A neurocomputational amygdala model of auditory fear conditioning: A hybrid system approach

Nicolás Navarro-Guerrero
Knowledge Technology Group
University of Hamburg
Department of Computer Science
Vogt-Koelln-Str. 30
22527 Hamburg, Germany
navarro@informatik.uni-hamburg.de

Robert Lowe
Cognition & Interaction Lab (COIN)
Informatics Research Centre
University of Skövde
PO Box 408
54128 Skövde, Sweden
robert.lowe@his.se

Stefan Wermter
Knowledge Technology Group
University of Hamburg
Department of Computer Science
Vogt-Koelln-Str. 30
22527 Hamburg, Germany
wermter@informatik.uni-hamburg.de

***Abstract*—In this work, we present a neurocomputational model for auditory-cue fear acquisition. Computational fear conditioning has experienced a growing interest over the last few years, on the one hand, because it is a robust and quick learning paradigm that can contribute to the development of more versatile robots, and on the other hand, because it can help in the understanding of fear conditioning and dysfunctions in animals. Fear learning involves sensory and motor aspects [1] and it is essential for adaptive self-protective systems. We argue that a deeper study of the mechanisms underlying fear circuits in the brain will contribute not only to the development of safer robots but eventually also to a better conceptual understanding of neural fear processing in general. Towards the development of a robotic adaptive self-protective system, we have designed a neural model of fear conditioning based on LeDoux's dual-route hypothesis of fear [2] and also dopamine modulated Pavlovian conditioning [3]. Our hybrid approach is capable of learning the temporal relationship between auditory sensory cues and an aversive or appetitive stimulus. The model was tested as a neural network simulation but it was designed to be used with minor modifications on a robotic platform.**

I. Introduction

Pavlovian fear conditioning is a form of emotional learning in which a neutral or innocuous stimulus (conditioned stimulus or CS) such as a sound or light, is paired with an aversive stimulus (unconditioned stimulus or US) such as an electric shock. Animals are evolutionarily hard-wired to rapidly acquire, consolidate, and generalize fear memories. After only a few trials, animals quickly learn to anticipate the aversive US using the CS information and elicit behavioral defense responses and associated autonomic and endocrine adjustments [4].

The amygdala (AMG) plays a crucial role in affective and self-protective systems. The AMG represents the affective/emotional valence of a situation, a "state value" necessary for coordinating physiological, behavioral and cognitive responses. Furthermore, recent evidence suggests that the human amygdala, in addition to its important role in cue fear conditioning, contributes to many reward-based decision-making tasks [5]. Understanding and modeling these mechanisms is of great interest not only for the interpretation of neuropsychological findings, but also for computational modeling and building safer robot assistants [6], [7].

A detailed literature review reveals that the most meaningful and closely related publications in recent years are based on mathematical models of the cue-dependent fear conditioning dynamics of *acquisition* and *extinction*. Many of these models are based on the dual-route hypothesis (cortical and subcortical) proposed by LeDoux [2], [8]–[14], which explains parallel processing of stimuli at different degrees and temporal response improvements. Often, they use simplified binary or abstract numerical inputs [3], [10], [12]–[15] neglecting unforeseen sensory and temporal relationships that may be relevant for fear learning dynamics.

An anatomically constrained model has been suggested by Armony et al. [8] to investigate information processing in the two afferent fear conditioning pathways indicated by LeDoux [2]: One originating in the auditory cortex (cortical pathway), the other in the auditory thalamus (subcortical pathway). The model, however, has been revealed to have several shortcomings [16] when considered as a standalone model of fear circuitry. It also fails to deal with the temporal association of stimuli.

A more comprehensive model for emotional learning including acquisition, extinction, habituation and blocking was presented by Balkenius and Morén [10]. The model focuses on the interaction between the amygdala and the orbitofrontal cortex (OFC), where the former serves as the locus for acquisition and the latter for inhibition of emotional responses as produced by the amygdala. However, since some modules only consist of linear units there is a need to extend and refine the proposed model for reasons of biological plausibility and computational flexibility.

Most recent research [12], [13] has included models of context fear conditioning. Vlachos et al. [12] presented a biologically plausible spiking neural model of the basal amygdala (BA) for fear memory encoding. Despite the detailed model of the basal nucleus, sensory input pathways for cue and context information as well as interaction with downstream structures were neglected in this work. Krasne et al. [13] presented a firing rate-coded model of three amygdala sub-nuclei. In contrast to Vlachos' work, Krasne's research addresses fear

conditioning in a more integrative manner, modeling not only one amygdala's input nucleus but also nuclei involved in the expression of fear responses. Despite the completeness of this model in terms of the broad dynamics captured such as fear acquisition, consolidation, extinction, etc., they used an abstraction of sensory and contextual information. Furthermore, no model of other areas involved in fear conditioning, such as the thalamus, was included.

In general, there is a lack of research on amygdala modeling with realistic sensory input taken from a realistic physical environment. In one rare example, however, Mannella et al. [15] addressed cue conditioning in a simulated robot experiment. The model is able to reproduce and demonstrate, with a simulated rat, experiments of first and second order conditioning and devaluation. Alexander and Sporns [17] and Zhou and Coggins [18] conducted research on prediction learning and conditioning with real Khepera robots, but only from a normative, rather than a neurocomputationally realistic, viewpoint. These models consist of feed-forward networks with a very abstract timing model, only coarsely mapped to neurobiological circuits and do not capture as rich a variety of dynamics as other works [12], [13], [15], but the embodied approach makes them attractive and their relative success encourages the development of more sophisticated and biologically plausible embodied models.

