This shows that an adaptation of neuron positions based on forces exerted between units is appropriate for achieving WLM.

Given the ability of our model to reduce the weighted wiring length between units, we now demonstrate that WLM is suitable for improving topology preservation. Therefore, we compare the mappings which have been developed by the two simulation runs. To do so, we first calculate the neighborhood relations between the excitatory units using Delaunay triangulation of their positions after training. We additionally calculate the positions of the centers of the de-

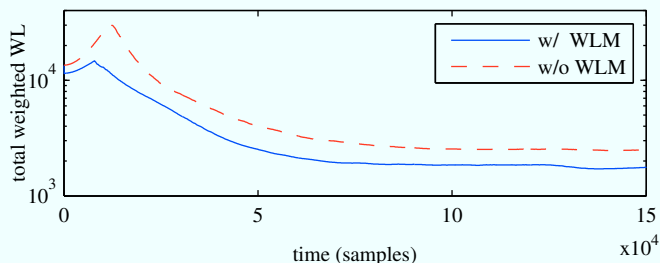


Fig. 12. The evolution of the normalized total weighted wiring length for a simulation using WLM as well as a simulation not using WLM.

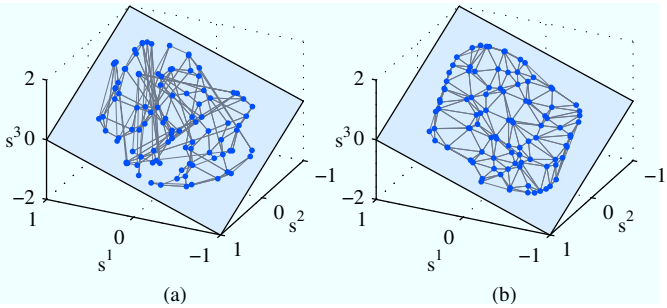


Fig. 13. The center positions of the developed receptive fields are plotted. Connections are drawn between those receptive fields, whose neurons are adjacently positioned in the output plane. (a) shows the result of a simulation where WLM has not been used, whereas the result depicted in (b) has been obtained using WLM.

veloped receptive fields. The resulting receptive field positions are plotted in Fig. 13, where we overlaid connections between them according to the calculated neighborhood relations. For a topology preserving mapping this would result in a plot where neighboring receptive fields are connected (due to the adjacent positions of their corresponding neurons). As shown in Fig. 13 (b), this is the case for the neural field which has been trained using WLM. In contrast, not using WLM results in significant topological defects (see Fig. 13 (a)).

These qualitative results can be confirmed by a quantitative analysis using the topographic function [49]. This widely used measure characterizes the topology preservation of mappings by analyzing the degree of topological defects on varying scales: from local to global ones. The results are plotted in Fig. 14. There, the normalized rank k determines the effective neighborhood range, i.e. small $|k|$ correspond to a local neighborhood, whereas large $|k|$ correspond to a global one. The results show that WLM decreases the number of topological defects on both a local scale and particularly on a global scale.

F. Application to Speech Processing

We finally apply our network model in the domain of speech processing. Therefore, we present results of simulations where the neural field has been trained using continuous speech input. The model consequently should develop a mapping where individual neurons specialize to specific sounds, i.e. phonemes. By incorporating the WLM process, the mapping should further maintain topology, i.e. similarly sounding phonemes should be mapped onto neighboring neurons.

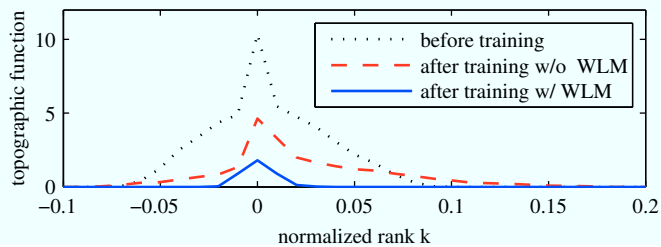


Fig. 14. The topographic function is plotted for neural fields which have been trained with WLM or without WLM.

