Mind the (Computational) Gap

Matthew Casey, Athanasios Pavlou, and Anthony Timotheou

***Abstract*— Despite many advances in both computational intelligence and computational neuroscience, it is clear that we have yet to achieve the full potential of nature inspired solutions from studying the human brain. Models of brain function have reached the stage where large-scale models of the brain have become possible, yet these tantalising computational structures cannot yet be applied to real-world problems because they lack the ability to be connected to real-world inputs or outputs. This paper introduces the notion of creating a computational hub that has the potential to link real sensory stimuli to higher cortical models. This is achieved through modelling subcortical structures, such as the superior colliculus, which have desirable computational principles, including rapid, multisensory and discriminative processing. We demonstrate some of these subcortical principles in a system that performs real-time speaker localisation using live video and audio, showing how such models may help us bridge the computational gap.**

I. Introduction

Despite many advances in both computational intelligence (CI) and computational neuroscience (CN), it is clear that we have yet to achieve the full potential of nature inspired solutions from studying the human brain. As a consequence, a broader reference point for intelligent behaviour which encompasses artificial agents is now being advocated [1]. However, before we abandon human intelligence as our reference, have we fully exploited the available neuroscience?

To understand this question, let us remind ourselves of the approach adopted by Turing in proposing how to build an *intelligent machine* [2]. Turing focused on building an artificial cortex, which he described as an "unorganised machine" [2, p6]; a machine which starts in a random state and becomes organised through learning. In essence, Turing identified the computational principle of learning in the cortex as a key component of developing intelligent behaviour. Extracting such principles has become the foundation of CI and the development of biologically-inspired and plausible algorithms has reached the stage where large-scale models of the brain have become possible [3], [4], bringing us closer to Turing's original ideal. However, while such large-scale models are impressive CN tools, they are not connected to real-world inputs or outputs, yet this is an essential step if we wish to exploit such models and develop "intelligent machines".

Although Turing and contemporary followers of his ideas were less concerned with "circuits required for quite definite purposes", such as controlling respiration or eye movements [2, p12], focusing on the cortex appears too narrow. For example, the nervous system of a human is far more complex than just the cortex and even non-nervous system biological processes appear to exhibit cognitive behaviour, such as the immune system [5]. One potential solution is to focus on other areas of the nervous system that exhibit interesting computational properties and which are more closely associated with sensory inputs and motor control. Such areas can be found in subcortical sensory processing. To focus on one example from Turing, saccadic eye movements are controlled by at least 10 cortical and subcortical regions of the brain, with the superior colliculus (SC) in the midbrain playing a pivotal role [6]. Although performing a 'definite purpose', the SC is known to develop becoming mature in mid gestation [7] and may adapt [8], hence it shows some interesting computational principles. It is also directly connected to the optic tract and projects to the brain stem, and it is therefore many steps closer to real-world inputs than the cortex. With other similar examples, can focusing on subcortical areas help us bridge the gap?

In this paper, we attempt to answer this question by demonstrating how we can extract computational principles from sensory subcortical structures and demonstrate their practical application on a simple real-time speaker localisation task (a more intelligent version of eye saccades). This allows us to show how a CN model can be applied to a real-world CI problem. Our aim is to demonstrate that there is still potential left in modelling the nervous system (but not just the cortex) to inspire adaptive, real-time, neuro-inspired solutions to problems, and that this potential lies in exploiting often overlooked structures that directly process inputs and outputs, as well as mediate with the (still important) cortex. In section II we present our candidate set of principles. In section III we describe the nature-inspired method we will use in the demonstration of these computational principles. We demonstrate the results from this model and evaluate its effectiveness in section IV, while we discuss the implications of this approach in section V.

II. Extracting Computational Principles

Following the example set by Turing on eye movements, we focus in this paper on subcortical visual processing. This focus is also overtly motivated by our own previous work on developing CN models of relevant structures, including the SC [9] and amygdala [10], which provide a foundation upon which we can extract computational principles.

