A Computational Model For Audio Triggered Emotions

計算認知神經科學 期末書面報告

電機三 b06901184 謝佑芃

電機三 b06901007 戴子宜

2020.1.10

1. Introduction

1.1 The necessity to studying emotion

Artificial Intelligence and machine learning have progressed rapidly in recent years. A large majority of current researches focus on how to create, or to teach a machine that outperforms human in different tasks. However, although these models seem to be "artificial intelligence", the underlying architectures are actually far from human brains. One important mechanism they lack is emotion, which is crucial to how humans act and make decisions since humans are not always rational. Hence, we would have to have a full understanding of emotions in human brains in order to apply the mechanisms to artificial intelligent agents.

One might ask: why should we create machine with emotions if emotions are the reason that humans make irrational decisions? However, this question is actually based on a misunderstanding of the functions of emotions in the human brain. One intuitive idea is that: while cognition is often equated as rationality and logic, emotion, on the other hand, is often mistaken as the cause of irrational behaviors. This idea that cognition and emotion are separable and competitors for dominance of the human brain may sound reasonable and intuitive at first. However, lots of evidence has shown that emotion and cognition are both required for a human being to make the correct decision. For instance, a pathological absence of emotions leads to profound impairment of decision making [1]. Hence, emotions actually play an important role in the human brain to make the correct decision, or response.

1.2 Framework of Emotion Circuits : How are stimuli processed in the human brain?

When an emotional stimuli (such as a scary picture or an angry tone) is presented, the stimuli would be transformed from sensory neural circuits into semantic pointers, which then would be sent into the limbic system, the region in human brain that processes emotional information and outputs emotional response. For example, when a person sees an emotional scene, the visual information is encoded by the neural circuits consisting of retina ganglion cells and visual cortex and transformed into a

semantic pointer. This semantic pointer would later be sent to the limbic system, which would further decide the reaction to the emotion stimuli.

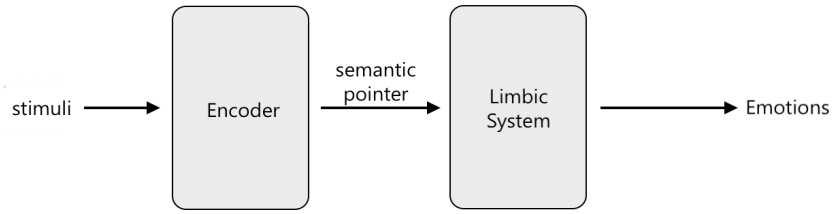

Fig. 1-1 The framework of the emotion circuits in the brain. [3]

1.3 Simulation of a complete pathway for emotion stimuli: Audio-Triggered Fear

As mentioned above, all emotion circuits share a simplified framework: a sensory network acts as an encoder, encoding the emotional stimuli into a semantic pointer, which can be viewed as a high-dimensional vector. This semantic pointer would then be sent to the limbic system. In this project, we aimed to simulate one of these networks. For the encoder, we chose the auditory periphery, so as to simulate the hearing process. The limbic system is constructed in order to simulate the conditioned fear response, which would be described in detail in section 4. We chose "fear" out of all 6 basic emotions of human beings: angry, disgust, fear, happy, sad, and surprise [2], since the biology neural circuits as well as the psychology basis of fear response are the most studied and understood. The hearing model is described in section 2 and section 3, the emotion model is described in section 4 to 6, and the results are in section 7.

2. The Mechanism of Hearing in Human Ears

To model the hearing process in brain, the transformation of sound to electric signals, which can be transmit by neurons, should be conducted first. First of all, according to Fig. 1-1 [4], for the sound transduction in human, fluctuations in air pressure travel down the auditory canal and cause the tympanic membrane to vibrate. These vibrations result in a tiny bone called the stapes rapidly moving in and out of the oval window of the cochlea.

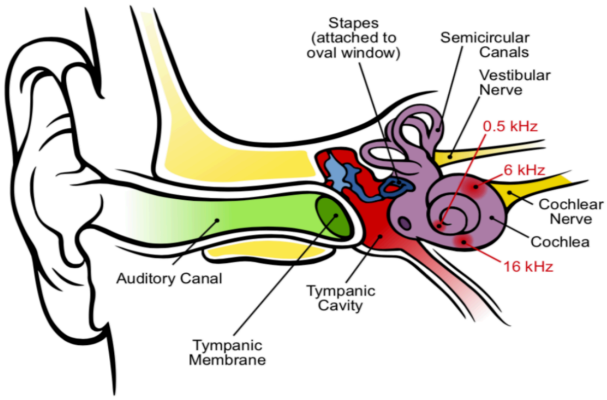

Fig. 2-1 The overview of the mechanism of human ears

Electrical activity in the form of action potentials are produced in the cochlea by inner hair cells, projected through spiral ganglion, and enter the brain through the cochlear nerve (Fig. 1-2) [5].

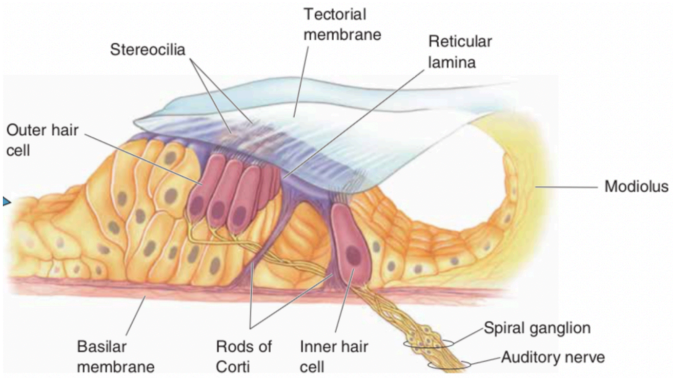

Fig. 2-2 The basilar membrane

The disturbances of the liquid in the cochlea deflect the basilar membrane inside, with high frequency disturbances causing deflections at the base of the cochlea, and low frequency disturbances causing deflections at the apex of the cochlea. From Fig. 1-3 [5], we can see that the place code for frequency on the basilar membrane of uncoiled cochlea represents in log scale (Mel scale).

