Temporal Context Effects on Attention in an Auditory Oddball Task

Eunice Pang

COGS 402: Research in Cognitive Systems

Supervisor: Max Jativa

December 10, 2020

Studying Attention and Memory

Attention is an important part of learning and cognition. Without making a conscious effort to focus on something, our attention will drift and we will have a more difficult time making sense of incoming stimuli because there is so much of it in our environment. Attention and other aspects of cognition are quickly becoming popular topics of study as technology advances and the development of artificial intelligence pushes for a deeper understanding of cognitive processes such as memory, learning, planning, and perception.

Memory is a crucial part of attention and we cannot keep track of our memories without the concept of time. Many models have been explored to try to uncover the internal workings of memory but “it is far from clear how event durations and temporal sequences are encoded in memory” (Teki et al., 2017). In addition, expectation also plays a role in memory formation and attention because “much of what a person can remember is based on their expectation of the information they will need to recall” (“Expectation may be essential to memory formation”, 2016). Expectation, time, memory, and attention are all intricately linked in cognition. The scientific community still has a long way to go in terms of researching and understanding how time fits into the bigger picture of memory and attention because “while the notion of working memory extends naturally for sensory information like visual or auditory signals, it is not clear how it applies to temporal information, even though conceptually, time and memory are closely interlinked.” (Teki et al., 2017). In other words, we are still trying to figure out exactly how the temporal components of memory contribute to attention.

Modelling Attention

Since attention cannot be measured directly, it is often measured indirectly, by obtaining the electroencephalography (EEG) data or recording behavioural reaction time data of a participant doing a specific task. In other words, attention can be operationalized as the change in brain waves as the participant completes a task, or the time it takes to do a specific action within the task. After obtaining data from an experimental task, hypotheses about the brain can be tested using the data collected.

One way to understand the various forms of cognition is to test theories about our understanding by modelling what we think to be true about a specific process and then testing the model against real data. Models of cognition allow us to develop and test hypotheses about how the brain perceives, processes, stores, and retrieves information. Furthermore, by fitting models to data, we can test whether or not the model is providing accurate assumptions about the underlying cognitive processes that we can then continue to confirm or dispute with further research.

Sequential Search Asymmetry Study

An aspect of attention that is specific to learning is sustained attention. Professor Blundon and her colleagues at the University of British Columbia carried out multiple oddball task studies. Oddball tasks are a type of psychological paradigm that are used to test attention. In oddball tasks, participants are instructed to respond to “deviant” or “oddball” stimuli presented in the midst of “regular” stimuli. This task aims to test sustained attention through recording EEG and reaction time data as the task is completed.

In this study, participants hear sequences of tones in “runs”. There are two types of runs — flat runs and change runs. Flat runs consist of five identical tones played consecutively and change runs consist of four identical tones, followed by one tone of a different frequency. At the start of the experiment, participants hear the common run 30 times to get them accustomed to it. After that, the rare run that participants are instructed to search for would be dispersed within common runs — specifically, “the rare run was presented on a random 20% of occasions among 80% of common runs” (Blundon et al., 2017). As pictured in Figure 1, there are four variations of sequence runs in total: feature present (change up), feature absent (flat low), feature present (change down), and feature absent (flat high).

Within the sequence runs, the participants had to indicate when they heard a deviant tone — the fifth tone on a rare run — by clicking a mouse. The recorded data is averaged across all the trials and it is the only thing that is common across the trials so the change in brain waves and the reaction times can be assumed to be a result of the participant’s reaction to the oddball task, and not due other factors such as the participant’s mind wandering.