Our paper presents a biologically motivated model of auditory-cue fear conditioning. The model neurocomputationally describes the known thalamic and auditory cortex routes plus reward learning based on phasic dynamics of dopamine, previously described by Lowe et al. [3]. We propose to study fear conditioning taking into consideration bio-plausible sensory pathways and interpretable real-world sensory input. As a midterm goal we are aiming to develop a computational robotic model able to process real sensory input to endow robots with an adaptive method for self-protective action selection. This could represent an important step towards more biologically plausible computational models of fear processing and allow testing on a humanoid robot demonstrator. Applications of this learning mechanism may be used in artificial self-protective systems to predict both appetitive and aversive behavioral outcomes, or in the modulation of complex behaviors such as autonomous battery recharging [19].

II. Model Architecture and Learning

The amygdala is a brain region in the medial temporal lobe composed of diverse nuclei. Since the amygdala is not a single brain structure or region it has historically been defined on the basis of connection density, chemical signature and configuration. An initial coarse division may consist of the basolateral complex (BLA) and the central nucleus (CE) [4]. The BLA is the main input structure in the amygdala and receives sensory information from many cortical and subcortical regions. The BLA consists of three nuclei: the lateral (LA), basolateral (BL) and basomedial (BM) also known as accessory basal (AB). Almost 80% of BLA neurons are glutamatergic cells (GLU) having multiple projections to neighboring cells, amygdala

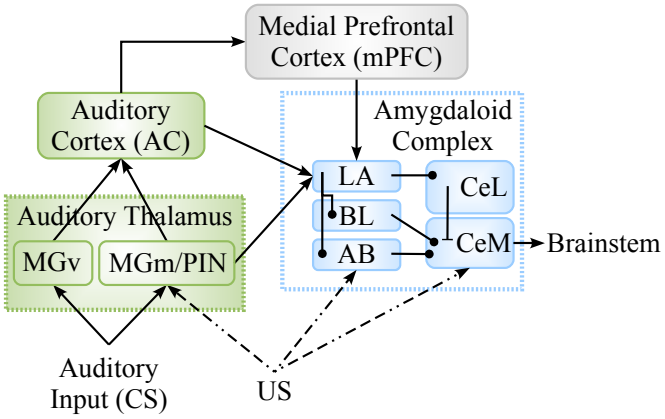

Fig. 1. Main inputs to the amygdala and intranuclear pathways of the amygdala involved in cue fear conditioning. MGv, ventral division or lemniscal component of the medial geniculate body; MGm, medial division of the medial geniculate body; LA, lateral nucleus; BL, basolateral nucleus; AB, accessory basal nucleus; CeL, lateral division of the central nucleus; CeM, medial division of the central nucleus; US, unconditioned stimulus. Adapted from [1], [4], [20].

nuclei, and other brain structures. The remaining 20% are GABAergic cells (GABA) of short axons regarded as local-circuit neurons [1]. In contrast, the CE is recognized as the main output component from the amygdala, modulating both cortical and subcortical structures and controlling the selection of passive and active fear reactions [20]. The CE mainly GABAergic in nature can be divided into a lateral (CeL) and a medial (CeM) part [1]. Many comprehensive reviews on the structure, connectivity and influences of amygdaloid and fear conditioning dynamics can be found, e.g. [1], [4], [21].

Although fear conditioning is ubiquitous to all sensory modalities, most progress has been made on auditory-cue fear conditioning, which is why we based our model on this paradigm. The standard dual-route hypothesis suggested by LeDoux [2] identifies the medial geniculate body (MGB) of the thalamus as the subcortical auditory pathway to the amygdala. Specifically, the medial division of the MGB (MGm) and the posterior intralaminar nucleus (PIN) project to the primary and association areas of the auditory cortex, and also to the lateral nucleus of the amygdala. The MGm/PIN complex is considered as an auditory and somatosensory relay to the LA [22]. The MGm is highly multimodal responding to auditory, tactile, thermal and nociceptive stimulation. With respect to auditory input, MGm lacks tonotopic organization. PIN is also multimodal. In contrast, the ventral division of the MGB (MGv) specializes in auditory stimuli, it has a tonotopic organization and it is identified as the main subcortical route to the primary auditory cortex [22]. More precise information about the auditory CS seems to indirectly reach the LA via the primary auditory and the associative cortex [23]. It is likely that this information includes fine frequency tuning, abstraction of pitch and pattern discrimination, among other possible functions [24]. The medial prefrontal cortex (mPFC) has also important projections to the amygdala. Although the mPFC projects to all amygdaloid nuclei, the connections to LA

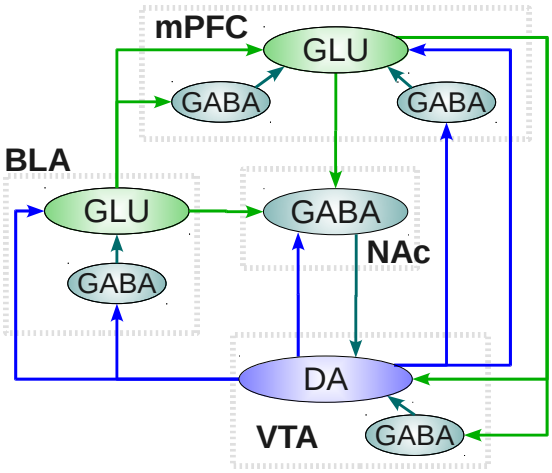

Fig. 2. Schematic illustration of stress-responsive projections between NAc, mPFC and BLA that are involved in fear conditioning. NAc, nucleus accumbens; mPFC, medial prefrontal cortex; VTA, ventral tegmental area; BLA, basolateral complex of the amygdala; GLU, glutamatergic cells; GABA, GABAergic cells; DA, dopaminergic cells. Adapted from [25].