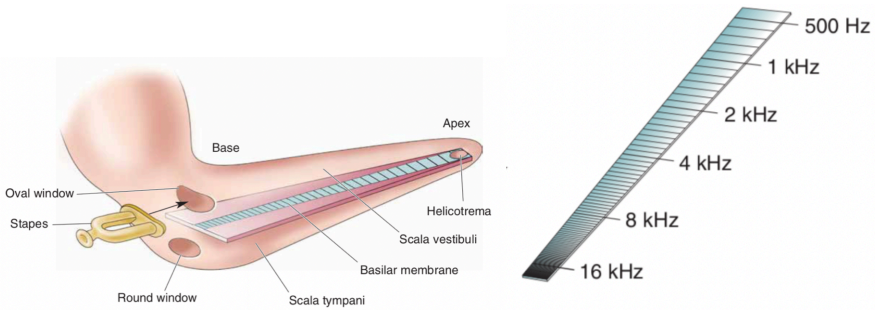

Fig. 2-3 The basilar membrane in an uncoiled cochlea and frequency producing maximum amplitude

3. A Model for Hearing: Neural Cepstral Coefficients (NCCs) Neural Network

3.1 Overview

According to the thesis [6], Neural Cepstral Coefficients (NCCs) can yield significantly better training and test correctness rates in speech classification than Mel-Frequency Cepstral Coefficients (MFCCs), which are the most commonly used feature vector in ASR systems. Therefore, we decided to implement NCCs to extract features from audio files instead of MFCCs. The following content is the implementation of NCCs neural network.

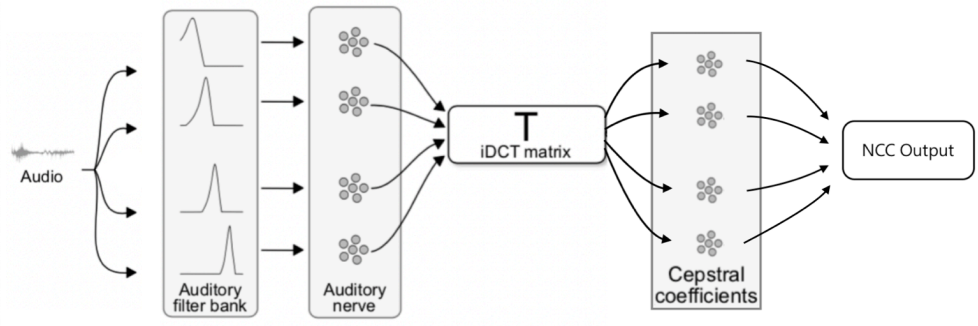

Fig. 3-1 Construction of NCCs neural network

3.2 Mel Scale

Mel Scale is a perceptual scale of pitches judged by listeners to be equal in distance from one another. The reference point between this scale and normal frequency measurement is defined by assigning a perceptual pitch of 1000 Mel to a 1000 Hz tone, 40 dB above the listener's threshold. Above about 500 Hz, increasingly large intervals are judged by listeners to produce equal pitch increments. As a result, four octaves on the hertz scale above 500 Hz are judged to comprise about two octaves on the Mel scale. The transformation formulas between Mel (m) and Hertz (f) are

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right), \qquad f = 700(10^{\frac{m}{2595}} - 1)$$

After filtering a designated number of frequencies to be observed, which are equally spaced samples in the interval 20Hz ~ 8000Hz, the following steps are, then, the further analysis and transformations based on the waveforms of these frequencies. In this work we choose four frequencies equally spaced in the Mel Scale of the interval [20, 8000] Hz.

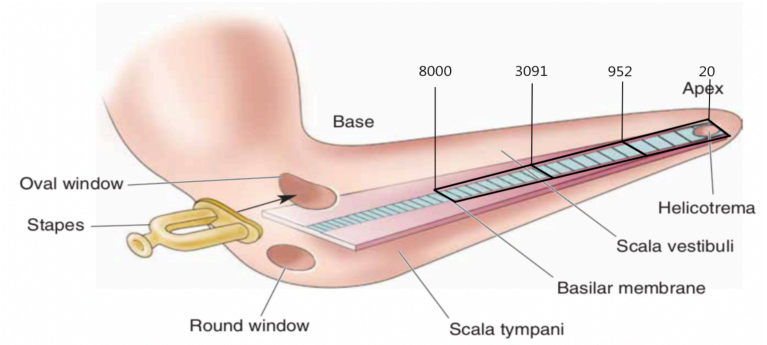

Fig. 3-2 A schematic notation on the basilar membrane in an uncoiled cochlea of the four frequencies extracted in this work

3.3 Auditory Filter Bank

Auditory Filter Bank processes incoming sound waveforms of the designated frequencies. After going through the filter banks, the output signals are then connected to the ensemble array of Auditory Nerve.

3.3.1 Middle Ear

The Middle Ear function implements the middle ear model, which is a linear bandpass filter with two pole pairs and one double zero [7]. The gain is normalized for the response of the analog filter at 1000Hz.

3.3.2 Gammatone Filter

The Gammatone Filter is a linear filter described by an impulse response that is the product of a gamma distribution and sinusoidal tone. It is a widely used model of auditory filters in the auditory system. The Gammatone Filter has an impulse response of the form

$$IR(t) = t^{n-1}\cos(2\pi f t)\, e^{-2\pi b ERB(f) t},$$

where n is the order of the filter, b is a parameter determining the filter bandwidth, and ERB(f) is the equivalent rectangular bandwidth of the filter centered at f.

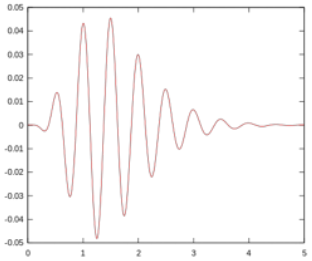

Fig. 3-3 Impulse response of the Gammatone Filter.

