frontiers

Learning Structure of Complex Sensory Inputs with Plasticity in Neural Microcircuits

Joseph Chrol-Cannon [1] **and Yaochu Jin** [1,*]

[1] *Department of Computing, University of Surrey, Guildford, UK*

Correspondence*:
Yaochu Jin
Department of Computing, University of Surrey, Guildford, GU2 7XH, UK,
yaochu.jin@surrey.ac.uk

ABSTRACT

Synaptic plasticity is explored as a form of unsupervised adaptation in cortical microcircuits to learn the structure of complex sensory inputs and thereby improve performance of classification and prediction. The question of whether the structure of specific input patterns is encoded in the structure of neural networks has been largely neglected. Existing studies that have analyzed structural adaptation have used simplified, synthetic inputs in contrast to complex and noisy patterns found in real-world sensory data.

In this work, class-specific structural changes are analyzed for three empirically derived models of plasticity applied to three temporal sensory classification tasks that include complex, real-world visual and auditory data. Two forms of spike-timing dependent plasticity (STDP) and the Bienenstock-Cooper-Munro (BCM) plasticity rule are used to adapt the recurrent network structure before training is performed for pattern recognition tasks.

It is shown that plasticity does learn highly input-specific structure in its adaptation. However, it does not improve the performance on sensory pattern recognition tasks due to synaptic interference between consecutively presented input samples. The changes in synaptic strength produced by one stimulus are reversed by the presentation of another, thus largely preventing synaptic changes from being retained in the structure of the network.

To solve the problem of interference, we suggest that models of plasticity be extended to restrict neural activity and synaptic modification to a subset of the neural circuit, which is increasingly found to be the case in experimental neuroscience.

Keywords: Synaptic plasticity, spiking neurons, recurrent neural networks, interference, spatiotemporal patterns

1 INTRODUCTION

Recurrent neural networks consisting of biologically based spiking membrane models have only recently been applied to real-world learning tasks under a framework called reservoir computing **Buonomano and Maass** (2009); **Maass et al.** (2002). The models of this framework use a recurrently connected set of neurons driven by an input signal to create a non-linear, high-dimensional temporal transformation of the input that is used by single layer perceptrons **Rosenblatt** (1958) to produce desired outputs. This restricts the training algorithms to a linear regression task, while still allowing the potential to work on temporal data in a non-linear fashion.

Given an initially fixed, unchanging connectivity, reservoir computing is based on the principle of random projections of the input signal in which the network structure is completely independent of the input

patterns. In these models, the only features learned by the trainable parameters of the perceptron readout are correlations between the randomly projected features and the desired output signal.

We believe that learning in neural networks should go farther than supervised training based on error from the output. All synapses should adapt to be able to encode the structure of the input signal and ideally, should not rely on the presence of a desired output signal from which to calculate an error with the actual output. The neural activity generated by the input signal should provide enough information for synapses to adapt and encode properties of the signal in the network structure. By applying unsupervised adaptation to the synapses in the form of biologically derived plasticity rules **Bi and Poo** (1998); **Bienenstock et al.** (1982); **Wittenberg and Wang** (2006) it is hoped to provide the means for the recurrently connected neurons of the network to learn a structure that generates more effective features than a completely random projection that is not specific to the input data.

In this work we will explore the impact of applying several biologically derived plasticity mechanisms on three temporal sensory discrimination tasks. Two forms of spike-timing dependent plasticity (STDP) **Bi and Poo** (1998); **Wittenberg and Wang** (2006) will be tested, along with the Beinenstock-Cooper-Munro (BCM) rule **Bienenstock et al.** (1982). The sensory tasks will include real-world speech and video data of human motion. Synaptic plasticity will be applied in an unsupervised pre-training phase, before the supervised regression of the perceptron readout occurs. We will compare the impact that plasticity has on the performance in these tasks and also analyze the specific structural adaptation of the weight matrices between each of the classes of input sample in each task. A method will be introduced to evaluate the extent to which the synaptic changes encode class-specific features in the network structure. Interference between different samples is a well-established phenomenon in sequentially trained learning models **French** (1999); **McCloskey and Cohen** (1989); **Ratcliff** (1990). As this directly concerns neural networks trained on sensory recognition tasks, a quantification of the interference between the synaptic parameters will be made for each plasticity model while applied to each task.

Existing studies report that adapting neural circuits with plasticity improves their performance on pattern recognition tasks **Xue et al.** (2013); **Yin et al.** (2012) but there is no analysis of how the adaptation of synaptic parameters leads to this result. On the other hand, work that does analyze the structural adaptation of the network does so using synthetic input patterns that are already linearly separable, completely unlike realistic sensory input **Toutounji and Pipa** (2014). For a review of work applying plasticity models to improve the function of neural networks, the reader is referred to **Chrol-Cannon and Jin** (2014a).

2 RESULTS

2.1 DESCRIPTION OF SENSORY INPUTS

Complex sensory signals are projected through a common set of nerve fibers to cortical regions that must learn to distinguish between them based on differences in their spatial-temporal features.

Three sensory recognition tasks are selected, among which two of them consist of real audio and video signals of human speech and motion. For all tasks, the neural network output is trained to respond uniquely to each of the different types of input sample and therefore be able to perform effective recognition between them. Also, sample specific synaptic adaptations are analyzed to determine if unique structure is learned within the network due to synaptic plasticity.

The auditory task is to distinguish between nine different speakers based on short utterances of the vowel **/ae/**. Each of the 640 samples consists of a frequency 'spectrogram' that plots frequency intensity over a sequence of audio time frames. Figure 1 plots an example sample from each of the nine speakers.

