

Decoding motion prediction using EEG pattern classification

Luuk Spronck

Abstract

Due to the processing delay of the visual system, we would expect the perception of moving stimuli to lag behind their actual location. This is not the case: we are able to react to the veridical location of moving objects. Evidence suggests that motion prediction is possible due to extrapolation, but not much is known about the neural mechanisms underlying this phenomenon. We used multivariate pattern classification on EEG data to decode motion prediction. Pattern classification has recently gained much traction as a method for studying time sensitive neural mechanisms in the brain. We trained a classifier on easy-to-decode EEG data. Using this classifier, we compared time-to-peak latencies in classifier performance between an apparent motion condition, where motion extrapolation is possible, and a scrambled condition, where motion extrapolation is not possible. We found faster time-to-peak latencies for the apparent motion condition. This is in agreement with results found in previous experiments, and adds to the body of evidence for motion extrapolation as mechanism for motion prediction. We also take a short look at why linear discriminant analysis was used as our classifier.

1. Introduction

For humans, the visual world is a large part of how we experience everything around us. We are constantly processing and reacting to visual information. This information enters the eyes, travels through the optic nerve, and reaches the brain, where it is processed. Each of these steps takes time, which causes a delay before this information is processed and can be reacted to. This visual response latency can be as high as 100 ms (Maunsell and Gibson, 1992). However, this conflicts with every day experience. If we had no way of accounting for this delay, moving objects would be perceived noticeably behind their actual location. This would cause problems when, for example, you're playing sports, driving a car, or any other situation where we have to interact with fast moving objects. In the time that it takes for the visual information of an object to reach the brain and be processed, that object is already at a different location. If you were responding to the observed location of the object, and not its veridical location, you would not be able to react accordingly. Luckily, this doesn't happen. The fact that we can play baseball at all proves this. Somehow, our brain compensates for the neural processing time of visual information.

Humans perceive a moving object at a prognostic position rather than the position it was at when the visual information trajectory started. (Brenner, Smeets, 2000; Jancke, Chavane, Naaman, & Grinvald, 2004; Whitney and Murakami, 1998; Hogendoorn² & Verstraten, 2013). While this function of the brain is very important in our daily life, and its effects have been studied (Nijhawan, 1994), not much is known about its neural mechanisms. One explanation is that the brain prevents mislocalization of moving objects by extrapolating the trajectory of those objects. By observing the previous location of an object, it predicts its future location, constructing a representation of the veridical location. (Hogendoorn, Carlson, & Verstraten, 2008; Jancke et al., 2004; Nijhawan, 1994; Roach, McGraw, & Johnston, 2011). In other words, the observed location of a moving object is corrected by extrapolating its location along its path of movement. This means that the object will be perceived on its expected, or predicted, location instead of the physical location it was at when it was observed. This effect can be demonstrated by looking at optical illusions such as the flash-grab illusion (Cavanagh, 2013), the flash-lag effect (Mackay, 1958; Nijhawan, 1994) and the Fröhlich effect (Fröhlich, 1929).

This explanation arises from behavioral and physiological experimental studies. Berry, Brivanlou, Jordan, & Meister (1999) showed that anticipation of moving objects begins in the retina. They report that a moving bar elicits a moving wave of spike activity in the population of retinal ganglion cells. The population activity does not lag behind the actual visual image, as we might expect. Instead, it travels near the leading edge of the moving bar, apparently to help compensate for the visual response latency. Hogendoorn and Verstraten (2013¹) showed that as early as 80 ms post-stimulus, neural activation in the middle temporal area of the visual cortex codes for the perceived position of a stimulus. Sharma, Dragoi, Tenenbaum, Miller, & Sur (2003) showed that a brain can form an internal representation of a visual stimulus based on previous information. When this occurs, neural responses in the primary visual cortex were modulated: sequential presentation of a stimulus affected responses to that stimulus, yet random presentation did not. Hogendoorn et al. (2008) showed that the brain creates a relatively high level representation of a moving object using both extrapolation and interpolation.

Motion extrapolation has been demonstrated not just in humans (Hogendoorn et al., 2008), Sharma et al., (2003), but also in animals. Animal physiology experiments have reported anticipatory coding in the lateral geniculate nucleus in cats (Sillito, Jones, Gerstein, & West, 1994) and in the retina in salamanders and rabbits (Berry et al., 1999; Schwartz, Taylor, Fisher, Harris, & Berry, 2007). Jancke et al. (2004) used real-time optical imaging to explore the cortical mechanisms that underlie illusory motion perception of non-moving objects in cats. They found that under the right conditions, illusory motion created cortical activity that is indistinguishable from that which is created by real motion. Furthermore, they found a significant reduction in time-to-peak latencies (18 ms) when comparing illusory motion with no motion. However, it is unknown if this same effect is found in humans. We don't know how far ahead, in both time and space, the human visual system extrapolates the position of a moving stimulus.

Applying pattern classification to time-resolved recordings of brain activity, such as EEG (Deng, Srinivasan, Lappas, & D'Zmura, 2010; Hogendoorn and Verstraten, 2013¹; Hogendoorn and Verstraten, 2013²; Ratcliff, Philiastides, & Sajda, 2009), MEG (Carlson, Hogendoorn, Kanai, Mesik, & Turret 2011; Carlson, Tovar, Alink, & Kriegeskorte, 2013; Isik, Meyers, Leibo, & Poggio, 2014; King, Gramfort, Schurger, Naccache, & Dehaene, 2014), and multi-unit neuronal recordings (Meyers, Freedman, Kreiman, Miller, & Poggio, 2008; Stokes et al., 2013; Zhang et al., 2011), is an effective method of studying the human brain (King & Dehaene, 2014). It allows us to take a close look at the time course of interesting phenomena in the brain. These studies also show that multivariate pattern analysis is capable of enhancing the analysis of time-resolved signals. Multivariate pattern analysis (MVPA) is a technique that is often used for fMRI research. However, while fMRI has very high spatial resolution (1mm), and can be used to determine where in the brain something takes place (Yoo et al., 2004), it has quite poor temporal resolution (1s). Therefore, to analyze motion extrapolation, which happens on a scale of millisecond, not seconds, MVPA applied to time-resolved signals is much more useful (King & Dehaene, 2014).

We designed an experiment to induce motion extrapolation, and used pattern classification to decode the EEG signal we measured from observers. We hoped to find evidence for motion prediction. Using apparent motion, we generated motion perception. Gestalt psychologists (Wertheimer, 1912; Kenkel, 1913) showed that under the right spatiotemporal circumstances, sequentially presented stationary stimuli can induce apparent motion. By using a classifier to decode when a stimulus was at a certain location, and comparing this to a control condition, in which no extrapolation was possible, we expected to find reduced time-to-peak latencies in the conditions where motion extrapolation was possible, compared to the conditions where no motion extrapolation was possible.