Some of the key functional areas in the visual pathway are shown in Figure 1, together with their afferent cortical areas. In cognitive psychology, human vision is often referred to in terms of the *retino-geniculate-cortical pathway* [11], placing emphasis on the retina, lateral geniculate nucleus (LGN) and the visual cortex. In general terms the LGN,

Matthew Casey, Athanasios Pavlou and Anthony Timotheou are with the Department of Computing, University of Surrey, Guildford, Surrey, GU2 7XH, UK (phone: +44 (0)1483 689635; fax: +44 (0)1483 686051; email: {m.casey, athanasios.pavlou, at00087}@surrey.ac.uk).

situated in the thalamus, is perceived as routing retinal stimuli to appropriate areas of the visual cortex. While large-scale models, such as Izhikevich and Edelman's [4] use a model of the visual thalamocortical circuit in their 1,000,000 neuron simulation, take care over how neurons and their connections are modelled, their view of the LGN and other subcortical structures appears too simplistic. For example, the LGN receives input from the retina and forms a layered topographic map of the whole visual space, which develops before the animal can open it eyes and hence before it can match binocular images [12]. Another area in the visual pathway which is topographic is the SC. This also receives direct input from the retina, and consists of a series of aligned topographic maps of the visual, auditory and somatosensory space, combining these into a multisensory representation to initiative eye movement [13].

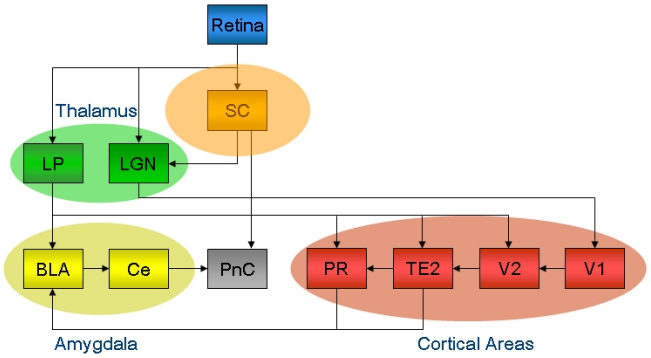


Fig. 1. Selected subcortical visual pathways leading to the visual cortex [14], [15]. Retinal neurons connect to the superior colliculus (SC), and the lateral posterior nucleus (LP) and lateral geniculate nucleus (LGN) in the thalamus. LP connects to the basolateral amygdala complex (BLA), which in turn connects to the central amygdaloid nucleus (Ce). The Ce and SC provide motor output via the pontine reticular nucleus (PnC). Cortical areas include the primary and secondary visual cortices (V1, V2), temporal cortical area (TE2) and the perirhinal cortex (PR).

The lateral posterior nucleus (LP) and pulvinar, which are linked to the LGN, also perform more complex functions than they are often credited with, such as being orientation and motion selective [16], properties often associated with the visual cortex. If we travel further up the visual pathways to the limbic system, then the amygdala also performs complex sensory processing to discriminate. For example, the amygdala has been implicated in the crude processing of visual emotional stimuli [17]. This crude discrimination is also adaptive as demonstrated through conditioning [18].

These subcortical structures therefore demonstrate key properties we normally associate with many different areas of the brain, and particularly those in the cortex. However, they also have other key properties associated with their lower level of operation. First, subcortical sensory processing occurs rapidly. For example, the SC is implicated in "express saccades" which can occur in as little as 80ms [6]. Second, this rapidity may occur because they are more directly connected to sensory inputs and motor outputs (this is true not just for vision, but also for olfactory connections to the amygdala [19]). Third, while some structures operate on single sensory modalities, such as the LGN, LP, pulvinar for vision [16] or the inferior colliculus (IC) for audition [20], key structures are multisensory [21]. Fourth, subcortical structures appear essential not just for sensory to cortex routing [22], but also for cortical-to-cortical communication [12], while they also receive cortical feedback to moderate their operation [23].

Although we have biased our argument on just a few of the many subcortical structures, the properties that are exhibited by just these are sufficiently interesting to warrant exploitation. To summarise these principles, our selected subcortical structures *develop* representations through processes such as self-organisation, *adapt* such as with conditioning and cortical feedback, *discriminate*, albeit on crude stimuli, are *rapid* in operation by being closely *connected to sensory inputs and motor outputs*, are often *multisensory* and are essential for cortical operation and *communication*.

III. Functional Implementation

To implement the subcortical computational principles we have identified, we take as inspiration system level models of the visual and audio pathways [18], [24], and in particular those modelling the SC [9], [10]. This approach allows us to modify existing CN models for practical use as we can select the key aspects we wish to embody through inputs, architecture and parameters. A systems level model simplifies the extensive processing undertaken by these structures, yet allows us to focus on key functionality and properties, particularly development, discrimination, rapidity resulting from simple, parallel processing of unisensory stimuli, combining these into a multisensory representation.

