Extracting duration information in a picture category decoding task using hidden Markov Models

This content has been downloaded from IOPscience. Please scroll down to see the full text.
2016 J. Neural Eng. 13 026010
(http://iopscience.iop.org/1741-2552/13/2/026010)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 136.152.142.47
This content was downloaded on 29/04/2016 at 01:37

Please note that terms and conditions apply.
Extracting duration information in a picture category decoding task using hidden Markov Models

Tim Pfeiffer¹, Nicolai Heinze², Robert Frysch¹, Leon Y Deouell³, Mircea A Schoenfeld⁴, Robert T Knight⁵ and Georg Rose¹

¹ Institute for Medical Engineering, Otto-von-Guericke-University Magdeburg, Germany
² Leibniz Institute for Neurobiology, Magdeburg, Germany
³ Department of Psychology, Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel
⁴ Institute for Neurology, University Medical Center Magdeburg, Germany
⁵ Helen Wills Neuroscience Institute, University of California, Berkeley, USA

E-mail: tim.pfeiffer@ovgu.de

Received 10 August 2015, revised 4 November 2015
Accepted for publication 18 December 2015
Published 9 February 2016

Abstract

Objective. Adapting classifiers for the purpose of brain signal decoding is a major challenge in brain–computer-interface (BCI) research. In a previous study we showed in principle that hidden Markov models (HMM) are a suitable alternative to the well-studied static classifiers. However, since we investigated a rather straightforward task, advantages from modeling of the signal could not be assessed. Approach. Here, we investigate a more complex data set in order to find out to what extent HMMs, as a dynamic classifier, can provide useful additional information. We show for a visual decoding problem that besides category information, HMMs can simultaneously decode picture duration without an additional training required. This decoding is based on a strong correlation that we found between picture duration and the behavior of the Viterbi paths. Main results. Decoding accuracies of up to 80% could be obtained for category and duration decoding with a single classifier trained on category information only. Significance. The extraction of multiple types of information using a single classifier enables the processing of more complex problems, while preserving good training results even on small databases. Therefore, it provides a convenient framework for online real-life BCI utilizations.

Keywords: brain–computer-interfaces, classification, hidden-Markov-models, electrocorticography, magnetoencephalography, support-vector-machines

(Some figures may appear in colour only in the online journal)

1. Introduction

Classification of acquired brain signals is a central part of brain–computer-interface (BCI) research and implementation.

A broad variety of algorithms have been studied, covering different kinds of neural networks [1–3], decision trees [4, 5] and support vector machines (SVM) [6–8]. These classifiers feature easy architecture and implementation along with a manageable set of parameters. This guarantees high robustness for difficult data, but these static classifiers are unable to assess temporal dynamics [9]. This can be critical for specific applications such as speech decoding for clinical purposes and everyday life solutions (e.g. smartphones, computers,
The established gold-standard method for these temporal decoding problems is represented by hidden Markov models (HMMs) [10–12]. HMMs have been rarely used in BCI. The broad variety of possible configurations provided by HMMs may contribute to this underutilization. The particular setup needs to be chosen carefully in order to adapt to the given problem. Also, most of these configurations have an extensive amount of free parameters that have to be estimated during training. With a limited amount of training data, the classifier, without further restrictions to the model is likely to end up in a scenario that is known as the ‘curse of dimensionality’ [9, 13]. The curse of dimensionality refers to the case when the amount of training data is small compared to the dimension of the feature space. This leads to improper description of the different classes and therefore, often results in poor decoding results. Despite these issues, it has been shown that HMMs can serve as a viable classifier, if appropriate adaptations are applied. In a previous study [14] we investigated HMMs as an alternative method for BCI decoding purposes. We showed for a simple finger tapping paradigm that HMMs can provide similar decoding accuracies as SVMs, which represent the gold-standard static classifier. Combining suitable features and rigorous restrictions to the model structure as well as optimized and problem-adapted initialization we were able to show that HMMs can compete with SVMs regarding decoding accuracies. This is especially true for high accuracy cases, which might seem as a limitation at first glance. However, real-life BCI solutions demand low error rates in order to guarantee reliable function even in risky applications and to avoid user frustration [15].