To properly classify EEG data, we had to make sure that the classifier which we used could differentiate between two stimuli at different locations. We designed a preparatory experiment to test this. By making the difference between stimuli at two different locations as large as possible, we attempted to allow the classifier to more easily discern when the stimulus is at a particular location.

2. Pattern classification

Deciding which classifier to use for pattern classification is an important choice, as it can influence experimental results. In machine learning, the goal of classification is high prediction accuracy. However, this is not the main goal for classifiers used in cognitive neuroscience, where instead they focus on simplicity and ease of interpretation. Nonlinear classifiers have better performance in some applications, while linear classifiers are simpler in nature, and have equal performance in many other applications. Linear classifiers are more often used for brain decoding studies (Misaki, Kim, Bandettini, & Kriegeskorte, 2010; Müller, Anderson, & Birch 2003; Schwarzkopf & Rees, 2011), where high accuracy is not required and simplicity is a plus. Linear discriminant analysis (LDA), Gaussian Naïve Bayes (GNB), and support vector machines (SVM) are the most common classifiers used for decoding EEG, MEG, and fMRI data as experiments show that these three have the best overall performance (Grootswager et al., 2016).

Typically, the default classifiers used to decode fMRI data are SVM, as SVM are generally better with the large number of variables that fMRI data has. While a high-resolution fMRI scan will have thousands, sometimes hundreds of thousands of variables (Pereira et al., 2009), our EEG data only has 64 variables (we used 64-channel EEG). The classifier accuracy of SVM is often equal to that of LDA and GNB, but it is a more complex classifier, which affects the computational requirements. In our experiment, we repeated classification many times (per location and per time point). This argues in favor of other less complex, and thus faster, classifiers. Naïve Bayes is often called ‘the punching bag of classifiers’ (Lewis, 1998). It is fast and easy to implement, and as such is often used for text classification. However, compared to other classifiers, it is often lacking in performance (Yang & Liu, 1999; Zhang & Oles, 2001).

When choosing between LDA and GNB, it is important to take into consideration the assumptions that each classifier makes about the dataset. These assumptions differ between LDA and GNB, and whether or not the data satisfies these assumptions influences classifier performance. Classifier performance for both LDA and GNB improves when the assumptions are strictly followed. If the assumptions are not followed, performance will suffer. However, LDA performs well when there are only small deviations from some assumptions (Pohar, Blas, & Turk, 2004), while Naïve Bayes models have assumptions which severely affect the quality of its result (Rennie, Shih, Teevan, & Karger, 2003).

The GNB model assumes that variables, in our case the 64 electrodes, are independent of each other; they are not correlated. LDA makes the assumption that the correlation structure of the 64 variables is the same between groups. Groups, in our case, are the locations that the stimulus can be presented at. Figure 1 shows correlations between all 64 electrodes of four random groups from one observer at one time point. It shows that there is high multicollinearity between variables, and, while some groups have an overall higher multicollinearity than others, they all have the same structure; the same electrodes have a high correlation in each group. As EEG data measures electrical activity in the brain, it stands to reason that if one electrode measures increased electrical activity, other nearby electrodes will measure this same increase: this causes a high correlation between electrodes.

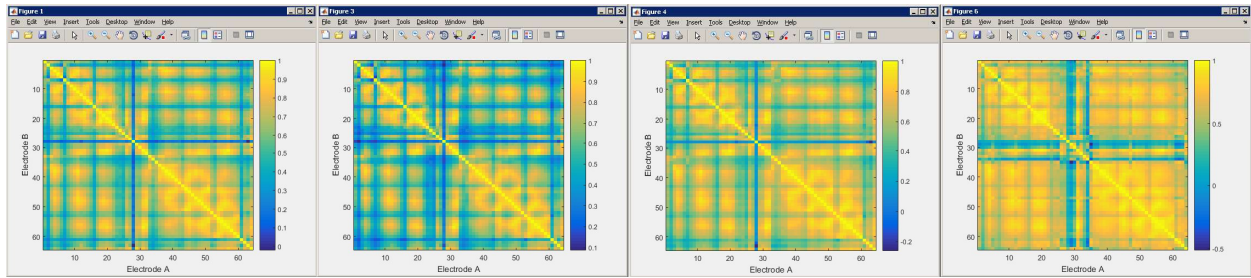


Figure 1: Correlation matrices between electrodes, four random groups from one observer at a single time point.

LDA requires that the size of the smallest group is larger than the number of predictor variables. We have at least a thousand trials per group, so groups are much larger than the number of predictor variables. Both GNB and LDA assume a multivariate normal distribution. However, Pohar et al. (2004) showed that LDA still performs well as long as distributions do not deviate much from normality. Figure 2 shows data from one group and one observer. This data seems to follow a normal distribution. While no test for normality can confirm this, this might be because EEG data has too much noise for these tests.

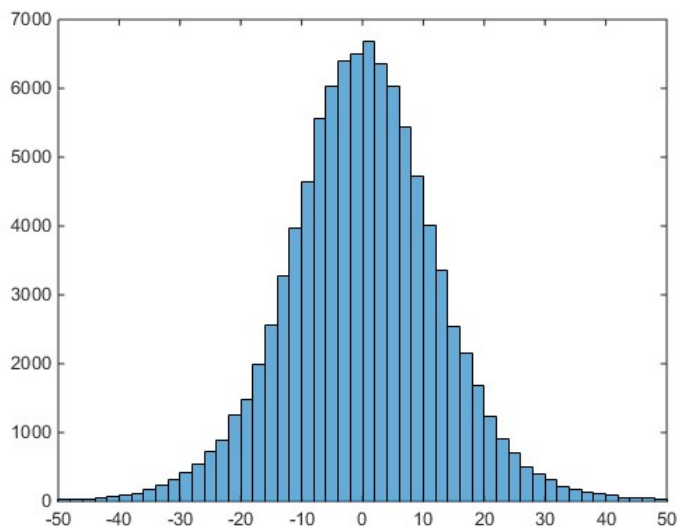


Figure 2: A histogram of EEG data from one observer and one group.

3. Method

3.1 General method

3.1.1 Observers

Six observers were used, with roughly six hours of EEG data being measured from each observer. The observers' ages ranged from 21 to 58, and all six were male. They all had normal or corrected to normal vision, and after completing the experiment they received a monetary compensation.