The visual task is to distinguish between six types of human behavior; boxing, clapping, waving, walking, running and jogging. The 2391 samples are video sequences of many different subjects performing

those six motions. There is a simple pre-processing stage that converts the video data into a sparse representation before being used as input to the neural network. Extracted still frames and processed features are plotted in Figure 2 for one subject performing each of the six behaviors.

A synthetic data set is generated to model a low spatial dimension but very high frequency temporal structure, in contrast to the previous two sensory tasks. Three functions generate time-varying single dimensional signals that the network learns to distinguish between. A complete description and method for generating the data is described in **Jaeger** (2007).

The auditory and visual tasks are described in **Kudo et al.** (1999) and **Schuldt et al.** (2004), respectively, with data availability also provided.

2.2 LEARNING INPUT-SPECIFIC STRUCTURE USING PLASTICITY

We wish to test the hypothesis that synaptic plasticity is encoding a distinct structure for input samples of different labels. For the speech task, these labels consist of different speakers and for the video recognition task the labels consist of different human behaviors.

The data sets are divided evenly into two. Each subset is used to train a recurrently connected network for 10000 iterations, selecting a sample at random on each iteration. The changes to the weight matrix due to plasticity are recorded for each sample presentation. This is then used to create a class-specific average weight change for each of the class labels in both of the sample subsets. Finally, we calculate the Euclidean distance between each class in one set and each class in the other according to the following formula:

$$\text{Dist}(C^X, C^Y) = \sum_{i=1}^{N} |\Delta W_i(C^X) - \Delta W_i(C^Y)| \tag{1}$$

Where C denote class labels, X and Y distinguish the separated sets of samples, ΔW is the change in weight matrix for a presented sample, N is the number of synapses, and i the synapse index.

This effectively produces a confusion matrix of similarity in the synaptic weight change for different classes of input. Having lower values on the descending diagonal means that there is structural adaptation that is specific to the class of that column compared with the similarity between structural adaptations of two different classes.

Figure 3 shows the 'weight change confusion matrices' described above, for each plasticity model applied to all sensory tasks (nine experiments in total). All of the experiments show at least some stronger similarity in the descending diagonals and most are stark in this manner. It is certainly a strong enough pattern to show that through the many iterations of training, each of the plasticity models have become sensitive to the particular structure of the sensory input signals so that each different class of sample will give rise to changes in synaptic strength that are distinct from other classes compared with the similarity to themselves. We re-iterate that the class labels were not used in any way in the plasticity models themselves and so the differences in the weight change arise from the input signals alone.

There are a few exceptions to the strong diagonal patterns in Figure 3. This means that some classes are not effectively distinguished from each other; speakers 8/9 with bi-phasic STDP, behaviors 1/2 with BCM, behaviors 1/2/3 and 4/5/6 with tri-phasic STDP. The latter confusion corresponds to the behaviors of boxing/clapping/waving and walking/running/jogging. From the similarity of those input features shown in the lower panes of Figure 2, it is evident why this confusion might occur.

Table 1. Classification Error Rates

	Static	BCM	STDP	TP-STDP
Tri-func	0.153	0.157	0.204	**0.138**
KTH	**0.283**	0.3	0.333	0.383
Vowels	0.089	**0.086**	0.092	**0.086**

Table 2. Synaptic Interference

	BCM	STDP	TP-STDP
Tri-func	0.82	**0.8**	0.88
KTH	**0.92**	0.93	0.96
Vowels	0.96	**0.58**	0.9

2.3 CLASSIFICATION PERFORMANCE WITH PLASTICITY

Perhaps the ultimate goal of neural network methods when applied to sensory tasks is the ability to accurately distinguish different types of input sample by their patterns. We compare the error rates achieved by our neural network on the three sensory tasks, with and without the different forms of plasticity used in this work. Table 1 lists the error rates achieved for each of the learning tasks with the different plasticity rules active in a pre-training phase in addition to a static network with fixed internal synapses.

From the error rates in Table 1 it is evident that pre-training the network with synaptic plasticity can make insignificant improvements in lowering the error rate. However, the results here indicate that it can have a greater negative impact than a positive one. In the KTH human behavior data set, all three plasticity models increase the error rate by between 1.7% and 10%. Conversely, the best improvement was found on the tri-function signal recognition task with tri-phasic STDP at only 1.5%.

It is clear from the network output that pre-training with synaptic plasticity is not a suitable method for this class of model, This does not contradict the result that plastic synapses are learning useful, input-specific structure. However, it does suggest that the structure being learned is not effectively utilized in the generation of a network output. We next investigate interference between synaptic changes to determine if the structural learning is retained in the network or if interference is a barrier for effective application of synaptic plasticity.

2.4 SYNAPTIC INTERFERENCE

When a model adapts incrementally to sequentially presented input, existing patterns that have been learned by the model parameters are prone to be overwritten by learning new patterns. This is known as interference. The work that has studied this effect **French** (1999); **McCloskey and Cohen** (1989); **Ratcliff** (1990), test the ability to recognize previously presented input after the model has been trained on new ones in order to estimate how much learning has been undone. When new training leaves the model unable to recognize old patterns, it is said there has been catastrophic interference and forgetting.

We introduce a method of measuring interference directly in synaptic parameters instead of the model output. Our measure is described in detail in the Methods section. I^{total} directly quantifies all synaptic changes that are overwritten.

The interference for each of our experiments is listed in Table 2. In all but one of the experiments the interference level is between 82% and 96%. Most of the learned structure for each class of input is forgotten as consecutive samples overwrite each other's previous changes. Bi-phasic STDP applied to speaker recognition has the lowest level of interference at 58%.