Machine learning algorithms rely on information from large training ensembles. This becomes particularly important for noisy data and easily confusible classes. Small databases and noisy signals are typical in BCI research. Many studies try to compensate for this by long trial durations as well as restricted and highly distinguishable classes (see table 1) that favor less complex classifiers. These tasks are focused on pure decoding of class information in a condition where all trials of a class have the same properties (i.e. each class has only a single token). In these tasks dynamic classifiers are not exploited to their full potential. Here, we increase the complexity of the problem by adding an additional attribute to the stimuli. Besides decoding the type of a stimulus (‘quality’), we also aim at detecting their durations (‘quantity’). Dynamic classifiers might be particularly suitable for this task as they model the time course of given signals by default. There is no direct dynamic classifier analogue for SVM static classifiers. The aim of this work is to assess whether HMMs—as dynamic classifiers—after being trained for the quality classification, are also able to perform the task of quantity classification without any additional training effort. The routine we introduce here is stable and ‘nearly’ unsupervised. In section 2 of this paper, we present methods and material used for this study followed by the results and discussion in sections 3 and 4, respectively. Finally, we conclude our work in section 5.

### Table 1. Examples of BCI studies using HMM classification.

| Study      | Task                                                   | # Classes | Trial duration (sec) |
|------------|--------------------------------------------------------|-----------|----------------------|
| Obermaier  | Left/right-hand imagination                           | 2         | 5                    |
| Chiappa    | Left/right-hand imagination, mental generation of words | 3         | 1–3                  |
| Lederman   | Left/right-hand imagination                           | 2         | 6                    |
| Rezaei     | Left/right-hand imagination                           | 2         | 5.5                  |
| Wissel     | Finger tapping (4 fingers)                            | 4         | ~0.5                 |

### 2. Material and methods

#### 2.1. Data acquisition and paradigm

The study is based on datasets from four different subjects, recorded with electrocorticography (ECoG) and magnetoencephalography (MEG). ECoG data were recorded from two volunteering patients (right-handed males) within a larger study that will be reported separately [20]. Both patients have been implanted with subdural electrode grids for clinical purposes in the course of pre-surgical planning of epilepsy treatment at Stanford, CA, USA. Electrode grids were solely placed based on clinical criteria and covered a variety of cortical areas including lateral occipital and medial temporal areas (details see [20]). The patients gave their informed consent in advance of the recordings. The ECoG was recorded with a sampling frequency of 3051.7 Hz. For pre-processing, a high pass filter (cut-off: 0.5 Hz) as well as a notch filter around the power line frequency (60 Hz) were applied to the data and all electrodes were re-referenced to the common-average-reference. Afterwards, artifacts (e.g. epileptic activity, machine noise, loose contacts) were manually rejected by visual inspection. The resulting time series were epoched into individual trials covering the interval from −100 to 2000 ms with respect to picture onset times (see figure 1). MEG acquisition was performed with a whole-head BTi Magnes system (4D-Neuroimaging, San Diego, CA, USA). Data from 248-sensors have been recorded from two healthy volunteers (male, age 26–28) with a sampling frequency of 1017.25 Hz. Artifacts (e.g. blinking, eye movement) were removed from

TVs, customer service etc.).
Table 2. Number of recorded and rejected trials per subject and class (objects/faces/watches).

| Subject | Recorded trials | Rejected trials | Final class breakdown |
|---------|-----------------|-----------------|-----------------------|
| ECoG 1  | 312             | 6/4/3           | 98/100/101            |
| ECoG 2  | 18/19/19        | 86/85/85        |                       |
| MEG 1  | 225             | 10/5/6          | 65/70/69              |

Table 3. Number of states $Q$ used in the HMMs, threshold $\theta$ for duration decoding and corresponding accuracies (chance levels in brackets) for all data sets. For comparison, accuracies of supervised decoding of duration classes are shown in the last column. A detailed description and discussion of these results can be found in the supplementary material.

| Subject | $Q$ | $\theta$ | RMSE (ms) | duration (%) | superv. (%) |
|---------|-----|----------|-----------|--------------|-------------|
| ECoG 1  | 4   | 4        | 279 [600] | 53.3 [20.0]  | 30.8        |
| ECoG 2  | 4   | 4        | 123 [600] | 82.8 [20.0]  | 37.2        |
| MEG 1   | 5   | 4        | 232 [585] | 68.5 [33.3]  | 65.9        |
| MEG 2   | 5   | 4        | 187 [585] | 76.5 [33.3]  | 71.9        |