3.4 Auditory Nerve

To model the auditory nerve, an ensemble array of leaky integrate-and-fire (LIF) neurons is constructed. The neurons serve as inner hair cells, inputting the signals formerly processed, and are assembled into ensembles, which serve as spiral ganglion cells. Each ensemble inputs a filtered signal output from auditory filter banks. The ensembles are then gathered into an array, who simulates the activity of spiral ganglion cells projecting down the auditory nerve.

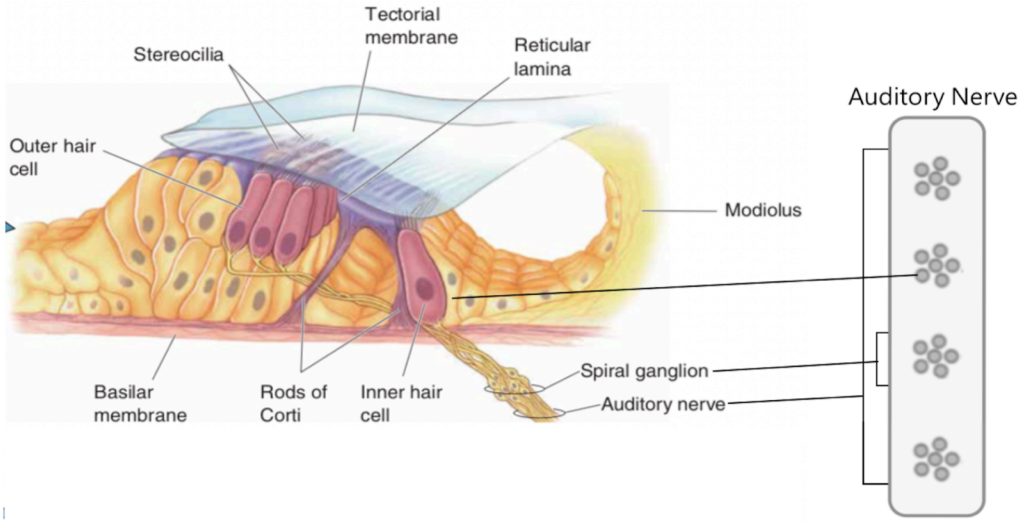

Fig. 3-4 A schematic diagram of the model of Auditory Nerve corresponding to the basilar membrane

3.5 Cepstral Coefficients

3.5.1 iDCT Matrix

Similar to MFCCs, NCCs applies the inverse discrete cosine transform (iDCT) to the output of auditory nerve before connecting to the inputs of Cepstral Coefficients.

3.5.2 Cepstral Coefficients

After passing the linear T matrix, which implements the iDCT, the ensemble array of auditory nerve is connected to the ensemble array of Cepstral Coefficients.

3.6 The Comparison Between MFCCs And NCCs Neural Networks

In comparison with another feature extracting method, MFCCs, Fig.5-1 shows the pipelines of both methods. Typical MFCCs extraction algorithms break the audio signal into overlapping fixed-length windows called frames, and each frame is processed independently. Then in the smoothing stage, the discontinuities due to frame

boundaries are minimized. However, different from MFCCs, the NCCs algorithms processed the incoming audio in a continuous fashion, rather than in a frame-based fashion. No explicit smoothing and normalization step is necessary, as signals will be smoothed and normalized in the NCCs model by virtue of being represented by spiking neurons, and filtered by the synaptic filters between neurons. The pipeline maintains internal state at each processing stage, and updates that internal state for each incoming audio sample. As each internal state update depends on the current state, changes occur smoothly through time, which eliminates the need for a smoothing layer. While the NCCs pipeline appears to be simpler than the MFCCs pipeline, the missing components from the MFCCs pipeline are actually already incorporated into parts of the NCCs pipeline; e.g., the logarithmic compression is in the first part of the auditory periphery model.

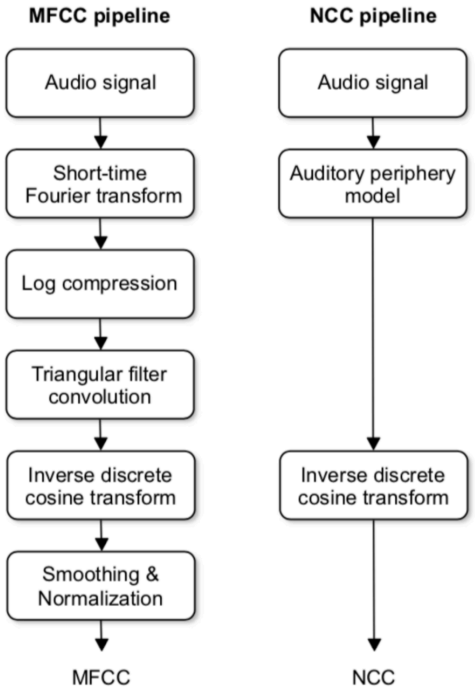

Fig. 3-5: Pipelines for generating NCCs and MFCCs

3.7 Brian.hears Library

Brian.hears library is an auditory modeling library for Python, which is part of the neural network simulator package, Brian. By importing this library, several sound processing functions can be used, such as loading .wav files, simulation of the sound processing in human middle ears, and constructing filter banks.

After the simulation of NCCs on a sound file, the output of NCCs are provided as the input of the model for limbic system described in the section 6.

4. Theories for Conditioned Fear

4.1 Classical Conditioning

Classical conditioning (also known as Pavlovian conditioning) is learning through association and was discovered by Pavlov, a Russian physiologist. In the process of learning, two stimuli are paired up together while only one of them actually triggers a response, which we called as an unconditioned response (UR). This stimuli is called the unconditioned stimuli (US), while the other neutral one is called the conditioned stimuli (CS). The response could be in any type, emotional responses or behaviors etc. However, as the two stimuli are constantly paired up, the human brain would eventually associate the two stimuli and produced a learned response which we call as the conditioned response (CR). That is to say, when only the conditioned stimuli is shown, it would actually be able to trigger the same response. This learning process of pairing up the US and CS is called classical conditioning. If the learned response of classical conditioning is a fear response, the process is also known as "conditioned fear".