To further explore interference and visualize the impact of plasticity, synaptic changes will be analyzed directly. Figure 4 is an illustrative example for the speaker recognition task with BCM plasticity (similar figures for the other experiments are given in S1–S8). It shows the adaptation of the synaptic weight matrix produced by each speaker in the voice recognition task. This is plotted against the activity level for each neuron, **S**, and the readout weights, **R**, that are trained to generate an output that is sensitive to that given speaker. Each of these sub plots is the average response taken over all sample presentations from that speaker. This makes a whole chain of effect visible: from the synaptic change of an internal network connection, to the average neuron state for a given speaker, to the selective weights of the readout for that speaker. For all to be working well in a cohesive system, we expect that a positive weight change should correspond with a neuron activation unique to the class which would in turn improve the recognition ability of the readout to identify that class.

The sections of the class weight matrix highlighted in green in Figure 4, highlight an example where synaptic interference is occurring between different types of pattern. Directly opposing features in the weight matrix adaptations show the samples negating each other's changes. However, the same features are also most distinctively class specific.

Any synapse can only change in two directions: positively or negatively, which means that a single synapse can only adapt to distinguish between two mutually exclusive kinds of input pattern. If n synapses are considered in combination, then the number of input patterns that can be discriminated becomes 2^n in ideal theoretical conditions. Figure 4 illustrates this principle in practice with regards to the nine speaker recognition tasks. The adapted synapses labeled (a) can clearly distinguish speaker $\{\#1\}$ from speakers $\{\#2, \#3\}$ but cannot distinguish $\{\#2\}$ from $\{\#3\}$. Similarly, the adapted synapses labeled (b) can distinguish speakers $\{\#1, \#6, \#8\}$ from speakers $\{\#3, \#4, \#9\}$ but cannot distinguish speakers within either of those sets. However, if the synapses (a), (b), (c) and (d) are considered in combination, then all speakers can be distinguished by synaptic plasticity changes alone.

Figure 4 also shows the weight changes are not correlated with the neural activity or readout weights. For plasticity to improve the accuracy of sensory discrimination, it would be expected that synapses would strengthen for class specific neural activity and weaken for common neural activity. This is not the case in our results.

3 DISCUSSION

3.1 UNSUPERVISED PLASTICITY LEARNS LABEL SPECIFIC STRUCTURE

Both STDP and BCM models adapt the synapses of a network in distinctive patterns according to which type of sample is being presented to the network. We can conclude that presenting a training signal with the sample label is not required for plasticity to learn specific information for complex sensory inputs from different sources. This result holds for the speech, visual and benchmark pattern recognition tasks. To achieve this feat, we hypothesize that plasticity drives the synaptic parameters to a structure that represents an average between all input samples. Once converged, any further input stimulus will drive the synaptic parameters in a unique direction away from this average structure. On balance, scrambled presentation of random inputs keeps the network in this sensitive state.

3.2 UNIFORMLY APPLIED PLASTICITY LEADS TO SYNAPTIC INTERFERENCE

We show synaptic plasticity spends most of its action counter- acting previous changes and overwriting learned patterns. The same patterns of synaptic adaptation that distinctly characterizes each class of input are the same ones that reverse adaptations made by other inputs.

Plasticity is applied uniformly to all synapses. All neurons in a recurrent network produce activity when given input stimulus. Combined, these factors mean that any input sample will cause the same synapses to change. This leads to synaptic competition, interference and ultimately, forgetting.

3.3 LOCALITY OF PLASTICITY REQUIRED TO OVERCOME INTERFERENCE

To overcome the problem of interference, the mechanisms of plasticity need to be restricted to adapt only a subset of the synapses for any given input stimulus. There is much existing research that supports this conclusion and a number of possible mechanisms that can restrict the locality of plasticity.

It has been shown *in vivo* (using fMRI and neurological experiment) that synaptic plasticity learns highly specific adaptations early in the visual perceptual pathway **Karni and Sagi** (1991); **Schwartz et al.** (2002). Simulated models of sensory systems have demonstrated that sparsity of activity is essential for sensitivity to input-specific features **Barranca et al.** (2014); **Finelli et al.** (2008). In fact, in a single-layer, non-recurrent structure, STDP is shown to promote sparsity in a model olfactory system **Finelli et al.** (2008). Conversely, in recurrent networks, STDP alone is unable to learn input specific structure because it 'over-associates' **Bourjaily and Miller** (2011). Strengthened inhibition was used to overcome this problem and combined with reinforcement learning to produce selectivity in the output **Bourjaily and Miller** (2011). By promoting sparsity, the lack of activity in most of the network will prevent activity-dependent models of plasticity in adapting those connections.

Reward modulated plasticity has also been widely explored in simulated **Darshan et al.** (2014); **Gavornik et al.** (2009) and biological experiment **Lepousez et al.** (2014); **Li et al.** (2013). Input-specific synaptic changes are shown to be strongest in the presence of a reward signal **Gavornik et al.** (2009); **Lepousez et al.** (2014). Lasting memories (synaptic changes not subject to interference), are also seen to rely on a process of re-consolidation consisting of fear conditioning **Li et al.** (2013). A reinforcement signal based on either reward or fear conditioning can be effectively used to restrict synaptic changes in a task dependent context such as sensory pattern recognition.

Another way to restrict synaptic changes in a task dependent way is to rely on a back-propagated error signal that has well established use in artificial neural networks. This might be achieved in a biologically plausible way through axonal propagation **Kempter et al.** (2001) or top-down cortical projections sending signals backwards through the sensory pathways **Schfer et al.** (2007). Top-down neural function in general is thought to be essential in determining structure in neural networks **Sharpee** (2014), providing a context for any adaptations.

Neural cascades **Polat and Sagi** (1994) and synchronicity **Hoppensteadt** (1989) have also been suggested as factors that enable input-specific plasticity.