3.1.2 Materials and stimuli

A 27" ASUS ROG Swift LCD monitor running at 120Hz with a resolution of 2560x1440 was used to present the stimuli. This monitor was driven by a Dell Precision T3610 computer, running Matlab with the PsychToolbox extensions (Brainard, 1997; Pelli, 1997). A chinrest was used to keep observers at a constant 55cm from the screen.

Black, wedge shaped stimuli were presented on a uniform gray background. Stimuli were 200x250x300 pixels (roughly 1.8 cm at the narrowest, 2.8 cm at their widest, and 2.3 cm tall) and were presented in a regular octagon pattern, circling the fixation dot at a distance of ten centimeters. Monitor contrast was at 50% of maximum.

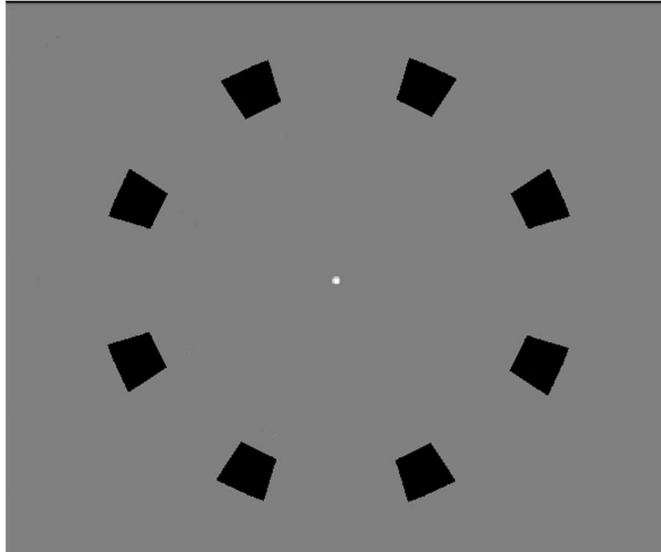


Figure 3: wedge-shaped stimuli at all eight possible locations.

3.1.3 Procedure

Observers were asked to rest their head on the chin rest and focus on the fixation dot. They were instructed to perform a simple target finding task to keep their attention on the screen. One trial was one presentation of the stimulus at one location. Each trial was on the screen for 8 frames, or 66.67 milliseconds. Following each trial was an inter-stimulus interval that differed from 52 frames and 433 ms in part one of our experiment, to four frames and 33.33 ms in part two. Target trials were discarded for the analysis.

Trials were displayed in sets, each set lasting roughly 32 seconds. Thus, each set contained 64 trials in part one and 256 trials in part two. Sixteen sets formed a single block of trials, each block lasting eight to ten minutes. Every set of trials, 3 or 4 random trials were red instead of white, and observers were asked to press a button whenever they saw the red trial. A set of trials was started by the observers pushing another button. They could pause for a few seconds between each set of trials and take longer breaks after each block.

3.1.3 EEG acquisition

To measure the observers' EEG activity, a BioSemi™ Active2 EEG system was used. EOG electrodes were placed above and below the left eye and at the outer edges of both eyes to measure eye movements and blinks. Reference electrodes were placed on both mastoids. The

EEG data was captured on a different computer, connected to the electrodes via a BioSemi™ AD converter.

3.2 Training the classifier

Our first goal was to train a classifier and test whether it was capable of differentiating between stimuli at two different locations. As such, apparent motion was not required and the stimuli were presented in a scrambled, not continuous, order. Twelve blocks were measured over two sessions on different days. This gave us a total number of 12,288 trials; 1,536 trials per location. A total of one and a half hours of EEG data was measured from each observer.

3.2.1 EEG analysis

Sixty-four-channel EEG data was recorded at 2048Hz, and was resampled offline to 512 Hz. It was referenced against the reference electrodes (to filter out noise) and epochs of EEG activity were created from 100 ms pre-stimulus to 500 ms post-stimulus, containing one trial per epoch.

A linear discriminant classifier was trained on these epochs, using all 64 non-reference electrodes. This classifier was trained on half of the trials (chosen randomly) and tested on the other half, then trained on the second half and tested on the first. For every trial, the classifier compared location A with location B, deciding which one was more likely. This was done for every combination of locations, except for comparisons which were superfluous (comparing location one with location one), or had already been done (comparing location one with two, then comparing location two with one). Comparing the decisions of the classifier with the actual location of the stimulus, and averaging these answers over trials, gave us a percentage of how often the classifier was correct for all 28 combination of locations at each time point (figure 4).

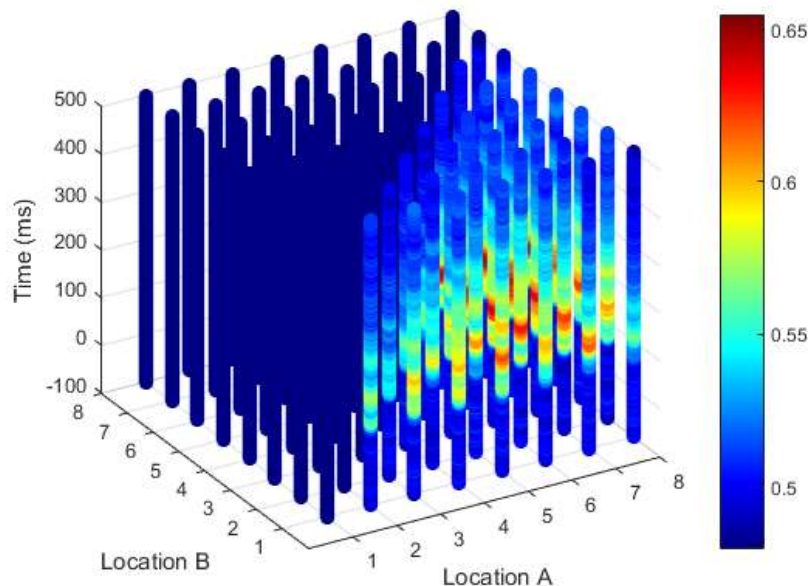


Figure 4: Classifier performance over time for each combination of location A with location B (except for when A equals B), averaged over observers. Peaks in performance for every comparison are around 120-150ms post-stimulus presentation. Dark blue lines are combinations of locations that gave no extra information and thus were not computed.

3.2.2 Results part one

Collapsing over location gave us classifier performance over time (figure 5). We found that, while classifier performance differed between observers, all observers' classifier performance reached peaks of 55% or more, with the most accurate reaching 66% accuracy. We then averaged over observers and found that peak performance was 59%. For every observer, classifier performance was roughly chance until 70-80 ms post-stimulus, after which it rose to its max at around 120-150 ms post-stimulus. Accuracy then slowly went down, until, at 400-500 ms post-stimulus, performance was back at chance level. These results confirm that the classifier is capable of classifying our EEG data at an above-chance level; we could now attempt to decode motion prediction.