The MEG data was referenced and epoched into trials with the same interval (i.e. [−100, 2000 ms]) as for the ECoG data. The stimuli of the paradigm consisted of varying pictures from four categories (objects, faces, watches and clothes). Pictures were presented to the subjects for durations chosen randomly from five different time spans (300, 600, …, 1500 ms). The inter-trial interval varied in duration (600, 750, …, 1200 ms; see figure 1). To control attention subjects responded to the presentation of a piece of clothing with a button press. These clothing target stimuli accounted for approximately 10% of the total trial count. The ECoG data is obtained from another study, for which the paradigm was originally designed. To avoid any bias in the decoding routine, neural response to all stimulus durations needs to be fully covered. This is ensured if trial lengths are adapted to the longest appearing stimulus. With respect to the data (compare figure 4, bottom right, red line), we therefore choose to extend trial segments up to two seconds after stimulus onset. In this setup, combinations of brief picture presentation (e.g. 300 ms) with short inter-stimulus intervals (e.g. 600 ms) lead to additional stimuli at the end of a trial (see figure 1). As a consequence of that and our former findings [21] some parameters were slightly modified for MEG recording. We changed presentation durations to 500, 1000 and 1500 ms and increased ISIs to (1500, 1650, …, 2100 ms). In order to focus on visual information only, target trials from the ECoG data have been omitted for the study, as they include motor responses. Consequently, they have been omitted in the MEG paradigm. Stimuli were shown either on a projection screen (MEG data) positioned 1 m away from the subject or on a notebook screen (ECoG data) within the patients reach. Detailed information on the number of recorded trials and rejections per class is listed in table 2.

2.2. Feature extraction and selection

In this study, features of two different types have been extracted. Low frequency time domain (LFTD) features were generated by low-pass filtering in Fourier domain with a cutoff frequency of 30 Hz (MEG: 10 Hz) and subsequent down sampling of the time series. The resulting feature vectors consist of 98 sample points which corresponds to a re-sampling rate of approximately 48 Hz. High gamma (HG) features were extracted by computing the...
spectral power of the signals using a sliding Hann-window approach (window length = 250 ms). The square root of the power spectrum was computed by fast Fourier transform for each window and the resulting coefficients were averaged in the frequency band of 70–200 Hz. We computed HG features for 100 time points, which leads to a window overlap of approximately 92.5%. Consequently, the resulting features vary slowly over time (compare figure S1). Both feature extraction routines are described in detail in a previous study [14]. Channel selection is performed on training data only, using a two stage approach. First, an algorithm based on the Davies–Bouldin index [14] is applied to select a predefined number of most informative channels for the picture category separation problem. The corresponding labels of the training data are used to select the channels. In a second step, channels containing information on the picture duration are selected using an unsupervised method as follows: raw data amplitude values of all trials are averaged separately for each channel. A channel is selected if the mean amplitude in the first 250 ms of the time series (150 ms after image onset respectively) exceeds a defined threshold. That particular time frame has been chosen manually to match the typical interval in which visual activation appears (compare figure S2). If more than the requested number of channels fulfills this condition, the threshold is iteratively increased, until the channel count fits. Note that this algorithm does not depend on any labels. The number of channels selected by both methods has been fixed for all datasets (7 duration and 3 category channels). The total count of 10 selected channels is motivated by previous findings ([14], table 3) showing a superiority of that particular channel count in a comparable HMM setup, especially for LFTD features. We decided to reserve 3 channels for category decoding, as there are three
Bottom right: LFTD features of a typical ECoG 2.

For the shortened version, containing no multi-stimulus issues, the relevant time segment is used, i.e., transitions are limited to single-step forward jumps. Both models are forced to start with the first state.

Figure 3. Topology of the HMMs. For ECoG data a four state HMM is used (top). Transitions are allowed between a state and its predecessor and successor. For MEG data a simple five state left-to-right model (bottom) is used, i.e. transitions are limited to single-step forward jumps. Both models are forced to start with the first state.