3.4 LEARNING INPUT STRUCTURE DOES NOT NECESSARILY IMPROVE PERFORMANCE

Structural adaptation with plasticity in the pre-training phase, while specific, may not be utilized by the output produced by the network readout. This could be due to the following reasons. Firstly, there is a disparity in the neural code. The output from a recurrent spiking network model is currently decoded as a rate code. In contrast, synaptic plasticity updates structure in a way that depends on the precise temporal activity of neural spikes. Secondly, information content is reduced. While creating associations between co-activating neurons, Hebbian forms of plasticity may also increase correlations and reduce information and separation. These can determine the computational capacity of a recurrent network model **Chrol-Cannon and Jin** (2014b). Both discrepancies could be barriers for the effective application of plasticity

to improve pattern recognition. Therefore new frameworks of neural processing should be based directly on the adapting synapses. This will lead to functional models of neural computing that are not merely improved by synaptic plasticity, but that rely on it as an integral element.

4 MATERIAL & METHODS

4.1 NEURAL MICROCIRCUIT

The neural network model used in this work is illustrated in Figure 5. Recurrently connected neurons, indicated by L are stimulated by the inputs directly as injected current, I, that perturbs the membrane potential modeled with a simple model **Izhikevich** (2003). This method for modeling the spiking activity of a neuron is shown to reproduce most naturally occurring patterns of activity **Izhikevich** (2004). The real-valued inputs are normalized between 0 and 1, which are multiplied by a scaling factor of 20 before being injected as current into L. Input connections number $0.2 \cdot network\,size$, projected randomly to the network nodes. The network activity dynamics are then simulated for $30ms$. Then, the resulting spike trains produced by each of the neurons is passed through a low-pass filter, f, to produce a real valued vector used to train a linear readout with the iterative, stochastic gradient descent method.

For our experiments the network consists of 35–135 spiking neurons with the ratio of excitatory to inhibitory as 4:1. Neurons are connected with static synapses i.e. the delta impulse (step) function. Connectivity is formed by having $N^2 \cdot C$ synapses that each have source and target neurons drawn according to uniform random distribution, where N is the number of neurons and C is the probability of a connection between any two neurons. Weights are drawn from two Gaussian distributions; $\mathcal{N}(6, 0.5)$ for excitatory and $\mathcal{N}(-5, 0.5)$ for inhibitory. When plasticity adapts the reservoir weights, w_{max} is clamped at 10 and w_{min} at -10. All parameters for excitatory and inhibitory neuron membranes are taken from **Izhikevich** (2003). The equations for the membrane model are as follows:

$$v' = 0.04v^2 + 5v + 140 - u + I \tag{2}$$

$$u' = a(bv - u) \tag{3}$$

With the spike firing condition:

$$\text{if} \quad v > 30mV \quad \text{then} \quad \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \tag{4}$$

To generate a real-valued output from the discrete spiking activity, the spike train from each neuron is low-pass filtered according to Equation 5. The vector of values produced is then weighted with the readout weight matrix and summed to produce a single output value, shown in Equation 6.

$$x_i = f(S(t)) = \sum_{t=1}^{T} \exp\left(\frac{-t}{\tau}\right) \cdot wmax \tag{5}$$

$$y = \sum_{i=1}^{n} x_i \cdot w_i \tag{6}$$

The state vector for a neuron is denoted by x_i, the filter function is $f()$ and the spike train is $S(t)$. The maximum number of time-steps in $S(t)$ is T, in this case 50. The decay constant τ is 6.

This output weights are updated according to the iterative, stochastic gradient descent method: Least Mean Squares, given in Equation 7.

$$w_i \longleftarrow w_i + \mu(d - y)x_i. \tag{7}$$

Here, d is the desired output, y is the actual output, x_i is the input taken from a neuron's filtered state, and μ is a small learning rate. The weight from x_i to the output is w_i.

4.2 SYNAPTIC PLASTICITY MODELS

Three synaptic plasticity mechanisms are employed in this study, each of them based on the Hebbian postulate **Hebb** (1949) of "neurons that fire together, wire together". Each mechanism is outlined as follows:

BCM Plasticity: The BCM rule **Bienenstock et al.** (1982) is a rate based Hebbian rule that also regulates the post-neuron firing rate to a desired level. It works on a temporal average of pre- and post-synaptic activity. The BCM rule is given in Equation 8. The regulating parameter is the dynamic threshold Θ_M, which changes based on the post-synaptic activity y and the desired level y_0 in the following relationship: $E[y/y_0]$, where $E[\cdot]$ denotes a temporal average. There is also a decay parameter ϵw_i for additional stability, that slowly reduces connection strength and so provides a mechanism for uniform weight decay, irrespective of the level of activity or correlation. A plot of the BCM weight change is presented in Figure S9 in Supporting Information.

$$\frac{dw_i}{dt} = y(y - \theta_M)x_i - \epsilon w_i \tag{8}$$

Bi-phasic STDP: The STDP rule depends on the temporal correlation between pre- and post-synaptic spikes. The synaptic weight change is computed based on the delay between the firing times of the pre- and post- neuron. This is described in a fixed 'learning window' in which the y-axis is the level of weight change and the x-axis is the time delay between a pre- and post-synaptic spike occurrence. The bi-phasic STDP rule consists of two decaying exponential curves **Song et al.** (2000), a positive one to potentiate in-order spikes, and a negative one to depress out-of-order spikes. This rule was derived from experimental work carried out on populations of neurons *in vitro* **Markram et al.** (1997)**Bi and Poo** (1998). Bi-phasic STDP is given in Equation 9.

$$\Delta w(t) = \begin{cases} A_+ \cdot \exp\left(\frac{-\Delta t}{\tau_+}\right) & \text{if } t > 0 \\ A_- \cdot \exp\left(\frac{\Delta t}{\tau_-}\right) & \text{if } t \leq 0 \end{cases} \tag{9}$$

Tri-phasic STDP: A tri-phasic STDP learning window consists of a narrow potentiating region for closely correlated activity but depressing regions on either side: for recently uncorrelated activity, and for correlated but late activity. This learning window has been observed *in vitro*, most notably in the hippocampi, between areas CA3 and CA1 **Wittenberg and Wang** (2006). The tri-phasic STDP is given in Equation 10.

$$\Delta w(t) = A_+ \cdot \exp\left(\frac{-(x-20)^2}{200}\right) - A_- \cdot \exp\left(\frac{-(x-20)^2}{2000}\right) \tag{10}$$

Both STDP learning windows are plotted in Figure S10 in Supporting Information.