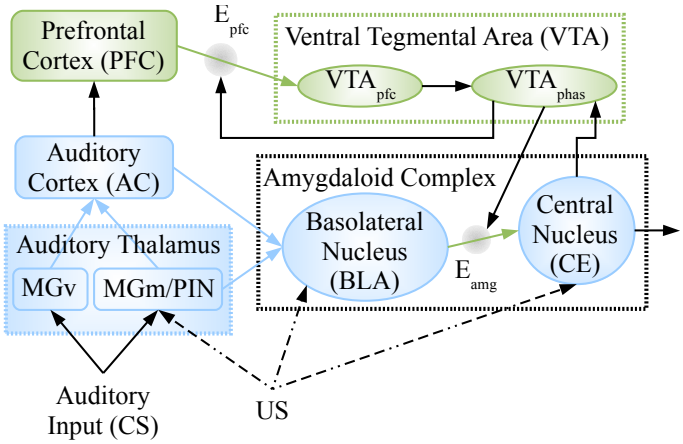

Fig. 3. System's architecture overview, based on Armony et al. [8] auditory fear conditioning model and Lowe et al. [3] dopamine modulated Pavlovian conditioning. Black arrows represent fixed weight values. Blue arrows represent weights updated using the Stent-Hebb rule [28]. Weights values represented by green arrows are updated using the Hebbian learning rule and an eligibility trace.

seem to be the more important [1]. In turn the BA projects back to the mPFC. Moreover, the mPFC plays a key role in extinction of fear conditioning affecting ITCm cells blocking the excitation of CeM neurons through the BA [1].

Dopamine dynamics (DA) are thought to be involved in the coordination of different stress responses. Stress-induced dopamine release allows animals to relocate attention, prioritize perceptual processing and is involved in appropriate action selection [25]. A broad body of research links stress-responsive dopamine projections from the ventral tegmental area (VTA) to the basolateral complex of the amygdala (BLA) with fear conditioning [25], [26]. In turn glutamatergic projections from the amygdala to the nucleus accumbens (NAc) and medial prefrontal cortex (mPFC) regulates dopamine stress responses in Nac, mPFC and VTA. The amygdala also mediates further autonomic, endocrine and behavioral responses to emotionally significant stimuli [21]. Fig. 2 shows the interplay between mPFC, NAc, VTA and BLA during stress-responsive dopamine release.

We are aiming at developing a computational architecture to endow humanoid robots with an adaptable self-protective mechanism. At this stage, the overall architecture, shown in Fig. 3, is intended to capture both phasic dynamics of dopamine and input and output pathways underlying fear conditioning learning. This hybrid architecture combines Hebbian components (blue modules) for association, and a recurrent component (green modules) for reward prediction. It was designed to be portable to a real NAO robot [27] working in a home-like environment, see Fig. 4.

In order to emulate auditory-cue fear conditioning experiments, the model is fed with synthetic audio signals which consist of single tones plus a very low amplitude (about 2.5%) noise floor from measurements in our home lab, which are sampled at 48 KHz by the NAO robot. We process the incoming signal in frames of approx. 21 ms (1024 samples), which

Fig. 4. Conceptual scenario for further testing and development.

corresponds to the physiological construction of receptive fields [8]. In this way, the system response to a given input may be interpreted as the time-averaged response of a cell to a tone presented in this 21 ms window or time step. For each window, we compute the spectral amplitude of the signal using a short-time Fourier transform (STFT). The entire available frequency range of 20 Hz to 20 KHz, which corresponds to the NAO's microphones' characteristic, is divided into 24 intervals. Each interval is represented by one neuron in an auditory input layer. The neural activation corresponds to the sum of all spectral amplitudes in the corresponding frequencies' interval, and is then normalized to $[0, 1]$. We implemented a simple signal detector module that detects signal onset and ending based on the root mean square (RMS) value of the incoming signal and RMS value of the noise floor. This information is used to generate the unconditioned stimulus (US) just for the desired neutral conditioned stimulus (CS).

The auditory thalamus is modeled based on Armony et al.'s model of the medial geniculate body (MGB), which is also

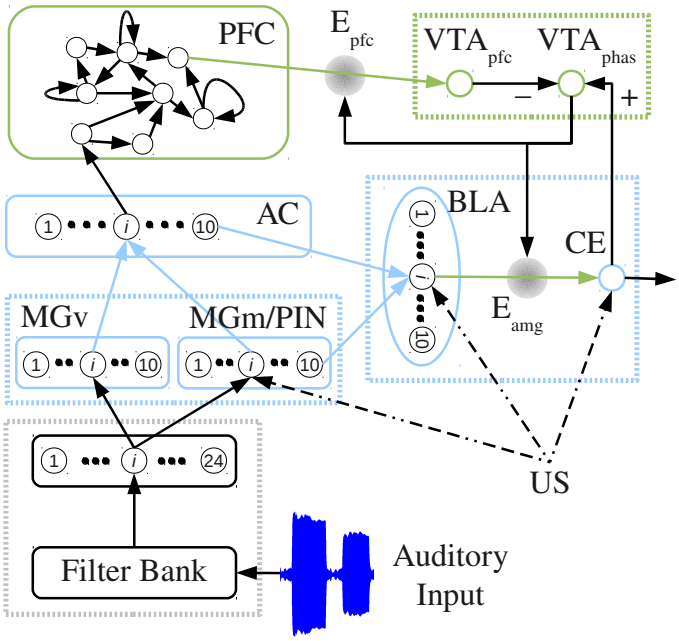

Fig. 5. Neural network architecture overview. Based on an echo state network (PFC), online learning algorithm for echo state network readout layer and amygdala internal connections indicated with E_{pfc} and E_{amg}, and single-layer feed-forward neural networks. For simplicity only one connection per layer in shown. Black lines represent fixed weight values.