Our architecture is a variation of a model of the SC [9], which uses Hebbian learning for the development of a layered, modularised model based on principles defined by [18]. It comprises three modules each of size 2 by 128 neurons: two modules process the visual and auditory inputs separately, the output of which is then combined in a third integration map. The visual input size is 32 by 128 pixels, while the auditory input is 5 by 100 pixels wide (we use the term pixels here, even though we are dealing with sound locations). To train the model we use virtual inputs, prior to testing it on live video and audio. The activation y of a neuron at location (i,j) in a map given an m dimensional input x is given by:

$$
\begin{aligned}
u_{ij} &= \sum_{k=1}^{m} x_k w_{kij}(t) \\
y_{ij} &= \begin{cases} f(u_{ij}) & \text{if } \|c_{ij} - c_{win}\| < h(t) \\ f(u_{ij} - y_{win}) & \text{otherwise} \end{cases} \\
f(u) &= \begin{cases} 1 & u \geq 1 \\ u & 0 < u < 1 \\ 0 & u \leq 0 \end{cases}
\end{aligned}
$$

The weight from an input k to a neuron (i,j) on the map at time step $t \geq 0$ is represented by $w_{kij}(t)$. The model implements circular winner areas that introduce lateral inhibition in order to topographically organise during the

training process, similar to Kohonen's SOM [25]. At each time step the neuron which produces the maximum activation $y_{win} = max_{ij} f(u_{ij})$ on the map is labelled as the winner. Any neuron, c_{ij}, is considered to fall within the winning neuron's, c_{win}, area if its location (i, j) lies within the current radius $h(t)$. Training occurs by epochs, each of which is a random presentation of the full training set. At each epoch the winning neuron's area is reduced following a Gaussian neighbourhood:

$$h(t) \quad = \quad r_{min} + (r_{max} - r_{min}) e^{-(\frac{(t/t_e)^2}{2r_s^2})}$$

Here t_e equals the number of training samples, r_{min}, r_{max} representing the minimum and maximum radii for the neighbourhood and r_s representing the bandwidth. This results in a gradual tuning of neurons' weights to respond to inputs depending on their location in the input space. The maximum radius r_{max} is always the larger dimension of a map ($r_{max} = 128$) and the minimum radius was set to $r_{min} = 1$. The bandwidth was set to $r_s = 100$. Each weight is updated separately and we normalize weights to avoid exponential growth:

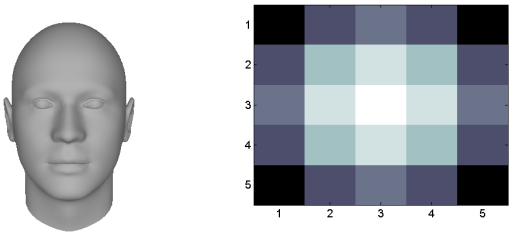
$$w'_{kij}(t+1) \quad = \quad w_{kij}(t) + \epsilon(t) x_k y_{ij}$$
$$w_{kij}(t+1) \quad = \quad \frac{w'_{kij}(t+1)}{\sum_{l=1}^{m} w'_{lij}(t+1)}$$

The learning rate $\epsilon(t)$ is constant for each map ($\epsilon(t) = 0.001$ for the visual map, 0.01 for the auditory map, and 0.0001 for the integration map). Each of the modules is trained for 400 epochs. The parameters used were derived from an in depth analysis described in [9].

IV. Speaker Localisation

The aim of our experiments is to demonstrate that the model can develop topographic maps with integrative capabilities that result in real-time detection of coincident visual and auditory stimuli. The context of the experiments is speaker localisation where higher responses are to be induced upon coincident detection of a human head and sound. Here we are not aiming for state-of-the-art speaker localisation, which can be achieved reliably using specific voice and head tracking algorithms [26], rather we are attempting to demonstrate that generic, adaptive solutions can be applied to such specific problems. A stepwise evaluation approach is followed and we start by examining the model responses to virtual stimuli to determine how well the model has developed. Next we evaluate the model against real-time visual and auditory inputs from a camera and a microphone.