4.3 SYNAPTIC INTERFERENCE MEASURE

We wish to quantify interference directly between synaptic changes of plasticity. Our formulation of synaptic interference is based on the synaptic changes from sequentially presented samples. Changes for

a given type of sample are called ΔW_t and changes for all others are ΔW_o. Interference must be calculated individually for each type of sample, I_t^{class}, and averaged together to get the overall interference, I^{total}. The equations are as follows:

$$I_t^{\text{class}} = \sum_{i=1}^{N} \frac{1}{N} [\Delta W_{ti} \cdot \Delta W_{oi} < 0][|\Delta W_{ti}| < |\Delta W_{oi}| \cdot C] \tag{11}$$

$$I^{\text{total}} = \sum_{t=1}^{C} \frac{I_t^{\text{class}}}{C} \tag{12}$$

Where I is interference, N is the number of synapses, C is the number of competing sample types and ΔW is a vector of synaptic changes. Subscript i denotes the parameter index, subscript t denotes samples of a given type and subscript o denotes samples of all other types.

4.4 SYNTHETIC SIGNAL DATA

A synthetic benchmark task is taken from a study performed with Echo State Networks **Jaeger** (2007), a similar type of network model to the one we employ, but using artificial neurons instead. The task is to predict which of three signal generating functions is currently active in producing a time-varying input signal. To generate a sample of the signal at a given time step, one of the three following function types is used; 1) A sine function of a randomly selected period, 2) A chaotic iterated tent map, 3) A randomly chosen constant. The generator is given some low probability, 0.05, of switching to another function at each time-step. The full method of generating the data is described in **Jaeger** (2007).

4.5 SPEAKER RECOGNITION DATA

A speaker recognition task is a classification problem dealing with mapping time-series audio input data to target speaker labels. We use a data set taken from **Kudo et al.** (1999) which consists of utterances of 9 male Japanese speakers pronouncing the vowel **/ae/**. The task is to correctly discriminate each speaker based on the speech samples. Each sample is comprised of a sequence of 12 feature audio frames. The features of each frame are the LPC cepstrum coefficients. The sample sequence ranges between 7-29 frames. The dataset is divided into training and testing sets of 270 and 370 samples each, respectively. Note that unlike the benchmark data used in this report, the samples are not in a consecutive time-series, yet each sample consists of a time-series sequence of audio frames.

4.6 PRE-PROCESSING OF THE HUMAN MOTION DATA

A visual task is selected to test high dimensional spatial-temporal input data. The KTH data set **Schuldt et al.** (2004) consists of 2391 video files of people performing one of six actions; boxing, clapping, waving, walking and jogging. There are 25 different subjects and the samples cover a range of conditions that are described in more detail in **Schuldt et al.** (2004). Each video sample is taken at 25 frames per second and down sampled to a resolution of 160x120 pixels. We process the raw video sequences according to a formula shown in the following equations:

$$M(t) = \|[\Delta(I_1, I_2), ..., \Delta(I_{N-1}, I_N)]\| \tag{13}$$

$$M(t,i) = \begin{cases} 1 & \text{if } M(t,i) \geq 0.2 \cdot max(M(\cdot)) \\ 0 & \text{else} \end{cases} \tag{14}$$

The final input matrix M is indexed by time-frames, t and spatial samples i. Column vectors I_n are individual frames, re-shaped into one dimension. Each sample contains up to a total of N frames. In plain language, this process essentially further down samples by a factor of 0.2 and calculates the difference between pixels in consecutive frames, which are then used as the new input features. Each frame is then re-shaped into a single dimensional column vector then appended together to form an input matrix in which each column is used as the neural network input at consecutive time steps. Figure 2 shows frames extracted from an example of each type on motion along with the corresponding processed features.

DISCLOSURE/CONFLICT-OF-INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

Conception and design of the work was by YJ and JCC. Experiments were performed by JCC. Analysis and interpretation were undertaken by JCC and YJ. Manuscript was written by JCC and revised by YJ.

ACKNOWLEDGMENTS

Funding: JCC was supported by Engineering and Physical Sciences Research Council's Doctoral Training Grant through University of Surrey.

REFERENCES

Barranca, V. J., Kovai, G., Zhou, D., and Cai, D. (2014), Sparsity and compressed coding in sensory systems, *PLoS Comput Biol*, 10, 8, e1003793

Bi, G. Q. and Poo, M. M. (1998), Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type, *Journal of Neuroscience*, 18, 24, 10464–72

Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982), Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex., *Journal of Neuroscience*, 2, 1, 32–48

Bourjaily, M. A. and Miller, P. (2011), Synaptic plasticity and connectivity requirements to produce stimulus-pair specific responses in recurrent networks of spiking neurons, *PLoS Comput Biol*, 7, 2, e1001091

Buonomano, D. V. and Maass, W. (2009), State-dependent computations: spatiotemporal processing in cortical networks., *Nat Rev Neurosci*, 10, 2, 113–25

Chrol-Cannon, J. and Jin, Y. (2014a), Computational modeling of neural plasticity for self-organization of neural networks, *Biosystems*, 125, 0, 43–54