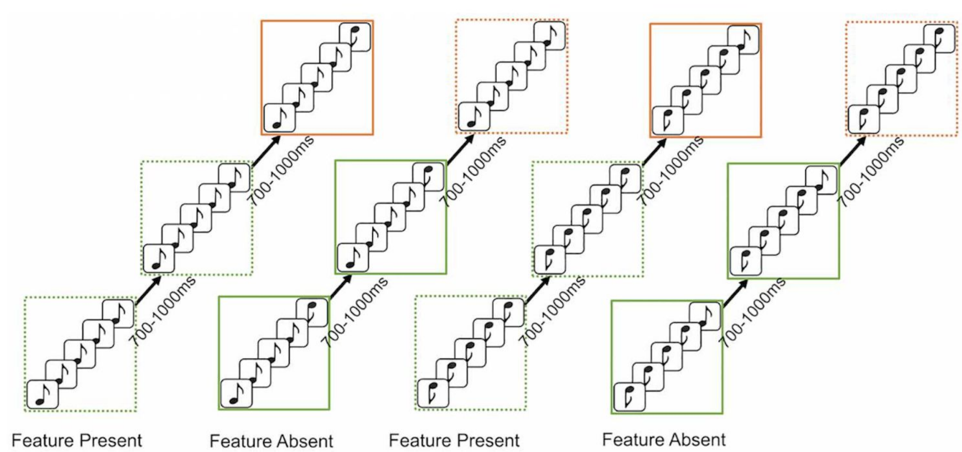

Fig. 1: Examples of sequence runs and how they are structured in the oddball experiment. (Blundon et al., 2017).

Modelling the Data

As mentioned earlier, it is difficult to model memory without including some component of time. One way to include time in models is to borrow the idea of incorporating a drift equation where “the new state of context, t_i, is derived from the previous state t_{i-1}, and the input t_i^{IN}” (Howard & Kahana, 2002). In other words, previous outputs are inputted into the equation along with the new input to incorporate a sense of “memory” into the model.

In order to model data collected from the auditory oddball task, existing models of attention and learning were modified to include a temporal component and fitted to data to test if they could successfully model what was happening in the oddball task. In the model, the high tones were represented by a value of 1, the low tones by a value of -1, and the silences were represented by zeroes. These sequences of numbers would be input into the modified version of existing models. The models were coded and plotted using Matlab software.

Drift Diffusion Model: Modelling Attention

The drift diffusion model is a model of attention describing how someone decides between two options during an activity called the two-alternative forced choice task. In this task, the participant must choose one out of two options presented to them. According to the drift diffusion model, as participants work through a task where they are forced to choose between two options, they accumulate “evidence” of which choice they are going to choose. Once enough evidence is collected, they settle on a choice.

It was found that a diffusion decision and memory model hybrid that was developed by Ratcliff in 1981 could successfully “model for the representation of letter strings in short-term memory was able, when combined with the diffusion decision model, to correctly predict the full range of accuracy and RT data” (Ratcliff, 2008, p. 912). Although this model was for an experiment using strings of letters, the idea of incorporating memory into attention models was successful for visual stimuli implying that having a memory aspect to attention models is worth exploring for other senses such as auditory stimuli.

Rescorla-Wagner Model: Modelling Learning

The Rescorla-Wagner model is a model of learning that describes how we learn about our environment through making predictions and error correction. The idea behind the Rescorla-Wagner model is that our brains are hypothesis generating machines and we use predictive coding to learn about our environments. We generate hypotheses about the world, match our predictions about stimuli with actual input, and then error corrections are made. This model is an error correction model describing classical conditioning.

The Rescorla-Wagner model is a model of learning — not attention. However, when participants are learning about when they will hear the oddball tones, they are forming their own hypotheses whenever the fourth note ends and the fifth note is about to play. In other words, the participants are learning when to expect a rare run, so learning is very much involved and this model is a possible contender to be modified and fitted to the data.

The Modified Models

The modified versions of the drift diffusion model and the Rescorla-Wagner model are the base models for what the models could look like with a temporal component added to it.

The modified version of the drift diffusion model is an expression of the accumulation of evidence with an additional term in order to include values from the previous timesteps into the current output:

$$x(t) = x(t-1) + dx$$

where $dx = A(dt) + noise$

$x(t)$ is the total accumulation of evidence up until timestep t, A is the rate parameter, and $noise$ accounts for any noise that is found in the data. For example, if the model was fitted to EEG data, this would be noise in the brain waves of the participants. On the other hand, if the model was fitted to behavioural reaction time data, this would be any random movements of the participants that caused a recorded reaction, but in reality, wasn't an appropriate action for the task at that point in time.