Video and audio data for the experiments were captured using a Java implementation of the model[1]. This was run on a Dell Latitude D630 laptop with an Intel Core 2 Duo T7250 (2.00GHz) processor, 3.5GB of RAM, a Logitech Live! Cam voice USB camera and two Logitech USB desktop microphones, running Windows XP Professional SP3, Live! Cam driver version 1.1.2.410, JRE 1.6.0_17, JMF 2.1.1e. The models were trained and evaluated using Matlab version 7.8.0.347. The camera was positioned 0.7m above the plane of the microphones and 0.54m centrally behind. The microphones were separated by 1.34m. The heads of the subjects were 1.66m from the camera. To produce a loud and sustained sound, a sound source positioned 0.42m from the microphones. This was a pure tone produced from a laptop that could be moved independently of the head positions to allow testing of non-coincident audio and visual stimuli.



a) Visual Map Input b) Auditory Map Input

Fig. 2. a) Generic face used to train the visual map and b) the Gaussian activation pattern used to represent a localized sound used to train the auditory map.

For the model's training we use virtual visual and auditory inputs (Figure 2). Such stimuli ensure consistency amongst training examples and act as generic templates of the real stimuli. In particular, for the training of the visual map we use a generic non-textured head produced by the FaceGen application [27]. The head is grey scaled, normalised between 0 to 1 and rescaled to 32 by 32 pixels. This size allows computational efficiency while at the same time maintaining the characteristics of the original virtual head. For the creation of the visual map training data, we place the virtual head at consecutive locations within a 32 by 128 visual space. This results in 96 visual training examples. The hypothetical visual space and examples represent a stripe of a visual scene within which a head is moving across from one end to the other.

The training data for the auditory map are Gaussian activation patterns that represent the location (and not the type) of a sound within the defined auditory space. We define the location of a sound as a 5 by 5 grid with the Gaussian pattern centred at (c, d):

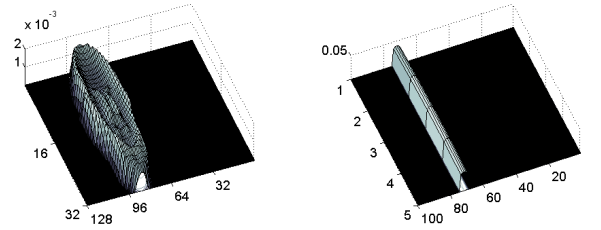
$$x_{ij} \quad = \quad \lambda e^{-(\frac{(i-c)^2+(j-d)^2}{2\sigma^2})}$$

where $\lambda = 1$ for the amplitude and $\sigma = 4$ for the bandwidth. Since we have 96 visual examples we need the same number of auditory inputs so that we can achieve a one to one correspondence (a head and a sound at the same location). Therefore, we place 96 consecutive examples of sound locations in a similar manner to the visual inputs resulting in the 5 by 100 auditory space.

[1]A Java demonstration of the system will be made available at `http://www2.surrey.ac.uk/computing/people/matthew_casey/`.

A. Model Responses to Virtual Visual and Auditory Stimuli

The aim of this experiment is to demonstrate the key properties of the model. Each of the unisensory maps is trained using the 96 virtual samples described above. After the maps have finished training, their outputs were presented with coincident audio-visual samples in order to train the integration map. Figure 3 illustrates that the trained unisensory maps have learned to develop receptive fields that respond preferentially thus demonstrating discriminative properties to certain inputs at particular locations.