Since human speech perception relies to a large extent on vocal tract resonance frequencies and their variation in time [50], we decided to use formant frequencies as input to our model. Thereby, formants refer to the energy concentrations in the spectro-temporal domain (which are correlates of the underlying vocal tract resonance frequencies). It has been shown that formant trajectories can be robustly extracted from continuous speech [51]. In the present experiment, however, we use the hand-labeled trajectories provided by the VTRFormant database [52]. As a subset of the widely used TIMIT corpus, this database comprises a total of 516 utterances from which we used all 322 utterances spoken by male speakers.

We use a population code of 128 neurons to represent the formant frequencies at each time frame. Thereby, the response patterns of the input neurons correspond to the transfer functions of a 128-channel Patterson-Holdsworth auditory filter bank [53]. This filter bank is based on neurophysiological findings on the human auditory system and models the peripheral processing as carried out by the cochlea, where sound is transformed into spatio-temporal response patterns on the auditory nerve. In our setup the filter bank is composed of Gammatone filters whose logarithmically arranged center frequencies cover the range from 80 Hz to 8 kHz. An exemplarily selected speech utterance is shown in Fig. 15. There, (a) depicts the time-domain signal, (b) the corresponding formant trajectories coded by the population of input neurons, and (c) an example input pattern at the specific time frame marked in (b). We constructed the input samples using a sampling rate of 1 kHz. The remaining setup of our network model equals the one described in section V-A.

To illustrate the benefit of using WLM we once again trained two neural fields on the speech data. The positions of neurons in the first field have been fixed to a 10x10 grid, whereas neurons of the second field could change their positions using WLM. After training we analyzed the response of the fields to different input stimuli. More precisely, we calculated the mean formant frequencies for each phoneme transcribed in the VTRFormant database and recorded the neuron responses to the corresponding input patterns. Finally, this allowed us to label the neurons with the symbols of the phonemes which evoke the largest response of the neuron. In Fig. 16 we plot the final layout of the excitatory neurons as well as their labels. For ease of interpretation we restricted the labels in the plot to vowels and semivowels. The results depicted in (b) illustrate that the incorporation of WLM produces a topology preserving mapping where neurons with

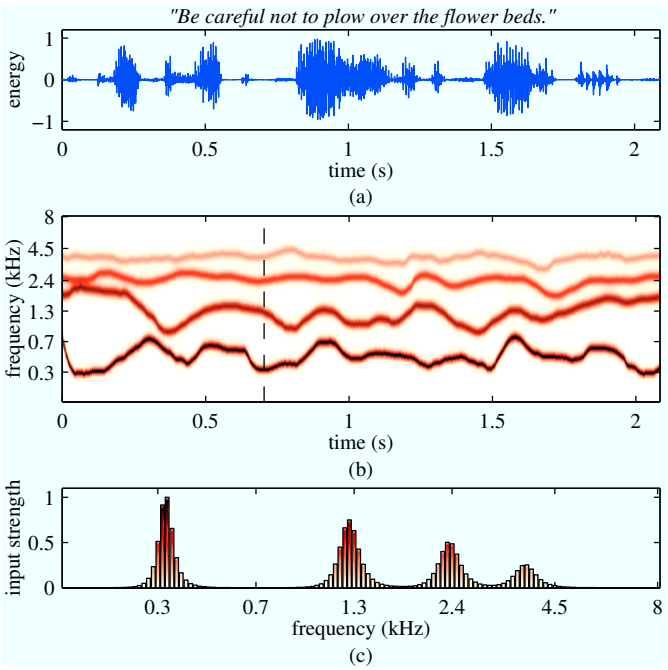


Fig. 15. In (a) the time-domain signal for a speech utterance is shown. The corresponding formant trajectories are depicted in (b), whereas (c) shows the population code of the formant frequencies at a specific time frame.

