MASTER THESIS

Artificial Intelligence

Author

Aikaterina Tompoidi

Supervisors

dr. Chris Janssen

dr. Daniel Oberski

Can we predict susceptibility to audio signals in autonomous driving by using EEG and HMM?

Utrecht University

July 2018

Abstract

The aim of this study is to investigate whether susceptibility to audio stimuli can be predicted during autonomous driving. Previous studies have shown that the susceptibility to unexpected audio stimuli is intrinsically related to attentional resources being allocated during driving. Namely, the more attentiveness is required during driving, the less susceptible a driver is to unexpected audio stimuli. In our study, we trained Hidden Markov Models in attempt to find hidden state(s), which could be indicative of the susceptibility to potential audio signals played in autonomous cars. Retrieved transitional probabilities of hidden states showed that transitions between states are very unlikely and the driver tends to remain at the same state for some time. On the other hand, by considering only 100ms before stimulus onset of data, could not provide us significant information in regards to attentiveness or expected susceptibility to the audio stimuli, as the most frequent state's mean value was around zero.
Additionally, by comparing before and after stimulus onset states, no significant results could be retrieved. Finally, we discuss multiple reasons which might have contributed to inability of identifying the potential information in EEG signal using HMMs.

Keywords: Hidden Markov Models, EEG, autonomous driving, attention

Acknowledgements

To start with, I would like to express my gratitude to dr. Chris Janssen and dr. Daniel Oberski, my research supervisors, for their patient guidance and useful critiques of this research work. I would also like to thank especially dr. Chris Janssen for his advice and assistance in keeping my progress on schedule. Special thanks to dr. Daniel Oberski for his detailed explanations on HMMs.

My grateful thanks are also extended to drs. ing. Remo van der Heiden for giving his feedback on EEG processing.

Finally, I want to thank my husband Vasilis and my son Kanelos for their continuous support and understanding. This work would not be accomplished without their everyday help and encouragement.

Contents

Introduction

Levels of automation in cars

The idea of a vehicle being able to drive without human intervention has been described by John McCarthy back in 1970 when he first introduced an “automatic chauffeur”. Since then much has been accomplished towards realisation of autonomously driving cars. And, although there is still no fully autonomous car in massive production, the society of automobile engineers (SAE,2014) proposed to distinguish six levels of automation. These levels differ in how much control the human driver has of the dynamic driving tasks, compared to the autonomous systems. Specifically, levels 0-2 are the most commonly encountered on the roads in which there is either no assistance at all (level 0), either cruise control or lane centering is incorporated (level 1), or both are incorporated (level 2). The levels 3-5 which fall under the name *automated driving systems* are capable of either conditional autonomous driving such as driving without human intervention in specific road context (e.g., within a speed limit or specific road types; level 3), driving without human intervention in wider road conditions (level 4), or driving autonomously anywhere (level 5).

At the time of this study, the automation level in cars which are massively produced, can vary from level 1 to level 3. It is important to stress at this point that in level 3 automation, the vehicle can keep track of its environment through different types of sensors, but the driver remains essential due to the restrictions this level entails. Ergo, the driver is expected to have her attention focused on the road to take over the control of the vehicle whenever it is indicated by the autonomous system. Some form of communication from the car to the driver to alert the driver of such a transition of control is then needed.

One of the options to accomplish communication with the driver can be an audio signal. Audio alerts might be prefered over other types of communication, because of their “omni presence” compared to visual signals which can be easily overlooked. Furthermore, as the automation level gradually rises, drivers might tend to engage into other activities inside an autonomous car (de Winter et al., 2014), rather than constantly keep supervising the road conditions. The problem which arises then is, if the driver is not able to hear the audio signal then she will not act on it. This in turn, can cause safety-critical situation which can lead to an accident. It is therefore essential that we understand under what conditions drivers are more or less susceptible to audio alerts in (autonomous) driving conditions. Therefore, the question that we tried to answer is whether driver’s susceptibility to unexpected stimuli such as vehicle sounds can be predicted.

Attentional Measurements

Susceptibility to stimuli is intrinsically related to attention and essentially expresses the degree of ease for a subject to direct the concentration towards various types of stimuli. In the context of voluntary and involuntary allocation of attention (Theeuwes, 2010), two types of attentions have been defined which account to goal-directed attention and stimulus-driven attention (Corbetta et al., 2002). For example, during driving, the attention allocated by the driver to keep the speed limit can be characterised as goal-directed voluntary attention. The driver intentionally directs his attention towards the speed indicator to check the speed. This process is also known as *top-down* stimuli selection. On the other hand, mental response to a sound within the car is stimulus-directed attention and is involuntary. There is a different mechanism responsible for reacting to unexpected stimuli and directing the attention towards them. This mechanism is also known as *bottom-up* stimuli selection which can also interrupt the top-down processes. According to Corbetta et al.(2002) these two mechanisms can cooperate with each other and comprise the attentional system.

In an attempt to measure attention, various methods have been employed which can be divided into three categories: behavioral, subjective experience, and physiological metrics. In particular, the behavioural measures of attention require speeded response to stimuli without accuracy-speed trade-off (Johnson et al., 2004). To behavioral measurements belong such metrics as the reaction time to stimuli (e.g., Shulman et al., 1979; Sperling et al., 1980; Prinzmetal et al., 2005) and accuracy in a task. The subjective experience metrics such as Nasa TLX questionnaire (Hart & Staveland, 1988), assess various aspects of the physical and mental demands of a task or workload. Finally, physiological methods measure attention from neurophysiological point of view. Such metrics are for example eye tracking (e.g., Theeuwes, 1991;Poole et al., 2006; Tsai et al.,2012; Blair et al.,2009), pupil size dilation (e.g.,Hess et al.,1960; Kahneman, 1966; Hoeks et al.,1993; Partala et al., 2003), heart rate (e.g., Porges et al.,1969; Laumann et al., 2003), skin conductance (e.g., Frith et al.,1983), electroencephalography (EEG) (e.g., Anllo-Vento et al.,1998; Monastra et al.,1999; Lenartowicz, et al., 2014, van der Heiden et al., submitted), functional magnetic resonance imaging (fMRI) (e.g., O'Craven et al.,1997; Coull et al.,1998; Vuilleumier et al., 2001) and magnetoencephalography (MEG) (e.g., Downing et al., 2001; Shtyrov et al., 2003).

Not all of the previously discussed measurements can be used to measure human behaviour in real-time, in a way that can predict poor performance before it occurs. Namely, traditional behavioral measures are not suitable, as by the time the poor performance is detected, it might be too late. The subjective experience metrics are typically used in hind-sight and therefore, cannot be used in online predictions. Then, we are left with physiological measures. Of these, some are useful in car setting. For instance, MEG and fMRI require expensive technical setup which is not possible to be incorporated into a car system. For instance, MEG requires appropriate magnetic shielding which can prevent any other magnetic field (including Earth's

magnetic field) from intervening with brain's produced magnetic signals. On the other hand, a physiological measurement that can be incorporated into a portable device and be used in online predictions, is the EEG method. The EEG represents an unobtrusive method that has been used to measure susceptibility to sounds (Debener et. al, 2005; Wester et al., 2008; Bulthoff et al.,2016; van der Heiden et al., submitted) before and therefore can be regarded as an appropriate option.

EEG - a brain imaging technique

EEG represents a noninvasive method to measure amplified voltage changes in electromagnetic waves produced by the brain, by placing electrodes on the scalp. Some of the most prominent features that EEG exposes is its 3N (Klonowski, 2009) - non-stationary, nonlinear and noisy attributes. Namely, non-stationarity expresses the fact that EEG-signal changes its statistical characteristics over time. Nonlinearity, on the other hand, is intrinsically related to that human brain is a complex system comprised of complicated non-linear properties. Finally, EEG is often contaminated by noise, unrelated to cerebral activity. This noise can be caused by two types of factors, physiologic and extraphysiologic. Physiological artifacts are caused by the body and can be related to ocular, cardiac and other muscle activity. The extraphysiologic artifacts on the other hand, can occur due to equipment instability. For instance, electrodes when not applied well on the scalp can cause noise.

Even so, this measure has been extensively used in different scientific fields, due to its high temporal resolution and low cost. In the field of cognitive science, there are two approaches to use EEG data: ERP component and frequency bands. Both of these techniques have been invented in order to solve the problem that EEG represents a combination of conglomerative neural activities. This, in turn, makes it extremely difficult to attribute raw EEG signal to individual cognitive processes (Luck, 2015). And this is the main reason why raw EEG is hardly used in cognitive science.

A frequently used approach in EEG related studies involves frequency-based signal transformation which describes the brain's rhythmic activity. In the frequency-based approach, the signal is translated into a number of frequencies and a particular space of frequencies, i.e. frequency band, is examined separately. For example, the lowest frequency band, delta band lies around 0.5-4Hz, while the highest band, gamma ranges around 36-90 Hz, although usually due to filtering of EEG signal it might reach not higher than 50 Hz (Michel,1992). This method has been used in for example the diagnosis of abnormal cerebral activity (Tatum, 2014) and also in attempt to describe brain functions (Klimesch,1998).

