Computational investigation of visually guided learning of spatially aligned auditory maps in the colliculus

TIMO OESS [1,*], MARC O. ERNST[1], HEIKO NEUMANN[2]

[1] *Applied Cognitive Psychology, Ulm University, D-89081 Ulm, Germany*

[2] *Institute of Neural Information Processing, Ulm University, D-89081 Ulm, Germany*

The development of spatially registered auditory maps in the external nucleus of the inferior colliculus in young owls and their maintenance in adult animals is visually guided and evolves dynamically. To investigate the underlying neural mechanisms of this process, we developed a model of stabilized neoHebbian correlative learning which is augmented by an eligibility signal and a temporal trace of activations. This 3-component learning algorithm facilitates stable, yet flexible, formation of spatially registered auditory space maps composed of conductance-based topographically organized neural units. Spatially aligned maps are learned for visual and auditory input stimuli that arrive in temporal and spatial registration. The reliability of visual sensory inputs can be used to regulate the learning rate in the form of an eligibility trace. We show that by shifting visual sensory inputs at the onset of learning the topography of auditory space maps is shifted accordingly. Simulation results explain why a shift of auditory maps in mature animals is possible only if corrections are induced in small steps. We conclude that learning spatially aligned auditory maps is flexibly controlled by reliable visual sensory neurons and can be formalized by a biological plausible unsupervised learning mechanism.

INTRODUCTION

Identifying a location of a visual or auditory event in our environment is advantageous for an organism to orient in space. Localizing a visual stimulus is a rather easy task, since the receptor array (sensorial neurons on a sensory organ) is topographically ordered. Hence, the location of a stimulus can be directly read out by its position on the retina. However, in the auditory domain localizing a stimulus is a complicated task that involves intensive computational steps to overcome two major obstacles. The challenge in sound source localization begins at the stage of the cochlea. The tonotopically organized receptor array represents neighboring frequencies but not adjacent spatial locations. Consequently, the location of a sound source cannot be directly inferred from its array position but needs to be computed using cues created by the head shadow, the distance between the ears or their shape. However, these cues lack associations to absolute locations in space. In order to establish such associations the brain utilizes vision as a guidance signal (Knudsen and Knudsen,

[*] Corresponding author: timo.oess@uni-ulm.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 7: Auditory Learning in Biological and Artificial Systems, August 2019, Nyborg, Denmark. Edited by A. Kressner, J. Regev, J. C.-Dalsgaard, L. Tranebjærg, S. Santurette, and T. Dau. The Danavox Jubilee Foundation, 2019. ©The Authors. ISSN: 2596-5522.

1989). Whenever a visual stimulus appears in temporal registration with an auditory stimulus, an association is established between the spatial location of the visual stimulus and computed auditory localization cues to form a topographically ordered representation of auditory sound source locations. Such associations can be found in the external subdivision of the inferior colliculus of barn owls in the form of a topographically aligned map of auditory space (Knudsen and Konishi, 1978).

Here, we present a neural model of auditory map formation in owls utilizing an unsupervised learning rule that evaluates correlations of sensory input streams. Map structures are defined as 1-dimensional arrays of topographically arranged single compartment neurons. Activations of model neurons are defined by rate-based changes of first-order conductance-based membrane dynamics and their transformation into firing rates. The model architecture is based on physiological findings in barn owls (Knudsen and Brainard, 1991; Hyde and Knudsen, 2000; Linkenhoker and Knudsen, 2002) and incorporates parts of external nucleus of the inferior colliculus (ICx) and the optic tectum (OT) to represent auditory and visual inputs, respectively. Unsupervised learning is defined by a neoHebbian 3-factor rule that incorporates an eligibility control signal, a co-activation plasticity mechanism, and a temporal trace of post synaptic activation (Gerstner *et al.*, 2018). Together, these components enable stable map alignment of auditory space instructed by visual guidance signals.

