Discovering Processing Stages by combining EEG with Hidden Markov Models

Jelmer P. Borst (jelmer@cmu.edu)
John R. Anderson (ja+@cmu.edu)
Dept. of Psychology, Carnegie Mellon University

Abstract

A new method is demonstrated for identifying processing stages in a task. Since the 1860s cognitive scientists have used different methods to identify processing stages, usually based on reaction time (RT) differences between conditions. To overcome the limitations of RT-based methods we used Hidden Markov Models (HMMs) to analyze EEG data. The HMMs indicate for how many stages there is evidence in the data, and how the durations of these stages vary with experimental condition. This method was applied to an associative recognition task in which associative strength and target/foil type were manipulated. The HMM-EEG method identified six different processing stages for targets and re-paired foils, whereas four similar stages were identified for new foils. The duration of the third, fifth and sixth stage varied with associative strength for targets and re-paired foils. We present an interpretation of the identified stages, and conclude that the method can provide valuable insight in human information processing.

Keywords: EEG; HMM, processing stages.

Introduction

One of the main goals of cognitive science is to understand how humans perform tasks. To this end, scientists have long tried to identify different processing stages in human information processing. The first to do this in a systematic manner was probably Franciscus Donders. Almost 150 years ago, Donders proposed a method to measure the duration of cognitive stages (1868). By subtracting the RTs of two tasks that were hypothesized to share all but one processing stage, the duration of that stage could be calculated. A strong – and often problematic – assumption of Donders' subtractive method is the idea that it is possible to add an entire stage without changing the duration of other stages. To test whether different stages exist in the first place, Sternberg proposed the additive-factor method (1969). Although Sternberg overcame a limitation of Donders' method, the additive-factors method has its own drawbacks: it can only indicate the minimum number of stages in a task and it does not yield duration estimates of the stages. To improve on these inherent problems of RT-based methods and get better insight in stage existence and duration we propose a new method that uses HMMs (e.g., Rabiner, 1989) to analyze EEG data.

The basic idea of our method is to fit HMMs with different numbers of states to the EEG data (note that we use 'processing stages' and 'HMM states' interchangeably throughout the paper). The optimal number of states can then be determined by comparing the log-likelihoods of the fitted HMMs. Subsequently, the durations of the different states can be inspected, as well as how these durations vary with condition. Using this information, and by comparing EEG signatures between states and experimental conditions, one can interpret the functional characteristics of the identified processing stages.

Our approach is based on a similar method that was used to analyze fMRI data (Anderson & Fincham, in press; Anderson et al., 2010). For instance, Anderson and Fincham (in press) applied the method to mathematical problem solving, and discovered four stages: encoding the problems, planning a solution strategy, solving the problems, and entering a response. Although these results were promising, the temporal resolution of fMRI is severely limited, both by having scans that typically last one to two seconds and by the sluggish nature of the hemodynamic response. EEG, on the other hand, has a millisecond resolution, allowing for the discovery of processing stages in fast-paced tasks.

We applied the HMM-EEG analysis to an associative recognition task. During the study phase of this task, subjects were asked to learn word pairs. In a subsequent test phase – during which EEG data were collected – subjects were again presented with word pairs, which could be the same pairs as they learned previously (targets), rearranged pairs (re-paired foils), or pairs consisting of novel words (new foils). Subjects had to decide whether they had seen the pair during the study phase or not. Successful discrimination required remembering not only that the words were studied (item information), but also how the words were paired during study (associative information).

A conventional EEG analysis and a classifier analysis of this study were reported elsewhere (Borst et al., submitted). Currently, we are interested in finding out how many stages the subjects went through while determining a correct response.

Methods

Subjects

Twenty individuals from the Carnegie Mellon University community participated in a single 3-hr session for monetary compensation (9 males and 11 females, ages ranging from 18 to 40 years with a mean age of 26 years). All were right-handed and none reported a history of neurological impairment.

Design

The experiment consisted of a study phase in which subjects learned word pairs and a test phase in which they were tested on these word pairs. In addition to probe type (targets, re-paired foils, or new foils), we manipulated word length and associative strength. Words could either be short (4 or 5

letters) or long (7 or 8 letters). Associative strength was manipulated by varying the number of word pairs a particular word occurred in. This is referred to as *associative fan*, and is known to have a strong effect on RT and accuracy (for a review, see Anderson & Reder, 1999). Words in our experiment could have a fan of 1 or 2, that is, they could occur in one or two word pairs. Both words in a word pair always had the same associative fan. New foils (foils consisting of words that were not presented in the study phase) always had an associative fan of 1, they only appeared in a single word pair. Thus, there were 10 conditions: 2 (Probe: target or re-paired foil) × 2 (Word Length: short or long) × 2 (Fan: 1 or 2) + short and long new foils.