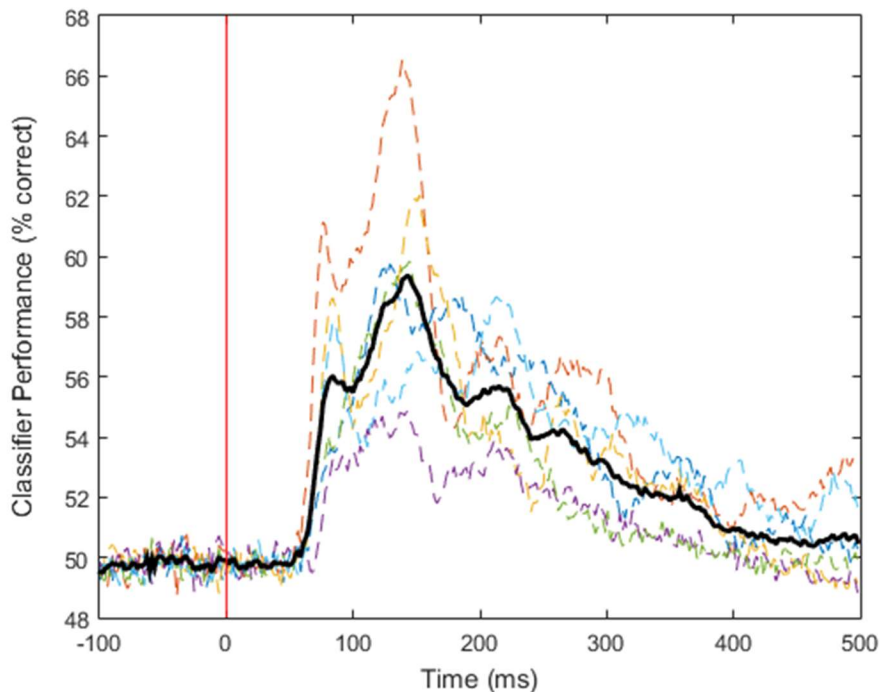


Figure 5: Classifier performance (percentage of trials classifier correctly) over time. The vertical red line is the stimulus appearance, the black line is an average of all observers, the dotted lines are individual observers. Performance is chance until ~70 ms post-stimulus, reaching its peak at 120-150 ms post-stimulus.

Out of curiosity, we calculated how accurate the classifier was based on how far away the compared locations were. We averaged classifier performance over time, then looked at how accurate the classifier was for every combination of locations that were next to each other, every combination of locations that were one location apart, and so on (figure 6). We found that classifier performance increased when the distance between locations increased. There is a larger difference between two locations that are farther away than two locations that are near each other. This is in agreement with the results Carlson et al. (2011) found when decoding object position using MEG.

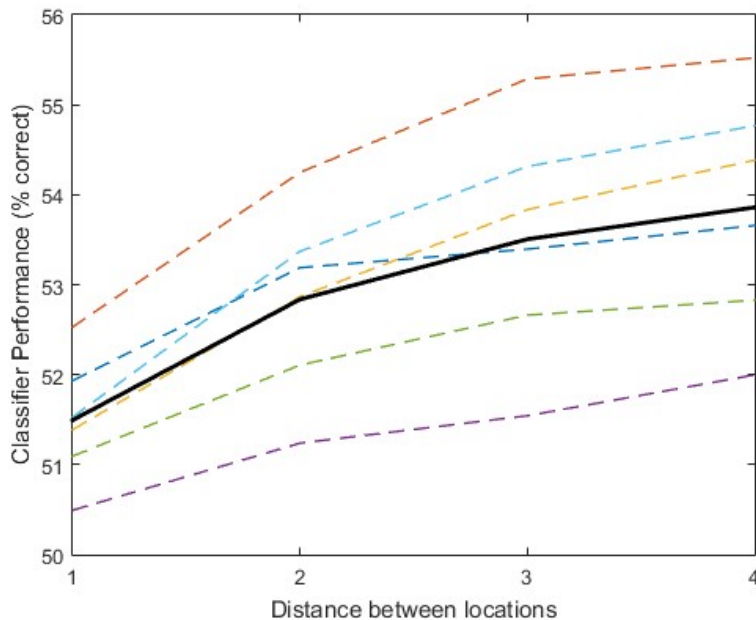


Figure 6: Averaged classifier performance based on distance between compared locations. The black line is an average of all observers, the dotted lines are individual observers. A distance of one means locations are next to each other (one step between them).

3.3 Decoding motion prediction

Assured that our classifier was capable of decoding stimulus location with reasonable accuracy, we proceeded with testing our hypothesis. Our goal was to find a difference in time-to-peak latencies between a continuous, apparent-motion condition and a control condition.

3.3.4 Procedure

We used the same basic setup as in part one of the experiment. Observers were asked to perform a simple target finding task while looking at the fixation dot. This time, a single set of trials consisted of 256 trials: the stimulus was presented at each location 32 times. Again, sixteen of these sets together formed one block. Twenty blocks were measured over three sessions, all on different days. A total of 81.920 trials was measured.

For the apparent-motion condition we used a clockwise and counterclockwise condition. As control conditions, we used scrambled versions of the continuous conditions, which we called the clockwise scrambled and counterclockwise scrambled conditions. All four conditions appeared equally often, so for each observer 20.480 trials per condition were measured. The order of the continuous conditions was always the same, starting at the first location and looping 32 times. In the scrambled conditions the order in which the stimuli appeared was constant during a single set of trials, but it differed between sets of trials to prevent observers from learning the order.

3.3.5 EEG analysis

The EEG data was, just as in part one, resampled down to 512 Hz and referenced against the reference electrodes. However, epochs differed from part one: eight trials were grouped together to form one epoch, starting at the moment the first trial was shown and lasting until 900 ms and 7 trials later. One epoch contained one presentation of each location, for the continuous

conditions, this meant one epoch contained one full rotation of the stimuli. The order of the scrambled locations differed between sets of trials; this was corrected when creating the epochs, so that the order in which the stimulus was presented at each location was the same for each epoch.

We attempted to train a linear discriminant classifier on this data, but found that classifier performance was lower than in part one. Because stimuli remain active in the brain for up to 500 ms after they are no longer visible (Carlson et al., 2011), the short inter-stimulus interval lead to interference from previous stimuli. The classifier in part one is able to classify our EEG data with above chance accuracy. By training the classifier on the data from the first experiment, then testing it on the data from the second experiment, we managed to increase classifier performance. Because the training set and the testing test were now two separate data sets, we could train the classifier on the full data set from part one, then test it on the full data set from part two. In part one we discovered that classifier performance peaks at 120-150 ms post-stimulus. Training the classifier on this time point gave us the most powerful probe to measure neural representation.