Simulation results demonstrate the ability of the model to resemble behavioral and physiological findings in barn owls and explain the difference in remapping of auditory space in juvenile and adult owls when their visual field is shifted. Specifically, we show that the ability of remapping critically depends on the receptive field size of auditory neurons. This explains why a gradually induced prismatic shift enables remapping of auditory space in mature animals whereas a single large shift only works for juvenile owls.

METHODS

The alignment of auditory space maps in barn owls occurs at the level of the midbrain between the central nucleus of the inferior colliculus (ICC) and ICx. This alignment is guided by retinotopic visual inputs from the OT. The ICC comprises tonotopically ordered neurons that are responsive to certain values of localization cues such as interaural time or level differences. By combining these cues over different frequency channels and associating them with visual information provided by the OT, the ICx forms a topographical map of auditory space.

Inputs to ICx model neurons at location j arise from ICC (audio) and OT (vision) and are denoted s_j^A and s_j^V, respectively. Each input is a one-dimensional vector of $N = 40$ entries that describes the input conductances to the ICx neuron population of size N. This number is chosen to achieve high input resolution while keeping the computational cost in a feasible range. Visual inputs from the OT are topographically structured whereas inputs from the ICC are tonotopical but show a topographical organization of localization cues (Feldman and Knudsen, 1997). Therefore, independent auditory $s_j^A(x_t)$ and visual $s_j^V(x_t)$ sensory input at location j can be modeled by

$$s_j^{\{A,V\}}(x_t) = \exp(\frac{-(j-x_t)^2}{2 \cdot \sigma_{\{A,V\}}^2}), \qquad \text{(Eq. 1)}$$

where x_t is the location at time t and $\sigma_{\{A,V\}}$ the spatial extent of an auditory and visual stimulus on the receptor array, respectively. To ensure similar input strengths the input vectors are min-max normalized. An input location x is randomly chosen between 0 and N and for each such a location a Wiener process (initial setting: $\sigma_{WP} = 1, \mu = x$) is applied that randomly selects five locations in the vicinity of x. Finally, these five locations are consecutively presented, each for 100 time steps (sufficient time to reach a steady state of the membrane potential), before choosing another initial location x. This process is repeated N times.

One model assumption is an increasing energy level in the visual map of space over time due to maturation of the visual system. This incremental increase of energy of visual inputs is modeled by filtering the visual input signal with a temporal high-pass filter according to $\hat{s}_j^V = s_j^V * f$ ($*$: convolution operator) ,where

$$f(t) = \frac{1}{(1+\exp(-(10^{-4} \cdot t - 1)))}. \qquad \text{(Eq. 2)}$$

This leads to increased plasticity of the learning for more reliable visual input signals. The auditory input signal is fed to an ICx model neuron r_j, whose membrane potential change is defined by:

$$\tau \dot{r}_j = -\alpha \cdot r_j + (\beta - r_j) \cdot \sum_{i=0}^{N} w_{ji} \cdot s_i^A, \qquad \text{(Eq. 3)}$$

where s_i^A describes the auditory input, weighted by connection weight w_{ji} over all input locations. Parameter $\tau = 0.1$ defines the membrane time constant, $\alpha = 1.0$ is a passive membrane leakage rate and $\beta = 1.0$ describes a saturation level of excitatory inputs (standard neuron parameters). To create a firing rate, membrane potential r_j at time t is transformed by an output function $g(r_j(t)) = [r_j(t)]_+ = max(r_j(t), 0)$. Note, that visual inputs do not drive the neuron auditory map neurons in accordance with neurophysiological findings (Knudsen and Knudsen, 1989).