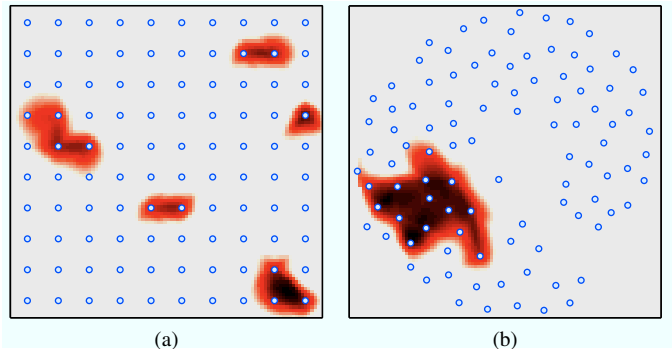


Fig. 17. The field activity as evoked by the phoneme "i". The field depicted in (a) has been trained without WLM, whereas the field in (b) has been trained with WLM.

similar labels cluster together. Furthermore, the labels are distributed over the map such that similar sounding phonemes are close to each other. In contrast, a training of the field without using WLM produces a map, where similar sounding phoneme symbols are widely distributed.

The difference in using WLM and not using WLM also becomes evident in the response patterns to single phonemes. This is illustrated in Fig. 17 where we plot the field activity following the presentation of an "i". More precisely, the field constructed without using WLM produces multiple loci of activity, whereas the use of WLM forces the formation of a map exhibiting single activity bubbles.

The development of topology preserving mappings is particularly advantageous for the processing of speech. This is because a continuously changing stimulus (e.g. formant trajectories) evokes a continuous activity trace in the field. This is illustrated in Fig. 18, where we plot the trace of neurons which exhibit the largest response to the word "money" [m-ʌ-n-i]. In (a) we see that a mapping with many topological defects does not produce a continuous activity trace, whereas the topology preserving mapping in (b) does. We further added jitter to the plot. This allows us to estimate the time course at which the activity bubble moves from one position to another. From the plot it becomes evident that the activity bubble remains at positions, which correspond to the phoneme cores, for a relatively long period of time. In contrast, a fast movement of the activity bubble is observed for transitions between phonemes.

A higher-level processing of speech, e.g. a phonetic transcription [54] or a speech synthesis [55], could be based on such traces of activity.