4.2 Conditioned Fear: Acquisition, Extinction, Reacquisition, and Renewal

There are different types of conditioned fear, depending on how the response is learned, which are namely fear acquisition, fear extinction, fear reacquisition, and fear renewal. Fear acquisition refers to the increased expression of the CR as a result of CS-US pairing. Fear extinction refers to the reduction of CR expression when the CS is no longer paired with the US. For example, consider a mouse in a box. A tone appears every time when the mouse experiences a foot shock and thus freezes. Eventually the mouse freezes upon hearing the tone even without a foot shock, implying a fear acquisition process. If foot shocks stop appearing after the tone, eventually the mouse will not freeze upon the tone, which is a fear extinction process.

4.3 Contextual and Cue Conditioning

So far, the fear conditioning we have discussed from the mouse model is called "cue conditioning", where the originally neutral stimuli (CS) is the tone. Studies have shown; however, the tone is not the only stimulus that the mouse associates with the foot shock. If the mouse was placed in a box with a certain odor and some other

features, the mouse might link the foot shock to the environment of the box as well, showing identical freezing fear responses when placed in the same box again. In this situation, the originally neutral stimuli also include the features and odors in the box, which we refer to as “context”. The conditioned fear triggered by “context” (the features of the box) rather than “cue” (tone) is called “contextual conditioning”.

5. Neurobiology Basis Behind Conditioned Fear

5.1 The Amygdala, Prefrontal Cortex, and the Hippocampus

As mentioned in the introduction, the limbic system is the region of the brain where emotions are processed and evaluated. The limbic system consists of the amygdala, hippocampus, cingulate cortex, hypothalamus, thalamus, and the neocortex, which contains the orbitofrontal cortex (OFC) (Fig. 5-2).

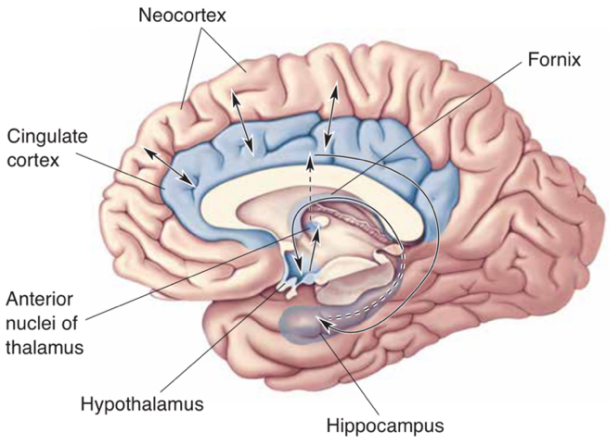

Fig. 5-1 The components of the limbic system [12]

Evidence has shown that three different brain areas of the fear response have been implicated in fear conditioning: amygdala, hippocampus, and the ventral-medial prefrontal cortex (vmPFC). The connections between these brain regions are depicted in (Fig. 5-2) [13]. This remaining part of this section would focus on the connections between each brain region in depicted below.

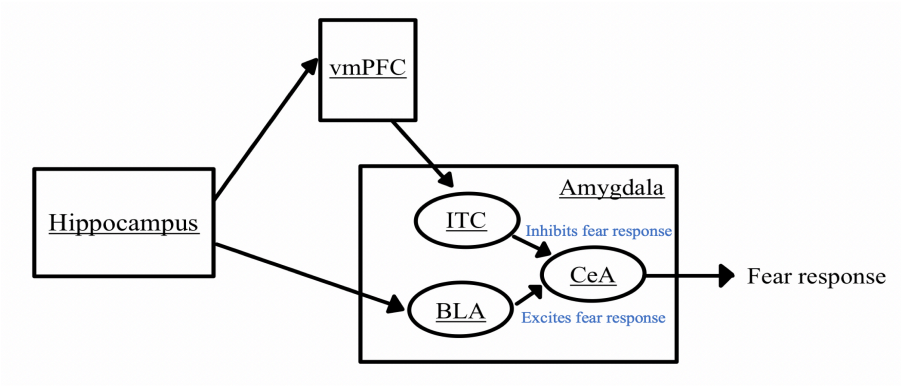

Fig.5-2 Connectivity between the hippocampus, vmPFC, and the amygdala

5.2 The Amygdala Subsystems

The amygdala is a collection of different nuclei, including the central nucleus (CeA), lateral and anterior basolateral nuclei (BLA). CeA is responsible for the initiation of fear response, as the neurons projects to parasympathetic nervous system, hypothalamus, and brainstem motor areas. These areas controls heart rates, freezing, and release of stress hormones, which are all categorized as fear responses. CeA receives projections from both BLA and intercalated cells (ITC), which is also part of the amygdala. BLA neurons are excitatory and facilitates fear responses from the CeA while ITC inhibits fear responses. Hence, it can be said that the BLA is responsible for fear acquisition while the ITC is responsible for fear extinction.

5.3 The Role of Hippocampus

Conditioned fear can be categorized into “contextual” and “cue” conditioning. Studies have shown that only contextual conditioning is hippocampus dependent, and both cue and contextual conditioning is amygdala dependent [10]. This implies not all conditioned fear is related to the hippocampus and that the hippocampus plays a role in contextual conditioning by storing memories for different contexts.

5.4 Ventral Medial Prefrontal Cortex

Since fear extinction involves the formation of new memories rather than erasure of older fear memories, researchers have been looking for the brain region that holds the extinction memories. Studies from mice [11] with ventral medial prefrontal cortex lesions have implied that the vmPFC is what we are looking for. In fact, the output

neurons of vmPFC projects to ITC in the amygdala, which is considered as the inhibition site for fear response. This is consistent with the suggestion that the vmPFC is in charge of fear extinction.