More directly related to our aim, the frequency-based method has been used by Simon et al. (2012) to analyse alpha band power and gamma band power in a driving experiment. They

showed that alpha spindles as well as alpha band power are positively correlated with a secondary auditory task. The increased occurrence of alpha spindles was assumed to be positively correlated to attentional shifts which were required during the experiment. In particular, these attentional shifts were interpreted as the process of inhibition of visual processing mechanism and increased processing of auditory stimuli. This study also demonstrated that different conditions generated different levels of alpha spindles with auditory stimuli causing higher activity of alpha spindles compared to visuomotor stimuli. Theta (4-8Hz) and alpha (8-13 Hz) oscillations have also been recorded by Yu-Kai Wang et al. (2018) in a driving experiment where the participants had to do mathematical computations as an intervention task.

The frequency-based methodology although useful, might impose some restrictions during data analysis. Namely, when examining only particular bands, the results are restricted only to that bands and all the rest frequencies are usually either discarded or examined apart. Consequently, if during frequency-based analysis only some of the available frequencies are used, there is a probability to miss valuable information.

Another widely used technique is Event-Related Potential procedure, or ERP. As its name implies, ERP is coupled to a particular event (or stimulus) presented during an experiment. In particular, in ERP procedure, after occurence of an event (e.g. playing of a sound) EEG activity which follows this event, is measured. Typically, the event is repeated in multiple trials. ERP is then used to attenuate electromagnetic signals which are not related to a target stimulus or a psychological event during an experiment. In particular, to find related to a stimulus or an event electromagnetic activity, the signal at time of the stimulus (or the event) is averaged over multiple trials and subjects. In this way, the irrelevant information in EEG can be cancelled out (Luck, 2015).

In the context of attentional shifts, one way to manipulate participant's attention in an experiment is to use an oddball paradigm. While performing a task, the subjects are being exposed to visual or auditory stimuli, which are either irrelevant to the primary task, or unexpected. These oddball stimuli can trigger cognitive processes. For instance, one such cognitive process could be the shift of goal-driven attention during primary task such as driving to stimulus-driven attention caused by unexpected auditory signals.(Wester et al., 2008; Van der Heiden et al, submitted). As these cognitive processes represent electromagnetic activity, they can be tracked by EEG and consequently be measured in the ERP procedure.

There are many types of ERP components which have been identified over time. (Luck, 2015). In the context of oddball paradigm, concrete ERP components have been discovered. These ERP components can be divided based on three-stage model which describes the attentional changes caused by unexpected stimuli (Correa-Jaraba et al., 2016). Namely, the first stage represents pre-attentive change detection, the second level is the involuntary orientation of attention. While the third level accounts to voluntary reorientation of attention. Given these stages, ERP component which occurs in the first stage (peaks around 150–250 ms) is called

MMN and it is a negative wave which is proposed to reflect a detection mechanism for unexpected changes. In the second stage we have a positive ERP component which peaks around 300 after stimulus onset and is called P3a, fP3 or novelty P3 component. This mechanism represents the involuntary attentional shift towards the unexpected stimulus. Finally, the ERP component assigned to the third stage is called RON (peaks around 400–600 ms after stimulus onset) and this is also a negative wave which accounts to attentional reorientation towards primary task (Escera et al., 2001).

Although widely used, this methods imposes a restriction that is, it can only be used in relevance to some external stimuli or cognitive event. It therefore cannot evaluate more extended or constant cognitive states. Such extended, constant cognitive state assessment is valuable however in the context of autonomous driving where we might want to anticipate driver's fluctuations in attention so as to alert them at the right moment.

Bottom-up theory establishment

Most of the studies referred so far used top-down approach: based on a theory, a specific prediction is made regarding what to measure, and this theory is tested using hypothesis testing procedure. This procedure is probably the most optimal when fine-grained theory is developed. For example, in the ERP approach, theory predicts the presence or absence of specific signals at specific time intervals in response to an event. This can then be tested. In other words, the value of the top-down approach in what concerns ERP related experiments, is that it can often result into refined theory. On the other hand, as only the specific part of the continuous signal is used, it might discard information less relative for theory establishment, which can be proved insightful, though.

The standard top-down methods are less useful for a dynamic driving scenario, where attentional state is not always tied to a specific *event* but rather to a continuous phenomenon. Moreover, the state might fluctuate over time exactly due to various environmental stimuli, such as sounds. Therefore, we suspect that in attempt to predict the attentional fluctuations of a driver in a real driving context, the estimations should be based on continuous driving rather than tied to an event. A general state detection method can be achieved in theory using bottom-up, or data-driven methods. These do not start necessarily with theoretical assumptions, but instead start with the raw data .

Recently, the bottom-up methodology has been used in attempt to refine theoretical frameworks. For example, the full potential of the bottom-up approach was revealed in studies by Anderson et al. (2014, 2015) where to analyse EEG recordings during memory retrieval experiment, they used the bottom-up approach. To discover different stages that occur during memory retrieval process in the brain, they applied multivariate pattern analysis in combination

with Hidden Semi Markov Models (HSMM). These cognitive stages were represented by short sinusoidal peaks in the continuous EEG signal. In particular, the number as well as the duration of those peaks were estimated by using HSMMs. After identifying the peaks which represented neural signatures of stages as well as their durations, the authors were able to further refine theoretical background of memory retrieval process. In particular, they applied the acquired knowledge to enrich modules in the ACT-R (Adaptive Control of Thought-Rational) cognitive framework with information concerning not only time and the number of cognitive processes, but also the actual function of each cognitive process.

Furthermore, concerning the oddball sound paradigm during driving, a bottom-up approach has been used in combination with ERPs in a driving scenario by Bulthoff et al. (2016) where they used mass univariate analysis to reveal changes in ERP components before and after audio stimuli onset. Next, the results were found to be in agreement with a three stage destruction framework within which those are further discussed.

All the previous studies showcased the power of bottom-up approaches in EEG studies. But, they also raise a question whether such methods could be successfully applied on raw EEG data in a dynamic experimental setup when no theoretical framework exists to describe cognitive processes during autonomous driving.

Next, I would like to introduce the backbone study this work will use to investigate the bottom-up approach on. The understanding of the experimental setup and the main findings will help us in understanding the data. Once we know what type of data we possess, it is more easy to decide on the actual algorithm.

Study of susceptibility to audio signals during autonomous driving

In this study, I used data which were obtained in the study performed by Van der Heiden et al. (submitted), in which a three-stimulus-oddball paradigm was applied in a driving set up. Subjects had to perform two types of driving in a driving simulator, namely manual driving, or autonomous driving. This was also compared with a baseline, stationary condition. In parallel, the oddball task was conducted. For this task, three types of sounds were played: standard sound (a regular tone of 1000Hz), deviant sounds (a slightly higher tone of 1100 Hz) and novel sound (unique environmental sounds). These tones were randomly played with 80% being standard tones, while deviant and novel sounds were equally represented 10% of the time. The interval between two consecutive sounds was 2.34 seconds.

Two groups of participants were used. One group (half of the participants) was the “active group”, that had to press a button when they heard a deviant sounds. The other half was the “passive group” and did not need to press a button. The idea behind this manipulation was that

for the active group the sounds were more relevant, and thus they might pay more attention to them and be more susceptible to novel sounds (Wester et al., 2008; Kenemans, 2015). Additionally, the requested response to the deviant sounds has been shown to increase the brain response to unexpected novel sounds during the involuntary orientation of attention stage.(Wester et al., 2008)

The main findings of the study showed that the susceptibility to different tones depends on the task's attentional requirements. In particular, the amplitudes of ERP response on the difference wave between novel and standard tones were estimated by using FCz electrode for time interval of 325 - 375 ms after stimulus onset. The estimated ERPs showed significant differences across driving conditions. Namely, it has been shown that the amplitude of ERP component during manual driving is lower compared to the one acquired during autonomous driving and even more reduced compared to stationary condition. Therefore, it has been shown that during autonomous driving attentional requirements are less and therefore the ERP response is higher to unexpected stimulus, compared to active driving.

The finding of study of susceptibility to audio signals revealed an experimental setup which is also suitable for the current study, as within this oddball paradigm fluctuations of susceptibility to audio signals can be measured. However, it also raises a question whether such fluctuations can be predicted *before* the stimulus onset using a bottom-up approach. In other words, in this study we would like to investigate whether the cognitive load during manual or autonomous driving can be identified not only as a comparison of ERP amplitudes but also during continuous driving act.

HMMs in EEG analysis

In order to be able to answer the previous question we need to select an algorithm which is appropriate for the type of data we have, in combination with the question we are trying to answer. In particular, we are interested in exploring feasibility of raw EEG in predicting the susceptibility to audio stimuli during driving. An EEG signal represents temporal data and at the same time, we do not possess explicit information about the subject's susceptibility in each data point. This observation leads us to favour so called unsupervised machine learning (ML) algorithms which can handle temporal data. Unsupervised ML algorithms are algorithms which are able to identify common features in order to cluster data into groups (i.e. clustering algorithms) without requiring annotated data. The main restriction which is imposed then is that such algorithms usually cannot handle the temporal dependencies in the data.

From the bottom-up studies we have seen that HSMMs performed well with unlabeled EEG data (Anderson et al., 2016), and this could be a good option in our case. But, as we are not

interested in identifying the duration of different levels of susceptibility but merely in presence of such, we have decided to use a simpler version, which is HMMs. What's more, HMMs have been applied in a study by Solhjoo et al. (2005), where they could successfully classify the data from imagery movement task. Next, I would like to introduce the reader to HMMs.