The essential part of the model is a 3-component neoHebbian learning algorithm (Gerstner *et al.*, 2018), that facilitates learning for spatially and temporally aligned inputs. Empirical exploration was used to choose best learning parameter values. Weights are initialized with a large receptive field kernel, $w_{ji} = \exp(\frac{-(i-j)^2}{2 \cdot 20^2})$, so that each receptive field ranges over the entire input space to replicate juvenile owls' receptive fields. The weight adaptation Δw for each learning step is governed by the 3-component structure

$$\Delta w_{ji} = \eta \cdot [post_j(t) \cdot pre_i(t) \cdot fb_j(t) - stabilizer_j(t)], \qquad \text{(Eq. 4)}$$

with $\eta = 0.005$ denoting the learning rate, $post_j(t) = \bar{r}_j(t)$ is the temporal trace value of the map activity r_j, $pre_i(t) = s_i^A$ is the activity of the auditory input neuron and $fb_j(t) = F_j^V(t)$ is an eligibility signal. This eligibility signal is driven by the activity of a spatially coincident visual neuron and the overall energy in the visual map, $F_j^V(t) = \hat{s}_j^V(t) \cdot E(t)$ where is $E(t)$ is the normalized total energy in the visual map and is calculated by $E(t) = \sum_j^N \hat{s}_j^V(t) / \sum_j^N s_j^V(t)$. Here, $\hat{s}_j^V(t)$ is the current activation of the visual input after filtering (see Eq. 2) and $s_j^V(t)$ the maximal input strength before filtering. This mechanism guarantees that learning takes place only for reliable visual signals and for spatially aligned auditory and visual events (Brainard and Knudsen, 1993). The temporal trace value $\bar{r}_j$ is calculated by $\bar{r}_j(t+\delta t) = (1-\lambda) \cdot \bar{r}_j(t) + \lambda \cdot r_j(t)$ to achieve more robust map formation. $\lambda = 0.5$ denotes the trace parameter that defines the influence of previous values of r_j on its current state $\bar{r}_j$.

The term $stabilizer_j(t)$ is added to counterbalance the correlative 3-component learning and is defined by $stabilizer_j(t) = \bar{r}_j^2 \cdot w_{ji}$ to pull weights towards constant energy by a rate proportional to the energy of the map neuron activation. By collecting all terms in the weight adaptation mechanism we arrive at a modified Oja learning rule (Oja, 1989) that incorporates an eligibility control signal to regulate the map learning process.

Connections from one neuron to another can degrade over time which is modeled by a decay function that is applied in each time step: $w_{ji}(t+\delta t) = (1-10^{-6}) \cdot w_{ji}(t)$. Here, $w_{ji}(t)$ can become very small but never 0. However, it is possible that connections eventually vanish completely. We model this process by $w_{ji} = 0$, for $w_{ji} < 0.01$. To compensate for this degeneration the model is endowed with a process that allows for reestablishment of already vanished weights. This is achieved by increasing weight values randomly in close vicinity to weights which values exceed a given threshold:

$$\hat{w}_{ji} = exp(\frac{-(i - \text{argmax}_i(j,i))^2}{\sigma_{syn}^2}),$$

$$\Delta w_{ji} = \begin{cases} \hat{w}_{ji} \cdot |\mathcal{N}(0, \hat{w}_{ji})| \cdot \phi_{syn}, & \hat{w}_{ji} \ \geq t_{syn} \\ 0, & otherwise \end{cases} \qquad \text{(Eq. 5)}$$

where $\sigma_{syn} = 2.8$ defines the range in which new weights can be created, $t_{syn} = 0.5$ is a threshold a weight has to exceed to initiate the process and $\phi_{syn} = 0.1$ defines a scaling factor of the noise. This process is repeated every 1000th time step.

RESULTS

In the following, we present simulation results that, first, show learning abilities of our model for spatially and temporally correctly aligned visual and auditory inputs. This learning of normal responses is used in a second experiment as a reference for comparison with learning for shifted visual inputs. In a third experiment, we show

how the regained ability of shifting an auditory map for incrementally shifted visual inputs in adult owls depends on receptive field size of auditory neurons.