The classifier was trained on location A versus location B, using data from part one. This classifier was then tested on all data from part two. It decided which location was more likely for every time point of every trial. This was repeated for every combination of locations. Because each epoch contained eight stimulus presentations, we could not compare the result of the classifier with the actual location of the stimulus: the actual location was hard to determine as it changed very rapidly (every 100 ms) and its location at a certain time point differed between conditions. Instead, we looked at how many trials the classifier classified as location A, and how many trials it classifier as location B. For every combination of location A and location B, we calculated a ratio of how often the classifier chose location A over location B by averaging over trials.

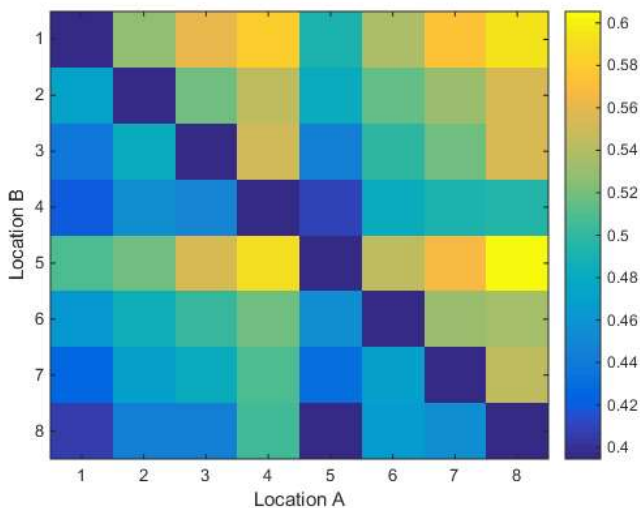


Figure 7: How often the classifier chose location A for each comparison of location A with location B.

Using this data, we calculated classifier preference for each location. We calculated how often the classifier chose a certain location when the chosen location was location A, then subtracted how often the classifier chose the other location when the chosen location was location B. Or, to put it another way, looking at figure 7, we subtracted the average of row A from the average of column A to calculate classifier preference for every location, shown in figure 8. One epoch contained one presentation of each stimulus location. We knew the order in which the stimulus locations were presented, so we were able to shifted the data in figure 8 over time, so evidence for each location began at the same time point.

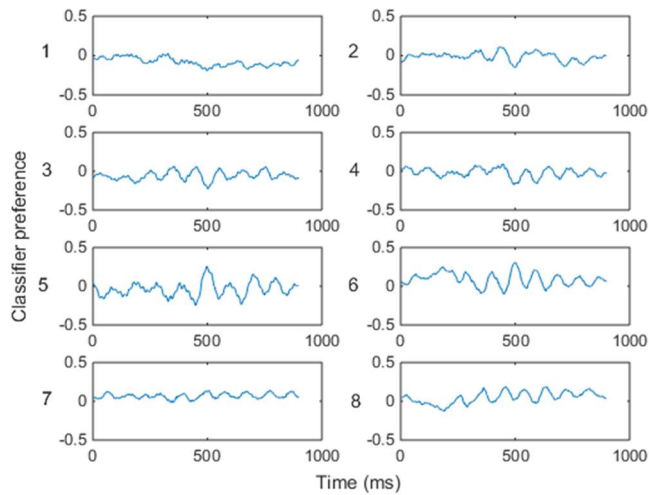


Figure 8: Classifier preference for all eight locations, for one observer, one condition, and one timepoint.

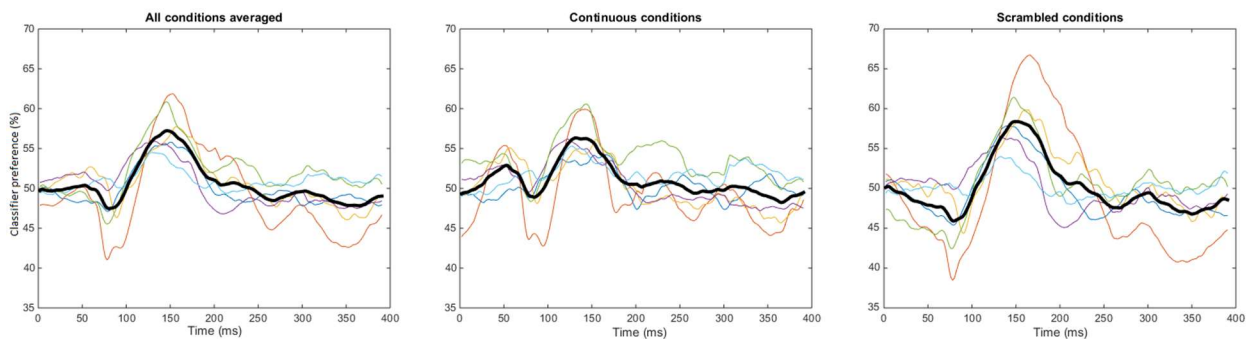
4. Results

4.1 Accuracy of the classifier

With classifier preference being time-corrected, collapsing over location showed an average preference over time for every condition. Figure 9.1 shows classifier preference of all four conditions, figure 9.2 shows classifier preference for the continuous conditions, and figure 9.3 shows classifier preference for the scrambled conditions. While figure 9.1 shows that classifier preference is, just as in part one, roughly chance level until 80 ms post-stimulus, figure 9.2 shows that there is small peak at ~50 ms post-stimulus, and in figure 9.3 performance is below chance until ~100 ms post stimulus.

There are some differences between classifier performance and classifier preference (figure 10). While some aspects, such as peaks, cannot be compared between part one and part two, as in part one we measured classifier accuracy and in part two we measured classifier preference, there are aspects that should be the same or similar. For example, when classifier performance and preference first rise above chance level. Whereas in part one this was at 70-80 ms post stimulus, preference did not rise above chance level until around 100 ms post-stimulus. The time that the classifier remains above chance level is also different than in part one. While in part one it took 300 ms to go from the peak back down to chance levels, in this experiment

classifier performance is back down to chance levels in only 60-100 ms. Looking at just classifier preference, peaks of the continuous conditions are lower than in the scrambled conditions.



Figures 9.1, 9.2, and 9.3. Classifier preference for all conditions, continuous conditions, and scrambled conditions. The black line is an average of all observers. The colored lines are individual observers.