Chrol-Cannon, J. and Jin, Y. (2014b), On the correlation between reservoir metrics and performance for time series classification under the influence of synaptic plasticity, *PLoS ONE*, 9, 7, e101792

Darshan, R., Leblois, A., and Hansel, D. (2014), Interference and shaping in sensorimotor adaptations with rewards, *PLoS Comput Biol*, 10, 1, e1003377

Finelli, L. A., Haney, S., Bazhenov, M., Stopfer, M., and Sejnowski, T. J. (2008), Synaptic learning rules and sparse coding in a model sensory system, *PLoS Comput Biol*, 4, 4, e1000062

French, R. M. (1999), Catastrophic forgetting in connectionist networks., *Trends in Cognitive Sciences*, 3, 4, 128–35

Gavornik, J. P., Shuler, M. G. H., Loewenstein, Y., Bear, M. F., and Shouval, H. Z. (2009), Learning reward timing in cortex through reward dependent expression of synaptic plasticity, *Proc Natl Acad Sci*, 106, 16, 6826–31

Hebb, D. (1949), The Organization of Behavior. (Wiley: New York)

Hoppensteadt, F. C. (1989), Intermittent chaos, self-organization, and learning from synchronous synaptic activity in model neuron networks, *Proc Natl Acad Sci*, 86, 9, 2991–5

Izhikevich, E. M. (2003), Simple model of spiking neurons, *IEEE Transactions on Neural Networks*, 14, 6, 1569–72

Izhikevich, E. M. (2004), Which model to use for cortical spiking neurons?, *IEEE Transactions on Neural Networks*, 15, 5, 1063–70

Jaeger, H. (2007), Discovering multiscale dynamical features with hierarchical echo state networks, Technical report, Jacobs University Bremen

Karni, A. and Sagi, D. (1991), Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity, *Proc Natl Acad Sci*, 88, 11, 4966–70

Kempter, R., Leibold, C., Wagner, H., and van Hemmen, J. L. (2001), Formation of temporal-feature maps by axonal propagation of synaptic learning, *Proc Natl Acad Sci*, 98, 7, 4166–71

Kudo, M., Toyama, J., and Shimbo, M. (1999), Multidimensional curve classification using passing-through regions, *Pattern Recognition Letters*, 20, 11–3, 1103–11

Lepousez, G., Nissant, A., Bryant, A. K., Gheusi, G., Greer, C. A., and Lledo, P.-M. (2014), Olfactory learning promotes input-specific synaptic plasticity in adult-born neurons, *Proc Natl Acad Sci*, 111, 38, 13984–9

Li, Y., Meloni, E. G., Carlezon, W. A., Milad, M. R., Pitman, R. K., Nader, K., et al. (2013), Learning and reconsolidation implicate different synaptic mechanisms, *Proc Natl Acad Sci*, 110, 12, 4798–803

Maass, W., Natschlager, T., and Markram, H. (2002), Real-time computing without stable states: a new framework for neural computation based on perturbations., *Neural Computation*, 14, 11, 2531–60

Markram, H., Luebke, J., Frotscher, M., and Sakmann, B. (1997), Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps, *Science*, 275, 5297, 213–5

McCloskey, M. and Cohen, N. (1989), Catastrophic interference in connectionist networks: The sequential learning problem., *The Psychology of Learning and Motivation*, 24, 109–64

Polat, U. and Sagi, D. (1994), Spatial interactions in human vision: from near to far via experience-dependent cascades of connections., *Proc Natl Acad Sci*, 91, 4, 1206–9

Ratcliff, R. (1990), Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions., *Psychological Review*, 97, 285–308

Rosenblatt, F. (1958), The perceptron: a probabilistic model for information storage and organization in the brain, *Psychol. Rev.*, 65, 6, 386–408

Schfer, R., Vasilaki, E., and Senn, W. (2007), Perceptual learning via modification of cortical top-down signals, *PLoS Comput Biol*, 3, 8, e165

Schuldt, C., Laptev, I., and Caputo, B. (2004), Recognizing human actions: a local svm approach, in ICPR Proceedings of the 17th International Conference on Pattern Recognition, volume 3, volume 3, 32–6

Schwartz, S., Maquet, P., and Frith, C. (2002), Neural correlates of perceptual learning: A functional mri study of visual texture discrimination, *Proc Natl Acad Sci*, 99, 26, 17137–42

Sharpee, T. O. (2014), Function determines structure in complex neural networks, *Proc Natl Acad Sci*, 111, 23, 8327–8

Song, S., Miller, K. D., and Abbott, L. F. (2000), Competitive Hebbian learning through spike-timing-dependent synaptic plasticity, *Nat. Neurosci.*, 3, 9, 919–26

Toutounji, H. and Pipa, G. (2014), Spatiotemporal computations of an excitable and plastic brain: Neuronal plasticity leads to noise-robust and noise-constructive computations, *PLoS Comput Biol*, 10, 3, e1003512

Wittenberg, G. M. and Wang, S. S. (2006), Malleability of spike-timing-dependent plasticity at the ca3-ca1 synapse, *Journal of Neuroscience*, 26, 6610–17

Xue, F., Hou, Z., and Li, X. (2013), Computational capability of liquid state machines with spike-timing-dependent plasticity, *Neurocomputing*, 122, 0, 324–9

Yin, J., Meng, Y., and Jin, Y. (2012), A developmental approach to structural self-organization in reservoir computing, *IEEE Transactions on Autonomous Mental Development*, 4, 4, 273–89