The modified version of the Rescorla-Wagner model is an expression of error correction values with an additional term to include the error correction values from previous timesteps in the current output:

$$v(t) = v(t-1) + dV(t)$$

where $dV(t) = \alpha(\lambda(t) - v(t))$

In the Rescorla-Wagner model, $\lambda(t)$ is the maximum amount of learning that can occur and $v(t)$ is the amount of learning that actually occurred. The subtraction of the two terms represents the "error correction". α is a rate parameter that would normally be learned from multiple data sets in machine learning. For the purposes of this base model, it was set to a number between 0 and 1.

Fitting the Models to Data

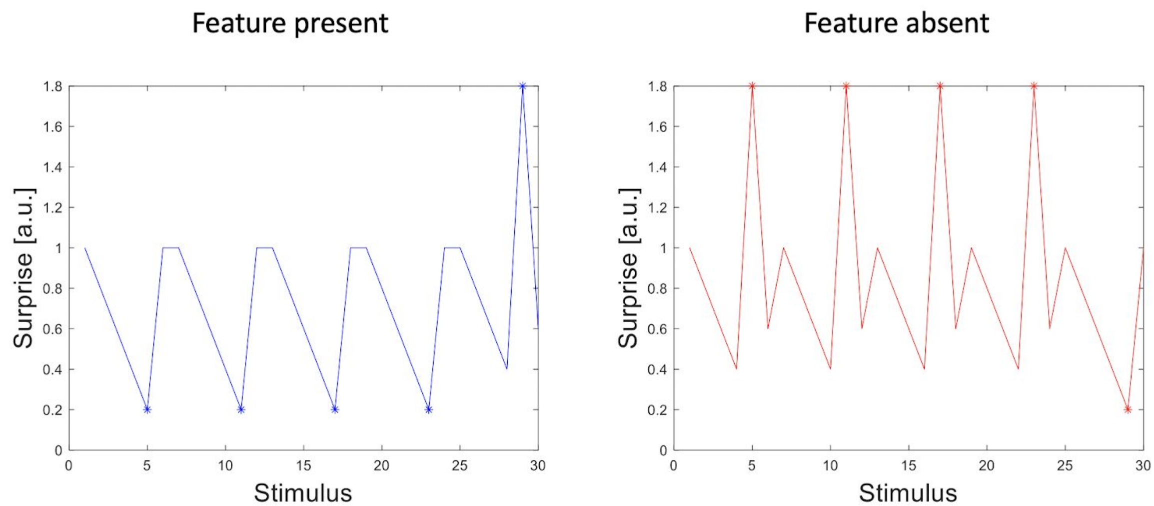

Fig. 2 Drift Diffusion Model: Surprise vs. Time for Feature Present and Feature Absent Conditions