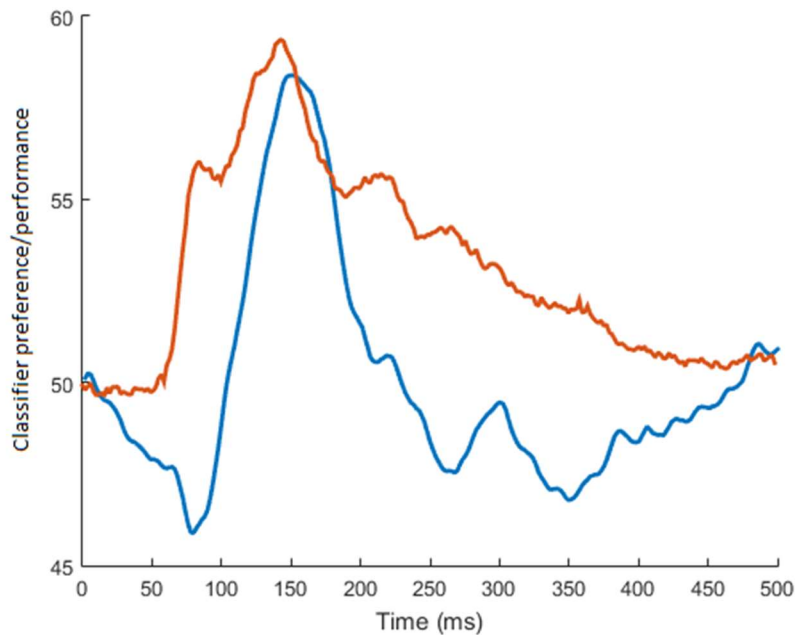


Figure 10: Classifier performance (orange line) and classifier preference scrambled conditions (blue line). While both have the same global shape, there are some differences.

4.2 Time-to-peak latency

Figure 11 compares classifier preference from the non-scrambled conditions with classifier preference from the scrambled conditions. The scrambled conditions have a higher peak, which is to be expected, as the preceding and following locations are further away and thus interfere less with the current location. When comparing time-to-peak latencies (90% of the maximum accuracy is considered the peak), we found that the two non-scrambled conditions were on average 15.6 ms faster than the scrambled conditions. In the continuous conditions, classifier preference reaches above chance levels at 90-95 ms post stimulus. This is also delayed in the scrambled conditions; they don't reach above chance levels until 105-110 ms post

stimulus. Figure 12 shows clearly when the continuous and scrambled conditions are at their peak.

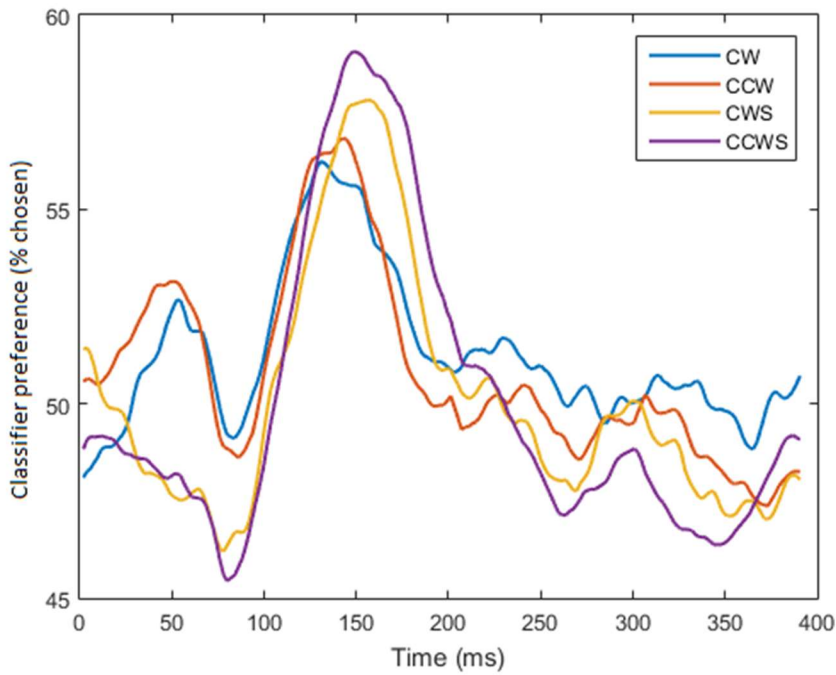


Figure 11: Classifier preference over time. Blue and red are continuous, yellow and purple are scrambled. Continuous conditions reach their peak sooner. Scrambled conditions have a higher peak.

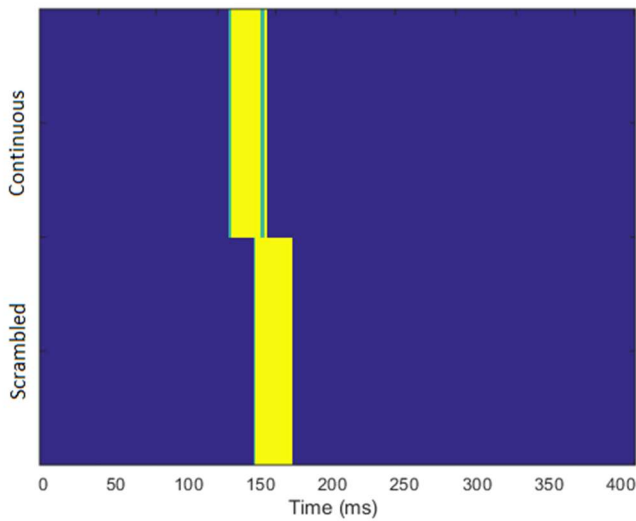


Figure 12: Blue is below peak performance, green is one of two conditions at peak performance, yellow is both conditions at peak performance. The continuous condition is both at peak and back below peak before the scrambled condition.

To test for significance, we performed permutation testing on the classifier data. We shuffled the trials over conditions, so any systematic differences that can be attributed to conditions disappear. We shuffled the data 10,000 times; a histogram of these shuffled runs is shown in figure 13. We found a maximum time-to-peak latency of 6.8 ms in these shuffled runs. Our result of 15.6 ms has a p-value of 0.0000; it didn't occur once in 10,000 shuffled runs. As such, it is highly unlikely to have arisen by chance.