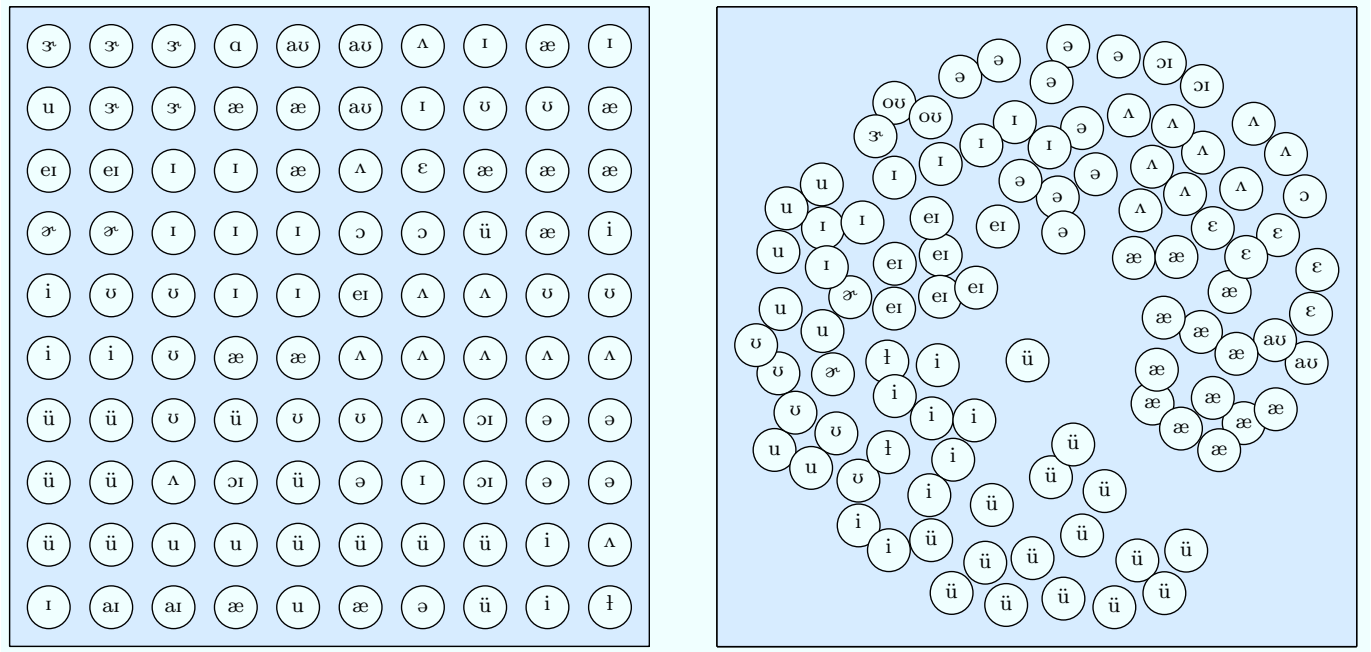
VI. Summary

The self-organized formation of representations – one of the key abilities of developing systems – has seldomly been discussed in the context of dynamic neural fields. The recurrent neural network presented in this paper thus constitutes an important advancement in this respect. In contrast to previously published approaches the network features a fully plastic connectivity. In particular, Hebbian learning is not only used to adapt the synaptic weights of afferent projections to the field, but also those of lateral connections within the field. To circumvent network degradation and maintain stable operation we proposed to incorporate biologically inspired homeostatic mechanisms, i.e. synaptic scaling and intrinsic plasticity. This locally operating self-regulation aims at changing the activity levels of individual neurons towards some target rate. Finally, our model consists of an additional process for the enhancement of topology preservation during map formation. This process adjusts the positions of neurons based on the objective of minimizing wiring length.

Due to the self-regulation applied within our network, the number of free parameters that have to be controlled reduces to the time constants and the target activity level of individual neurons. Here, particularly the latter is very interesting. More precisely, in our experiments we could confirm the hypothesis that the target activity level of individual neurons defines the overall activity in the field, the overlap between the receptive fields of neurons, and therewith also the sparsity of the developed representation.

The low number of controllable parameters further eases the use of the network in applications of various domains. Since the formation of appropriate input representations is self-organizing and intrinsically regulated, the network is particularly interesting for developing systems. This is because learning and adaptation during development causes a continuous change in system internals. It is thus feasible to use network models which develop representations solely based on the distribution of the applied input patterns. The proposed network provides a method for achieving this.

We evaluated the performance of our network model in different tasks. Firstly, artificially generated data has been used to address multi-modal association learning in the domain of reference frame transformation. Secondly, we investigated the



(a) (b)

Fig. 16. The positions of the excitatory units, when the field has been trained (a) without WLM and (b) with WLM. Phonemes (vowels and semivowels), to which the neurons exhibit the largest response, have been used as labels for the neurons. Phoneme labels are from the IPA phonetic alphabet.

development of phoneme representations following continuous speech input. Our results demonstrated that the network self-organizes without any external supervision, develops appropriate representations of the input, and even adapts to sudden changes in the strength or distributions of input patterns. Finally, we could show that wiring length minimization significantly enhances the quality of the developed mappings with respect to topology preservation.

REFERENCES

[1] S. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological Cybernetics*, vol. 27, pp. 77–87, 1977.

[2] S. Coombes, "Waves, bumps, and patterns in neural field theories." *Biological Cybernetics*, vol. 93, no. 2, pp. 91–108, 2005.

[3] H. R. Wilson and J. D. Cowan, "A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue." *Kybernetik*, vol. 13, no. 2, pp. 55–80, 1973.

[4] M. A. Giese, *Dynamic Neural Field Theory for Motion Perception*. Norwell, MA, USA: Kluwer Academic Publishers Group, 1998.