supported by Weinberger [22]. The ventral division (MGv) of the MGB with a tonotopic organization feeds the auditory cortex module. The medial division (MGm) and the posterior intralaminar nucleus (PIN) are merged in a single module, which makes an initial association between CS and US and then forwards its output to the amygdala's basolateral complex (BLA) and the auditory cortex modules.

The "auditory thalamus" (MGv and MGm/PIN), "auditory cortex" (AC) and "basolateral complex" (BLA) modules are based on the model described by Armony et al. [8]. Each structure is modeled by a single-layer neural network and the modules' connectivity is done in a feed-forward manner, see Fig. 5. The output of each of these modules is proportional to the output of the sending layer and normalized through both a squashing function and a winner-take-all algorithm that serves to laterally inhibit the activation of less active or "loser" neurons. From this point on we use f and g to denote a linear squashing function that trims neural activation to the interval $[0, 1]$ and $[-1, 1]$ respectively. For all equations, time dependence (t) is omitted and it is just indicated when it is different from the current time step. The activation of the winning unit a_{win} in the receiving module is computed as follows:

$$a_{win} = f\left(\sum_{j \in S} a_j \cdot w_{ji}\right), \tag{1}$$

where S are all units in the sending layer(s) and w_{ij} is the weight between the sending unit j and the current unit i.

TABLE I
SUMMARY OF PARAMETERS USED IN *MGv, MGm/PIN, AC, BLA* AND US SIGNAL.

Variable name	Value	Description
ϵ	0.2	common learning rate value
w_{us}	0.4	fixed weight value for US connections
μ for MGv	0.1	lateral inhibition in MGv module
μ for MGm/PIN	0.3	lateral inhibition in MGm/PIN module
μ for AC	0.6	lateral inhibition in AC module
μ for BLA	0.1	lateral inhibition BLA module
MGv size	10	number of units in module
MGm/PIN size	10	number of units in module
AC size	10	number of units in module
BLA size	10	number of units in module

Connection weights between modules are randomly initialized. The activation for each unit i in the receiving module r is calculated as follows:

$$a_i = f\left(\sum_{j \in S} a_j \cdot w_{ji} - \mu_r \cdot a_{win}\right), \tag{2}$$

where μ_r is the strength of the lateral inhibition in module r. Connection weights are updated after each cycle or epoch using the Stent-Hebb rule [28], which prevents weights saturation:

$$w_{ji}^{'} = \begin{cases} w_{ji}(t-1) + \epsilon \cdot a_i \cdot a_j, & \text{if } a_j > \overline{a} \\ w_{ji}(t-1), & \text{otherwise,} \end{cases} \tag{3}$$

and

$$w_{ji} = \frac{w_{ji}^{'}}{\sum_{j \in S} w_{ji}^{'}}, \tag{4}$$

where $\overline{a}$ is the mean activation of the sending layer and ϵ is the learning rate. Table I summarizes the parameters used for the auditory thalamus, the auditory cortex and the basolateral complex modules [29], which were determined with empirical trials and based on Armony et al.'s [8] results.

The modules representing the "prefrontal cortex" (PFC), "ventral tegmental area" (VTA) and amygdala's "central nucleus" (CE) are based on the model described by Lowe et al. [3]. The interactions of these three modules capture the basic functionality of biological reward prediction learning. This part of the model is based on an echo state network (ESN) approach [30], [31]. ESNs are three-layered recurrent neural architectures that have demonstrated to be particularly effective at processing temporal stimuli. Their main particularity is that only the *readout* weights are updated, which in this case are the weights connecting PFC units with a VTA_{pfc} unit. PFC is the reservoir of our ESN. This reservoir is sparsely connected with randomly generated weights. The reservoir has to satisfy the so-called echo state property that guarantees damping reverberations of the input signals, for details see Jaeger's report [31]. The input layer of our ESN corresponds to the auditory cortex units, which are connected to the PFC

TABLE II
SUMMARY OF PARAMETERS USED IN *PFC*, *VTA* AND *CE* MODULES.

Variable name	Value	Description
Reservoir size	40	number of units in module
Reservoir connectivity	25%	random weights w_{dr} in $[-1, 1]$
Spectral radius	0.95	reservoir spectral radius
Input connectivity	25%	random weights w_{in} in $[0, 1]$
κ	0.1	learning rate
η	0.075	learning rate

using fixed weights, randomly and sparsely generated. Table II summarizes ESN parameters.