Discrete Hidden Markov Models

HMMs represent a temporal statistical model of sequential data. Statistical models imply that data can be parameterized by some random process, such as a Gaussian processes, which in turn can be well approximated by a model (Rabiner,1989). The term temporal, on the other hand implies that the data is treated as a sequence where the order of data observations in time is important.

In particular, HMMs try to statistically describe a latent variable which produces outputs that can be observed. For example, in speech processing, the sound of speech is considered a sequence of observations. If we assume that a language is a stochastic process which can be modelled statistically, then hidden states can represent parts of speech such as phonemes, syllables or words.In the case of EEG signal modelling, the latent variable could represent a cognitive processes which produce the EEG signal.

Additionally, EEG is a signal data which means that the data points are temporal sequences and HMM can handle the dependencies that exist in temporal data. Namely, the first order HMMs make the assumption that the probability *P* of a latent state *q* which produced an observation O at time *t* can be predicted only based on the state at time *t-1* (Markov assumption). This dependency of latent states *q* is captured in equation 1:

$$P(q_t \mid q_1 ... q_{t-1}) = P(q_t \mid q_{t-1}) \quad (1)$$

Further, the assumption that HMMs make is that the current observation o_t can be emitted only by current state q_t and does not depend on any other state or other observations (Output independence). This is captured in equation 2:

$$P (o_t \mid q_1 ... q_t ..., q_T, o_1 \ ... \ o_t \ ..., o_T) = P (o_t \mid q_t) \quad (2)$$

In our case, there is a discrete latent variable Q which may represent the susceptibility to an audio signal. This variable can take on multiple values, i.e., a number of states. And, in perfect case each state would represent a different level of susceptibility to a sound during driving. A simple representation of this process with only two states of high and low susceptibility is depicted in Figure 1 below.

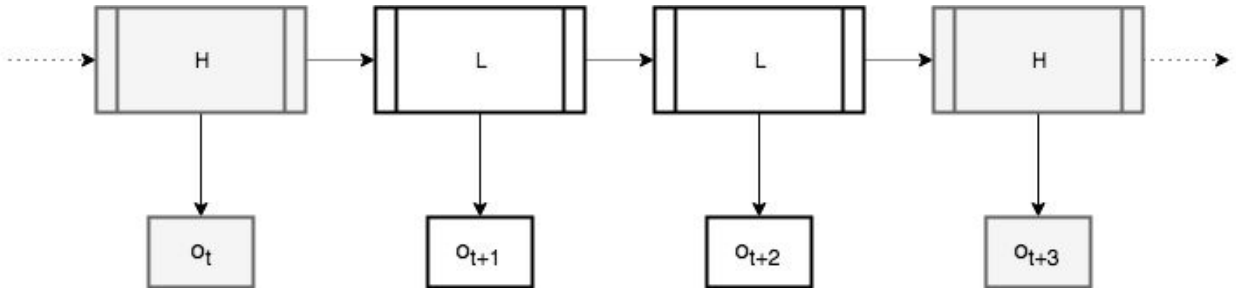

Figure 1. An ideal case of a HMM process with two states of susceptibility to audio signal producing observations o at different time points.. The model transitions from a “high” (H) susceptibility state to the “low” (L) state and stays there until t+3 time point, where it transitions back to the “high” state. Absence of an arrow represents an assumption of conditional independence.

From the previous, we see that generally there are two distributions, described by equation 1 and equation 2 above, which we try to approximate using a HMM. Because we use Gaussian distribution the equation 2 is then equal to estimating $N(\boldsymbol{\mu}_q,\boldsymbol{\sigma}_q)$, while equation 1 then represents a multinomial distribution with parameters $P_{qq'}$. In other words, multinomial distribution is used to describe the probabilities to transition from one hidden state to another(or stay at the same state) while Gaussian distribution is used to describe each state’s observations.

These two distributions are estimated during learning problem (Rabiner, 1989). Namely, given just the observed data as in our case, the learning problem is to estimate the model parameters $\lambda = (\pi, A, B)$ such that,

$$argmax\, P(O \mid \lambda) \qquad (3)$$

where π is the initial state distribution, *A* is the state transition probability distribution, and *B* is the observation symbol probability distribution.

Next, given the model parameters and the observed sequences, it is possible to solve decoding problem (Rabiner, 1989), during which the sequential data *O* is translated into the most likely sequence of the states, *X.*

$$argmax\, P(X \mid O,\ \lambda) \qquad (4)$$

In this study, we first used the forward-backward algorithm (Baum et al., 1970; Rabiner,1989) to solve the learning problem, and get the states transitional probabilities as well as the

estimations such as mean and standard deviation.More information about this process can be found in Appendix.
Finally, we performed the decoding of the sequences by using Viterbi algorithm (Forney, 1973) to obtain the sequences of states which correspond to the sequences of observations. Due to output independence assumption (equation 2), each time point in our data is assigned a corresponding state in the Viterbi's output.

Research Question

The question that is investigated here is whether it is possible to identify latent states of the driver's attention using Hidden Markov Models, which can help us in predicting driver's susceptibility to unexpected auditory stimuli.
By identifying differences in hidden states across driving conditions, we would be able to predict driver's susceptibility to potential auditory stimuli.

Methods

Data preprocessing

We used data from the experiment by Van der Heiden et al. (Submitted). This data contained EEG recordings of 18 participants. Out of the 64 electrodes that were recorded in the study, the *FCz* electrode was selected for analysis, as Van der Heiden et al. (submitted) used this electrode for computation of the fP3 component. According to Friedman et al. (2001), frontal and central scalp area is the area where reaction to novelty sounds is earlier and therefore is usually preferred over other scalp sites. As FCz electrode is located in that area, we preferred to use it out of 64 available electrodes in data.

The data were first offline preprocessed using MNE v0.16.1 python package. To attenuate artifacts from EEG signal, the data were subject to filtering, re-reference as well as EOG rejection and baseline correction. In particular, a 50 Hz notch filter was applied to remove noise from the mains. Next, a 0.16-0.3 Hz band-pass filter was used to remove heart potentials. Data were then referenced to the average of two mastoid electrodes. Finally, baseline correction was applied for the interval 100 ms before the stimulus onset to remove drifts and shifts in the data that is caused by skin hydration and static charges in electrodes.

The experimenters noticed a delay of 50ms when logging the stimulus onset. To overcome this problem, when extracting events 50 ms were added to the time which corresponded to the stimulus onset. The data were epoched based on events found in recorded EEG signal (i.e., registration of the onset of a sound). Each epoch started at 100 ms before stimulus onset and stopped 2 seconds after it. Epochs contaminated with blinks were automatically removed in MNE package by using EOG electrodes to locate and reject this type of artifacts. As a last step, the data were down sampled from 2048 Hz to 200 Hz. The down sampling was performed by attenuating every N^{th} observation. Consequently, the time step in our data is equal to 5ms.

Training and Model selection

In order to identify the sequences of hidden states in the decoding problem, it is necessary to define the model. A HMM is defined by its parameters $\lambda = (A, B, \pi)$, i.e., the transitional probabilities *A*, the emission probabilities *B* and the initial state distribution. Therefore, we needed to start from the learning problem, which solution is estimated during the training of HMM. As an input, we used our epoched data and a number of states. The training algorithm which is the most frequently used is the forward-backward algorithm or as it is alternatively known, Baum-Welch algorithm (Baum, 1972). This algorithm represents a special version of Expectation Maximization algorithm (Dempster et al., 1977) which iteratively during training improves the estimated probabilities *A* and *B* until it reaches a local optimum.

Sometimes, the number of states in HMM can be implied from the theoretical background. In our case though the number of states cannot be easily predefined. Therefore, we trained and tested the models for maximum 8 states.(Anderson et al., 2015)

To select the best model, LOOCV (leave one out cross validation) was performed during which the model was trained leaving one of the sequences out. When training the model, we are interested in its performance on unseen data, therefore, by leaving one sequence out and training on the rest, we can estimate how well the model generalizes. The measure used to compare different models is log-likelihood which is the log probability of test data being generated by the model λ (where λ are estimated using training data) .

$$log\,(\mathrm{P}(\mathrm{O}_{\mathrm{test}}|\,\lambda)) \qquad (5)$$

Usually in ML the model with the highest log-likelihood is preferred, because then it is expected to generalize on unseen data. In case of HMM, it is not that straightforward. As we deal with a generative model, the log-likelihood will tend to increase as we keep increasing the number of states. If we take into consideration only the test log-likelihood, we can end up with the number of states equal to the number of different values of EEG signal. As the number of states rises, the model's log-likelihood rises as well, but if we continue increasing the number of states at

one point the model will get extremely tight to the training data and will not generalize on the new data sequences. Consequently, to select the optimal number of states, we take into consideration the mean improvement of the log-likelihood as the number of states increases. This heuristics was also applied by Anderson et al.(2016) To compute the mean improvement over the number of states, we first compute the mean log-likelihood for each model.

$$Mean\ log\text{-}likelihood = \sum_{i=1}^{N} log(P(O_i | \lambda)) / N \quad (6)$$

Then, the mean improvement is estimated by subtracting the score of model with *j+1* states from the score of model with *j* states, i.e.,

$$Mean\ log\text{-}likelihood_{j+1} - Mean\ log\text{-}likelihood_j \quad (7)$$

As can be seen at Figure 2, the mean log-likelihood improvement from state 2 to state 3 is the highest, therefore we propose to start from model with 3 state.