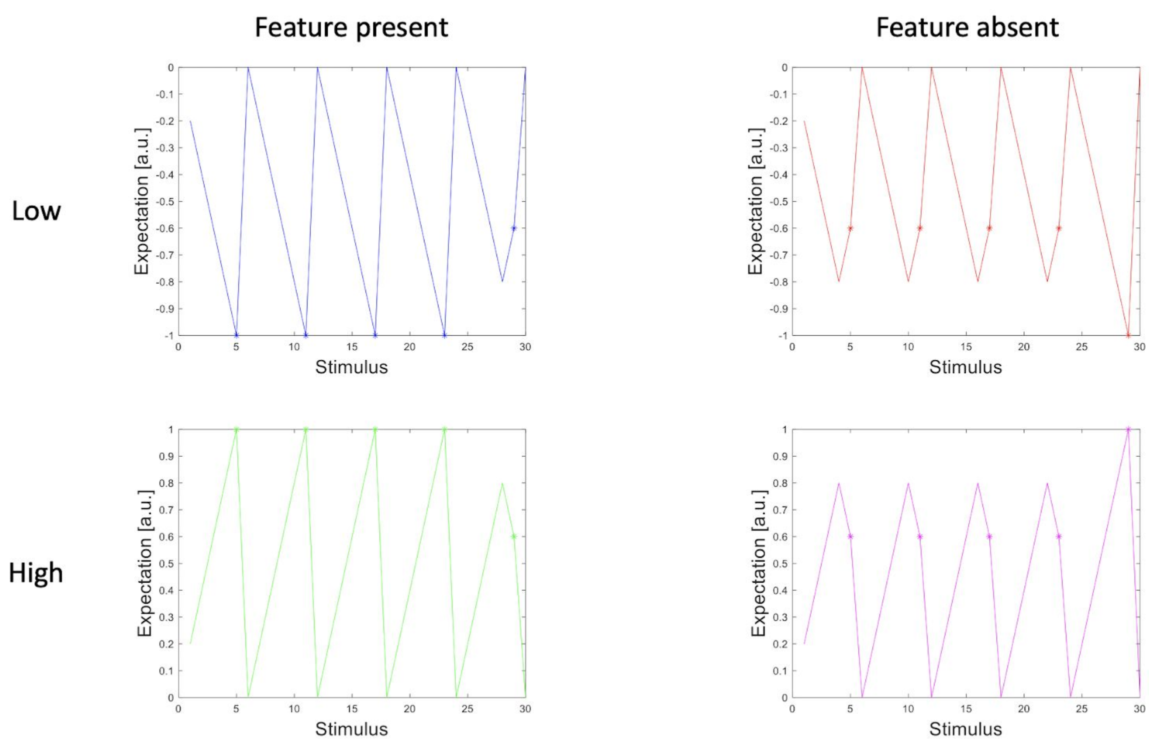

Fig. 3 Drift Diffusion Model: Expectation vs. Time for Feature Present and Feature Absent Conditions

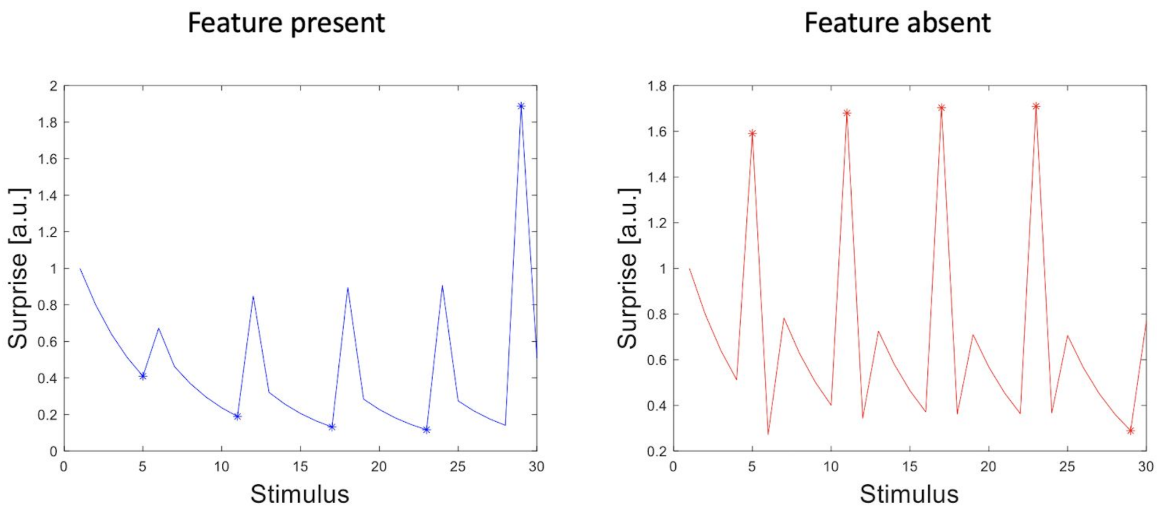

Fig 4. Rescorla-Wagner Model: Surprise vs. Time for Feature Present and Feature Absent Conditions