[5] A. R. Schutte, J. P. Spencer, and G. Schner, "Testing the dynamic field theory: working memory for locations becomes more spatially precise over development." *Child Development*, vol. 74, no. 5, pp. 1393–1417, 2003.

[6] W. Erlhagen and E. Bicho, "The dynamic neural field approach to cognitive robotics." *Journal of Neural Engineering*, vol. 3, no. 3, pp. R36–R54, 2006.

[7] J. G. Taylor, "Neural 'bubble' dynamics in two dimensions: foundations," *Biological Cybernetics*, vol. 80, pp. 393–409, 1999.

[8] I. Mikhailova and C. Goerick, "Conditions of activity bubble uniqueness in dynamic neural fields." *Biological Cybernetics*, vol. 92, no. 2, pp. 82–91, 2005.

[9] J. Sirosh and R. Miikkulainen, "Cooperative self-organization of afferent and lateral connections in cortical maps," *Biological Cybernetics*, vol. 71, pp. 65–78, 1994.

[10] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59–69, 1982.

[11] G. G. Turrigiano and S. B. Nelson, "Homeostatic plasticity in the developing nervous system." *Nature Reviews Neuroscience*, vol. 5, no. 2, pp. 97–107, 2004.

[12] W. Zhang and D. J. Linden, "The other side of the engram: experience-driven changes in neuronal intrinsic excitability." *Nature Reviews Neuroscience*, vol. 4, no. 11, pp. 885–900, 2003.

[13] J. Law, "Modeling the development of organization for orientation preference in primary visual cortex," Ph.D. dissertation, University of Edinburgh, 2009.

[14] J. Triesch, "Synergies between intrinsic and synaptic plasticity in individual model neurons," *Neural Computation*, vol. 19, no. 4, pp. 885–909, 2007.

[15] C. Gläser, F. Joublin, and C. Goerick, "Homeostatic development of dynamic neural fields," in *Proc. Int. Conf. on Development and Learning*, 2008, pp. 121–126.

[16] ——, "Enhancing topology preservation during neural field development via wiring length minimization," in *Artificial Neural Networks - ICANN 2008, Part I*, ser. LNCS, V. Kurkova, R. Neruda, and J. Koutnik, Eds. Springer Berlin / Heidelberg, 2008, vol. 5163, pp. 593–602.

[17] S. D. V. Hooser, J. A. F. Heimel, S. Chung, S. B. Nelson, and L. J. Toth, "Orientation selectivity without orientation maps in visual cortex of a highly visual mammal." *Journal of Neuroscience*, vol. 25, pp. 19–28, 2005.

[18] K. Ohki and R. C. Reid, "Specificity and randomness in the visual cortex." *Current Opinion in Neurobiology*, vol. 17, no. 4, pp. 401–407, 2007.

[19] E. Oja, "Simplified neuron model as a principal component analyzer," *Journal of Mathematical Biology*, vol. 15, no. 3, pp. 267–273, 1982.

[20] G. G. Turrigiano and S. B. Nelson, "Hebb and homeostasis in neuronal plasticity." *Current Opinion in Neurobiology*, vol. 10, no. 3, pp. 358–364, 2000.

[21] N. S. Desai, "Homeostatic plasticity in the CNS: synaptic and intrinsic forms." *Journal of Physiology – Paris*, vol. 97, no. 4-6, pp. 391–402, 2003.

[22] E. Marder and J.-M. Goaillard, "Variability, compensation and homeostasis in neuron and network function." *Nature Reviews Neuroscience*, vol. 7, no. 7, pp. 563–574, 2006.

[23] V. Santhakumar and I. Soltesz, "Plasticity of interneuronal species diversity and parameter variance in neurological diseases." *Trends in Neurosciences*, vol. 27, no. 8, pp. 504–510, 2004.

[24] E. Marder and A. A. Prinz, "Modeling stability in neuron and network function: the role of activity in homeostasis." *Bioessays*, vol. 24, no. 12, pp. 1145–1154, 2002.