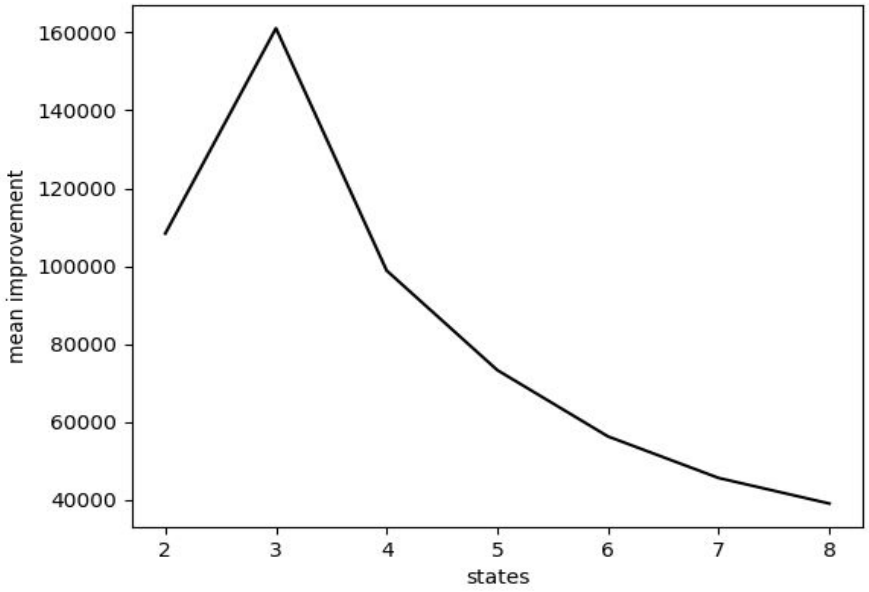

Figure 2: Mean scores improvement over number of states.The highest improvement is observed in 3-state model

Results

3-state model analysis

After training the 3-state HMM, we obtained the transitional probability matrix A as well as the parameters μ,σ of each of three states. From the state transition probability matrix (Table 1A) we deduced that each state is more likely to prolong, or in other words the probability for a state to transition to the same state is 0.97 (see values on main diagonal on Table 1A).Next, the probability of the model to transition from state 0 to state 1 or to state 2 is the same and equal to 0.2. On the other hand, for the model to transition from state 1 to state 0 as well as from state 2 to state 0 is equal to 0.3. Finally, we did not observe the transitions from state 1 to state 2 as well as from state 2 to state 1.

Next, given the values in Table 1B we see that state 0 represents values in EEG signal which are around zero(*M*=-0.27μV and *SD*=7μV), state 1 represents high negative values (M=-26.7μV and SD = 17μV) and state 2 represents high positive values (M=-25.8μV and SD=18μV).

Following that we were interested in predicting the susceptibility to auditory stimuli, we used 100 ms (which accounted to 20 data points in downsampled data) before stimulus onset to retrieve states probabilities. We did that by applying Viterbi algorithm (Forney, 1973) on each data sequence containing only 100ms before each stimulus onset. After obtaining the states, we were able to estimate the probability of each state to occur in each of three driving conditions.

As we can see,(Table 1C) state 0 is the most frequent state in each driving condition with probability to occur being around 0.76. On the other hand, states 1 and 2 are significantly less frequent in every driving condition, with probabilities around 0.11 and 0.13 accordingly.

states	0	1	2
0	0.97	0.02	0.02
1	0.03	0.97	0.00
2	0.03	0.00	0.97

A

states	Mean	SD
0	-0.3	7
1	-26.7	17
2	25.8	18

B

states	driving	autonomous	stationary
0	0.76	0.76	0.74
1	0.11	0.11	0.12
2	0.13	0.13	0.14

C

Table 1 A. Transitional probabilities from states to states show that there is a strong tendency for each state to prolong. **B** Mean and SD for each state in μV. **C**. Probabilities of each state to occur in each driving condition.

In general, we see that the prevailing state for all three conditions to be state 0. This is expected because according to our knowledge about the FCz position, the area is responsive to

unexpected stimuli and as we are using data which corresponds to 100 ms before stimuli onset, electromagnetic response should be minimized. Additionally, minor differences between the states and conditions are observed. These differences, although insignificant, raise the question whether by grouping the participants into groups defined by a fP3 component over all trials, can provide any significant observations. That is, are there differences in state identification and state transition when we separate participants with a relatively large fP3 component from those with a lower fP3 component?

Results per driving condition and ERP component

We obtained the fP3 components per participant per condition, the values and groups of which can be found in Appendix (Table 7). Next, we assigned the participants whose fP3 component was higher than the mean fP3 into the high susceptibility group, while the rest were assigned to low susceptibility group. Namely, the mean fP3 value in driving condition was 6μV which divided the participants into low driving and high driving groups, with 11 (7 passive, 4 active) and 7 (2 passive, 5 active) participants accordingly. In the autonomous condition, the mean fP3 value was 8μV which divided 12(7 passive, 3 active) of participants into low autonomous group and 6 (1 passive, 5 active) into high autonomous group. Finally, in the stationary condition, mean fP3 value was equal to 10μV which distributed the participants evenly with 9 in the high group (7 passive, 2 active) and 9 in the low group (2 passive, 7 active).

Table 2 shows the probabilities of each state based on the high and low susceptibility groups in each driving condition. We do not see significant differences in probabilities inside each of the driving conditions. But, if we are to compare the probabilities across conditions, we observe that in the driving-high group, the state 0 is approximately 5% more likely to occur than in the stationary high condition, and 2% more likely than in autonomous high condition. At the same time, the probability of state 2 (M=25.8μV and SD=18μV) is higher in the stationary-high group compared to driving-high by 2% and as likely as in autonomous high group.

States	Driving		Autonomous		Stationary	
	High	Low	High	Low	High	Low
0	0.78	0.76	0.73	0.76	0.73	0.74
1	0.10	0.11	0.12	0.11	0.13	0.12
2	0.12	0.13	0.14	0.13	0.14	0.14

Table 2. The probabilities per condition for high and low susceptible groups based on van der Heiden et al. analysis.

Discussion fP3 component groups

Although these differences are subtle, we hypothesize that this can be consistent with conclusions made by van der Heiden et al., (submitted), namely that the susceptibility to stimuli is lower when participants were driving in manual mode compared to autonomous and stationary conditions. Specifically, the probability of state 2 is lower in the driving high condition compared to autonomous high and even more compared to stationary high condition. On the other hand, this is not exactly the case for low susceptibility groups. Namely, we do not observe any difference in likelihood of state 2 between driving low and autonomous low groups.

Further, to investigate whether the slight differences among fP3 component groups persist and can be seen in regards to different sounds played afterwards, we computed the probabilities of each state to occur in each driving condition per type of sound. In other words, the question is, whether the susceptibility to particular sound types can be predicted before the actual stimulus onset. Such that, we hypothesized that attention allocated already to the task actually affects the susceptibility to external stimuli.

Analysis per sound type and driving condition

Table 3 Left, depicts the probabilities of each of three states to occur before each of the stimulus under different driving conditions. By comparing the probabilities across different conditions we can see that the highest difference is observed in novel sounds. In particular, state 0 in manual driving is prevailing with 0.80 against 0.73 in stationary in the high group and 0.76 in the autonomous group. Overall, the results show that in active driving than in any other, the EEG signal remains most of the time settled in state 0.

In attempt to estimate the actual susceptibility before the stimulus onset, it is important to compute the probabilities of each state in different conditions after the stimulus.
The time interval between 300 ms and 400 ms was used to estimate the sequences of states per condition, and the exact results can be observed in Table 3, Right. The reason for choosing the particular time interval (300-400 ms) after stimulus onset is because around that time the fP3 component increases demonstrating cognitive response to the stimulus (Van der Heiden et al., submitted).

On the heatmap (Table 3, Right) we see that the probabilities of states 1 and 2 are now higher. For instance, we see that in the autonomous high group the probability for state 2 to occur after deviant sound is equal to 0.39 while the corresponding probability for before stimulus onset is

equal to 0.15. We suspect that this rise of probabilities corresponds to cortical response to the unexpected stimuli. In order to investigate further these results, and see whether more concrete type of relationship can be identified between the states before and after stimulus, we ran a correlation test.