Figure 4. Comparison of mean Viterbi paths for full-length ECoG trials (top) and shortened versions (bottom) using LFTD features only. Decoding of full-length trials leads to nearly identical Viterbi paths in the relevant time segment (full lines), making duration identification impossible. Dashed lines show the behavior of the paths in the interval where additional stimuli typically appear (compare plot bottom right). For the shortened version, containing no multi-stimulus issues (refer figure 1), Viterbi paths differ clearly for different presentation durations. Bottom right: LFTD features of a typical ‘duration’ channel (subject ECoG 2) averaged for ‘900 ms’-trials without additional stimuli (dashed black line) and with an additional stimulus at 1500 ms (full red line).

different classes, and use the rest (i.e. 7 channels) for duration information. This choice represents a tradeoff between optimization of category decoding performance (which would benefit from a higher category channel ratio) and incorporation of the desired additional duration information. However, it has been made manually and cannot be justified theoretically.

2.3. Classification

The full procedure (see figure 2) contains a training of HMMs for each category class. Consequently, three HMMs (one for objects, faces and watches respectively) are built using the training dataset. A standard Baum–Welch algorithm is used for the estimation of all HMM parameters (i.e. transition matrices, state means and variances as well as priors). After training, classification is performed on the test set using a simple maximum likelihood approach [14]. For each trial in the test set, the so called Viterbi path of the HMM with the highest category likelihood is calculated. The Viterbi path is the state sequence of the corresponding HMM, that is most likely to explain the given time series of features. Based on the Viterbi paths, duration information is extracted as follows. Using a pre-defined state threshold the sample point, at which the Viterbi path first reaches this threshold, is extracted. This point is mapped to the picture presentation duration $d$ using a linear relation:

$$d = m \cdot s + n.$$ 

The slope $m$ in this relation is known, as it refers to the time increment from one feature sample $s$ to the following. The offset $n$ in this equation is a parameter that is not known a priori at the present stage. It is introduced to compensate for the delay between picture offset and according HMM state changes. In order to determine it, we use the labels of one presentation duration only. Using this information, we calculate a representative (mean) sample point at which the Viterbi paths of trials with that particular duration typically reach the threshold. To compensate for fluctuations and outliers, we use a histogram-based method. We first calculate a rough histogram of the threshold crossings dividing the sample point range into ten equally sized boxes $b$. We then select the box with the highest count $(b_{\text{max}})$ and calculate the median of all sample points in the proximity (20 sample points) of its center $c$:

$$s_{\text{rep}} = \text{median}_{i \in [c - 10, c + 10]} [s_i].$$ 

This value is used as the representative point $s_{\text{rep}}$ of calibration class cal (duration) to calculate the offset as follows:

$$n_{\text{cal}} = d_{\text{cal}} - m \cdot s_{\text{rep}}, \quad \text{cal} = \{1, 2, 3\} \text{ ECoG, MEG}. $$

For comparison, category decoding is also performed with the gold-standard SVMs. Classification is executed in a one-versus-one mode using linear kernels. The parameter $C$ has been determined in an earlier study [21] using exhaustive search: $C_{\text{ECoG}} = 2^{-3}$, $C_{\text{MEG}} = 2^{-10}$.
2.3.1. Initialization and model constraints. For initialization of the HMMs we reused the algorithm presented in [14]. The routine is based on k-means clustering with several small extensions that allow for adaptations to specific needs originating by the nature of the data. The most important parameter for these adaptations is the ‘coupling constant’ \( \tau \) that introduces the time point of a feature sample as an additional dimension for clustering. High values for tau force the algorithm to strictly cluster feature values that are temporally close, while \( \tau = 0 \) leads to an unmodified k-means clustering\(^6\).

Since LFTD features are complex in their temporal structure compared to HG features different values have been used for the two feature types (\( \tau_{LFTD} = 25 \) and \( \tau_{HG} = 0 \)). Due to the limited amount of training data, introducing appropriate model constraints is critical to adapt the HMMs to the problem. We employed our former findings [14] introducing only small changes to adapt to the specific task. State transitions have been limited to one (instead of two) forward jumps to smoothen the resulting Viterbi paths. For HG features, a single-step backward jump has been introduced as it was found to compensate for the multi-stimulus issue mentioned above. This aspect is discussed in detail in section 4. Since back-jumping enables re-use of early states, it allows the reduction of the total number of states to four (down from five). For LFTD features, back-jumping turned out to be not-feasible, as it results in messed-up Viterbi paths. This is most likely due to the complex structure of the LFTD features. A detailed discussion is also provided in section 4. Both topologies are illustrated in figure 3.