[25] G. G. Turrigiano, K. R. Leslie, N. S. Desai, L. C. Rutherford, and S. B.

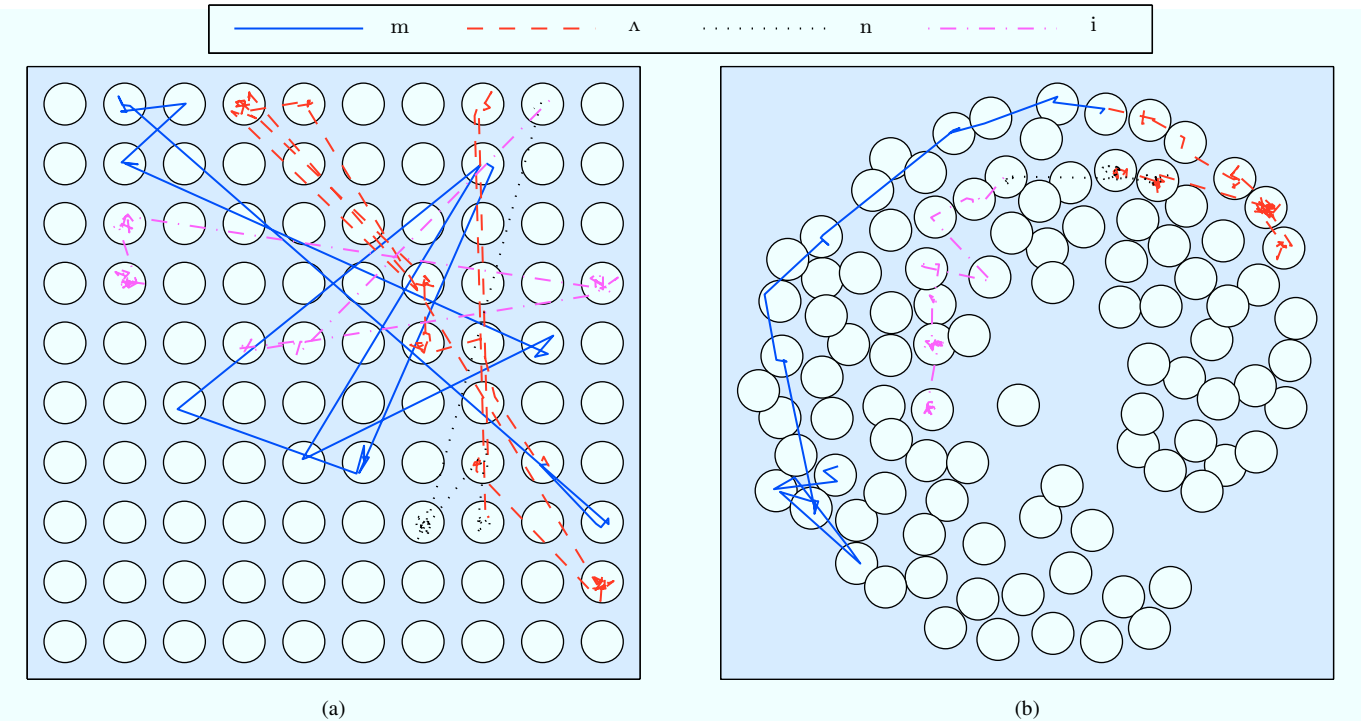


Fig. 18. The trace of neurons which exhibit the largest response to the word "money" [m-ʌ-n-i]. (a) shows a non-continuous trace for the map developed without WLM, whereas (b) depicts a continuous activity trace for the map developed with WLM.

Nelson, "Activity-dependent scaling of quantal amplitude in neocortical neurons." *Nature*, vol. 391, no. 6670, pp. 892–896, 1998.

[26] N. S. Desai, L. C. Rutherford, and G. G. Turrigiano, "Plasticity in the intrinsic excitability of cortical pyramidal neurons." *Nature Neuroscience*, vol. 2, no. 6, pp. 515–520, 1999.