6. A Fear Response Model for the Limbic System

A computational model for the interactions between the amygdala, prefrontal cortex, and the hippocampus is shown in (Fig.6-1). The input signals are the conditioned stimulus (CS), unconditioned stimulus (US), and the context. Since the context memories are stored in the hippocampus, the hippocampus decodes the input signals and outputs the encoded context as the form of a sematic pointer, which is sent to both the BLA and vmPFC. Both BLA and vmPFC receives the CS and US as input as well as the decoded context from the hippocampus. Together with the information, the BLA evaluates and outputs a signal encoding whether to excite a fear response, while the vmPFC does the opposite by evaluating whether to inhibit one. Both evaluations are projected to the CeA (the inhibition signal is sent to the CeA indirectly through the ITC), where the two signals compete and the stronger one decides if the CeA is to initiate a fear response. In a larger model, this fear response signal should be sent into the parasympathetic autonomic nervous system, hypothalamus, and the brainstem motor areas and thus performs the behaviors such as freezing. However, we omit this detail and focus on the fear circuit between the amygdala, prefrontal cortex, and the hippocampus for now.

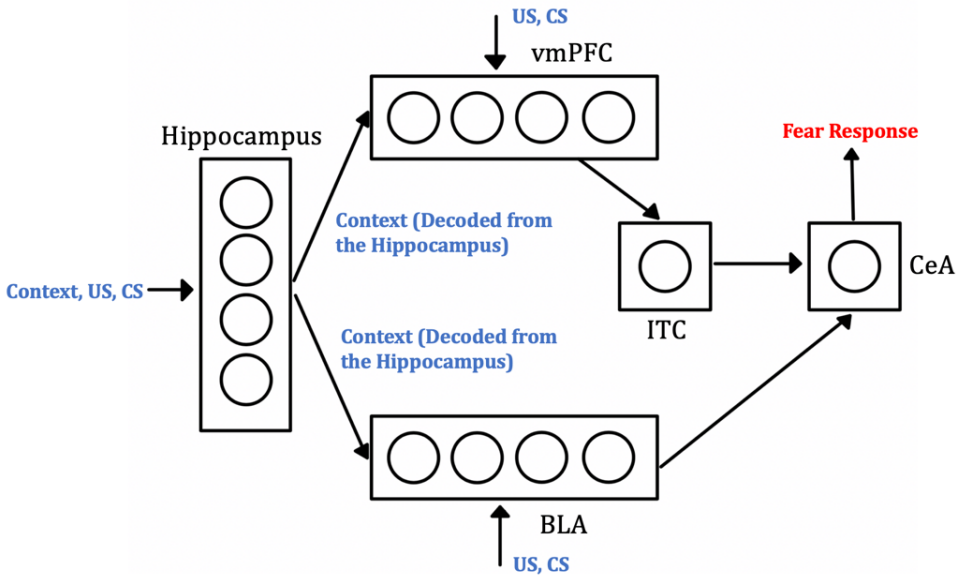

Fig. 6-1 Structure of the Computational Model

6.1 Encoding of the stimuli

There are three stimuli input to the model, a context, CS, and US respectively. The input to the model are output from the previous hearing model. Here we let each stimulus be a 160-bit vector, representing as semantic pointers. In our simulation, we would check the encoding of different stimuli or different contexts are orthogonal, as this condition is required in order to ensure that the dissimilarities between all stimuli is the same [12].

6.2 Hippocampus: Model structure and training rules

The hippocampus is a single layer of 20 nodes, each node receives projections from the input signal layer, which is a 24-bit vector that represents the context, CS, and US. The hippocampal layer and input signal layer is fully connected. The weights w_{ji} are initialized uniformly randomly from [0, 0.6] and then perturbed using Gaussian noise (i stands for the index of the nodes in the input layer and j stands for that of the output layer.). Hence the actual weights can be computed as $u_{ji} = w_{ji} + \delta_{ji}$, where δ_{ji} is a sample drawn from a Gaussian distribution with zero mean and variance 0.0025.

The hippocampal layer is trained as a Hebbian Network. Hence, the activation of node j in the hippocampal layer can be computed and updated as the following equations:

$$y_j(t) = f\left(\sum_{i=1}^{n} u_{ji}(t)x_i(t)\right), \text{where } f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

$$w_{ji}(t+1) = w_{ji}(t) + \alpha_{hipp}x_i(t)y_j(t) \quad (2)$$

The output $\vec{y}(t)$ is sent to both the BLA and vmPFC as the encoded context.

6.3 BLA and vmPFC: Model structure and training rules

Both the BLA and the vmPFC are single layers of 40 nodes as the input is a vector encoding both the context from the hippocampal layer and the original US and CS semantic pointers. The activation method of the units is the same as it is shown in the hippocampal layer in equation (1). However, the training process is different from the

hippocampal layer, where the Hebbian Network was used. This is due to the fact that the BLA and vmPFC act like competitors: the BLA is recruited during fear acquisition while the vmPFC is recruited during fear extinction. Research has shown that it is the prediction of the US that decides which of the two part of the brain wins the competition. Hence, the temporal difference (TD-algorithm) algorithm technique [13], a type of reinforcement learning, comes in handy.

The TD-algorithm can be described as follows. If a US is represented but the prediction of the BLA tells the opposite (a positive TD error signal), fear acquisition should be reinforced. On the other hand, if a US is not presented while the prediction of the vmPFC says otherwise (a negative TD signal error), fear extinction should be reinforced. Hence, the BLA and vmPFC should be able to predict the existence of the US based on the trials it has experienced. The TD error can be computed as follows:

$$TD(t) = R(t) + \gamma P(t) - P(t-1) \tag{3}$$

R(t) tells whether is US is actually presented at trial t, which takes on value 1 if the fear US is presented or value 0 if not. γ is the discount factor, which is set to 0.99 in later simulations. P(t) is the prediction made by the BLA and the vmPFC, which can be computed as equation (4).

$$P(t) = \sum_{i=1}^{n} w_i(t) x_i(t) \tag{4}$$

It should be noted that the weights $w_i(t)$ in equation (4) should not be confused with the weights w_{ji} in equation (1) and (2). These two are different weight matrices. The former only exists in the BLA and vmPFC to make predictions on the US in each trial, while the latter are weights that convert input to output in each layer. The weight updating rules are shown in equation (5)-(8).