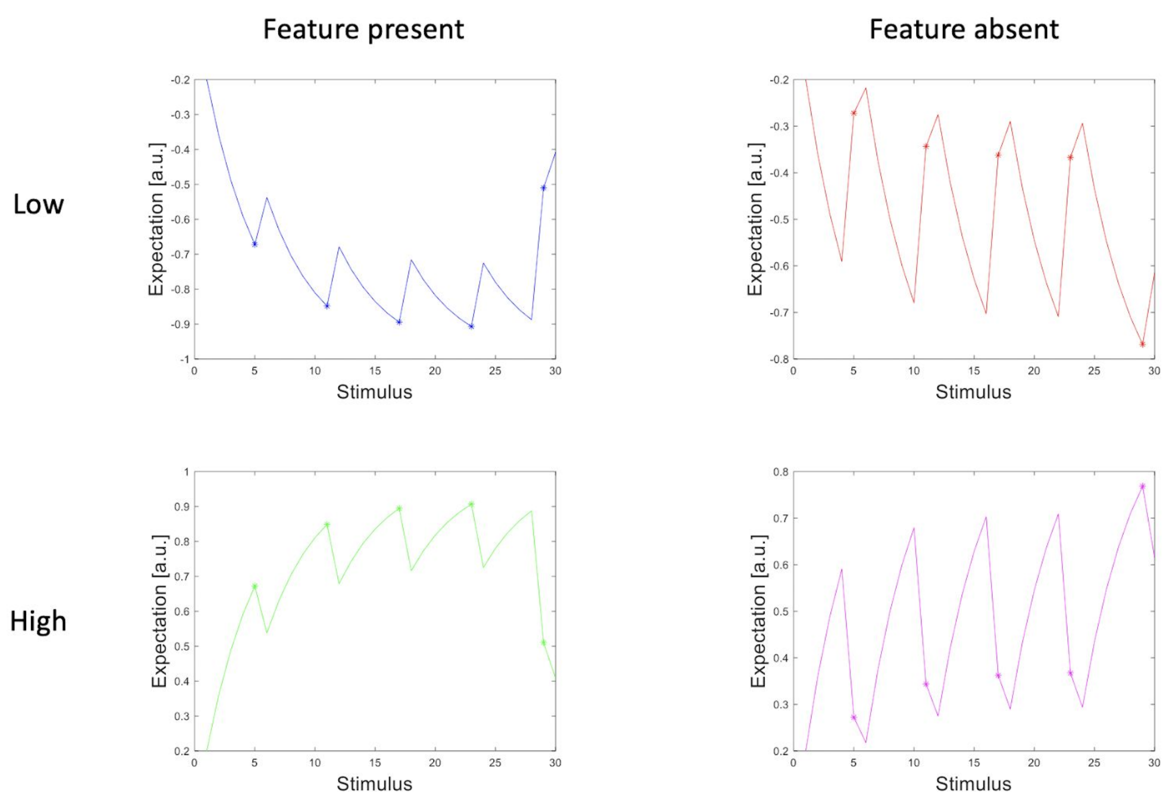

Fig. 5 Rescorla-Wagner Model: Expectation vs. Time for Feature Present and Feature Absent Conditions

All the graphs have arbitrary units on the y-axis, meaning the values do not mean anything specific, but rather, they show how the model performs at different time steps relative to other time steps. The stars on the graph indicate the fifth notes of the runs.

Figures 2 and 4 show surprise plotted against stimulus number (or time). As is seen in the feature absent graphs, after several common runs, the feature absent rare run should be surprising. However, due to how the model is coded, with high tones represented by a value of 1 and low tones represented by a value of -1, the model does not account for the surprise that the participant should feel. Had the model been able to predict surprise, the feature absent graphs for both models should see a spike instead of a dip on the fifth note of the sequence run.