Materials

Word pairs were constructed from a pool of 464 words selected from the MRC Psycholinguistic Database (Coltheart, 1981). Half of the words were nouns of 4 or 5 letters and composed the short word list. The other half of the words were nouns of 7 or 8 letters and composed the long word list. Word frequency and imageability ratings were matched between those lists. The 232 words of each length were divided randomly into two lists – a 24-word study list and a 208-word new foil list – such that the lists were matched on word frequency, imageability, and word length according to *t*-tests (all *p*s > .1).

The lists were used to create three sets of probes: targets, re-paired foils, and new foils. A set of 32 target word pairs was constructed from the study lists such that there were eight word pairs for each combination of length (short or long) and fan (1 or 2). Both words in short pairs were 4 or 5 letters and both words in long pairs were 7 or 8 letters. Each word in a fan 1 pair appeared only in that pair, whereas each word in a fan 2 pair appeared in two pairs. A corresponding set of 32 re-paired foil pairs was constructed in a similar manner by combining words from different target pairs of the appropriate length and fan. A set of 208 new foil word pairs was constructed from the new foil lists such that there were 104 word pairs for each length (all fan 1). The randomization of words and their assignment to conditions were unique for each subject.

Procedure

The study phase started with each target word pair presented onscreen for 5000 ms, followed by a 500-ms blank screen. Subjects were instructed to read each pair and make an initial effort to memorize it. Following target presentation, subjects completed a cued recall task designed to help them learn the word pairs. On each trial they were presented with a randomly selected target word and had to recall the word(s) paired with it (two-word responses were required for fan 2 words). The self-paced responses were typed and feedback (in the form of the correct response) was provided for 2500 ms following errors. If a target word elicited an error, it was presented again after all other target words had been presented. A block of trials concluded when all 48 target words had elicited a correct response. Subjects completed a total of three blocks of cued recall.

After the study phase, subjects entered the EEG recording chamber and completed the test phase. Each trial began with a centrally presented fixation cross for a duration sampled from a uniform distribution ranging from 400 to 600 ms. Following fixation, a probe word pair appeared onscreen (one word above the other) until the subject responded with a keypress to indicate whether the probe had been studied during the training phase. The probe was either a target, re-paired foil, or new foil. Targets required "yes" responses (indicated by pressing the J key with the right index finger) and foils required "no" responses (indicated by pressing the K key with the right middle finger). Subjects were instructed to respond quickly and accurately. Following the response, accuracy feedback was displayed for 1000 ms, after which a blank screen appeared for 500 ms before the next trial began. Subjects completed a total of 13 blocks with 80 trials per block. All 10 conditions occurred equally often in random order in each block, resulting in 104 trials per condition during the test phase. Targets and re-paired foils were repeated during the test phase (they each appeared once per block), but each new foil appeared only once in the entire experiment.

EEG recording

Subjects sat in an electromagnetically shielded chamber. Stimuli appeared on a CRT monitor placed behind radio-frequency shielded glass and set 60 cm from the subjects. The electroencephalogram was recorded from 32 Ag-AgCl sintered electrodes (10-20 system). Electrodes were also placed on the right and left mastoids. The right mastoid served as the reference electrode, and scalp recordings were algebraically re-referenced offline to the average of the right and left mastoids. The vertical electrooculogram (EOG) was recorded as the potential between electrodes placed above and below the left eye, and the horizontal EOG was recorded as the potential between electrodes placed at the external canthi. The EEG and EOG signals were amplified by a Neuroscan bioamplification system with a bandpass of 0.1 to 70.0 Hz and were digitized at 250 Hz. Electrode impedances were kept below 5kΩ.