[27] L. C. Rutherford, A. DeWan, H. M. Lauer, and G. G. Turrigiano, "Brain-derived neurotrophic factor mediates the activity-dependent regulation of inhibition in neocortical cultures." *Journal of Neuroscience*, vol. 17, no. 12, pp. 4527–4535, 1997.

[28] Z. Kokaia, J. Bengzon, M. Metsis, M. Kokaia, H. Persson, and O. Lindvall, "Coexpression of neurotrophins and their receptors in neurons of the central nervous system," *Proc. of the National Academy of Sciences USA*, vol. 90, no. 14, pp. 6711–6715, 1993.

[29] L. C. Rutherford, S. B. Nelson, and G. G. Turrigiano, "BDNF has opposite effects on the quantal amplitude of pyramidal neuron and interneuron excitatory synapses." *Neuron*, vol. 21, no. 3, pp. 521–530, 1998.

[30] N. S. Desai, L. C. Rutherford, and G. G. Turrigiano, "BDNF regulates the intrinsic excitability of cortical neurons." *Learning & Memory*, vol. 6, no. 3, pp. 284–291, 1999.

[31] M. J. Berridge, "Neuronal calcium signaling." *Neuron*, vol. 21, no. 1, pp. 13–26, 1998.

[32] D. DeSieno, "Adding a conscience to competitive learning," in *Proc. Int. Conf. on Neural Networks*, vol. 1, 1988, pp. 117–124.

[33] T. J. Sullivan and V. R. de Sa, "Homeostatic synaptic scaling in self-organizing maps." *Neural Networks*, vol. 19, no. 6-7, pp. 734–743, 2006.

[34] R. Phaf, P. Den Dulk, A. Tijsseling, and E. Lebert, "Novelty-dependent learning and topological mapping," *Connection Science*, vol. 13, no. 4, pp. 293–321, 2001.

[35] M. Herrmann, "Self-organizing feature maps with self-organizing neighborhood widths," in *Proc. Int. Conf. on Neural Networks*, 1995, pp. 2998–3003.

[36] K. Kiviluoto, "Topology preservation in self-organizing maps," in *Proc. Int. Conf. on Neural Networks*, vol. 1, 1996, pp. 294–299.

[37] J. S. Kirk and J. M. Zurada, "A two-stage algorithm for improved topography preservation in self-organizing maps," in *Proc. Int. Conf. on Systems, Man, and Cybernetics*, vol. 4, 2000, pp. 2527–2532.

[38] P.-M. Lledo, M. Alonso, and M. Grubb, "Adult neurogenesis and functional plasticity in neuronal circuits." *Nature Reviews Neuroscience*, vol. 7, no. 3, pp. 179–193, 2006.

[39] C. Cherniak, Z. Mokhtarzada, R. Rodriguez-Esteban, and K. Changizi, "Global optimization of cerebral cortex layout." *Proc. of the National Academy of Sciences USA*, vol. 101, no. 4, pp. 1081–1086, 2004.

[40] B. Chen, D. Hall, and D. Chklovskii, "Wiring optimization can relate neuronal structure and function." *Proc. of the National Academy of Sciences USA*, vol. 103, no. 12, pp. 4723–4728, 2006.

[41] D. Chklovskii, "Synaptic connectivity and neuronal morphology: two sides of the same coin." *Neuron*, vol. 43, no. 5, pp. 609–617, 2004.

[42] D. B. Fogel, "An introduction to simulated evolutionary optimization." *IEEE Trans. on Neural Networks*, vol. 5, no. 1, pp. 3–14, 1994.

[43] Y. E. Cohen and R. A. Andersen, "A common reference frame for movement plans in the posterior parietal cortex." *Nature Reviews Neuroscience*, vol. 3, no. 7, pp. 553–562, 2002.