The BLA layer is trained on positive TD error signals.

$$w_{ji}(t+1) = w_{ji}(t) + \alpha_{BLA} TD(t) x_i(t) y_j(t) \tag{5}$$

$$w_i(t+1) = w_i(t) + \alpha_{BLA} TD(t) x_i(t) y_j(t) \tag{6}$$

The vmPFC layer is trained on negative TD error signals.

$$w_{ji}(t+1) = w_{ji}(t) - \alpha_{vmPFC} TD(t) x_i(t) y_j(t) \quad (7)$$

$$w_i(t+1) = w_i(t) - \alpha_{vmPFC} TD(t) x_i(t) y_j(t) \quad (8)$$

6.4 Central Nuclei of the Amygdala (CeA): Model Output

The central nuclei of the amygdala is where fear responses are initiated. It receives direct input from the BLA and indirect input from the vmPFC, which are responsible for fear acquisition and extinction. There is only a node in our model of the CeA, and the output is the strength of the fear response, which is calculated as the difference between the activation of the BLA and vmPFC.

$$CeA(t) = \sum_{j=1}^{n} y_{j,BLA}(t) - \sum_{j=1}^{n} y_{j,vmPFC}(t) \quad (4)$$

This output stands for the intensity of our fear response. A higher value means that the fear-emotion is strongly triggered to the stimuli represented.

7. Results

7.1 A scenario for conditioned fear

In order to simulate the conditioned fear response, we would like to come up with a situation in which conditioned fear can be described well. Imagine you are a 10-year-old kid sleeping in your bed in the morning. The clock is ticking and you are already late for school. Your mom is waiting for you at the door, expecting you to have gotten dressed already and be ready for school in any minute. However, you are just so addicted to the bed that you constantly ignore her urging. After a few minutes, you are still in bed and haven't showed up. Your mom finally loses her patience and stomps all the way to your room, her footsteps so loud that you can hear in your sleep. When she finally gets to your room and finds out you are still in bed, she lost control and screamed in anger. The scream is so terrifying that you jump out of bed immediately.

The scenario above is actually a possible situation for the 10-year-old to learn a conditioned fear. The unconditioned stimuli (US), which is the actual stimuli that should trigger a fear response (which is jumping out of bed in this case), is the

mother's terrifying scream. The conditioned stimuli (CS), which happens along with or shortly before the US, is sound of the mother's anxious footsteps. The footsteps are neutral at first, since the child does not consider the sound of footsteps as terrifying. However, as the scenario happens more and more, the child would eventually develop a conditioned fear such that the sound of footsteps is able to trigger his fear response, even without the US presented. The context in this scenario could be anything the environment of the room. We choose the clock ticking sound in our simulation.

7.2 The simulation

In our simulation, we first gathered three audio files, a women's scream, a sound track of footsteps, and a clock ticking sound. Each of the audio files are about 2 to 3 seconds long. The files can be found on the website (https://www.soundjay.com/).

The below plots are based on a .wav file, in which a woman screaming for about three seconds of is presented. Four filtered frequencies are chosen to be observed, which are equally spaced in the interval [20, 8000] Hz after transformed into Mel scale.

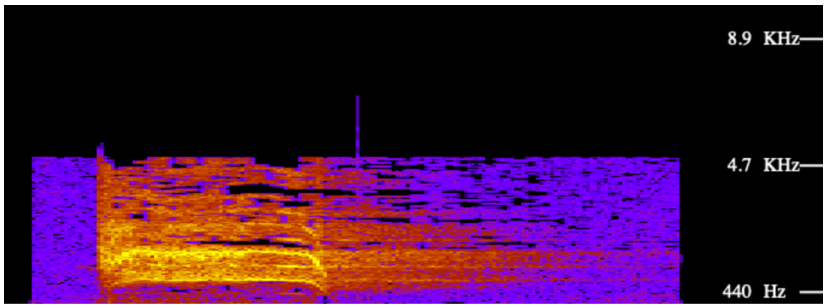

Fig. 7-1 The audio spectrum of the woman screaming .wav file

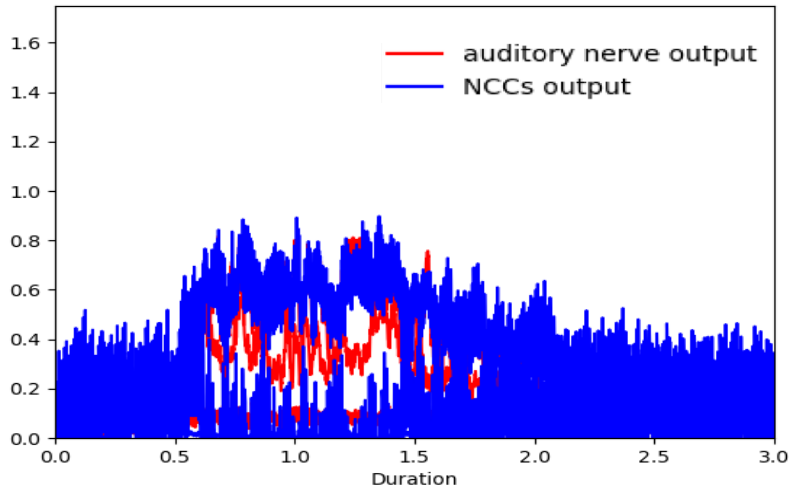

Fig. 7-2 The output of the ensemble array of Auditory Nerve, that of the ensemble array of Cepstral Coefficients.