EEG preprocessing

Recording artifacts in the EEG data were removed based on visual inspection. Following artifact rejection, the data were decomposed into independent components. Components associated with eye blinks were visually identified and projected out of the EEG recordings. A 0.5-30 Hz band-pass filter was applied to attenuate high-frequency noise. Trials were extracted from the continuous recording and baseline-corrected using a linear baseline, such that the 200 ms before stimulus onset and 80-160 ms after the response were on average 0 (visual inspection showed no condition difference at this interval after the trial). Incomplete trials due to artifact rejection were excluded, as well as trials containing voltages above +75 μV or below -75 μV. In

addition, all incorrect trials and correct trials with RTs exceeding three standard deviations (SDs) from the mean per condition per subject were removed. For the HMM-EEG analysis we also removed trials with RTs longer than 3000 ms. In total, 16.1% of the trials was excluded.

For efficiency, the EEG data were down-sampled to 50 Hz. Every four samples were then combined into a single 'super-sample', by quadrupling the number of channels. That is, from four 20-ms samples with 32 channels we created one 80-ms super-sample with 128 channels. A super-sample contained information about the mean voltage in each channel, as well as about whether this voltage increased or decreased over the 80 ms interval. Next, we normalized each channel to a mean of 0 and a SD of 1, and applied a principle component analysis (PCA) to the 128 channels. The results of the PCA were again normalized; the first 20 PCA components were used for the HMM-EEG analysis.

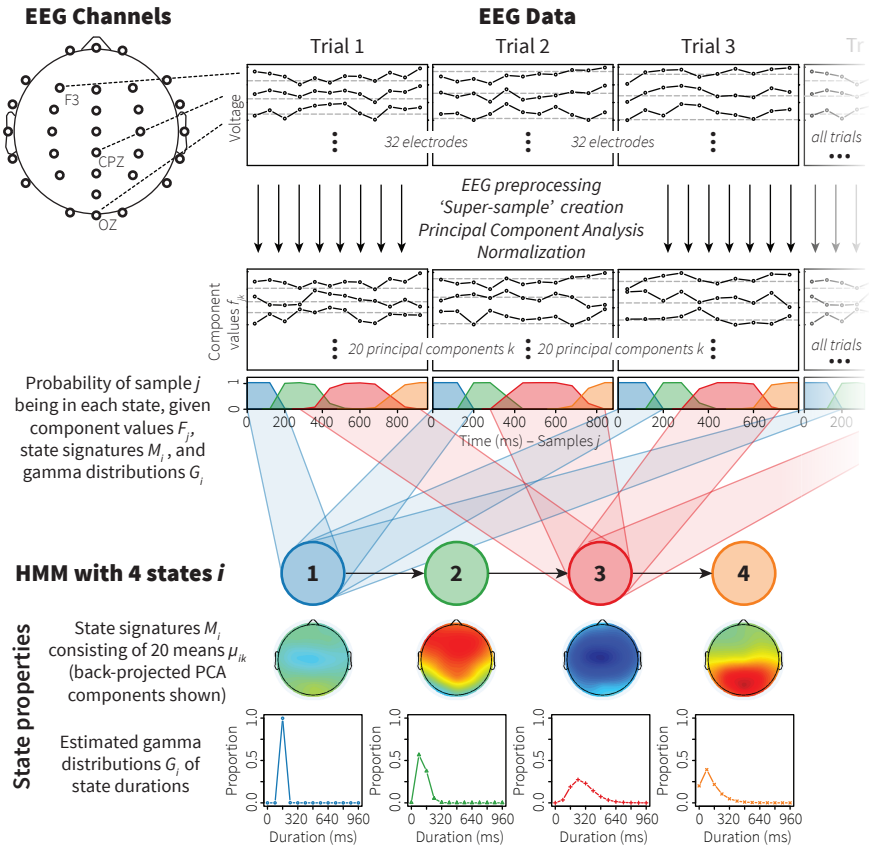

Figure 1. Overview of the HMM-EEG analysis. EEG data comes in at the top and is preprocessed into PCA components. At the bottom a fitted 4-state HMM is shown, with state signatures and gamma distributions. The center graph shows the probability of each sample *j* for each state given this HMM. The connections between sample likelihoods and states are shown for states 1 and 3.

The HMM-EEG Analysis

The HMM-EEG analysis consists of two main parts: (1) determining the optimal number of states and (2) computing the properties of the identified states. Both parts of the analysis depend on fitting HMMs to the preprocessed EEG data. We will therefore first discuss the structure and parameter estimation procedure that was used for the HMMs. We then explain how these HMMs were used to find the optimal number of states and how we computed the properties of these states.