2.3.2. Implementation. The entire framework of this study was implemented using MATLAB R2012b from mathworks\(^7\). HMM functionality is provided by the open source toolbox from Kevin Murphy from the University of British Columbia [22]. SVM decoding is performed using the LibSVM library [23] for MATLAB.

2.4. Testing setup

All findings have been achieved using a five-fold cross-validation (CV) routine. Trials are assigned to specific fold sets using uniform random permutations with a balanced amount of training samples per class. Balancing is required to assure unbiased classification. In order to average out fluctuations in the decoding accuracies originating in varying allocations of trials to the folds, the CV procedure is repeated 20 times.

3. Results

In previous studies we found that no informative HG features could be extracted from our MEG data [21]. Therefore, MEG

---

\(^6\) The choice of \( \tau \) becomes particularly important if identical feature values appear at different time points. For \( \tau = 0 \), these samples would be mapped to the same cluster and therefore, end up in the same Markov state for a HMM. Since HMMs are intended to model the time course without any discontinuities, such a clustering would be irrational. Choosing \( \tau > 0 \) prevents this scenario. However, if the value for tau is too high, clustering corresponds to the time domain only, i.e. all consecutive samples of a time segment form a cluster. This may result in unrepresentative mean values due to high intra-cluster variance of the features.

\(^7\) MATLAB and Parallel Processing Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.
decoding results are shown for LFTD features only. As mentioned earlier (section 2.1 and figure 1), the paradigm design used for ECoG acquisition leads to multiple stimuli within the time series of trials with briefer presentation durations. As a result of the complex structure of LFTD features (compare supplementary figure S1) and the limited amount of training data available, the HMMs are not capable of distinguishing the differences between true stimulus offsets and ‘ghost’ offsets resulting from the multi-stimulus issue. This can be seen easily when multiple stimuli are eliminated by restriction of data to a shorter time interval. In this setup, longer picture presentations (i.e. 1200 and 1500 ms trials) are not fully covered and hence, only brief durations stimuli can be analyzed. Considering this limitation, however, Viterbi paths exactly show the expected behavior (see figure 4). As a consequence of these findings, we decided to leave out LFTD results for ECoG data and focus on HG features only, which do not suffer from the above mentioned problem. This aspect will be discussed later.

3.1. Category decoding

For all four datasets, corresponding features have been extracted as described in section 2.2 and classification has been performed according to 2.3. The results for the category decoding task are shown in figures 5 and 6. Decoding accuracies for the ECoG data reach 80% and 85% for HMMs and SVMs, respectively. For MEG data results are about 20% lower for HMMs and 10%–15% lower using the SVM decoder. While the performance gap between SVM and HMM is about 12%–15% for MEG data, HMMs nearly reach the SVM results for the ECoG datasets (about 3%–6% difference). Detailed information on decoding results can be taken from the confusion matrices (figure 6). The results show a consistent behavior across all datasets. Face trials are detected best and the majority of classification errors results from mix-ups of object- and watch-trials.

After category classification, Viterbi paths are extracted for all datasets (figure 7). For purpose of better visualization and discussion of general behavior in figure 7, paths have been averaged over all trials separately for each of the presentation durations contained in the corresponding paradigm (i.e. 3 durations in MEG and 5 in ECoG data sets). Note that this average is not used for decoding. The paths’ standard deviations as well as all single trial paths are presented in supplementary material figures S3 and S4. Apparently, the paths show a consistent behavior across subjects and data acquisition modality even on a single trial basis. It is clearly

Figure 7. Mean Viterbi paths for the four datasets. Paths of all single trials have been averaged individually for each of the presentation durations. Plots in the top row show results from the ECoG datasets ECoG 1 and ECoG 2. MEG 1 and MEG 2 results are shown in the bottom row.
visible, that Viterbi paths are distinguishable for different presentation durations. Moreover, there is strong indication that the time points, at which the paths start to deviate from each other, directly correlate with the corresponding presentation duration. This is analyzed explicitly in the following.