Figures 3 and 5 showing expectation plotted against stimulus number are exactly what we hope to see for all four conditions in both models. In addition, change up and change down conditions as well as the flat low and flat high conditions mirror each other which is also what we would expect. As seen in the graph for the drift diffusion model, the change up run in the feature present condition shows that the fifth note is low where in previous runs it was high, and vice versa for the change down runs. This is exactly what is expected because the accumulation of the -1's for the low tones would cause the graph to go down for several runs until the last oddball run. Likewise, in the flat low run for the feature absent condition, the fifth note is low whereas in previous runs it was high, and vice versa for the flat high run. For the Rescorla-Wagner model, the graph depicting the feature present low (or change up) condition for expectation goes down during every run until the last run where it spikes up — exactly what is expected from that condition. The same pattern, but opposite direction, is seen for the feature present high (or change down) condition. For the feature absent runs, the flat low graph goes down and then up on the fifth note, and this pattern repeats for all the runs until the oddball run. This matches what we see in the actual sequence runs — four low tones followed by one high tone for every run until the last run where there are five consecutive low tones. Similarly, for the flat high graph, the same pattern is seen, but in the opposite direction.

Normally models would be fitted to the actual data in order to see if it can predict what the data shows. However, for this model, as the model did not work to predict what is expected of it using dummy data when modelling surprise, there was no need to also fit it to actual data.

Evaluation of the Models

Had the models been successful, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) could have been used to predict within-sample prediction error for the models. Both the AIC and the BIC take the maximum value of the likelihood function and the number of parameters as its input, and outputs the expected within-data error value. The equation for the AIC and BIC are:

$$AIC = 2k - 2ln(L)$$

$$BIC = kln(n) - 2ln(L)$$

where k is the number of parameters, L is the maximum value of the likelihood function, and n is the number of data points in the sample size. Since the AIC and BIC predict error, the lower the predicted value, the better. The first term in both equations includes k — the number of parameters. The larger k is, the larger the AIC or BIC values are because both equations penalize increasing the number of parameters. Any model can be fit to the data if enough parameters are present; the best models are ones where the model can predict the outcome or data it has never seen before with the least amount of parameters present in the model — in other words, the simpler the model, the better it is. The AIC and BIC do not give the absolute quality of a model, but rather, these equations are tools to evaluate the relative quality of models.

The value L refers to the output of the likelihood function which is the goodness of fit of a statistical model to a sample of data. The equation of the likelihood function consists of multiplying all the probabilities of the accumulation of evidence or learning after each tone in each sequence run, and then inputting that value into the AIC and BIC. The likelihood function is as follows:

$$\Lambda(A|X_1 \Lambda\, X_2 ... \Lambda\, X_n) = \Lambda(A|X_1) \cdot \Lambda(A|X_2) \cdot ... \cdot \Lambda(A|X_n)$$

where X_1 , $X_2, ..., X_n$ are independent events

Discussion

The base models of the drift diffusion and Rescorla-Wagner models developed here are not the only two modified models that could've been tested. These two base models were two of an infinite set of models that could have been developed and fitted to data. Further exploration of attention and learning models could possibly yield different results that are worth testing and reproducing with different sets of data or perhaps data from other sensory stimuli.

Since the goal of this project was to fit a modified version of existing models to data to test if they produce the desired results, the fact that the drift diffusion and Rescorla-Wagner models were found to be unsuccessful means the project overall was successful. The results prove that current models with a temporal component added cannot accurately model surprise. The way high tones, low tones, and silence are represented in these base models, as well as simply adding a temporal component to these models do not allow for the models to understand that a run with five identical tones should be surprising if it follows a run consisting of four identical tones paired with a higher or lower fifth tone. In order to model surprise in the feature absent runs, perhaps a model with a more integrated temporal context is needed.

References

Blundon, E. G., Rumak, S. P., & Ward, L. M. (2017). Sequential search asymmetry: Behavioural and psychophysiology evidence from a dual oddball task. *PLoS ONE 12*(3): e0173237. https://doi.org/10.1371/journal.pone.0173237

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269-299.

Penn State. (2016, March 28). Expectation may be essential to memory formation. *ScienceDaily*. Retrieved November 19, 2020 from www.sciencedaily.com/releases/2016/03/160328191855.htm

Ratcliff, R. & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation, 20*(4), 873-922. https://doi.org/10.1162/neco.2008.12-06-420

Teki, S., Gu, B., Meck, W.H. (2017). The Persistence of Memory: How the Brain Encodes Time in Memory. *Current Opinion in Behavioural Sciences*, 178-185. 10.1016/j.cobeha.2017.09.003