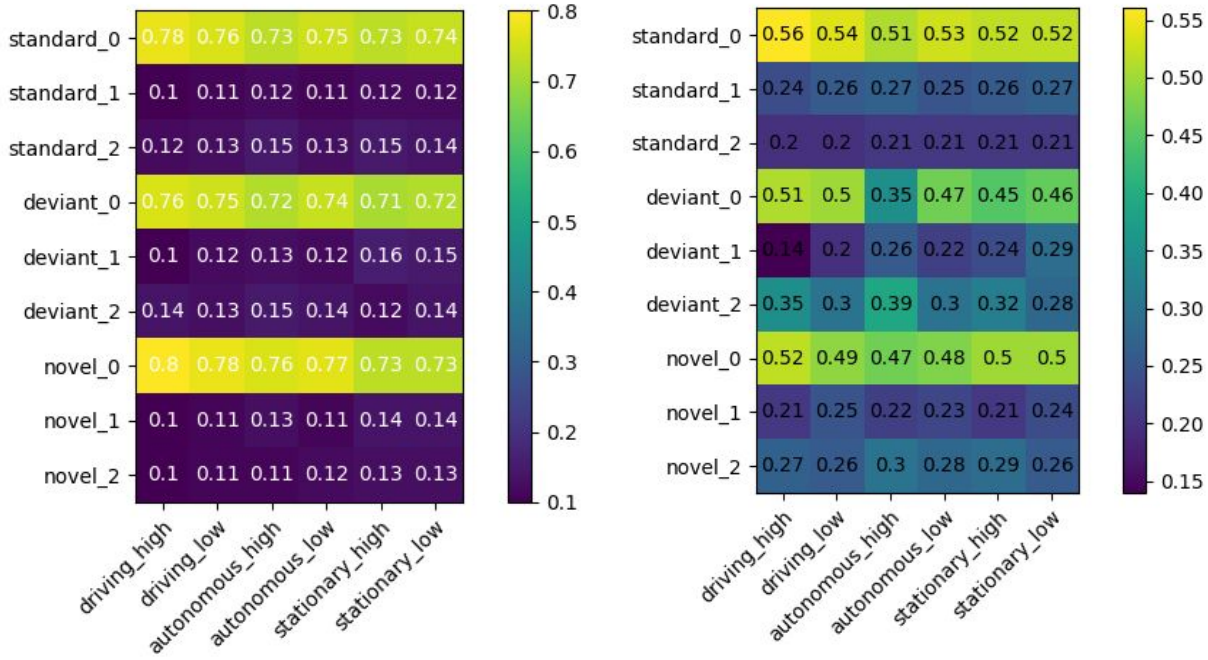

Table 3. Left: The probabilities per condition per stimuli to occur 100 ms before stimulus onset. **Right**: The probabilities per condition per stimuli to occur between 300-400 ms after stimulus onset.

Bivariate correlation test

We ran a Pearson correlation test to see whether there is any relationship between each state's probability before the stimulus and the same state's probability after the stimulus. For example, if we see that the probabilities before the stimulus for state 2 are positively correlated to probabilities of the same state after, then this would help us in predicting the susceptibility. Consequently, for each variable in the bivariate analysis we would have 18 observations (number of sound types x number of driving groups).

Then, given alpha=0.05 and df=16 for a state before stimulus to be correlated with any other state after (including itself), the correlation coefficient r should be higher than critical value 0.468 (Appendix, Table 2). The values of metric r are depicted per test in the table 4, below.

states	0		1		2	
	r	p-value	r	p-value	r	p-value
0	**0.486**	0.04	-0.36	0.138	-0.23	0.37
1	-0.36	0.14	0.33	0.18	0.31	0.20
2	-0.23	0.36	0.31	0.20	0.02	0.93

Table 4. Pearson's correlation test results.

Bivariate correlation results

The results from Pearson's correlation test are presented in the Table 4 below. We see that only the state 0 is significantly correlated with correlation coefficient to be equal to 0.486 and p-value=0.04. The rest of the states do not expose significant correlation.

Discussion 3-state model analysis

We modelled EEG data using 3-state HMM to investigate whether there was any consistent pattern of states' probabilities for different driving conditions. We used 100 ms before stimulus onset to compute the probabilities for each of the three states to occur. The results showed that the most likely state is state 0 before the stimulus onset which fluctuates around -0.38 μV .

The retrieved probabilities were then compared to probabilities of the same states to occur after stimulus onset in time interval from 300 ms to 400 ms. We saw a rise in probability for states (1 and 2) which represent more higher (negative and positive) values. We assume that the rise of likelihood of hidden state 1 (M=-26.7μV and SD = 17μV) can be related to voluntary attentional shift back to the primary task and corresponds to negative wave. While the rise of state 2 likelihood represents the involuntary orientation of attention which is represented by P3a ERP component. And therefore these findings can be regarded consistent with theoretical knowledge about the post stimuli responses to unexpected stimuli in the brain (Van der Heiden et al., submitted;Wester et al.,2008;Correa-Jaraba et al.,2016).

Finally, we investigated whether there was a correlation between the states before and after stimulus onset. We found that state 0 before stimulus onset is positively correlated to itself after stimulus onset. While no other state exposed significant correlation and given the fact that values of state 0 are around zero, it makes it hard to interpret the observed correlation.

The previous observation, in combination with that we could not observe any significant difference in probabilities across driving conditions and susceptibility groups, could be caused

by the model. In particular, we suspect that 3-state model is too general for the fine-grained analysis of our data. To overcome possible problem of too general states, we decide to use a model with more states. An arbitrary chosen model for this reason is the model with 5 states.

5-state model analysis

To try to overcome the problem of states representing wide range of values in HMM, we had to use a model with a larger number of states. As there is no particular reason to prefer either of the higher ranked models, I picked randomly a model represented by 5 states.

In Table 5, the results of analysis performed on 5-state model are reported. The main observations concerning the 5-state model is that the main state before stimulus onset is the state which accounts to values whose mean value is close to zero, and probabilities ranging around 45%. Then, we have a pair of states which are almost equally likely to occur, these are states 1 (M=13μV, SD=5μV) and 2 (M= -12μV, SD=5μV) with probabilities ranging around 25%. Finally, there are two states which account to outlier values, namely state 3 (M=-38μV, SD=18μV) and state 4(M=38μV, SD=21μV), these states are rarely observed in the data and range around 0.05%

states	mean	SD
0	0.1	4
1	13.1	5
2	-12.0	5
3	-37.8	18
4	38.3	21

state	driving	autonomous	stationary
0	0.46	0.47	0.45
1	0.24	0.24	0.25
2	0.25	0.24	0.25
3	0.02	0.02	0.03
4	0.03	0.03	0.03

states	0	1	2	3	4
0	0.90	0.05	0.05	0.00	0.00
1	0.06	0.92	0.00	0.00	0.02
2	0.06	0.00	0.92	0.02	0.00
3	0.00	0.00	0.04	0.96	0.00
4	0.00	0.04	0.00	0.00	0.96

Table 5. Left: mean and standard deviation of each state in 5-state model. **Right:** The probabilities of each state to occur per driving condition. The state 3 and 4 represent outliers which occur rarely. States 1 and 2 are positive and negative clusters of middle values in EEG signal. All the values are in μV. **Down:** transitional probabilities matrix shows that although ergodic Hidden Markov model some of the states are not connected.

Next, I used susceptibility groups obtained during the 3-model analysis, to investigate the probabilities of each state before and after stimuli onset.

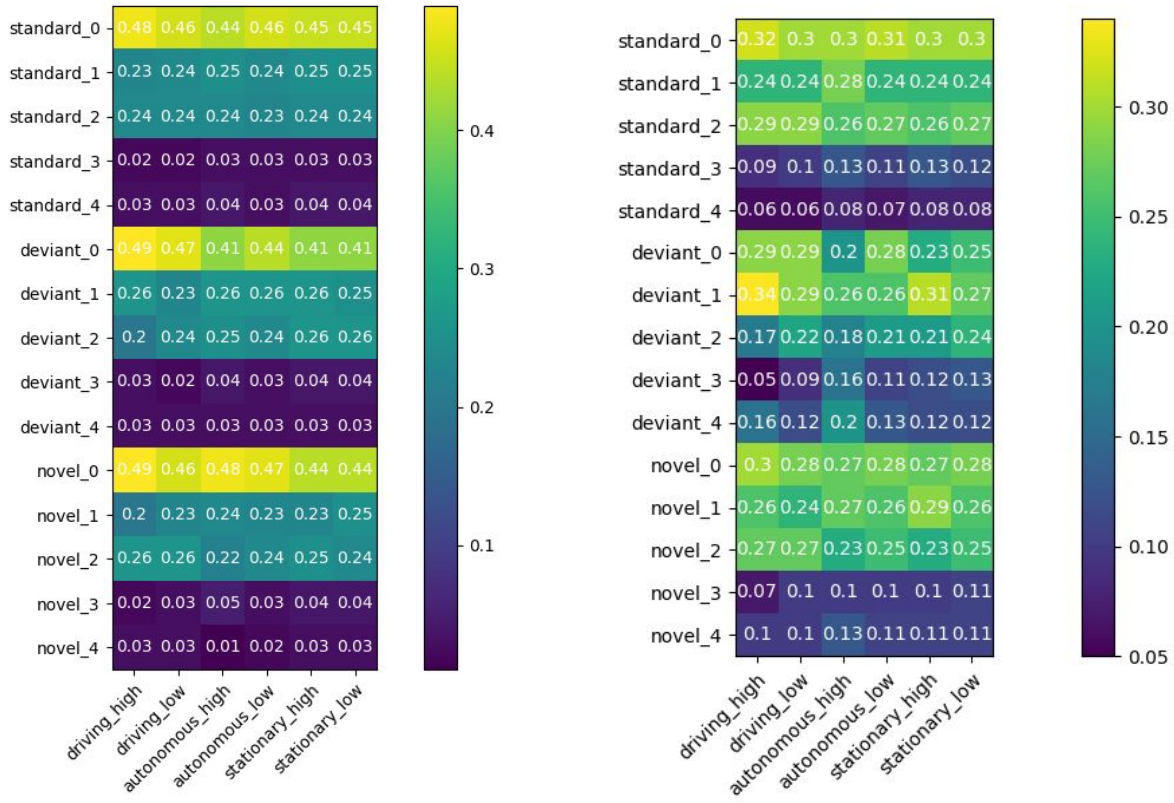

Table 6 Left. The states probabilities in 5 states model before the stimulus onset. **Right**. The states probabilities after stimulus onset.