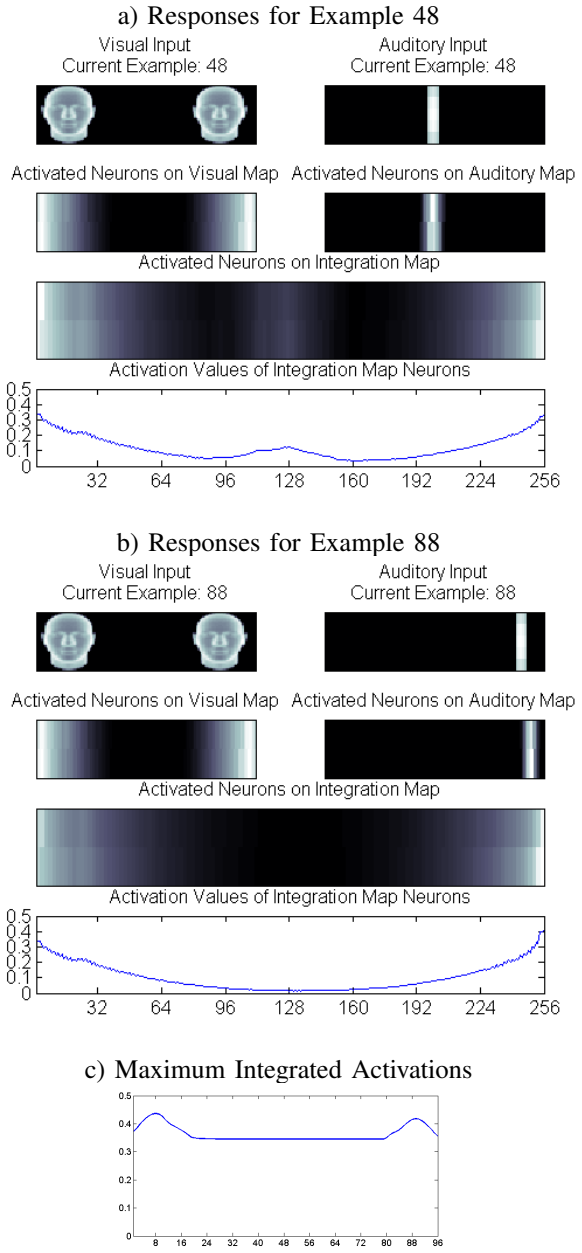
a) Visual Map Weights b) Auditory Map Weights

Fig. 3. Weight values of neuron $(1, 96)$ on the visual map and neuron $(1, 96)$ on the auditory map. The neuron weights resemble the virtual head and auditory location inputs that they are trained on.

To evaluate the model development of topographic maps we tested it with a visual input that consisted of two virtual heads at either end of the input (combining example 1 and 96), while the auditory input is a moving virtual sound from left to right within the auditory space (all 96 locations). Since the location of the sound changes, there will be only two auditory locations which are fully coincident with visual (head) locations. By observing the visual map activation for example 8 and 88 (Figure 4a, b) it is clear that the location of the heads activate the edges of the map since they are on the edges of the visual space. However, the sound location in trial example 8 is towards the centre of the auditory space whilst example 88 towards its end. This results in different areas of the auditory map being activated (Figure 4a, b) which in turn affects the activation strength of the integration map. Figure 4c shows that the integration map is highly responsive on its edges and this is attributed to the auditory stimulus being coincident with the visual stimuli towards the start and end of its movement within the auditory space. This demonstrates multisensory enhancement on coincident, multimodal stimuli, which is a well-studied property of the SC [14].

B. Model Responses to Real Visual and Virtual Auditory Stimuli

Having established the functionality of the model on artificial data we now evaluate it on video images and virtual moving sound. The image examples here depict two heads near the edges of the visual space. The images themselves have been extracted from a video sequence of two non-moving people. Therefore the only difference to the previous trial is the substitution of the virtual with real faces. As seen



c) Maximum Integrated Activations

Fig. 4. Model responses to virtual stimuli: a) and b) show the the visual and auditory inputs, visual and auditory map activations, and integration map activation for example inputs 48 and 88; c) shows maximum activations of the integration map for all 96 inputs.

in Figure 5a the visual map shows that it is still activated near its edges. Crucially, we also observe that the profile of activations in Figure 5b is similar to that of Figure 4c. This means that the model can successfully generalise its behaviour when given real visual data.

C. Model Responses to Real Visual and Auditory Stimuli

For our final experiment we present the model with both real visual and audio data and demonstrate its real-time deployment. For this trial we have used a single non-moving person on the right side of the visual space that activates