HMM structure and parameter estimation

An HMM simulates a system that is at any given time in one of a set of distinct states, between which it transitions at certain times (e.g., Rabiner, 1989). In our analysis, each state represents a processing stage in the task (e.g., encoding the stimulus, executing a response). A state is associated with a brain signature M_i that represents the average EEG-activation pattern during this processing stage, and with a gamma distribution G_i that represents the state's durations over the trials in the experiment. For current purposes, we only consider HMMs with a linear structure, that is, state 1 always transitions to state 2, state 2 to state 3, etc.

An example of a four-state HMM is shown in Figure 1. At the top of the figure EEG data is shown for three channels over three trials of the experiment, at the bottom the HMM with associated brain signatures and gamma distributions.

HMM algorithms can be used to find parameters M_i and G_i that yield the optimal interpretation of the data given an HMM with *r* states (see Anderson & Fincham, in press, for a more detailed explanation for the kind of HMMs that are used in this paper; Rabiner, 1989; Yu & Kobayashi, 2006). To calculate the solutions we adapted software that minimizes the summed log-likelihood of the HMM over all trials (Yu & Kobayashi, 2006).

Figure 1 shows the result of such an optimization procedure for a 4-state HMM. Given the optimal state signatures and gamma distributions, the probability that each sample *j* belongs to a state is depicted in the center of the figure. As expected, the first samples in each of the three trials probably belong to state 1 (blue), the next samples to state 2 (green), etc. In addition, state 1 is always two samples long in the three trials in the figure, matching the gamma distribution of this stage. State 3, on the other hand, is much more variable in duration.

For clarity the explanation above assumes a gamma distribution for each state. In the actual analysis we used separate gamma distributions for each condition and state, allowing for different duration estimates per condition.

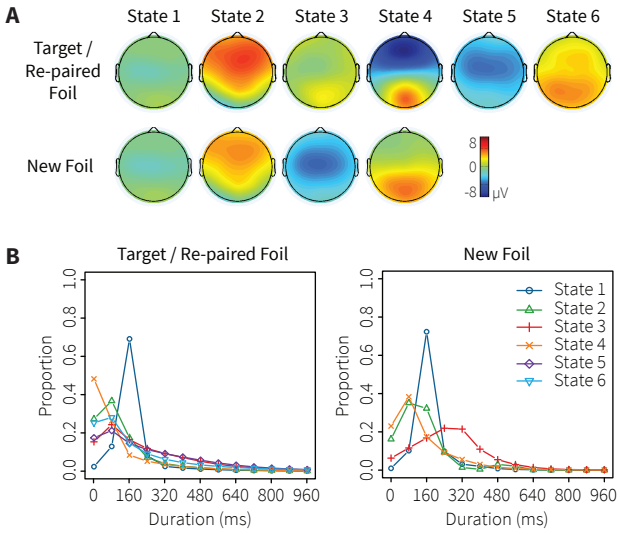

Figure 2. State signatures and gamma distributions.

Number of states and state properties

Above we explained how an *r*-state HMM can be determined that gives an optimal interpretation of the data. However, what we are really interested in is finding the optimal number of states to describe the data. A simplistic approach would be to compare the log-likelihoods of HMMs with different numbers of states. However, because HMMs with more states have more parameters to fit the data, they will typically yield a better fit. What we want to know is if the extra parameters explain enough extra variance to be warranted. To this end we applied leave-one-out cross validation (LOOCV).

Our LOOCV method estimated state signatures for *n*-1 subjects, and calculated the log-likelihood of the n^{th} subject given these signatures while allowing for different state durations for the n^{th} subject (to accommodate speed differences between subjects, unlike Anderson & Fincham, in press). This process was repeated for all subjects.

The LOOCV procedure was repeated for HMMs with different numbers of states. To select the best model we used a sign-test: if a *k*-state model fitted the data of *x* out of *n* participants better than all (*l*<*k*)-state models we choose it as the winner. The underlying idea is that while a (*k*+1)-state model will fit the data of *n*-1 subjects better in the estimation phase than a *k*-state model, it is at least as likely to fit the n^{th} subject worse (Anderson & Fincham, in press). According to a sign-test, a significant increase is reached when 15 out of 20 subjects improve (p = .04).