In Table 6 Left, we see the probabilities of each state to occur before stimulus onset. Here, again we observe the same pattern as before. Most of the signal is settled around zero while the states 1 and 2 are equally distributed across driving groups and sounds. The same applies on states with very high and very low values (states 3 and 4).

In Table 6 Right we see the probabilities of each state to occur after stimulus onset from 300 ms to 400 ms. Across all conditions, we see that state 0 (M=0.08μV, SD=4μV) has significantly reduced after stimulus onset. This can be again regarded as a result of response to stimulus. At the same time, we see that the rarely occurred states before stimuli, namely state 3 and state 4 are now at least doubled across all conditions. We also see that for deviant sound the probability of state 4 is the highest. This final observation provides us with the indication that

these two states could account to cognitive response to stimuli and more likely state 4 which ranges around high positive values. Here, again state 3 probably represents the cognitive phenomena which are measured by RON ERP component, while state 4 can represent the cortical reaction measured by fP3 component.

As we do not observe any concrete pattern directly in Table 6 Right, next question would be if there is any correlation between the states before and after stimulus onset.

Bivariate correlation test

To better investigate whether there is any significant correlation between the probabilities before and after stimulus onset, we again ran the bivariate two-tail correlation test.

Given alpha=0.05 and df=16 for a state before stimulus to be correlated with any other state after (including itself), the correlation coefficient r should be higher than 0.468. The values of metric r are depicted per test in table 7.

states	**0**		**1**		**2**		**3**		**4**	
	r	**p-value**	**r**	**p-value**	**r**	**p-value**	**r**	**p-value**	**r**	**p-value**
0	**0.70**	0.001	-0.03	0.92	0.27	0.27	-0.83	2.22	-0.28	0.26
1	-0.03	0.92	0.28	0.25	-0.54	0.021	**0.48**	0.041	0.39	0.11
2	0.27	0.27	**-0.54**	0.02	0.28	0.25	0.43	0.074	-0.11	0.65
3	**-0.83**	2.22	0.48	0.04	0.43	0.07	0.40	0.10	**0.56**	0.02
4	-0.28	0.26	0.39	0.11	-0.11	0.65	0.56	0.02	-0.31	0.20

Table 7. Pearson correlation for 5-state model.

The results from the correlation test show us that state 0 (r=0.70) is significantly correlated to itself, while no other state before stimulus is correlated to itself. But, contrary to the 3-state model correlations, here we can observe some more significant correlations. For example, we see state 0 being negatively correlated with state 3 (r=-0.827), meaning that when the probability of state 0 increases before stimulus, the probability of state 3 linearly decreases and the other way around. For this finding to be interesting, this relationship should be present only in before-after correlation test. But as further investigations showed, this correlation is present if we run the correlation test for these two states in before stimulus(r=-0.48) and after stimulus

(r=-0.52). In other words, the state 0 and 3 are correlated before stimulus onset, as well as after stimulus onset. Consequently, this does not represent an interesting finding.

Another correlation (r=0.48) observed, is between the states 1 (*M*=13μV and *SD*=5μV) and 3 (M=-38μV and SD=18μV). In this case the frequencies are positively correlated. What's more, these states are not correlated before stimulus as well as after stimulus.

Finally, the frequencies of states 3(*M*=-38μV and *SD*=18μV) and 4(*M*=38μV and *SD*=21μV) are positively correlated (r=0.56) while as complementary tests showed no correlation is observed between these states in before stimulus (r=-0.42,p-value=-0.07) as well as after stimulus (r=0.12,p-value=0.63) conditions.

Given these observations, it appears to be not that straightforward to predict the susceptibility to stimuli before the actual stimulus onset, although we can pinpoint more fine-grained details of our data by using 5-state model. The various reasons for that will be discussed in general discussion section.

Fp1 electrode analysis

Driven by the hypothesis that FCz electrode might not be the best candidate for estimation of attentional shifts, we performed the same analysis using a different electrode.
Here, we conducted the same steps as during the training and model analysis but instead of using FCz electrode we used Fp1 electrode. This electrode is located in a different area to FCz electrode, namely on left hemisphere, at anteriofrontal area (for concrete position see Appendix, Figure 1) and it was randomly preferred over other electrodes from the same area.The anteriofrontal area was also preferred randomly over other available brain areas.

The model with 4 states was briefly analyzed using partially the same methodology as the one applied for data from FCz electrode. We avoided using the ERP mean values per driving condition to divide the participants into high and low susceptibility groups instead we compared the driving conditions directly before and after stimulus. The details concerning the analysis can be found in Appendix.

The results showed similar pattern of states before and after stimulus onset as those observed in FCz electrode. Namely, we did not observe any significant difference across driving conditions while the states with low mean values were the prevailing ones both before and after stimulus onset. Finally, we also observed the rise of probabilities of states whose mean values were high.

The main conclusion that follows this analysis is that the anteriofronta area showed similar pattern with the fronto-central area where FCz electrode resides.

General Discussion

Summary of results and implications

In this thesis, I investigated whether susceptibility to auditory stimuli can be predicted using a HMM. Results showed that by using a 3-state model, states tend to prolong. Consequently, probabilities to transition from one state to a different state are low. What's more, I could not observe significant differences in probabilities of each state across different driving conditions, before stimulus onset. As a result, it was not possible to attribute any particular state to a general level of susceptibility or attention.

Next, by using a model with 5 states, we could see that one state could be a possible candidate to represent the cortical reaction to the unexpected stimuli (i.e., state 4 in model 5). But predicting it, turned out to be infeasible as we could not find any meaningful correlation of this state before and after stimulus onset.

The theoretical background which accompanied this thesis, was that people are expected to be more susceptible to unrelated sounds while performing a type of driving which requested less attentional commitment (van der Heiden et al., submitted). In other words, if people drive actively then we would expect a higher persistence of a particular state compared to when they were driven by a car in autonomous mode. Unfortunately, the performed analysis did not reveal such nuances: the probabilities of states are almost the same likely to occur across various driving groups and conditions.

The result raises the question whether there is such thing as a cluster of numerical values in EEG signal which can be an indication of high or low attentional commitment. If indeed there is no such group of values, then the processes occurring before the stimulus in the measured area, are irrelevant to the processes which occur after the stimulus onset. And, therefore ERP component analysis remains the most appropriate method to estimate the responses. Consequently, the signal before the stimulus represents pure noise for cognitive science research and the bottom-up exploratory methods such as HMM might not be helpful.

On the other hand, this cannot be categorically true. If we consider the experimental setup once again, we can imagine that when the tone is played with periodicity around 2 seconds, the response to unexpected stimulus might also be reduced as the participant after some time gets used to hearing the tones. Furthermore, the participants were divided into passive and active groups, with active to actually having the task of pressing a button when hearing deviant sound. Given that, we would expect that people in the active group should be more attentive to sounds, and anticipating the correct sound. The anticipation for a sound can affect other

cognitive processes which in turn can affect the attentional allocation to the primary task. Then, the states and their frequencies before the stimulus could be also influenced. Consequently, the inability to identify the hidden states could have been partially caused by the specificity of the experimental design.

What's more, the modelling of such a dynamic process as driving, can be a problem when restricting the model to have 3 or 5 states. On the other hand, having a very high number of states might not be robust enough, and lead to overfitting. In our case while performing the mean improvement analysis during training, a 3-state model was chosen as the best candidate based on the technicalities, as we did not have any theoretical explanation to show preference for one or the other model which is an important drawback of the method we preferred to follow.

Contrary, having some sort of expectation in regards to the number of states could prove to be the key in accepting the best model (Anderson et al., 2016). Of course, we purposely pursued the bottom-up approach, which in our case revealed an insufficiency of the method preferred. In particular, our bottom-up approach on the one hand provided us the possibility to analyse EEG signal at its raw form at any time point. On the other hand, drawing any strong conclusions about the findings was not always possible.

Limitations and future work

After using HMMs to model the EEG signal in a driving experiment with different types of driving, we attempted to find the states which can characterize best the data at hand. Apart from the mere estimation of the states, we were interested to see whether the findings could provide a valuable information of the attentional state the driver were in. Unfortunately, we could not observe any concrete statistically significant results that could adhere to some pattern which in turn would give a hint of increased or reduced susceptibility to the sound or more general attentional state.

One of many possible reasons could be that the actual attentional state is not in areas that are measured using FCz or Fp1 electrodes, but that this takes place in some other area of the brain. Therefore, using a different brain area could prove a better choice in identification of attentional states. What's more, maybe examining all areas at the same time could provide more significant results. Therefore, an improvement could be the estimation of the states given all available electrodes which could give a more spherical overview of the mental states(Anderson et al.,2016).

Another possible reason could be that the interval before the stimulus onset that was used in the current analysis (100 milliseconds) is too short in order to estimate the general attentional state. If we consider attention as a more constant state then taking into account a bigger time window before stimulus onset might provide more reliable results.