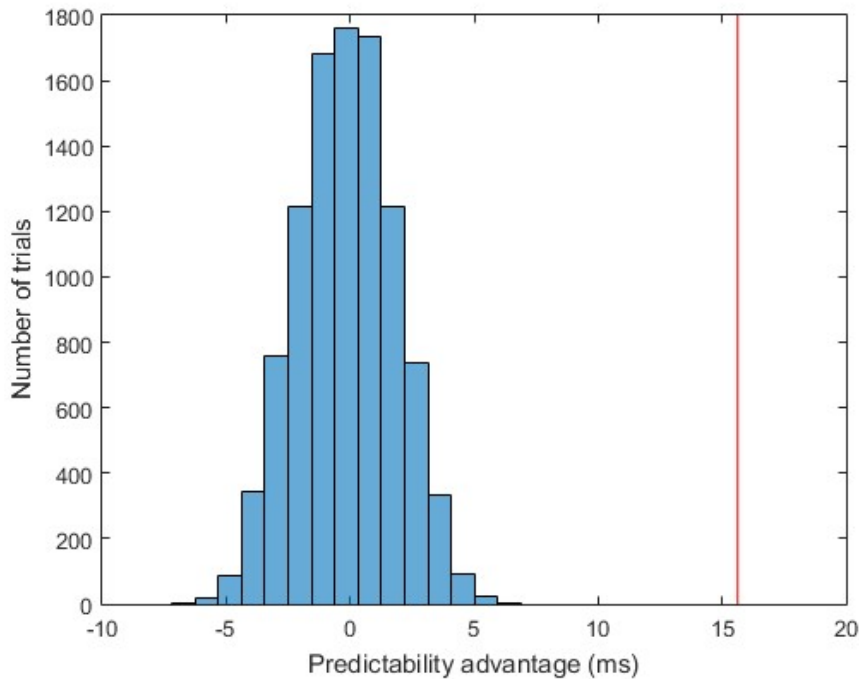


Figure 13: histogram displaying time-to-peak latencies of the shuffled trials. The red line is the real time-to-peak latency we found.

5. Discussion

5.1 Explaining the result

In this experiment, we tested whether or not we could decode motion prediction using EEG pattern classification. By using apparent motion to induce the impression of a moving object, we compared how long it took before we could decode stimulus location based on EEG data between an apparent motion condition and a scrambled condition. We trained a classifier on EEG data that had a larger delay between stimuli and used that classifier to decode motion prediction. At its peaks, this classifier reaches an accuracy of nearly 60%. As expected, stimuli become decodable at 80ms post-stimulus and classifier performance reaches its peak at 120-150 ms post stimulus.

There are some notable differences between classifier performance and classifier preference. The dip in performance preference before 80 ms post-stimulus in the scrambled conditions is likely due to a stimulus appearing at an opposite or far away location 100 ms ago. This will inhibit responses to other locations, explaining the decrease in performance preference. The small peak at 50 ms post-stimulus in the continuous conditions is likely also a response to

the previous stimulus, except, because it is close to the current stimulus, it provokes a similar response, leading to a small peak. The fast decrease in classifier preference after the peak in the fast condition can be explained by a new stimulus appearing soon after the current one has. The mental representation of the current stimulus is overwritten by the next stimulus, and accuracy drops.

To decode motion prediction, we looked at when classifier preference first reached 90% of its peak value, and found that the continuous conditions are, on average, 15.6ms faster than the scrambled conditions. We used permutation testing to ensure that this result was significant. Figure 12 shows that peaks both start and end earlier in the continuous condition. This is in line with our hypothesis: motion prediction allows for faster time-to-peak latencies. Jancke et al. (2004) found a similar result in cats: using real-time optical imaging, they found an 18 ms difference in time-to-peak latency between apparent motion and a non-moving stimulus.

There are two main trains of thought that could explain this result: extrapolation of moving stimuli allowing for motion prediction, or moving stimuli have a reduced visual processing delay. Purushothaman, Patel, Bedell, & Ogmen, 1998 and Whitney & Murakami (1999) argue that the visual system, instead of extrapolating and predicting movement, reduces the visual processing delay for moving stimuli. Their 'online' hypothesis proposes that stimuli that arrive in the visual cortex first are perceived before stimuli that arrive later. This view can explain some of the optical illusions that are also used as evidence for the extrapolation theory. However, the results of multiple studies make us question this hypothesis (Eagleman, 2010). They performed experiments using moving squares of differing brightness. If the online hypothesis were true, we would expect the brightest squares to be perceived ahead of the others, as they are more salient and thus reach the visual cortex first. Observers reported no such illusion; they did not perceive the brighter squares as moving in front of the others. This leads us to believe that the extrapolation hypothesis is more likely: we look at previous locations the stimulus was perceived at, and extrapolate its predicted location from that information.

5.2 Future research

While a ~16 ms delay is significant, the visual processing delay is much longer, up to 100 ms. So how can this 16 ms improvement account for a possibly much larger delay? Jancke et al. (2004) found that stimulus speed and direction affect the processing of moving objects. Higher stimulus speed leads to a larger reduction in time-to-peak latencies. This may account for some of the difference. There may also be other effects at play here which decrease the delay further beyond what motion extrapolation can do. It is likely that more research is needed in this area.

Following the results from Jancke et al. (2004), an interesting question is how speed and direction affect time-to-peak latencies in humans. By changing both how long a stimulus is presented and the inter-stimulus interval, it is possible to increase or decrease the speed of the apparent motion. By changing the placement of the fixation dot, apparent motion would be detected in a different part of the visual field. Both these changes are likely to affect time-to-peak latencies, and as such we would expect to find minute differences using these different conditions.

5.3 Pitfalls & discussion

There are a few caveats that should be considered when using time-series decoding methods. Double dipping happens when there is class-specific information in the test set, which

inflates classifier performance (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). Classifier performance being chance level until 80 ms post-stimulus, as our results show, is evidence against this. Another aspect of time-series of decoding that should be handled with care is that the strength of a signal (higher classifier performance peak) influences when decoding first reaches significance (Grootswager et al., 2016). Figure 11 shows that in both conditions, classifier performance starts to rise at 70-80 ms post-stimulus, while we would expect the scrambled condition to start later. This can be explained by the aforementioned: the scrambled conditions have a higher peak, which affects when classifier performance starts to rise. Finally, down sampling a signal smooths out the data, which may hide time-sensitive information. The raw EEG signal we obtained was 2048 Hz, but it was down sampled it to 512 Hz for computational reasons. Training and testing the classifier took multiple days with a sampling rate of 512 Hz so it was not feasible to keep the sampling rate at 2048 Hz. This decreases the number of timepoints by a factor of four, giving us one timepoint every two milliseconds. Because the effect that we found is on such a small scale (15.6 ms), decreasing the number of time points may have an effect on our results. However, permutation testing showed that the highest time-to-peak latency found by chance was 6.8 ms, so even in a worst-case scenario of down sampling affecting our result, the 15.6 ms difference we found is still significant.