At the beginning of each simulation experiment the auditory system is in its juvenile state (reduced energy in visual map, broad receptive fields of auditory neurons, no auditory map alignment) and develops over the course of $50,000$ time steps to its mature state (maximal energy in visual map, narrow receptive fields of auditory neurons, and supposedly aligned auditory map). To demonstrate correct functionality of the model, we present a learned auditory map for temporally aligned, non-shifted visual and auditory inputs (Fig. 1 **A**). It represents correct map alignment in healthy owls. The abscissa indicates the spatial offset of the alignment from the predicted normal. The predicted normal describes the location offset between auditory and visual signals and is 0° for a perfectly aligned auditory map (spatial coincidence). Data presented here is collected by repeatedly presenting different sound source locations and measuring the response of auditory map neurons.

Temporal coincidence between both inputs determines how well an auditory map can be aligned. For large temporal offsets the map alignment fails due to reduced activity of map neurons at the moment the visual signal would facilitate learning. Temporal coincidence is especially crucial if stimulus locations are randomly sampled from the environment (high σ_{WP} value of Wiener process). However, if consecutive auditory visual events are in spatial vicinity (they are spatially correlated, low σ_{WP} value) correct map alignment is still possible even for large offsets (Fig. 1 **E**).

Experiments with juvenile owls show that if a constant shift of visual inputs is introduced by prismatic goggles, the alignment of the auditory space map is shifted accordingly. This indicates a visually guided learning of auditory space (Brainard and Knudsen, 1993). We replicate this experiment by inducing a shift of the visual inputs by 10° or 20°, respectively, right at the beginning of the learning. Through its role as a guidance signal, the visually shifted input leads to a shifted alignment of the auditory space map (Fig. 1 **B**). However, this shift only occurs when introduced in juvenile owls, but remains ineffective when tested with adult owls (Linkenhoker and Knudsen, 2002). Our model exhibits the same behavior when presented with a non-shifted visual input during development (until time step $100,000$ to simulate sufficient adult experience) followed by a 15° shifted visual input. Map realignment fails since the shifted input is outside the receptive field range of auditory neurons (Fig. 1 **C**). Unlike the ineffective alignment modification in case of a single large shift, incremental small shifts in the visual input can lead to map realignment in adult owls. Simulations with multiple small incremental shifts of the visual signal demonstrate that our model is able to realign the auditory map of space in each consecutive step (Fig. 1 **D**).

Our results indicate that this phenomenon can be ascribed to the receptive field size of auditory map neurons. According to Hebbian learning, new connections and thereby realignment can only be learned if there is a temporally coincident pre- and postsynaptic activity of neurons (Hebb, 1949). If a large shift is introduced, the activity location of postsynaptic neurons in the auditory map and of visual input neurons do not coincide, since the stimulus location is outside of the spatially aligned auditory neuron's receptive field. In contrast, if small shifts are introduced the stimulus location might still be in the receptive field range of the auditory neuron which leads to an activation and therefore a relearning of connections. We tested this hypothesis by measuring the maximal shift step size depending on the receptive field size of the auditory neurons. For each receptive field size (range $[0.5\cdot\sigma_A, 7\cdot\sigma_A] \approx [4px, 40px]$) different shift steps (range $[5°, 25°] = [5px, 25px]$) are tested and the maximal possible shift is measured (a shift is successful if the activity location of the map neurons corresponds to the induced shift). In total, we ran 8 simulations and calculated the