FIGURES

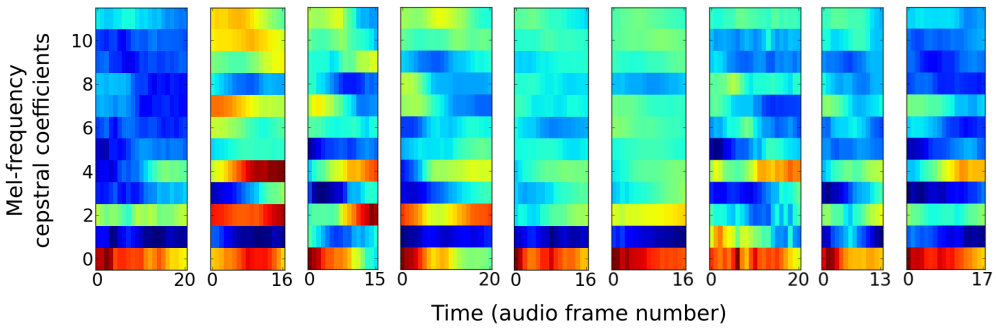

Figure 1. Vowel samples from the nine speakers in the speaker recognition task. The audio signals in the data set are pre-processed into 12 Mel-frequency cepstrum coefficients (MFCC) features. Samples from each speaker have variable time duration in the number of audio frames they consist of.

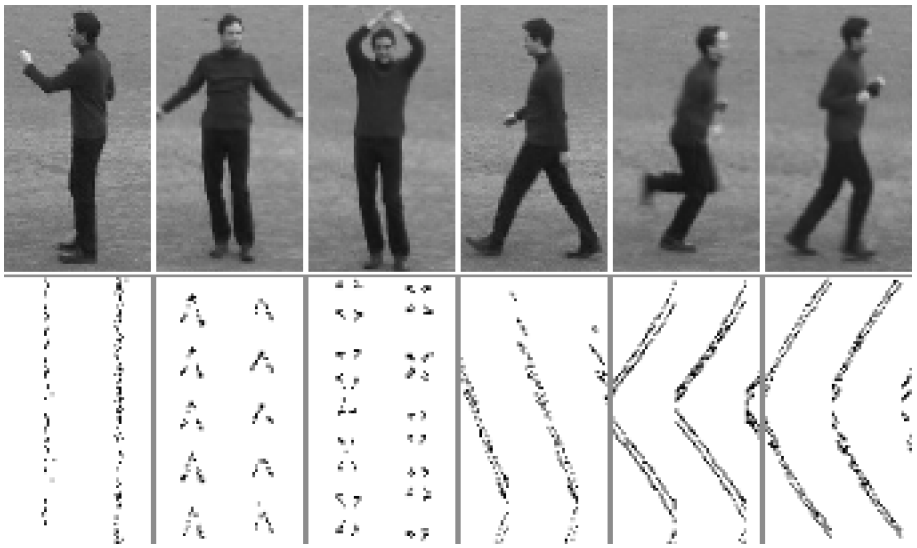

Figure 2. Human motion samples for the six types of behavior in the KTH visual discrimination task. This illustration consists of different behaviors from a single person, while the whole data set contains 25 persons. **Top row:** Still frames from example video samples; boxing, clapping, waving, walking, running and jogging. **Bottom row:** Features extracted corresponding to the samples above. Features are the raw time-series activity used as input to the neural network.