The VTA_{pfc} neuron corresponds to the readout layer of the ESN. To improve biological plausibility Lowe et al. [3] introduced two features in the use of ESN. First, they only allow non-negative activation within the reservoir. Second, they use a *"phasic dynamics of dopamine"* (DA) based online learning rule to update the readout weights, see Lowe et al. [3] for details. Weights of the ESN readout layer and amygdala's CE neuron are updated using the Hebbian learning rule and an eligibility trace E_{pfc} and E_{amg} respectively. E_{pfc} and E_{amg} are computed as follows:

$$E_k = max\left[\text{incoming signal}, \Omega \cdot E_k(t-1)\right], \tag{5}$$

where k substitutes for pfc and amg, Ω $(= 0.9)$ is a decay constant. The PFC readout weights w_{pfc_i} connecting the reservoir unit i to the VTA_{pfc} unit and the weights w_{bla_i} connecting BLA units to the CE unit are updated as follows:

$$w_{pfc_i} = \begin{cases} f\left(w_{pfc_i}(t-1) + \right. \\ \qquad \left. \kappa \cdot VTA_{phas} \cdot E_{pfc}(t-1) \cdot PFC_i\right), \\ \qquad\qquad \text{if } VTA_{phas} \geq 0 \\ f\left(w_{pfc_i}(t-1) + \kappa \cdot VTA_{phas} \cdot PFC_i\right), \\ \qquad\qquad \text{if } VTA_{phas} < 0 \end{cases} \tag{6}$$

$$w_{bla_i} = \begin{cases} f\left(w_{bla_i}(t-1) + \eta \cdot VTA_{phas} \cdot E_{amg} \cdot CE\right), \\ \qquad\qquad \text{if } VTA_{phas} \geq 0 \\ f\left(w_{bla_i}(t-1) + \eta \cdot VTA_{phas} \cdot E_{amg}\right), \\ \qquad\qquad \text{if } VTA_{phas} < 0 \text{ and } US = 0 \end{cases} \tag{7}$$

PFC_i is the current activation of reservoir unit i. VTA_{phas} is the output value of VTA module. CE is the output value of the amygdala module. The current activation value of PFC_i, VTA_{phas} and CE are computed as follows:

$$VTA_{phas} = g\left(CE - VTA_{pfc}\right), \tag{8}$$

$$PFC_i = tanh\left(\sum_{i,j} w_{dr_{ij}} \cdot PFC_j(t-1) + \sum_{ik} w_{in_{ik} \cdot AC_{out_k}}\right), \tag{9}$$

$$CE = f\left(US \cdot w_{us} + \sum_i BLA_i \cdot w_{bla_i}\right). \tag{10}$$

Eq. 9 shows the recursive nature of ESN. This property provides a short-term memory that is used for updating estimates of the value of the stimulus. This spatial-temporal relationship between input signals differs from the classical temporal difference learning rule but the system as a whole allows for temporal dynamics between stimuli to be captured.

III. EXPERIMENTAL RESULTS

For the experimental part the weights of Armony-based modules (MGv, MGm/PIN, BLA and AC) are randomly generated with values ranging in $[0, 1]$. The weights connecting the BLA units and the CE unit are set to 1 divided by the number of BLA units. PFC (ESN) readout weights are randomly initialized with values from $[0, 1]$. Although we try to process sensory input within bio-plausible time windows, we do not consider latencies in signal processing nor transmission. Instead, we only consider coincident convergence of subcortical and cortical information to the BLA, which seems to be around 15 ms as reported by Johnson et al. [23]. This is translated to the following dataflow in our current implementation: the auditory input is first preprocessed by the filter bank, then by the auditory thalamus followed by the auditory cortex. Both modules (AC and MGm/PIN) feed simultaneously the BLA module and an affective state is generated at the CE. The amygdala output in conjunction with the auditory cortex activation is then used to trigger a dopamine modulation in the amygdala via the PFC and VTA modules.

The experimental part was divided into two phases. The first phase, called *"development"*, allows the modules to define initial receptive fields for the frequencies (system's activation to a determined frequency), facilitates conditioning and reduces transient effects - that may emerge due to initial random initialization - during conditioning [8]. A number of randomly generated tones not paired with the US was presented to the Armony-based modules [8]. We added white noise to the generated signals to emulate a real robot's recordings. The dopamine circuit modules (PFC and VTA) were switched off. We repeated this procedure varying the frequency ranges, number of frequencies and signal lengths not detecting major changes. Based on Lowe et al.'s [16] findings we decided to use 300 randomly generated tones ranging from 100 Hz to 12 KHz for 5 time steps (approx. 100 ms per tone). Fig. 6 shows an example of a receptive field obtained after development. In this phase the CE output is characterized by a weak activation $(< 3 \cdot e^{-4})$ with similar activation profiles at all frequencies.

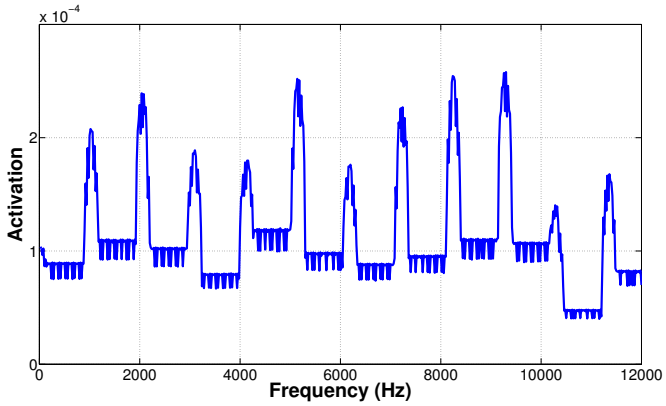

Fig. 6. Amygdala's (CE) receptive fields after development phase, i.e. no frequency has yet been paired with any US signal. The figure shows the CE activation after presenting single tones in the range [20, 12,000] Hz in a 20 Hz interval.