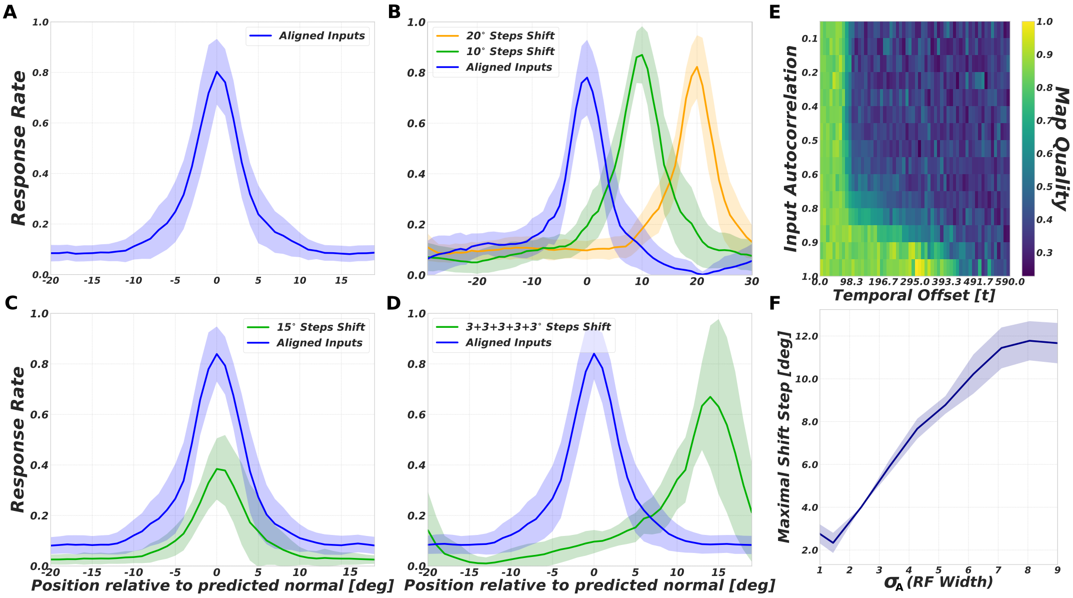

Fig. 1: Left and middle column. Population tuning curves of ICx neurons to different auditory stimulus locations relative to each neuron's predicted normal response are shown (as in Knudsen (1998)). Normal responses of a neuron are derived from the learned map for temporally aligned, non-shifted inputs. Lines depict mean of model neuron responses. Standard deviation in colored area. In all plots the blue line is the normal response plotted for easier comparison. **A** depicts model responses for temporally aligned, non-shifted inputs. **B** shows responses for various shifts of visual inputs. In **C** green line plots model responses for a shifted visual input of 15° in adult animals. **D** shows model responses for incremental shifted inputs in adult animals. Right column. **E** map alignment quality over autocorrelation value of inputs (ordinate) and their temporal offset (abscissa). Perfect map alignment is indicated in bright yellow, no alignment in dark blue. **F** depicts maximal shift over receptive field width when shift is introduced at a mature state.

mean maximal shift for each receptive field size. For increasing receptive field size the maximal shift step becomes proportionally larger (Fig. 1 **F**).

DISCUSSION & CONCLUSION

We presented a neuron model that learns aligned auditory maps of space by applying a 3-component neoHebbian correlative learning rule to its visual and auditory inputs. Previous investigations of auditory map alignment demonstrated that alignment is possible with a *map adaptation cue* in a spiking neuron model (Huo and Murray, 2009) or with a simple, unconstrained Hebbian learning rule for visual and auditory inputs (Witten *et al.*, 2008). Our model differs in that it directly uses visual inputs as a guidance signal for map alignment and is able to explain the unaffected map alignment in adult owls when a large prismatic shift is induced versus the ability to realign for incremental shift step size.

Simulation results demonstrated the model's ability to successfully learn aligned auditory maps of space for temporally and spatially aligned visual and auditory sensory inputs. Due to the introduced eligibility signal, visual inputs do not drive responses of auditory map neurons but merely elicit learning for coincident stimuli, which is important for close replication of biological findings.

The incorporated trace rule can compensate for small temporal offsets between the visual and auditory inputs. However, for large temporal offsets, results indicate that the success of learning strongly depends on the spatial autocorrelation of the input locations. That is, if auditory and visual inputs are not randomly sampled in space but are chosen according to a Wiener process with small σ_{WP}, correct map alignment is still possible. This implies that learning is enhanced for stimulus locations that show strong autocorrelation. When transferring our results to real world scenarios, this enhancement is of special interest since audio-visual events rarely happen to occur at just a single location, followed by other random locations but are likely to happen in spatial vicinity. Therefore, such an enhanced learning capability for strongly autocorrelated input locations seems to facilitate learning of real world stimuli.