Additionally, during preprocessing step of our data we used downsampling rate of 200Hz which essentially means that our data had as a time step 5 ms. Probably, by downsampling even more (Anderson et al.,2016) we could have more significant results for the states which occurred less frequently.

Finally, what we attempted can be considered a classification problem from ML perspective. But then the dataset used during training and testing set, would need to be finely annotated. While we also used fP3 component to categorise our data into groups, it is still not enough as the attention can be shifted many times back and forth across tasks and stimuli. Having in place such shifts annotated can provide a stronger predictability of a model(Solhjoo et al., 2005).

The further studies should entail taking into consideration these three factors in attempt to investigating the predictability of attentional state in drivers. Taken together, future work could include looking at the different brain locations, extending the prestimulus time for analysis, as well as having explicitly annotated data to provide more insight into attention fluctuations.

Conclusion

The attempt to predict susceptibility to unexpected auditory stimuli during active and autonomous driving, showed that modeling of such multifaceted task cannot be effectively accomplished using strictly data first approach. Therefore, prior theoretical assumptions, finely designed experimental set up as well as correctly annotated data can prove of a high importance in prediction of attentional allocation.

References

Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The discovery of processing stages: Extension of Sternberg's method. *Psychological review*, *123*(5), 481.

Anllo-Vento, L., Luck, S. J., & Hillyard, S. A. (1998). Spatio-temporal dynamics of attention to color: Evidence from human electrophysiology. *Human brain mapping*, *6*(4), 216-238.

Barua, S., & Begum, S. (2014). A review on machine learning algorithms in handling EEG artifacts. In *The Swedish AI Society (SAIS) Workshop SAIS, 14, 22-23 May 2014, Stockholm, Sweden*.

Bashivan, P., Rish, I., Yeasin, M., & Codella, N. (2015). Learning representations from EEG with deep recurrent-convolutional neural networks. *arXiv preprint arXiv:1511.06448*.

Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1196.

Borst, J. P., & Anderson, J. R. (2015). The discovery of processing stages: Analyzing EEG data with hidden semi-Markov models. *NeuroImage*, *108*, 60-73.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, *3*(3), 201.

Correa-Jaraba, K. S., Cid-Fernández, S., Lindín, M., & Díaz, F. (2016). Involuntary capture and voluntary reorienting of attention decline in middle-aged and old participants. Frontiers in human neuroscience, 10, 129.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience*, *18*(18), 7426-7435.

Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research*, *22*(3), 309-321.

De Winter, J. C., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation research part F: traffic psychology and behaviour*, *27*, 196-217.

Downing, P., Liu, J., & Kanwisher, N. (2001). Testing cognitive models of visual attention with fMRI and MEG. *Neuropsychologia*, *39*(12), 1329-1342.

Freeman, W., & Quiroga, R. Q. (2012). Imaging brain function with EEG: advanced temporal and spatial analysis of electroencephalographic signals. Springer Science & Business Media.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, *25*(4), 355-373.

Frith, C. D., & Allen, H. A. (1983). The skin conductance orienting response as an index of attention. *Biological psychology*, *17*(1), 27-39.

Forney, G. D. (1973). The viterbi algorithm. *Proceedings of the IEEE*, *61*(3), 268-278.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132*(3423), 349-350.

Hoeks, B., & Levelt, W. J. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*, *25*(1), 16-26.

Johnson, A., & Proctor, R. W. (2004). Attention: Theory and practice. Calif : SAGE Publications

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing* (Vol. 3). London: Pearson.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583-1585.

Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., & Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neuroscience letters*, *244*(2), 73-76.

Kiymik, M. K., Akin, M., & Subasi, A. (2004). Automatic recognition of alertness level by using wavelet transform and artificial neural network. *Journal of neuroscience methods*, *139*(2), 231-240.

Klonowski, W. (2009). Everything you wanted to ask about EEG but were afraid to get the right answer. *Nonlinear Biomedical Physics*, *3*(1), 2.

Laumann, K., Gärling, T., & Stormark, K. M. (2003). Selective attention and heart rate responses to natural and urban environments. *Journal of environmental psychology*, *23*(2), 125-134.

Lenartowicz, A., Delorme, A., Walshaw, P. D., Cho, A. L., Bilder, R. M., McGough, J. J.,& Loo, S. K. (2014). Electroencephalography correlates of spatial working memory deficits in attention-deficit/hyperactivity disorder: vigilance, encoding, and maintenance. *Journal of Neuroscience*, *34*(4), 1171-1182.

Liu, N. H., Chiang, C. Y., & Chu, H. C. (2013). Recognizing the degree of human attention using EEG signals from mobile sensors. *Sensors*, *13*(8), 10273-10286.

Luck, S. J. (2005). An introduction to the event-related potential technique MIT press. *Cambridge, Ma*, 45-64.

McCarthy, J. (1970). Computer Controlled Cars. Retrieved from http://www-formal.stanford.edu/jmc/progress/cars/cars.html.

Michel, C. M., Lehmann, D., Henggeler, B., & Brandeis, D. (1992). Localization of the sources of EEG delta, theta, alpha and beta frequency bands using the FFT dipole approximation. *Electroencephalography and clinical neurophysiology*, *82*(1), 38-44.

Mirowski, P. W., LeCun, Y., Madhavan, D., & Kuzniecky, R. (2008, October). Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. In *Machine Learning for Signal Processing, 2008. MLSP 2008. IEEE Workshop on* (pp. 244-249). IEEE.

Monastra, V. J., Lubar, J. F., Linden, M., VanDeusen, P., Green, G., Wing, W., ... & Fenger, T. N. (1999). Assessing attention deficit hyperactivity disorder via quantitative electroencephalography: An initial validation study. *Neuropsychology*, *13*(3), 424.

Oscillatory EEG Correlates of Arithmetic Strategies: A Training Study - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/Schematic-display-of-EEG-electrode-positions-For-statistical-analyses-ERS-ERD-was_fig2_233539681 [accessed 6 Jul, 2018]

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT–MST. *Neuron*, *18*(4), 591-598.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, *59*(1-2), 185-198.

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. *Encyclopedia of human computer interaction*, *1*, 211-219.

Porges, S. W., & Raskin, D. C. (1969). Respiratory and heart rate components of attention. *Journal of experimental psychology*, *81*(3), 497.

Rabiner, L. R., Lee, C. H., Juang, B. H., & Wilpon, J. G. (1989, May). HMM clustering for connected word recognition. In *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on* (pp. 405-408). IEEE.

SAE International. (2014). J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Retrieved from http://standards.sae.org/j3016_201401/.

Scheer, M., Bülthoff, H. H., & Chuang, L. L. (2016). Steering demands diminish the early-P3, late-P3 and RON components of the event-related potential of task-irrelevant environmental sounds. *Frontiers in human neuroscience*, *10*, 73.

Shtyrov, Y., Pulvermüller, F., Näätänen, R., & Ilmoniemi, R. J. (2003). Grammar processing outside the focus of attention: an MEG study. *Journal of Cognitive Neuroscience*, *15*(8), 1195-1206.

Sonnleitner, A., Simon, M., Kincses, W. E., Buchner, A., & Schrauf, M. (2012). Alpha spindles as neurophysiological correlates indicating attentional shift in a simulated driving task. *International journal of psychophysiology*, *83*(1), 110-118.

Solhjoo, S., Nasrabadi, A. M., & Golpayegani, M. R. H. (2005, September). Classification of chaotic signals using HMM classifiers: EEG-based mental task classification. In *Signal Processing Conference, 2005 13th European* (pp. 1-4). IEEE.

Tatum, W. O. (2014). Ellen r. grass lecture: Extraordinary eeg. *The Neurodiagnostic Journal*, *54*(1), 3-21.

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta psychologica*, *135*(2), 77-99.

Tsai, M. J., Hou, H. T., Lai, M. L., Liu, W. Y., & Yang, F. Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, *58*(1), 375-385.

Van der Heiden, R.M.A., Janssen, C.P., Donker, S.F., Hardeman, L.E.S., Mans, K., and Kenemans, J.L. (Submitted). Susceptibility to audio signals during autonomous driving. Submitted for review.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron*, *30*(3), 829-841.

Wang, Y., Jung, T. P., & Lin, C. T. (2018). Theta and alpha oscillations in attentional interaction during distracted driving. *Frontiers in behavioral neuroscience*, *12*, 3.

Wester, A. E., Böcker, K., Volkerts, E. R., Verster, J. C., & Kenemans, J. L. (2008). Event-related potentials and secondary task performance during simulated driving. Accident Analysis & Prevention, 40(1), 1–7. http://doi.org/10.1016/j.aap.2007.02.014

Wickens, C. D., & McCarley, J. S. (2007). *Applied attention theory*. New York, NY:CRC press.

Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *Journal of Neurophysiology*, *79*(3), 1574-1578.

Wright, R. D., & Ward, L. M. (2008). *Orienting of attention*. New York, NY:Oxford University Press.

Yeo, M. V., Li, X., Shen, K., & Wilder-Smith, E. P. (2009). Can SVM be used for automatic EEG detection of drowsiness during car driving?. *Safety Science*, *47*(1), 115-124.

Zhang, Q., van Vugt, M., Borst, J. P., & Anderson, J. R. (2018). Mapping working memory retrieval in space and in time: A combined electroencephalography and electrocorticography approach. *NeuroImage*, *174*, 472-484.