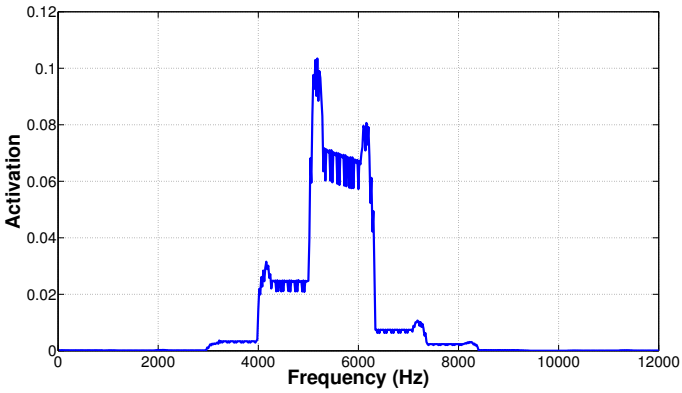

Fig. 7. Amygdala's (CE) receptive fields after conditioning phase, without US signal (6 KHz). Consistently with animal experiments [32], [33], amygdala activation decreases inversely with the distance to the CS (signal paired with the US). The figure shows the CE activation after presenting single tones in the range [20, 12,000] Hz in a 20 Hz interval.

In the second phase, termed *"conditioning"*, we selected an auditory input (CS) of 6 KHz (one eighth of the sampling rate), which was then paired with a binary US signal. Conditioning lasted 300 trials of 4 time steps each. CS and US were presented respectively at trial 75, 150, 225, and 300 and the US was delayed in two time steps. For the remaining trials a randomly generated tone without US signal was used. An example of receptive fields is presented in Fig. 7, where an enhancement of the system's response to the conditioned and neighboring frequencies can be seen. Similar results were observed in animal experiments, what is known as stimulus generalization [32], [33]. Stimulus generalization has been interpreted as crucial for survival since it can elicit fast defensive responses under ambiguous sensory stimulation [34]. We observed that the system's response is higher for frequencies in the range $[5100, 5500]$ Hz than for the CS frequency. This phenomenon is due to resolution lost when converting spectral intensity to neural activation. Since 6 KHz divides two intervals the spectral magnitude contributes to the activation of two neurons, i.e. intervals $[5000, 6000]$ Hz and $[6000, 7000]$ Hz respectively. This sort of ambiguous activation is encountered only for frequencies that divide two intervals. We believe that using a different discretization procedure, such as a gammatone filter along with a greater number of intervals, may help to address this issue in future implementations.

Fig. 8 shows the system's receptive fields when both the CS and US signals are presented. The CE activation is a combination of both the direct influence of the US signal (40%) and the dynamic magnification of the BLA response.

Fig. 9 shows the temporal changes of the system activation after conditioning. The maximal system activation is reached when both the CS and the US are presented at the same time. When only the CS is presented, the system's response is the result of the combination of the receptive field and the VTA modulation and a weak but consistent anticipatory system response is obtained, which is around 10% of the maximal possible activation. This output could easily be used to trigger a conditioned behavior.

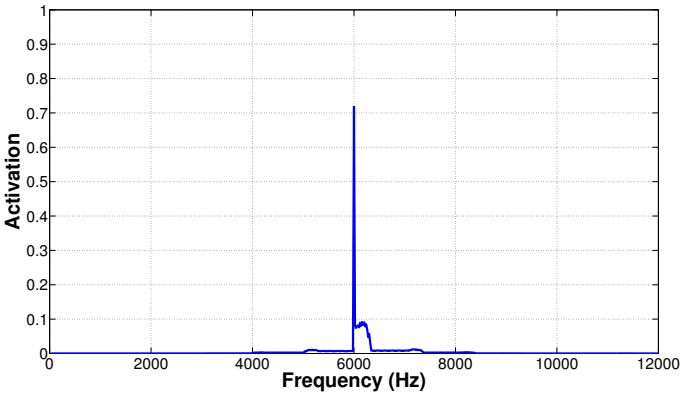

Fig. 8. Amygdala's (CE) receptive fields after conditioning phase, with US signal. The figure shows the CE activation after presenting single tones in the range [20, 12,000] Hz in a 20 Hz interval.

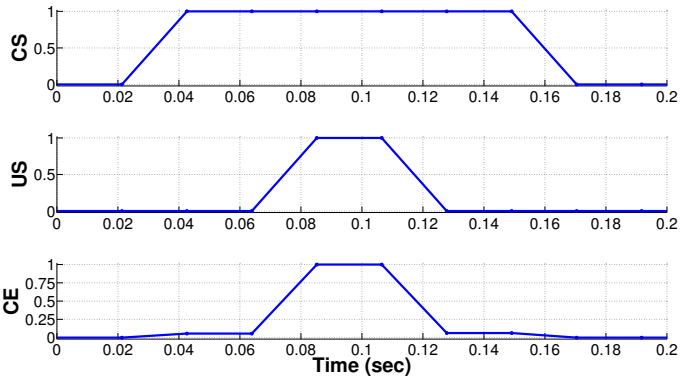

Fig. 9. Amygdala (CE) activation profile after conditioning when presenting no signal (first and last 2 time steps), CS (6 time steps) and US (2 time steps). 40% of total activation corresponds to a direct contribution of US signal. The activation related to reward prediction before and after US signal is about 0.1.

The feed-forward nature of the amygdala module allows the system to trigger a conditioned response independent of the US delay. We also observed that the number of trials does not have a major impact on the system activation. As few as one CS paired with the US and 200 trials suffice for conditioning, but a greater number of positive examples improve the overall response. The quick acquisition of fear memories is consistent with animal and human studies, where few trials account for a wider stimulus generalization [34]. Animal studies also support the fact that a greater number of trials increase stimulus discrimination, which improves inversely with the distance to the CS [32], [33].