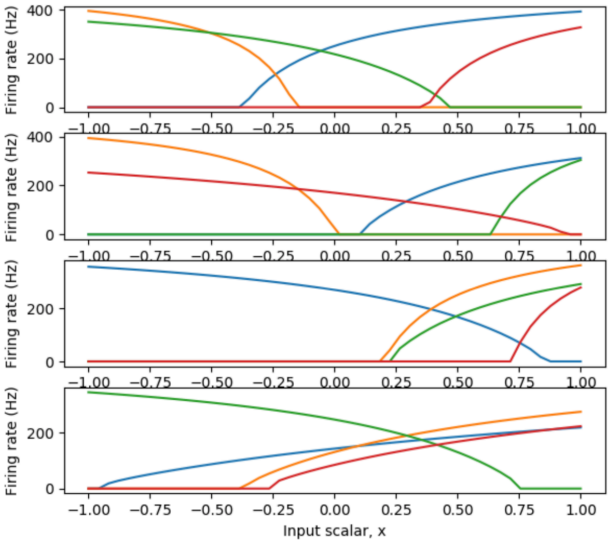

Fig. 7-3 The tuning curves of the ensemble array of Cepstral Coefficients

After the designated frequencies are extracted and processed through auditory periphery, a high-dimensional feature vector is then extracted through NCCs network, whose dimension depends on the duration of the audio file. In order to fit in the limbic system, dimension reduction is required. Additionally, the audio files acquired from the website often begin with a short period of non-characteristic sounds such as white noise or *blank*, worsening the differentiation between two audio files. In order to reduce the dimension and to trim out non-characteristic sounds, we chose 0.5 to 1.5 second of the audio file and segment it into frames with 50ms each, followed by averaging the features in each frame. Instead of selecting the features every 50ms, averaging the features in each frame prevents from bad selections such as ignoring characteristic moments of the audio file, while the sequence of time can be conserved as well. After framing the features and reducing the dimension, the new feature vector is then fed into the model for the limbic system for the following simulation of triggering fear response.

7.3 Fear Acquisition

Once we have the encoded semantic pointers of the three audio files, we can start simulate the fear acquisition process, which consists of constantly pairing CS(footsteps) and US(screaming) until the CS(footsteps) is able to trigger fear response itself. 50 trials were simulated. In each trial, we represent the context (clock-ticking), US (screaming), CS (footsteps) at the same time to the model. In order to test the validity of our model, we also obtained another audio file of a different footstep

sound. This audio file is also transformed into an encoded vector in the same way. We name the original footstep sound “CS+”, and the other as “CS-”. CS+ is the actual CS that is paired with the US during the training phase of conditioned fear, while CS- only act as a control group and is never paired with the US. After each trial, we would evaluate the results by only representing CS+, CS-, and the context to the model to see if each of them are able to trigger the fear response itself. Hence, in the evaluating phase, the US(scream) is not presented to the model. An encoded vector of an audio file of white noise is presented instead, indicating that there is no signs of any significant sound.

The fear acquisition process is depicted in (Fig.7-4). As the trials go on, the conditioned fear response triggered by CS+ only increases, as the blue line shown. The fear response triggered by the context is also acquired, but a bit slower than CS+. Since CS- and CS+ are both footstep sounds, although CS- is never paired up with US, it eventually would also develop a conditioned fear response as well, as the orange line shown.

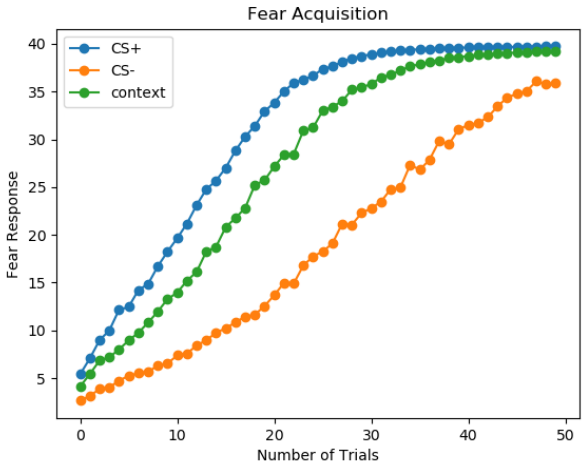

Fig. 7-4 The learning process of fear acquisition

7.4 Fear Extinction

After the model has learned conditioned fear, we can also put it through a fear extinction process. Fear extinction is simulated by repeatedly presenting CS+ without presenting US. At first, the model still shows fear response to CS+ itself. However, after a few trials, the fear response begins to decrease, as it is shown in (Fig. 7-5). The extinction process is rather slower, where the fear response gradually decreases after 400 trials and keeps reducing after 1500 trials.

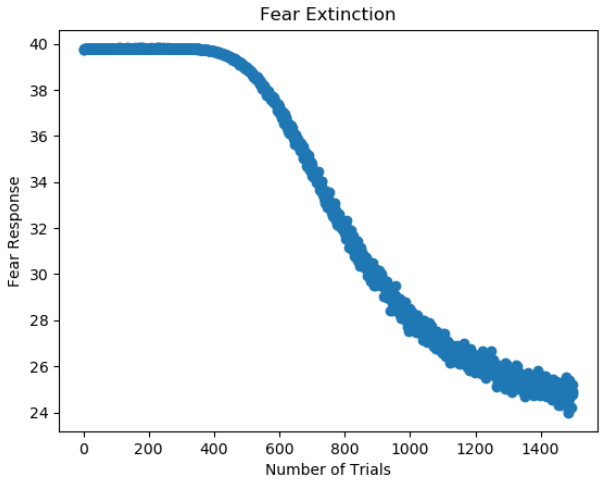

Fig. 7-5 The learning process of fear extinction

8. Discussion

8.1 Choosing feasible frequencies and sampling rate for the hearing model

The encoding of the stimuli differs with different frequencies and sampling rate directly affects the dimension of the feature vector. In this work, we chose four frequencies to be extracted from sound files, while in a more practical situation, a set of infinite frequencies should be observed simultaneously. Furthermore, the sampling rate in this work is chosen to be 50ms each frame, while in a real life scenario, the sampling rate should be as minimized as possible for continuity. One possible solution is frame shifting with half of the frame length, which may reduce boundary effects.

8.2 Our model does not learn to distinguish CS+ and CS- after training

As the results from Fig. 8-1 show, both of the footsteps sound are able to trigger fear response as the trials go on, even though one of them never actually shows up with the US. In a real life scenario, one would actually learn to tell the difference between the footsteps followed by a scream from the other. Hence, the fear response by CS- should decrease eventually, as the model learns to distinguish the difference and knows that this type of footstep does not precede a scream.