After determining the optimal number of states, we computed the properties of the identified states. First, we estimated an optimal HMM on the data of all subjects. We used the state signatures of this model to estimate optimal gamma distributions for each subject. These gamma distributions were used to calculate the average state duration for each subject and condition, which were used in subsequent ANOVAs to determine which states change in duration with condition. In addition, the subject-specific models give us the probability for each sample in the data to be in a certain state (center of Figure 1). This was used to calculate differences in EEG activation between conditions.

Results

For reasons of brevity we do not report behavioral results separately. RTs can be inferred from Figure 3. For targets and re-paired foils, Fan ($F(1,19) = 65.42$, $p < .001$), Probe ($F(1,19) = 45.10$, $p < .001$), and the interaction between Fan and Probe ($F(1,19) = 31.40$, $p < .001$) had a significant effect on RT, as indicated by a repeated measure ANOVA. In addition, new foils were responded to much faster than the other probe types, which was expected given that no associative information has to be retrieved for new foils.

Number of stages

Because new foils are very different from the other probe types – no associative information has to be retrieved for new foils – we decided to run separate analyses for new foils and targets/re-paired foils. For targets and re-paired foils a 6-state HMM turned out to be the winner. It was better for at least 16 subjects than HMMs with fewer states, and no HMM with more states had a higher log-likelihood for more than 9 subjects. The new foils also showed evidence for 6 states: 17 subjects fitted better with a 6-state HMM than with 4 states. However, the 4-state solution compares better to the 6-state solution of targets and re-paired foils.[1] Although there might be more stages in the data, we can be secure in the assumptions that there are at least 6 states for targets and re-paired foils and 4 states for new foils and in whatever conclusions these assumptions lead to. Thus, we will focus on the 4-state solution for new foils in this paper.

Stage properties

Figure 2 shows the gamma distributions and state signatures of the 6-state HMM for the targets/re-paired foils and the 4-state HMM for the new foils. Interestingly, the first two states of both solutions seem very similar. Correlations between the state signatures confirm that stage 1 and stage 2 in both HMMs resemble each other closely: 0.98 and 0.97.

The estimated gamma distributions in Figure 2b are averaged over conditions and subjects. They show that stage 1 has a very fixed duration, of two samples or 160 ms. The other stages are more variable. A duration of 0 means that the state is skipped, which happens most often (in 50% of the trials) for stage 4 of the targets/re-paired foils. For the other stages these percentages are under 30%.

Figure 3 shows the state durations in more detail, split out for conditions. We will only list major effects with *p*-values < .01 (repeated measure ANOVAs), as these are used below to interpret the results.

[1] The 6-state solution for new foils effectively splits up two of the stages into shorter stages. Although this might explain the new foils in themselves better, our interest is explaining associative recognition.

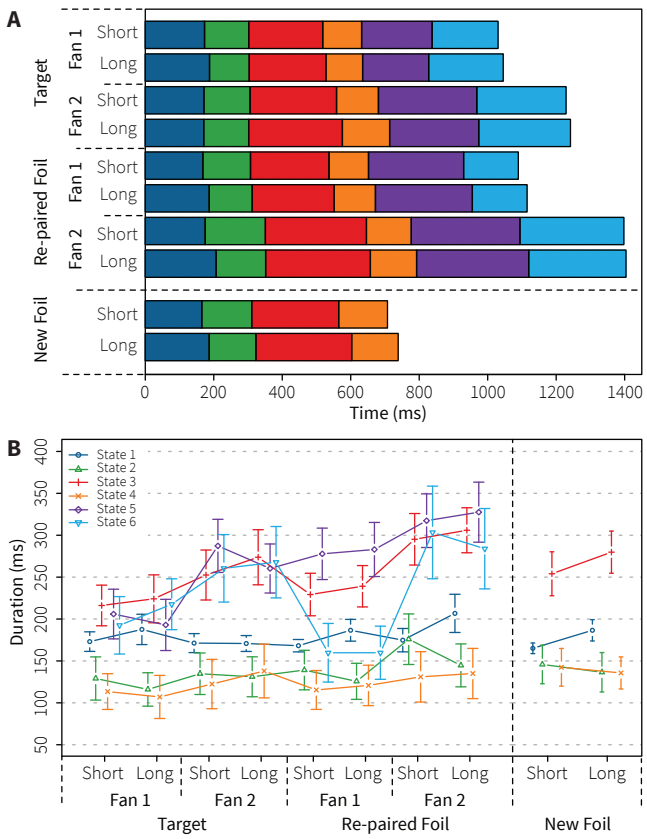

Figure 3. State durations. A shows how the state durations add up to form a complete trial, whereas B illustrates the effect of condition on state duration.