IV. Conclusion

A reservoir system approach for auditory-cue fear conditioning was presented. The hybrid architecture is able to quickly associate a CS with a US and to perform frequency discrimination and long-lasting fear memories. The current implementation can support acquisition reliably and it is consistent with animal and human studies in terms of stimulus generalization and discrimination [32]–[34]. Other related dynamics are still under development. The weak but consistent anticipatory response after conditioning can be used after amplification for triggering conditioned behaviors.

A difference between our implementation and Lowe et al.'s implementation [3] is the origin of the CS signal. Lowe et al.'s models use abstract CS signals that are connected directly to both the PFC and to the CE modules. Instead, we fed the PFC module with the output generated by the auditory cortex module and the CE with the output generated by the BLA module. The US signal is connected to the MGm/PIN, BLA and CE modules [1], [4]. Another difference to most models on fear conditioning, as explained in the introduction, is the auditory input dimensionality. Since we considered noisy input signals and a preprocessing layer, the number of active units and the amplitude of the activation vary between trials, which represent an important step towards more bio-plausible computational models on fear conditioning. Our model does not model detailed temporal contingencies and only convergent cortical and subcortical signals are considered.

One of the limitations of the system is the simple modulation made by the PFC module through the VTA. This pathway can be used to support not only acquisition, but also inhibition of ambiguous responses like that observed in Fig. 7, which originates when converting spectral amplitude to neural activation. The single output of the system limits the possible conditioned behavior that the model may trigger in a real-world scenario.

As future work, a biologically constrained preprocessing of the auditory signal is planned, i.e. incorporating different degree of processing and latencies for subcortical and cortical areas. An appropriate filter bank such as a gammatone filter may contribute positively to improve frequency discrimination and to develop a more biologically plausible thalamus and auditory cortex module. We believe that keeping a coarse division of the amygdala into two sub-modules may facilitate the use of a modulating signal coming from the PFC module. In addition, an improved amygdala model is required. The recurrent nature of the main input nuclei in the amygdala [23] encourages us to explore a reservoir approach for the future implementation of the BLA module. As we are aiming to develop an embodied model of auditory-cue fear conditioning a CE module with more output units may be necessary to encode a variety of different conditioned behaviors.

The successful implementation of auditory-cue fear conditioning motivates us to continue improving the model to have a fully embodied fear conditioning model on a humanoid robot, which may serve not only to improve robot assistance but also to contribute to a better understanding of fear circuits.

Acknowledgment

This research has been partly supported by the EU project RobotDoC [35] under 235065 ROBOT-DOC from the 7th Framework Programme, Marie Curie Action ITN and by the KSERA project funded by the European Commission under the 7th Framework Programme (FP7) for Research and Technological Development under grant agreement n° 2010-248085. The authors of this paper want to thank Elena Villanueva-Mendez and Sven-Alexander Elies for proof reading this material.

References

[1] H.-C. Pape and D. Pare, "Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear," *Physiological reviews*, vol. 90, no. 2, pp. 419–463, 2010.

[2] J. E. LeDoux, "Brain mechanisms of emotion and emotional learning," *Current Opinion in Neurobiology*, vol. 2, no. 2, pp. 191–197, 1992.

[3] R. Lowe, F. Mannella, T. Ziemke, and G. Baldassarre, "Modelling coordination of learning systems: A reservoir systems approach to dopamine modulated pavlovian conditioning," in *Advances in Artificial Life. Darwin Meets von Neumann. 10th European Conference on Artificial Life (ECAL)*, ser. Lecture Notes in Computer Science, G. Kampis, I. Karsai, and E. Szathmáry, Eds., vol. 5777. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 410–417.

[4] J. E. LeDoux, "The amygdala." *Current Biology*, vol. 17, no. 20, pp. R868–R874, 2007.

[5] R. Gupta, T. R. Koscik, A. Bechara, and D. Tranel, "The amygdala and decision-making," *Neuropsychologia*, vol. 49, no. 4, pp. 760–766, 2011.

[6] S. Wermter, M. Page, M. Knowles, V. Gallese, F. Pulvermüller, and J. Taylor, "Multimodal communication in animals, humans and robots: An introduction to perspectives in brain-inspired informatics," *Neural Networks*, vol. 22, no. 2, pp. 111–115, 2009.

[7] S. Wermter, G. Palm, and M. Elshaw, Eds., *Biomimetic neural learning for intelligent robots: Intelligent systems, cognitive robotics, and neuroscience*, 1st ed., ser. Lecture Notes in Computer Science. Springer, 2005, vol. 3575.

[8] J. L. Armony, D. Servan-Schreiber, J. D. Cohen, and J. E. LeDoux, "An anatomically constrained neural network model of fear conditioning," *Behavioral neuroscience*, vol. 109, no. 2, pp. 246–257, 1995.

[9] A. Pavlou and M. Casey, "A computational platform for visual fear conditioning," in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2009, pp. 2451–2458.

[10] C. Balkenius and J. Morén, "Emotional learning: A computational model of the amygdala," *Cybernetics and Systems: An International Journal*, vol. 32, no. 6, pp. 611–636, 2001.

[11] P. den Dulk, B. T. Heerebout, and R. H. Phaf, "A computational study into the evolution of dual-route dynamics for affective processing," *Journal of Cognitive Neuroscience*, vol. 15, no. 2, pp. 194–208, 2003.