[44] R. Morgan and P. Rochat, "Intermodal calibration of the body in early infancy," *Ecological Psychology*, vol. 9, no. 1, pp. 1–23, 1997.

[45] L. E. Bahrick and J. S. Watson, "Detection of intermodal proprioceptive-visual contingency as a potential basis of self-perception in infancy," *Developmental Psychology*, vol. 21, no. 6, pp. 963–973, 1985.

[46] F. Joublin, F. Spengler, S. Wacquant, and H. R. Dinse, "A columnar model of somatosensory reorganizational plasticity based on Hebbian and non-Hebbian learning rules." *Biological Cybernetics*, vol. 74, no. 3, pp. 275–286, 1996.

[47] P. W. Halligan, J. C. Marshall, D. T. Wade, J. Davey, and D. Morrison, "Thumb in cheek? Sensory reorganization and perceptual plasticity after limb amputation." *Neuroreport*, vol. 4, no. 3, pp. 233–236, 1993.

[48] E. Foeller and D. E. Feldman, "Synaptic basis for developmental plasticity in somatosensory cortex." *Current Opinion in Neurobiology*, vol. 14, no. 1, pp. 89–95, 2004.

[49] T. Villmann, R. Der, M. Herrmann, and T. M. Martinetz, "Topology preservation in self-organizing feature maps: Exact definition and measurement," *IEEE Trans. on Neural Networks*, vol. 8, no. 2, pp. 256–266, 1997.

[50] S. Furui, "On the role of spectral transition for speech perception," *Journal of the Acoustical Society of America*, vol. 80, no. 4, pp. 1016–1025, 1986.

[51] C. Gläser, M. Heckmann, F. Joublin, and C. Goerick, "Combining auditory preprocessing and Bayesian estimation for robust formant tracking," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 224–236, 2010.

[52] L. Deng, X. Cui, R. Pruvenok, Y. Chen, S. Momen, and A. Alwan,

"A database of vocal tract resonance trajectories for research in speech processing," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, 2006, pp. I – 369–372.

[53] R. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and A. M., "Complex sounds and auditory images," in *Proc. Symp. Hearing, Auditory Physiology and Perception*, 1992, pp. 429–446.

[54] T. Kohonen, "The neural phonetic typewriter," *Computer*, vol. 21, no. 3, pp. 11–22, 1988.

[55] M. Vaz, H. Brandl, F. Joublin, and C. Goerick, "Speech imitation with a child's voice: addressing the correspondence problem," in *Proc. Int. Conf. on Speech and Computer*, 2009.

Claudius Gläser was born in Gera, Germany, in 1982. He received a Dipl.-Inf. degree (with honors) in computer science from the Technische Universität Ilmenau, Ilmenau, Germany, in 2006. He is currently pursuing a Ph.D. degree at the Honda Research Institute Europe, Offenbach, Germany, focusing on the acquisition of sensorimotor competencies and their use for grounding language during social interaction.

Since 2006 he has been a Scientist at the Honda Research Institute Europe. His research interests include speech analysis, language acquisition, neural information processing, and machine learning.

Frank Joublin received the Dipl.Ing. degree in electrical engineering from the Ecole Universitaire D'Ingénieurs de Lille (EUDIL), Lille, France, in 1987 and the Ph.D. degree in neuroscience from the University of Rouen, Rouen, France, in 1993.

From 1994 to 1998 he was a Postdoctoral Research Fellow at the Institute for Neural Computation, Ruhr-Universität Bochum, Bochum, Germany. His research activity there focused on the modeling of cortical plasticity, stereo-vision algorithms, and place cell representations for robot navigation. From 1998 to 2001 he was a Customer Project Manager at Philips Speech Processing, Aachen, Germany, where he developed voice-activated telephony applications. Since 2001, he is a Principal Scientist at the Honda Research Institute Europe, Offenbach, Germany, and since 2008 a Board Member of the CoR-Lab Graduate School of the Bielefeld University, Bielefeld, Germany. His research interests include auditory scene analysis, language and semantic acquisition, and developmental robotics.