State 1 and 2 seem stable over the different conditions, even between the two different HMMs. This matches the observation that their state signatures are very similar. Stage 3 is longer for fan 2 items than for fan 1 ($F(1,19) = 15.14$, $p < .001$). Stage 4 seems to be an intermediate stage that is often skipped for the targets/re-paired foils, and it does not change with condition. For the new foils, stage 4 is the final stage. It does not change in duration with word length. Stage 5 varied strongly in duration with both Fan ($F(1,19) = 16.12$, $p < .001$) and Probe ($F(1,19) = 20.32$, $p < .001$). Stage 6, the final stage for targets/re-paired foils, is longer for fan 2 items than for fan 1 items ($F(1,19) = 21.55$, $p < .001$). In addition, there is an interaction between Fan and Probe ($F(1,19) = 16.75$, $p < .001$), with the fan effect being stronger for re-paired foils than for targets.

The HMM-EEG analysis aims to find states with similar brain signatures in the different conditions of the experiment. Although that is the case, there might still be differences between conditions within a stage. Figure 4A shows the differences between conditions for the 6-state HMM for targets/re-paired foils; Figure 4B for the 4-state HMM for new foils. These differences were calculated by estimating brain activity for each state, condition, and subject. The resulting values were subjected to t-tests for each electrode.

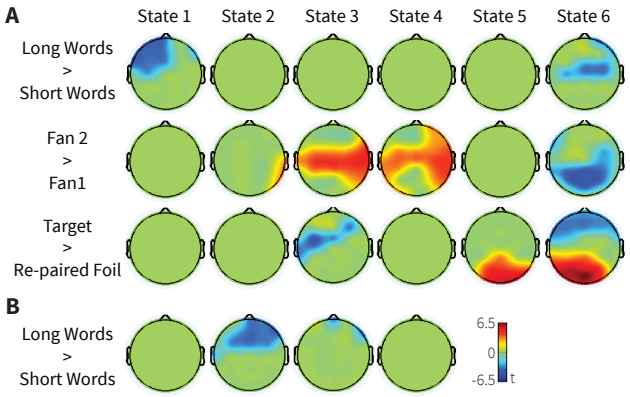

Figure 4. Differences between conditions in states for (A) targets/repaired foils and (B) new foils. The maps show *t*-values for FDR-corrected *p*-values < .05.

Figure 4A shows that long words resulted in less activity than short words in state 1 over left prefrontal electrodes, and in state 6 over central electrodes. Fan 2 items resulted in more activity than fan 1 items in states 3 and 4 over midline electrodes, whereas they showed less activity in state 6 over parietal regions. Finally, targets elicited a little less activity than re-paired foils in state 3, and more activity in states 5 and 6 over parietal and occipital sites. The largest effect for new foils was in state 2, where long words resulted in less activity than short words for frontal electrodes.

Interpretation of the Processing Stages

The underlying reason for wanting to identify processing stages is explaining how tasks are performed. In this section we will give our interpretation of the processing stages discovered by the HMM-EEG method.

The first two stages seem to reflect visually perceiving the two words on the screen. Both stages hardly varied with condition, and are very similar between targets/re-paired foils and new foils – implying that the words are not processed yet in relation to the experimental task. In addition, there are effects of word length on brain activity in stage 1 for targets/re-paired foils and in stage 2 for new foils. Although word length effects are typically strongest in occipital regions, Hauk et al. (2009) showed a left prefrontal effect that appears to match our observation.

We hypothesize that stage 3 reflects item retrieval, to determine whether the presented words were learned during the study phase of the experiment. First, the duration of stage 3 varies strongly with fan and there is also a strong effect of fan on brain activity in stage 3. Existing models of the fan effect assume that the effect originates in declarative memory, implying that this stage is memory related (e.g., Anderson & Reder, 1999). Second, for new foils this is the stage where information has to be retrieved about whether the words were studied or not. After the third stage there is only a short response stage, which is similar to stage 6 of the targets/re-paired foils. Given the matching time course, we assumed that stage 3 reflects an item retrieval stage for targets/re-paired foils as well.