It has been shown experimentally that the ability to shift the auditory map alignment is reduced for adult animals (Linkenhoker and Knudsen, 2002). Our model results predict that the receptive field size of the auditory neurons is responsible for this reduction and it can serve as an index of maximal shift step size for visual inputs that can still induce map realignment. This prediction could be tested by varying the prismatic shift step size and determining the receptive field size for adolescents of different ages. If our prediction is correct the two values should correlate.

Despite the presented results, at the moment our model is incapable to maintain already established maps and to quickly readjust to normal vision as it has been demonstrated for owls in (Knudsen, 1998). However, we argue that this could be achieved by adapting the learning algorithm and extend the input by elevation cues. Together with a N-methyl-D-aspartate (NMDA) receptor signal which in addition to the eligibility signal could control the learning, we expect the model to further extend its capability to resemble neurophysiological and behavioral studies.

REFERENCES

Brainard, M.S. and Knudsen, E.I. (**1993**). “Experience-dependent plasticity in the inferior colliculus: A site for visual calibration of the neural representation of auditory space in the barn owl,” J. Neurosci. doi:10.1523/JNEUROSCI.13-11-04589.1993.

Feldman, D.E. and Knudsen, E.I. (**1997**). “An Anatomical Basis for Visual Calibration of the Auditory Space Map in the Barn Owl’s Midbrain,” J. Neurosci. doi:10.1523/JNEUROSCI.17-17-06820.1997.

Gerstner, W., Lehmann, M., Liakoni, V., Corneil, D., and Brea, J. (**2018**). “Eligibility Traces and Plasticity on Behavioral Time Scales: Experimental Support of NeoHebbian Three-Factor Learning Rules,” Frontiers in Neural Circuits. doi:10.3389/fncir.2018.00053.

Hebb, D.O. (**1949**). *The Organization of Behavior: A Neuropsychological Theory* (Psychology Press). ISBN 978-1-135-63191-8.

Huo, J. and Murray, A. (**2009**). “The adaptation of visual and auditory integration in the barn owl superior colliculus with Spike Timing Dependent Plasticity,” Neural Netw. doi:10.1016/j.neunet.2008.10.007.

Hyde, P.S. and Knudsen, E.I. (**2000**). “Topographic projection from the optic tectum to the auditory space map in the inferior colliculus of the barn owl,” J. Comp. Neurol. doi:10.1002/(SICI)1096-9861(20000529)421:2⟨146::AID-CNE2⟩3.0.CO;2-5.

Knudsen, E.I. (**1998**). “Capacity for Plasticity in the Adult Owl Auditory System Expanded by Juvenile Experience,” Science. doi:10.1126/science.279.5356.1531.

Knudsen, E.I. and Brainard, M.S. (**1991**). “Visual instruction of the neural map of auditory space in the developing optic tectum,” Science. doi:10.1126/science.2063209.

Knudsen, E.I. and Knudsen, P.F. (**1989**). “Vision calibrates sound localization in developing barn owls,” J. Neurosci. doi:10.1523/JNEUROSCI.09-09-03306.1989.

Knudsen, E.I. and Konishi, M. (**1978**). “A neural map of auditory space in the owl,” Science. doi:10.1126/science.644324.

Linkenhoker, B.A. and Knudsen, E.I. (**2002**). “Incremental training increases the plasticity of the auditory space map in adult barn owls,” Nature. doi:10.1038/nature01002.

Oja, E. (**1989**). “Neural networks, principal components, and subspaces,” Int. J. Neural Syst. doi:10.1142/S0129065789000475.

Witten, I.B., Knudsen, E.I., and Sompolinsky, H. (**2008**). “A Hebbian Learning Rule Mediates Asymmetric Plasticity in Aligning Sensory Representations,” J. Neurophysiol. doi:10.1152/jn.00013.2008.