[12] I. Vlachos, C. Herry, A. Lüthi, A. Aertsen, and A. Kumar, "Context-dependent encoding of fear and extinction memories in a large-scale network model of the basal amygdala," *PLoS Computational Biology*, vol. 7, no. 3, pp. e1 001 104+, 2011.

[13] F. B. Krasne, M. S. Fanselow, and M. Zelikowsky, "Design of a neurally plausible model of fear learning," *Frontiers in Behavioral Neuroscience*, vol. 5, no. 41, 2011.
[14] G. Li, S. S. Nair, and G. J. Quirk, "A biologically realistic network model of acquisition and extinction of conditioned fear associations in lateral amygdala neurons," *Journal of neurophysiology*, vol. 101, no. 3, pp. 1629–1646, 2009.
[15] F. Mannella, S. Zappacosta, M. Mirolli, and G. Baldassarre, "A computational model of the amygdala nuclei's role in second order conditioning," in *From Animals to Animats 10*, ser. Lecture Notes in Computer Science, M. Asada, J. Hallam, J.-A. Meyer, and J. Tani, Eds. Berlin, Heidelberg: Springer-Verlag, 2008, vol. 5040, ch. 32, pp. 321–330.
[16] R. Lowe, M. Humphries, and T. Ziemke, "The dual-route hypothesis: Evaluating a neurocomputational model of fear conditioning in rats," *Connection Science*, vol. 21, no. 1, pp. 15–37, 2009.
[17] W. H. Alexander and O. Sporns, "An embodied model of learning, plasticity, and reward," *Adaptive Behavior*, vol. 10, no. 3-4, pp. 143–159, 2002.
[18] W. Zhou and R. Coggins, "Computational models of the amygdala and the orbitofrontal cortex: A hierarchical reinforcement learning system for robotic control," in *Advances in Artificial Intelligence*, ser. Lecture Notes in Computer Science, B. McKay and J. Slaney, Eds. Berlin, Heidelberg: Springer-Verlag, 2002, vol. 2557, ch. 37, pp. 419–430.
[19] N. Navarro, C. Weber, and S. Wermter, "Real-world reinforcement learning for autonomous humanoid robot charging in a home environment," in *Proceedings of the Annual Conference Towards Autonomous Robotic Systems (TAROS)*, ser. Lecture Notes in Computer Science, R. Groß, L. Alboul, C. Melhuish, M. Witkowski, T. Prescott, and J. Penders, Eds., vol. 6856. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 231–240.
[20] H.-C. Pape, "Petrified or aroused with fear: The central amygdala takes the lead," *Neuron*, vol. 67, no. 4, pp. 527–529, 2010.
[21] M. Davis, "The role of the amygdala in fear and anxiety," *Annual Review of Neuroscience*, vol. 15, no. 1, pp. 353–375, 1992.
[22] N. M. Weinberger, "The medial geniculate, not the amygdala, as the root of auditory fear conditioning," *Hearing Research*, vol. 274, no. 1-2, pp. 61–74, 2011.
[23] L. R. Johnson, M. Hou1, A. Ponce-Alvarez, L. M. Gribelyuk, H. H. Alphs, L. Albert, B. L. Brown, J. E. LeDoux, and V. Doyére, "A recurrent network in the lateral amygdala: A mechanism for coincidence detection," *Frontiers in Neural Circuits*, vol. 2, 2008.
[24] J. S. Bakin and N. M. Weinberger, "Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig," *Brain Research*, vol. 536, no. 1-2, pp. 271–286, 1990.
[25] C. W. Stevenson and A. Gratton, "Basolateral amygdala modulation of the nucleus accumbens dopamine response to stress: Role of the medial prefrontal cortex," *European Journal of Neuroscience*, vol. 17, no. 6, pp. 1287–1295, 2003.
[26] G. F. Koob and N. D. Volkow, "Neurocircuitry of addiction," *Neuropsychopharmacology*, vol. 35, no. 1, pp. 217–238, 2009.
[27] "Nao academics edition: medium-sized humanoid robot developed by Aldebaran Robotics," http://www.aldebaran-robotics.com/.
[28] G. S. Stent, "A physiological mechanism for hebb's postulate of learning," *Proceedings of the National Academy of Sciences*, vol. 70, no. 4, pp. 997–1001, 1973.
[29] J. L. Armony, D. Servan-Schreiber, J. D. Cohen, and J. E. LeDoux, "Computational modeling of emotion: Explorations through the anatomy and physiology of fear conditioning," *Trends in Cognitive Sciences*, vol. 1, no. 1, pp. 28–34, 1997.
[30] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer Science Review*, vol. 3, no. 3, pp. 127–149, 2009.
[31] H. Jaeger, "Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach," Fraunhofer Institute for Autonomous Intelligent Systems (AIS), Tech. Rep. 159, 2002.
[32] O. Desiderato, "Generalization of conditioned suppression," *Journal of Comparative and Physiological Psychology*, vol. 57, no. 3, pp. 434–437, 1964.
[33] H. S. Hoffman and M. Fleshler, "Stimulus factors in aversive controls: The generalization of conditioned suppression." *Journal of the experimental analysis of behavior*, vol. 4, pp. 371–378, 1961.
[34] J. Resnik, N. Sobel, and R. Paz, "Auditory aversive learning increases discrimination thresholds," *Nature Neuroscience*, vol. 14, no. 6, pp. 791–796, 2011.
[35] "The RobotDoC collegium: The Marie Curie doctoral training network in developmental robotics," http://robotdoc.org/.