The idea of an early item retrieval stage and a later associative retrieval stage (stage 5) resembles dual-process theories of recognition (e.g., Rugg & Curran, 2007). To judge whether a stimulus was experienced before, dual-process theories assume an early 'familiarity' process, followed by a functionally distinct recollection process. With respect to our experiment, the familiarity phase could correspond to stage 3 – in which it is determined whether the items are familiar – whereas stage 5 could correspond to the recollection stage in which associative information is retrieved.

Familiarity and recollection processes have been related to different ERP components (Rugg & Curran, 2007). Familiarity elicits a negative response between 300-500 ms over mid-frontal electrodes, with new items being more negative than studied items. This matches the observation that new foils in our experiment have a more negative brain signature over mid-frontal electrodes than targets/re-paired foils in stage 3. Recollection has been linked to the parietal old/new effect, which is more positive for old than for new items. If our stage 5 reflects recollection of associative information, it should show a parietal positivity for targets versus re-paired foils, which it does.

Stage 4 is skipped in 50% of the trials. We tentatively hypothesize that it reflects working memory consolidation of the items that are retrieved from memory in stage 3. This is not a necessary process, which might explain why it is skipped in 50% of the trials.

As explained above, we assume that stage 5 reflects associative retrievals. Not only does it show the parietal old/new effect, but it also varied in duration both with fan and probe type, which are known to influence the length of associative retrievals.

Stage 6 of the targets/re-paired foils and stage 4 of the new foils are the final stages in the task. We assume that they reflect response stages. The duration of stage 6 changes with fan, and shows an interaction between fan and probe type. For new foils this last stage is shorter than for the other conditions. We interpreted these duration differences as an effect of response confidence. Subjects responded faster and more accurate to new foils than to targets/re-paired foils, and faster and more accurate to fan 1 items than to fan 2 items – indicating they might have been more confident in those responses.

The effects on brain activity support this interpretation. There were differences over parietal electrodes between targets and re-paired foils (targets being more positive), between fan 1 and 2 items (fan 1 items being more positive), and between new foils and targets/re-paired foils (the signature of new foils is slightly more positive). We hypothesize that these effects resemble a P300, which is known to increase with response confidence (Wilkinson & Seales, 1978).

Discussion

In this paper we have presented a new method for identifying processing stages in a task, which uses HMMs to analyze EEG data. For the associative recognition task, the method yielded a 6-state solution for targets and re-paired foils and a 4-state solution for new foils. These solutions seem to be reasonable, and could be interpreted by using information about how the stages varied in length, and how the brain activity within stages differed between conditions. The results matched dual-process theories of recognition, both in expected stage duration and brain activity.

Naturally, other interpretations of these results are also conceivable. For instance, the duration of the last two stages could be explained with an accumulator model, which samples faster for the easier conditions.

That being said, especially in combination with earlier promising effects on fMRI data (e.g., Anderson & Fincham, in press; Anderson et al., 2010), we think that the HMM-EEG method shows great promise for investigating human information processing.

References

Anderson, J. R., Betts, S., Ferris, J. L., & Fincham, J. M. (2010). Neural imaging to track mental states while using an intelligent tutoring system. *PNAS USA*, *107*(15), 7018–7023. doi:10.1073/pnas.1000942107

Anderson, J. R., & Fincham, J. M. (in press). Uncovering the Sequential Structure of Thought. *Cognitive Science*.

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology General*, *128*(2), 186–197.

Borst, J. P., Schneider, D. W., Walsh, M. M., & Anderson, J. R. (submitted). *Journal of Cognitive Neuroscience*.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.

Donders, F. C. (1868). *De snelheid van psychische processen (On the speed of mental processes)*.

Hauk, O., Pulvermüller, F., Ford, M., Marslen-Wilson, W. D., & Davis, M. H. (2009). Can I have a quick word? Early electrophysiological manifestations of psycholinguistic processes revealed by event-related regression analysis of the EEG. *Biological psychology*, *80*(1), 64–74.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta psychologica*, *30*, 276–315.

Wilkinson, R. T., & Seales, D. M. (1978). EEG event-related potentials and signal detection. *Biological psychology*, *7*(1), 13–28

Yu, S. Z., & Kobayashi, H. (2006). Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden Markov model. *IEEE Transactions on Signal Processing*, *54*, 1947–51.