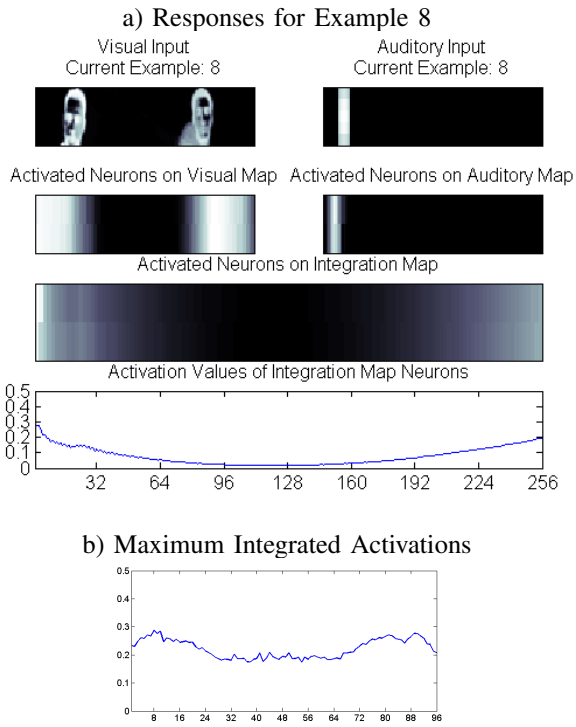


Fig. 5. Model responses to real non moving head stimuli and virtual moving sound: a) shows the the visual and auditory inputs, visual and auditory map activations, and integration map activation for example input 8; b) shows maximum activations of the integration map for all 96 inputs.

a sound in front of them for a duration of 10 seconds. The sound is a pure tone (16 bit, 800Hz). This is used, rather than talking, to provide a consistent and sufficiently loud sound for the microphones to capture. Although an appropriate localisation is given in response to talking, because of the low volume and reverberation in the room, the location varies with every frame unless noise is minimised. The pure tone was therefore used to provide consistency and to allow the sound to vary to different locations to that of the heads. The 325 examples used in this trial correspond to 13 seconds at 25 frames per second audio and video, with the sound played approximately after 2 seconds from the trial initialisation.

Sound localisation in mammals is achieved using both ILD and ITD, with the IC being the pivotal subcortical structure having a topographic representation of sound. For eye movement via the SC, ILD is used predominantly for azimuth localisation [20], which is our chosen task. We therefore use a simplified version of an ILD algorithm [28] which allows us to localise using two microphones. Here, the sound intensity for each microphone is calculated from a sample buffer of sound pressures. The intensity from each microphone is then used to describe the centre and radius of a circle which represents the possible locations of the sound given the relative position of the two microphones. With only two microphones it is not possible to localise any further than this circle without assumptions about the possible sound source locations. Here, we assume that the sound location falls at the point on the circle which is closest to the centre point between the two microphones.

Figure 6a shows the results for example input 100. We observed that all activation values above 0.22 correspond to coincident detections. On the other hand, non-coincident examples (such as example 300 illustrated in Figure 6b) have a lower activation in the integration map. We can therefore use this value to discriminate speakers as it corresponds to an activation of coincident head and sound localisation.

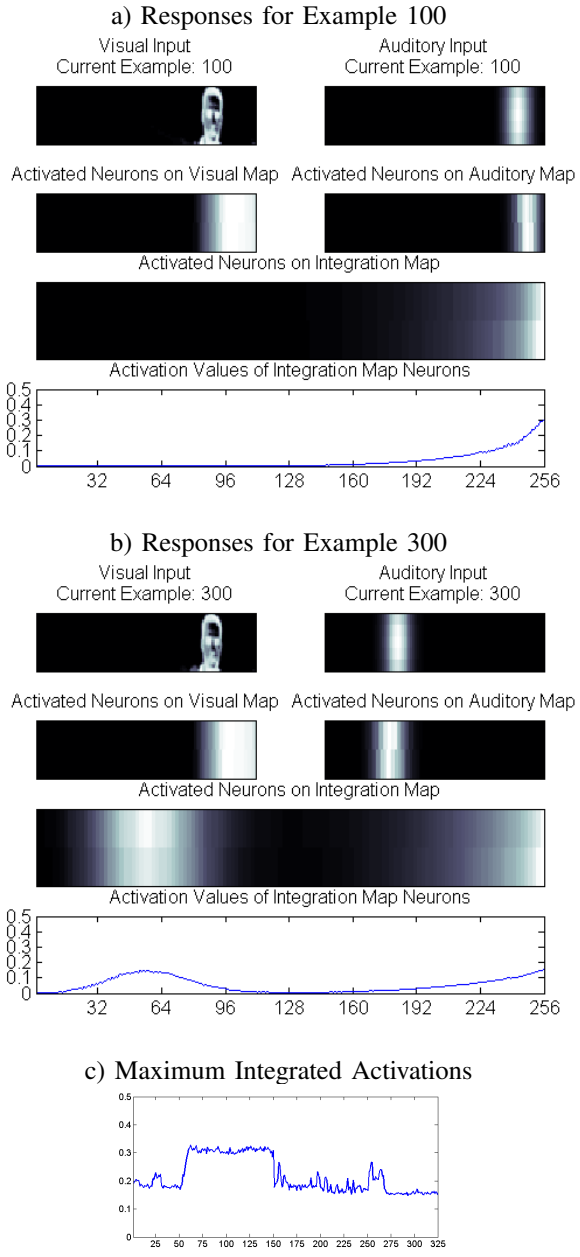


Fig. 6. Model responses to real visual and audio stimuli: a) and b) show the the visual and auditory inputs, visual and auditory map activations, and integration map activation for example inputs 100 and 300; c) shows maximum activations of the integration map for all 96 inputs.