Appendix

stationary	fP3	participant	**autonomous**	fP3	participant	**driving**	fP3	participant
Low	3,8	18	Low	2,3	10		Low	-1,4
	5,2	3		2,5	4			3,2
	5,2	10		4,4	5			4,2
	6,9	7		4,9	3			4,9
	7,9	14		5,3	13			4,9
	8,1	6		5,6	12			4,9
	9,3	4		7,1	6			5,0
	9,8	5		7,3	19			5,4
	10,4	12		7,6	1			5,4
High	11,0	16		7,8	18			6,3
	11,4	13		8,3	2			6,4
	12,4	9		8,4	16		High	7,1
	12,7	2	High	8,7	9			7,5
	13,4	17		8,8	14			7,7
	13,4	11		10,8	7			8,5
	14,2	19		12,7	11			8,9
	15,3	1		13,5	17			11,0
	15,7	15		13,8	15			11,4
Average	10		Average	8			Average	6

***Table* 1**: Groups of participants per mean ERP component which represents the susceptibility to audio stimuli

N	a=0.1	a=0.05	a=0.01
1	0.988	0.997	0.999
2	0.900	0.950	0.990
3	0.805	0.878	0.959
4	0.729	0.811	0.917
5	0.669	0.754	0.875
6	0.621	0.707	0.834
7	0.584	0.666	0.798
8	0.549	0.632	0.765
9	0.521	0.602	0.735
10	0.497	0.576	0.708
11	0.476	0.553	0.684
12	0.458	0.532	0.661
13	0.441	0.514	0.641
14	0.426	0.497	0.623
15	0.412	0.482	0.606
16	0.400	0.468	0.590
17	0.389	0.456	0.575
18	0.378	0.444	0.561
19	0.369	0.433	0.549
20	0.360	0.423	0.537
21	0.352	0.413	0.526

22	0.344	0.404	0.515
23	0.337	0.396	0.505
24	0.330	0.388	0.496
25	0.323	0.381	0.487
26	0.317	0.374	0.479
27	0.311	0.367	0.471
28	0.306	0.361	0.463
29	0.301	0.355	0.456
30	0.296	0.349	0.449
35	0.275	0.325	0.418
40	0.257	0.304	0.393
45	0.243	0.288	0.372
50	0.231	0.273	0.354
60	0.211	0.250	0.325
70	0.195	0.232	0.303
80	0.183	0.217	0.283
90	0.173	0.205	0.267
100	0.164	0.195	0.254
150	0.134	0.159	0.208
300	0.095	0.113	0.148

Table 2. Table of Critical Values: Pearson Correlation.N here represents degrees of freedom.

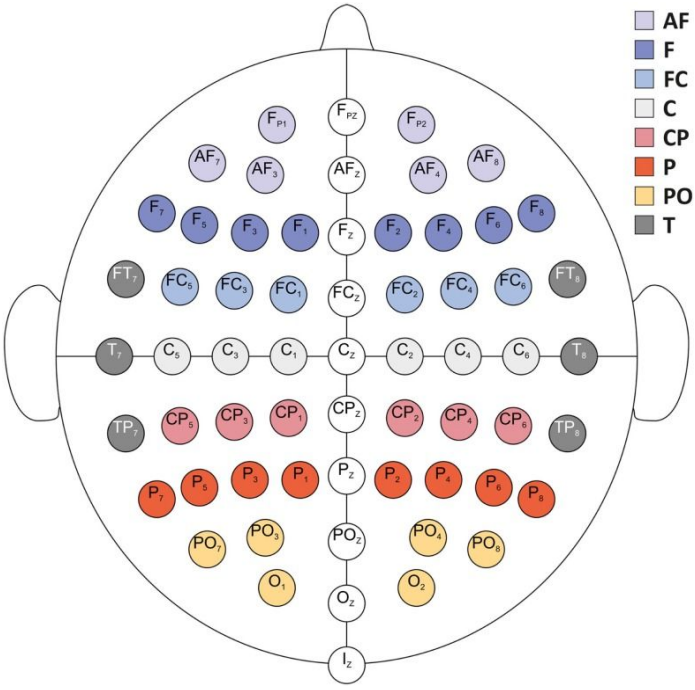

Figure 1: EEG electrodes positions. Retrieved from https://www.researchgate.net/Schematic-display-of-EEG-electrode-positions-For-statistical-analyses-ERS-ERD-was_fig2_233539681

Fp1 electrode analysis

Below you can find all the tables with the results we obtained during analysis of Fp1 electrode.

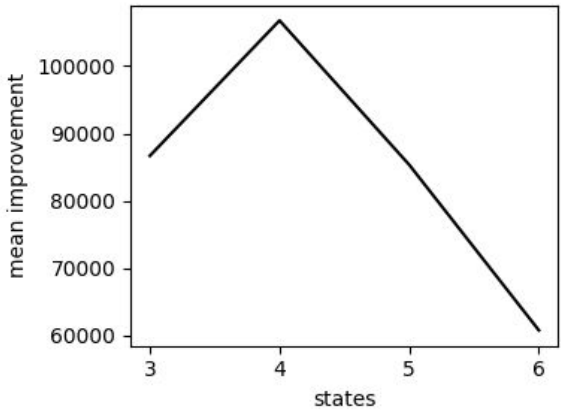

Table 3. Mean improvement over state increase during training. The highest improvement can be observed in state 4.

The mean improvement procedure showed that model with 4 states had the highest log-likelihood improvement. Therefore, we preferred model with 4 states over other to start the analysis. Table 4 shows matrix with transitional probabilities A, as well as each state's statistics.

states	0	1	2	3
0	0.95	0.00	0.01	0.03
1	0.00	0.98	0.00	0.02
2	0.02	0.00	0.98	0.00
3	0.04	0.01	0.00	0.95

states	Mean	SD
0	6	5
1	-41	27
2	39	25
3	-8	5

Table 4.Left, Transitional probabilities matrix. **Right**. Mean and SD of each state

Next, we took 100 ms before the stimuli onset in order to estimate the probabilities of each state to occur in that time interval. We also retrieved data from 300-400 ms post stimuli to estimate the probabilities of the states after the auditory stimuli. The results are shown in Table 5.

	driving		autonomous		stationary	
states	**before**	**after**	**before**	**after**	**before**	**after**
0	0.47	0.36	0.47	0.36	0.47	0.35

1	0.04	0.14	0.05	0.15	0.05	0.16
2	0.06	0.13	0.06	0.14	0.07	0.14
3	0.42	0.37	0.42	0.35	0.41	0.35

Table 5 The probabilities of each state to occur before and after the stimulus onset for every driving condition.

Given the probabilities before the stimulus onset, we do not observe any significant difference across driving conditions. We also do not see any significant difference in probabilities after the stimulus onset across driving conditions. But, here as well as in analysis of FCz electrode we see substantial increase in probabilities of state 1 and state 2, which is accompanied by the decrease of state 0 and state 3.

Forward - Backward Algorithm : The learning problem

In our study to identify the hidden brain processes we used Hidden Markov Models. During training of HMMs we used the forward-backward algorithm. In this section I would like to give a better understanding of the mechanics of this algorithm.

Our problem was: Given the EEG continuous data and a single value for the number of states, we need to learn the states transitional probabilities matrix A ,the emission probabilities B.

The forward-backward algorithm represents a special case of Expectation Maximization algorithm, whose main feature is that it is an *iterative* procedure. The iterative procedure is used to compute initial probabilities (A and B) and then by using the learned probabilities, it iteratively improves them.

In particular, to compute the transitional probabilities A, normally we can count the number of times that a transition from state i to state j occurs, and then normalize the result by dividing the total number of times that a transition from i occurs. But, as we do not know the states, this procedure is not possible.
To understand how the algorithm works, we need to introduce two important probabilities that contribute to the whole process.

First important probability for the algorithm is the forward probability. The forward probability is a probability for an observation o to be emitted by state i. If we have 3 states then we will have 3 such probabilities. The forward probability is computed by the forward algorithm, which sums over the probabilities of all possible hidden states paths.

$$\text{Forward probability} = P(q_t, o_{1:t}) \qquad (1)$$

Second important probability is the backward probability. The backward probability is the probability for an observation o to be observed from time t+1 till the end given that our model is at time t. Or, in other words, how likely is for an observation o to be observed in the future given estimated A and B at time t and state q_t.

$$\text{Backward probability} = P(o_{t+1:T}|q_t) \qquad (2)$$

The forward and backward probabilities are computed inductively based on the mechanics of the forward algorithm, which represents an example of dynamic programming.

Then, if we know the forward probabilities and the backwards probabilities we can compute the probability $P(q_t|O)$, where O is our sequence of observations. Namely,

$$P(q_t|O) \propto P(q_t,O) = P(o_{t+1:T}|q_t,o_{1:t})\, P(q_t, o_{1:t}) \qquad (3)$$

By claiming that $o_{t+1:T}$ is conditionally independent from $o_{1:t}$, we have

$$P(q_t|O) \propto P(q_t,O) = P(o_{t+1:T}|q_t)\, P(q_t, o_{1:t}) \qquad (4)$$

We see that equation 4 can be computed by multiplying the backward and forward probabilities. Having estimated the equation 4 for all timepoints $t \in T$, we can easily estimate the parameters A and B of HMM.

Finally, if we know $P(q_t|O)$, we can make any type of inference concerning our model.