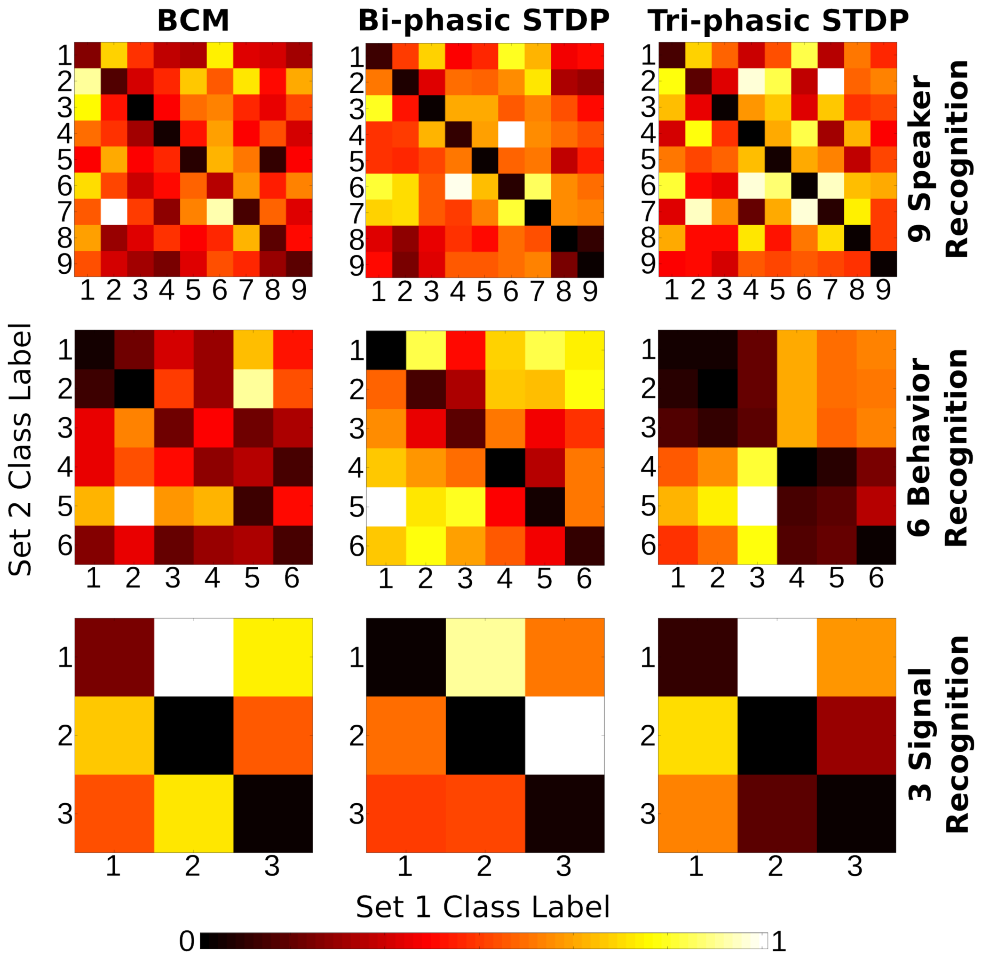

Figure 3. Class correlation of structural synaptic adaptation. Heat map plots indicate the structure learned on each class for the three tasks under each of the plasticity rules. Essentially, it is a confusion matrix of the geometric distance between the weight matrix adaptation of each class of sample. The training data for each task is divided into two sets. Class-average adaptation is found for each set. There is then a distance calculated between each class of the two sets. Lower values on the descending diagonal indicate higher correlation within a class adaptation and therefore strong class-specific structure learned.

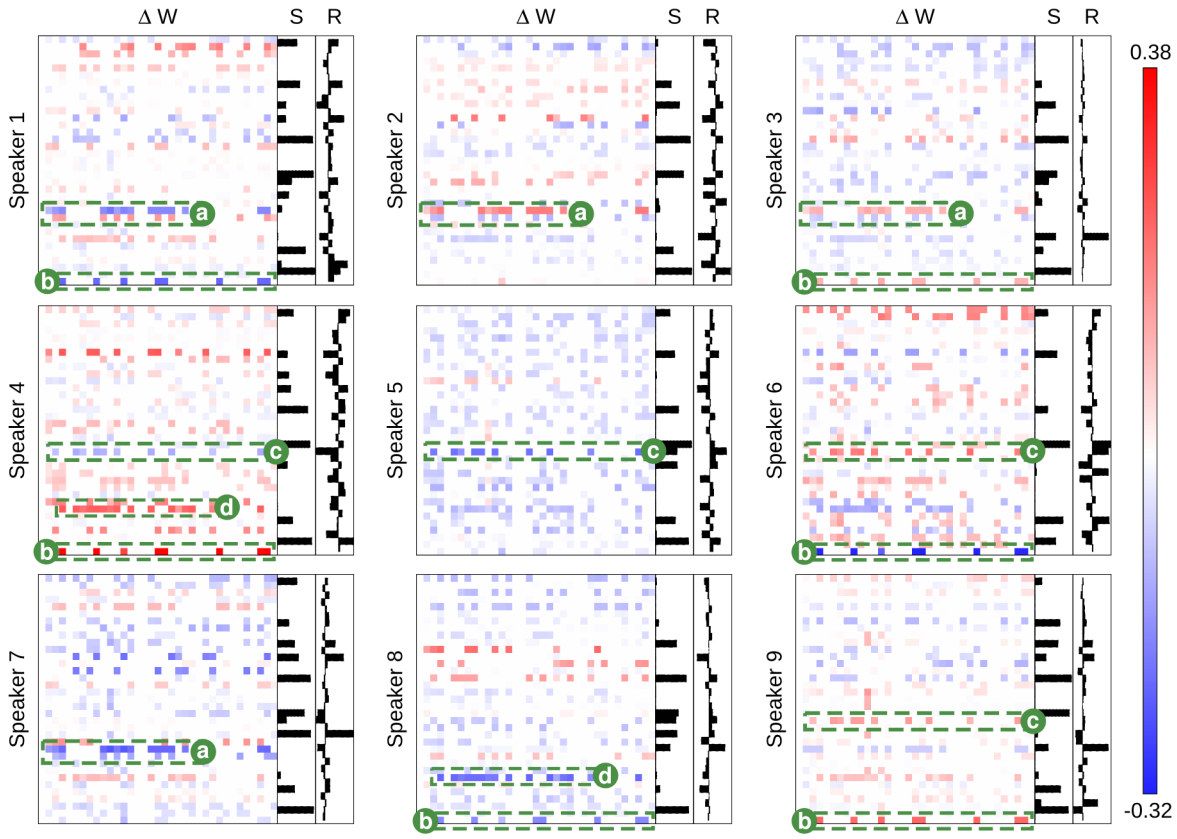

Figure 4. The class-specific synaptic adaptation for the 9 class speaker recognition task under BCM plasticity. The main heat maps in each subplot show the adaptation of the weight matrix (synapses) after the presentation of voice input data from each speaker. Blue values show a reduction in synaptic strength and red values show an increase. The bar-chart, **S**, shows the average neuron activation for each class. The bar-chart,**R**, shows the learned readout weights. Labeled synapses **a,b,c,d** indicate key structural changes that are selective between different speakers. Each label alone can distinguish between two sets of speaker. Taken all together, the labeled synapses adapt specifically to each speaker in a unique pattern, learning a distinct network structure for each one.

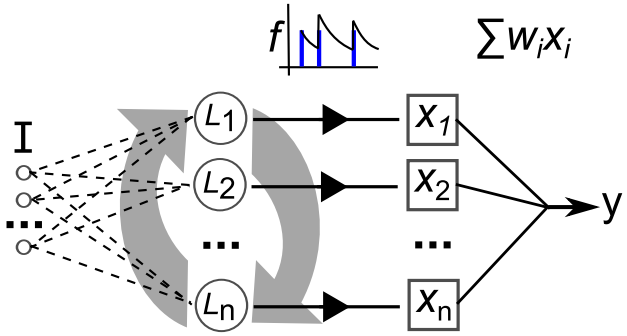

Figure 5. Depiction of the elements of our recurrent network model. ***I*** is a multi-dimensional input signal, ***L*** nodes constitute the recurrent network, the ***x*** vector is the neural activation state, ***f*** is the filtering of the spike trains and ***y*** is the output after weight and sum.