V. Conclusions

In this paper we have demonstrated how CN models of subcortical processing may be used to inspire CI solutions. We focused on key subcortical structures because they directly process sensory input and produce motor output, offering us insight into how more complex cortical models might be connected to the real-world. Although we focused on only a small number of subcortical structures, including the SC and amygdala, and ignored others, such as the hippocampus, we identified seven computational principles that we wish to emulate: *development*, *adaptation*, *discrimination*, *rapid operation*, *sensory input and motor output connectivity*, *multisensory integration* and *communication*. At least the first three are shared with cortical structures, yet in terms of real-world operation, rapid processing of crude multisensory stimuli appears key to our aim of demonstrating the potential for inspiring novel and relevant solutions to real-world problems. However, we recognise that this intuitive "reasoning by metaphor" [29] approach of extracting principles is not ideal, and we should be following a more analytic approach to understand the computational properties of these structures. Nonetheless, this is enough to initiate bridging the computational gap of higher cortical simulations and real sensory input.

To demonstrate the potential of subcortical approaches to CI, we selected the real-time task of speaker localisation. Here, we adapted a system level model of the SC to localise heads and sound sources. Through evaluating this on both virtual and real data, we demonstrated that the model develops topographic representations, discriminating and combining multisensory stimuli, all in real-time. Although this demonstrator is not as capable as state-of-the-art implementations of speaker localisation, it does demonstrate how CN models may be applied to CI problems.

Acknowledgements

The authors would like to thank Andre Grüning and Jon Timmis for listening to the argument presented in this paper.

References

[1] N. Cristianini, "Are we there yet?" *Neural Networks*, vol. 23, no. 4, pp. 466–470, 2010.

[2] A. M. Turing, "Intelligent machinery," National Physical Laboratory, Tech. Rep., 1948. [Online]. Available: http://www.alanturing.net/turing_archive/archive/l/l32/L32-001.html

[3] H. Markram, "The blue brain project," *Nature Reviews Neuroscience*, vol. 7, no. 2, pp. 153–160, 2006.

[4] E. Izhikevich and G. M. Edelman, "Large-scale model of mammalian thalamocortical systems," *Proceedings of the National Academy of Sciences of the USA*, vol. 105, no. 9, pp. 3593–3598, 2008.

[5] I. R. Cohen and D. Harel, "Two views of a biology-computer science alliance," in *Proceedings of the 2009 Workshop on Complex Systems Modelling and Simulation (CoSMoS 2009)*, S. Stepney, P. Welch, P. S. Andrews, and J. Timmis, Eds. Luniver Press, 2009, pp. 1–8.

[6] A. K. Moschovakis, C. A. Scudder, and S. M. Highstein, "The microscopic anatomy and physiology of the mammalian saccadic system," *Progress in Neurobiology*, vol. 50, no. 2-3, pp. 133–254, 1996.

[7] J. Qu, X. Zhou, H. Zhu, G. Cheng, W. S. Ashwell, and F. Lu, "Development of the human superior colliculus and the retinocollicular projection," *Experimental Eye Research*, vol. 82, no. 2, pp. 300–310, 2006.

[8] A. G. Constantin, H. Wang, and J. D. Crawford, "Role of superior colliculus in adaptive eye-head coordination during gaze shifts," *J Neurophysiol*, vol. 92, no. 4, pp. 2168–2184, 2004.

[9] A. Pavlou and M. C. Casey, "Simulating the effects of cortical feedback in the superior colliculus with topographic maps," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN) 2010*. IEEE, 2010.

[10] ——, "Identifying emotions using topographic conditioning maps," in *Advances in Neuro-Information Processing: Proceedings of the 15th International Conference on Neuro-Information Processing, Lecture Notes in Computer Science 5506*, M. Koeppen, N. Kasabov, and G. Coghill, Eds. Springer-Verlag, 2009, pp. 40–47.