We found large differences in classifier performance between observers. While classifier performance of all observers had the same basic shape – chance level until 80 ms post-stimulus, rising to a peak at ~120-150 ms post-stimulus before going back down to chance levels ~250 ms post-stimulus – peak accuracy ranged from 55% to 62%. These peaks also differed in time, from 128 ms post-stimulus to 156 ms post-stimulus. EEG data is always noisy. There is interference from the environment, from the observer moving or blinking, or even from the observers' hair. While we tried to filter out as much of this noise as possible, some is bound to slip through. This noise can explain some of the differences between observers. Of course, people are also different. Some people are very slow at perceiving things, while others are very fast. These two causes together might explain why the differences between observers can be quite large.

While we argued for using a linear discriminant analysis model, there are many other classifiers that could have been used. It is possible that using one of these other classifiers, such as a neural network or a k-nearest neighbour algorithm, would result in different results. Testing other classifiers and receiving the same result would improve the robustness of the results we found, and ensure that it is not a result of some innate preference of the classifier. We could have used eight-class classification instead of the two-class classification that we used. When we started our experiment, we were not aware that LDA was capable of performing multi-class classification. We believe that using eight-class classification would likely have improved the accuracy of our classifier, but that would not influence our results in any meaningful way.

5.4 Conclusion

While the exact mechanisms of motion prediction are still unknown, using EEG pattern classification we have managed to shine a light on some of its temporal aspects. We showed that classifiers can be used to decode motion prediction. The difference in time-to-peak latencies between apparent motion and no motion that we found is consistent with those of others (Jancke et al., 2004; Carlson et al., 2011), who have found evidence for motion prediction using EEG or

MEG. Our findings add to the body of evidence supporting motion extrapolation as mechanism for motion prediction.

6. References

- Berry II, M. J., Brivanlou, I. H., A., J. T., & Meister, M. (1999). Anticipation of moving stimuli by the retina. *letters to nature*.
- Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*.
- Brenner, E., & Smeets, J. B. (2000). Motion extrapolation is not responsible for the flash-lag effect. *Vision Research*.
- Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of Vision*.
- Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*.
- Cavanagh, P., & Anstis, S. (2013). The flash grab effect. *Vision Research*.
- Deng, S., Srinivasan, R., Lappas, T., & D'Zmura, M. (2010). EEG classification of imagined syllable rhythm using Hilbert spectrum methods. *IOP Science*.
- Eagleman, D. M. (2010). How does the timing of neural signals map onto the timing of perception.
- Fröhlich, F. (1929). Die Empfindungszeit. Ein Beitrag zur Lehre der Zeit-, Raum- und Bewegungswahrnehmung.
- Grootswager, T., Wardle, S., & Carlson, T. (2016). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging. *Journal of Cognitive Neuroscience*.
- Hogendoorn, H., & Verstraten, F. (2013). Decoding sensation and perception over time with EEG pattern cross-classification. *Perception*.
- Hogendoorn, H., Carlson, T. A., & Verstraten, F. A. (2008). Interpolation and extrapolation on the path of apparent motion. *Vision Research*.
- Hogendoorn, J., & Verstraten, F. (2013). Decoding the motion aftereffect in human visual cortex. *NeuroImage*.
- Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of Physiology*.
- Jancke, D., Frédéric, C., Naaman, S., & Grinvald, A. (2004). Imaging cortical correlates of illusion visual cortex. *Letters to nature*.
- Kenkel, F. (1913). Untersuchungen über den Zusammenhang zwischen Erscheinungsgröße und Erscheinungsbewegung bei einigen sogenannten optischen Täuschungen. *Z. Psychol.*
- King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in cognitive sciences*.

- King, J.-R., Gramfort, A., Schurger, A., Naccache, L., & Dehaene, S. (2014). Two distinct dynamic modes subtend the detection of unexpected sounds. *PLOS*.
- Kriegeskorte, N., Simmons, K. W., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience – the dangers of double dipping. *Nature Neuroscience*.
- Lewis. (1998). Naïve (Bayes) at Forty: The Independence Assumption in Information Retrieval. *Proceedings of ECML*.
- Mackay, D. M. (1958). Perceptual stability of a stroboscopically lit visual field containing self-luminous objects. *Nature*.
- Maunsell, J. H., & Gibson, J. R. (1992). Visual Response Latencies in Striate Cortex of the Macaque Monkey. *Journal of Neurophysiology*.
- Meyers, E., Freedman, D., Kreiman, G., Miller, E., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*.
- Misaki, M., Kim, Y., Bandettini, P., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage*.
- Müller, K.-R., Anderson, C. W., & Birch, G. E. (2003). Linear and nonlinear methods for brain-computing interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 165-169.
- Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*.
- Pelli. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*.
- Pohar, M., Blas, M., & Turk, S. (2004). Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study. *Metodoloski zvezki*, 143.
- Purushothaman, G., Patel, S. S., Bedell, H. E., & Ogmen, H. (1998). Moving ahead through differential visual latency. *Nature*.
- Ratcliff, R., Philiastides, M., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Science*.
- Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the Poor Assumptions of Naive Bayes Text Classifiers. *ICML*, 616-623.
- Roach, N. W., McGraw, P. V., & Johnston, A. (2011). Visual motion induces a forward prediction of spatial pattern. *Current Biology*.
- Schwarzkopf, D., & Rees, G. (2011). Pattern classification using functional magnetic resonance. *Wiley Interdisciplinary Reviews: Cognitive Science*, 568-579.

- Sharma, J., Dragoi, V., Tenenbaum, J. B., Miller, E. K., & Sur, M. (2003). V1 Neurons signal acquisition of an internal representation of stimulus location. *Science*.
- Stokes, M., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*.
- Wertheimer, M. (1912). Experimentelle studien über das sehen von bewegung. *JA Barth*.
- Whitney, D., & Murakami, I. (1998). Latency difference, not spatial extrapolation. *Nature America*.
- Yang, Y., & Liu, X. (1999). A re-examination of text categorization methods. *Proceedings of SIGIR*.
- Yoo, S.-S., Fairmeny, T., Chen, N.-K., Choo, S.-E., Panych, L. P., Park, H., . . . Jolesz, F. A. (2004). Braincomputer interface using fMRI: spatial navigation by thoughts. *Neuroreport*.
- Zhang, T., & Oles, F. (2001). Text categorization based on regularized linear classification methods. *Information retrieval*, 5-31.
- Zhang, Y., Meyers, E. M., Bichot, N. P., Serre, T., Poggio, T. a., & Desimone, R. (2011). Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*.