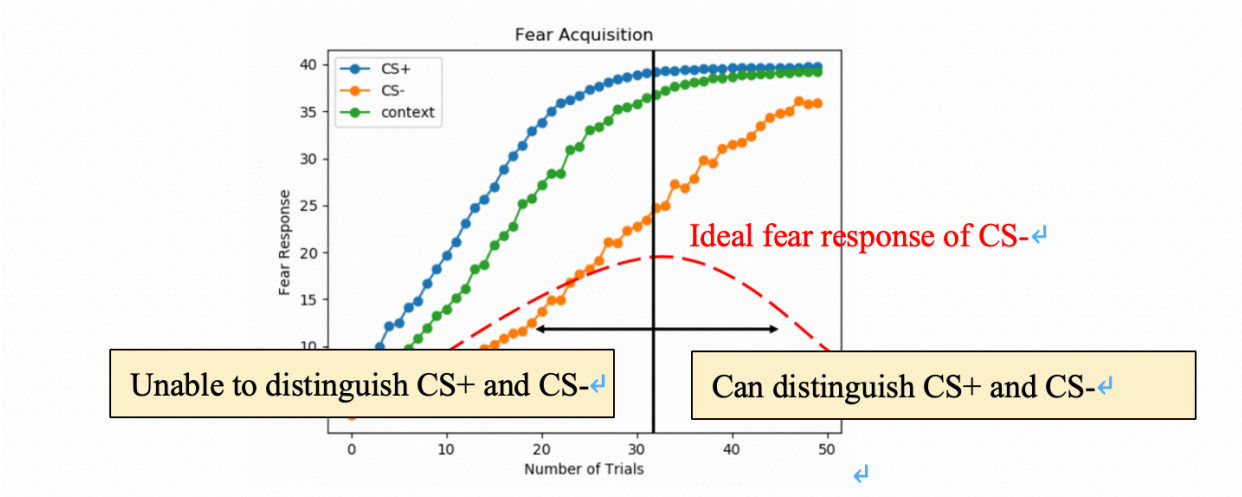

Fig. 8-1 The ideal response to CS-

Our model does not show this phenomenon, though. This is because our model lacks feedback from the limbic system to the NCCs neural network. A more realistic model should have this pathway in order to know when and how to distinguish the two similar sounds, and further learning should be conducted to distinguish between the two sounds.

9. Conclusions and Future Work

In this work, we first implemented the hearing process by following steps. First of all, the .wav file passes through a filter bank using Brian.hears package. The number of filters can be specified in the arguments, with a default of 4. The signals extracted are served as the signals read by inner hair cells. Then, the signal is projected to Auditory Periphery layer, in which the neurons simulate the activity of spiral ganglion cells projecting down the auditory nerve. After passing the inverse discrete cosine transformation, the signal is projected to the Cepstral Coefficient layer, in which another neural network is constructed, with the output of features extracted by NCCs. The feature vector is then provided for the simulation of limbic system in triggering fears.

The fear conditioned circuit is simulated and able to produce results of conditioned fear, including fear acquisition and extinction. Simulation results are consistent with the psychological results in Pavlov conditioning. The computational model proposed not only takes the interactions between the amygdala, the prefrontal cortex, and the hippocampus into account, but it also considers the differentiated nuclei in the amygdala rather than treats it as a black box.

For future work, lesions to different brain regions may be simulated to check if the results are consistent with the biological basis. We can also add feedback network from the limbic system model to the hearing model, since this two parts of the brain communicates bidirectionally rather than one-direction in our current model.

Another attemptable improvement to the model may be adding other basic emotions aside from “fear”, such as “happy”, “sad”, “angry”, “disgusted”, so that the model not only tells the intensity of emotional response, but is able to tell which emotion it is showing. This may be challenging since the pathways and circuits in the amygdala for those emotions other than fear still remains unclear. In order to achieve this, we would either have to treat the amygdala as a black-box, or we would have to have full understanding of the biological pathways and construct a new pathway for each of the emotions.

10. References

[1] Bechara, Antoine. "The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage." Brain and cognition 55.1 (2004): 30-40.

[2] Ekman, Paul. "Basic emotions." Handbook of cognition and emotion 98.45-60 (1999): 16.

[3] Lotfi, Ehsan, Saeed Setayeshi, and Saeed Taimory. "A neural basis computational model of emotional brain for online visual object recognition." Applied Artificial Intelligence 28.8 (2014): 814-834.

[4] Chittka L, Brockmann A (2005) Perception Space—The Final Frontier. PLoS Biol 3(4): e137. https://doi.org/10.1371/journal.pbio.0030137

[5] Bear, Mark F., author. Neuroscience: Exploring the Brain. Philadelphia :Wolters Kluwer, 2016.

[6] Trevor Bekolay (2016). Biologically inspired methods in speech recognition and synthesis: closing the loop. UWSpace.

[7] Tan, Q., and Carney, L. H. (2003). A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide. J. Acoust. Soc. Am. 114(4 Pt 1), 2007–2020. doi: 10.1121/1.1608963

[8] Bear, Mark F. et al. (2016) Neuroscience: Exploring the brain (4th edition)

[9] Knight, David C., et al. "Amygdala and hippocampal activity during acquisition and extinction of human fear conditioning." Cognitive, Affective, & Behavioral Neuroscience 4.3 (2004): 317-325.

[10] Anagnostaras, Stephan G., et al. "Scopolamine and Pavlovian fear conditioning in rats: dose-effect analysis." Neuropsychopharmacology 21.6 (1999): 731.

[11] Lebrón, Kelimer, Mohammed R. Milad, and Gregory J. Quirk. "Delayed recall of fear extinction in rats with lesions of ventral medial prefrontal cortex. " Learning & memory 11.5 (2004): 544-548.

[12] Moustafa, Ahmed A., Catherine E. Myers, and Mark A. Gluck. "A neurocomputational model of classical conditioning phenomena: a putative role for the hippocampal region in associative learning." Brain research 1276 (2009): 180-195.

[13] Sutton, Richard S., and Andrew G. Barto. Introduction to reinforcement learning. Vol. 2. No. 4. Cambridge: MIT press, 1998.