[11] T. Pasternak, J. W. Bisley, and D. J. Calkins, "Visual processing in the primate brain," in *Biological Psychology*, ser. Handbook of Psychology, M. Gallagher and R. J. Nelson, Eds. John Wiley and Sons, Inc., 2003, vol. 3, chapter 6, pp. 139–185.

[12] S. M. Sherman and R. W. Guillery, "The visual relays in the thalamus," in *The Visual Neurosciences*, L. M. Chalupa and J. S. Werner, Eds. A Bradford Book, The MIT Press, 2004, vol. 1, chapter 35, pp. 565–591.

[13] A. J. King, "The superior colliculus," *Current Biology*, vol. 14, no. 9, pp. R335–R338, 2004.

[14] B. E. Stein, R. F. Spencer, and S. B. Edwards, "Efferent projections of the neonatal cat superior colliculus: Facial and cerebellum-related brainstem structures," *The Journal of Comparative Neurology*, vol. 230, no. 1, pp. 47–54, 1984.

[15] C. Shi and M. Davis, "Visual pathways involved in fear conditioning measured with fear-potentiated startle: Behavioral and anatomic studies," *The Journal of Neuroscience*, vol. 21, no. 24, pp. 9844–9855, 2001.

[16] C. Casanova, "The visual functions of the pulvinar," in *The Visual Neurosciences*, L. M. Chalupa and J. S. Werner, Eds. A Bradford Book, The MIT Press, 2004, vol. 1, chapter 36, pp. 592–608.

[17] L. Pessoa, "To what extent are emotional visual stimuli processed without attention and awareness?" *Current Opinion in Neurobiology*, vol. 15, pp. 188–196, 2005.

[18] J. L. Armony, D. Servan-Schreiber, L. M. Romanski, J. D. Cohen, and J. E. LeDoux, "Stimulus generalization of fear responses: Effects of auditory cortex lesions in a computational model and in rats," *Cerebral Cortex*, vol. 7, no. 2, pp. 157–165, 1997.

[19] J. A. Gottfried, "Smell: Central nervous processing," in *Taste and Smell: An Update (Advances in Oto-Rhino-Laryngology)*, T. Hummel and A. Welge-Lssen, Eds. S. Karger AG, 2006, pp. 44–69.

[20] B. Delgutte, P. X. Joris, R. Y. Litovsky, and T. C. Yin, "Receptive fields and binaural interactions for virtual-space stimuli in the cat inferior colliculus," *J Neurophysiol*, vol. 81, no. 6, pp. 2833–2851, 1999.

[21] B. E. Stein and M. A. Meredith, *The Merging of the Senses*. Cambridge, MA.: A Bradford Book, MIT Press, 1993.

[22] D. G. Amaral, H. Behniea, and J. L. Kelly, "Topographic organization of projections from the amygdala to the visual cortex in the macaque monkey," *Neuroscience*, vol. 118, no. 4, pp. 1099–1120, 2003.

[23] J. C. Alvarado, T. R. Stanford, B. A. Rowland, J. W. Vaughan, and B. E. Stein, "Multisensory integration in the superior colliculus requires synergy among corticocollicular inputs," *The Journal of Neuroscience*, vol. 29, no. 20, pp. 6580–6592, 2009.

[24] R. Miikkulainen, J. A. Bednar, Y. Choe, and J. Sirosh, *Computational Maps in the Visual Cortex*. New York: Springer Science+Business Media, 2005.

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

[26] K. Hyun-Don, C. Jong-Suk, and K. Munsang, "Speaker localization among multi-faces in noisy environment by audio-visual integration," 2006, pp. 1305–1310.

[27] Singular Inversions Inc. (2010, Jun) Facegen - 3d human faces. [Online]. Available: http://www.facegen.com/

[28] S. T. Birchfield and R. Gangishetty, "Acoustic localization by interaural level difference," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2005)*, vol. 4, 2005, pp. iv/1109–iv/1112.

[29] S. Stepney, R. E. Smith, J. Timmis, A. M. Tyrrell, M. J. Neal, and A. N. W. Hone, "Conceptual frameworks for artificial immune systems," *International Journal of Unconventional Computing*, vol. 1, no. 3, pp. 315–338, 2005.