3.2. Extracting duration information from Viterbi paths

The procedure to extract duration information from the Viterbi paths is described in section 2.3. As a first step, the threshold needs to be specified. To optimally fit the actual scenario, chosen values differ slightly due to the different structure of the two feature types. All values can be found in table 3. Using these values, threshold crossings of all paths can be determined. The results are shown in figure 9. Afterwards, the offsets for duration decoding are estimated based on the threshold crossings of a single calibration class (‘training’). To ensure unbiased results, each duration class was used once for calibration in a CV-like manner. Listed results (table 3) are averaged over all appearing duration classes (i.e. three for MEG and five for ECoG). In order to measure the accuracy of the decoded durations, root mean square errors (RMSE) with respect to the true durations have been computed. Individual decoding errors are visualized as histograms in figure 8. Furthermore, all estimated offset values are shown in table 4. Chance values for duration decoding (as listed in table 3) have been determined assigning random durations from the interval $[0, 2000]$ ms to all trials and computing accuracies afterwards.

For means of easier comparability, predicted durations have been mapped to discrete classes via simple assignment to the nearest discrete duration values that appear in the paradigm (i.e. durations of $0–449$ ms are assigned to the 300 ms class, $450–749$ to the 600 ms class etc). Classification accuracies are illustrated as confusion matrices in figure 10.

---

Table 4. Estimated offset parameter $n^{cal}$ in ms for all calibration sets and subjects.

| Calibration set cal | Subject 1 | 2 | 3 | 4 | 5 | Mean |
|---------------------|-----------|---|---|---|---|------|
| ECoG 1              | $-148$    | $-110$ | $-90$ | $-89$ | $-107$ | $-108.8$ |
| ECoG 2              | $-148$    | $-147$ | $-165$ | $-164$ | $-163$ | $-157.4$ |
| MEG 1               | $-421$    | $-350$ | $-236$ |          |          | $-335.7$ |
| MEG 2               | $-293$    | $-286$ | $-257$ |          |          | $-278.7$ |

Figure 8. Histograms of duration decoding errors for all four datasets (f.l.t.r. top: ECoG 1, ECoG 2; bottom: MEG 1, MEG 2). Relative occurrence rates have been computed normalizing the histogram by the highest appearing value. Negative error values indicate the predicted duration is shorter than the actual.

Additional statistical information on threshold crossings is provided in supplementary figure S5.
Comparing figures 9 and 10, the confusion matrices can also be understood as a discretized form of the threshold crossings.

4. Discussion

The category decoding shows consistent behavior for all data-sets. Confusion matrices reveal frequent mix-ups of watches and objects images by the classifier. This might result from the similarity of both image types, since watches are a sub-category of objects. However, faces could be consistently decoded without prominent confusions with other categories, resulting in the highest accuracy among the three classes. This behavior is in accordance with other studies [24–26]. There is a slight superiority of the SVM decoder for MEG data (figure 5). For ECoG data, HMM and SVM provide comparable results. This is consistent with our former findings [14] suggesting that both classifiers are at the same level when dealing with high quality features (i.e. in the upper accuracy range). The different properties of the two feature types used introduce the necessity to make adoptions in model structure and decoding strategy (e.g. Viterbi threshold). LFTD features, as they are time domain representatives, contain the ongoing brain background activity with superimposition of task-related characteristics. Therefore, time segments following the offset of a picture are not expected to result in reproducible LFTD feature sequences. This complexity introduces increased difficulty for appropriate modeling of such feature sequences with HMMs. As a consequence, only very simple HMM topologies can be used considering the limited amount of available training data. We address this issue by choosing a pure left-to-right model. In contrast, HG features show a more convenient behavior (compare figure S1). This allows us to loosen the restrictions to the model and permit a backward jump without losing smoothness of the Viterbi paths. Periods of picture presence lead to a significant increase in HG activity and this activity quickly returns to its initial low value, as soon as the picture presentation terminates. In the HMM state sequence this is expressed in backward jumps to earlier states after termination of the picture. Due to the decoding strategy of identifying only the first crossing of the state threshold, a later change to higher states (and therefore, possible threshold crossings) has no influence on decoding quality. Hence, this routine is able to deal with the ‘multi stimulus issue’ in the ECoG paradigm. In the absence of multiple stimuli within one trial, the backward jump is optional, thus, exactly reproducing the setup from [14]. However, it would still lead to increased category decoding accuracy since the model topology is simplified. As expected from the varying model structure, Viterbi paths of ECoG (HG features) and MEG data (LFTD features) show different behavior (see figure 7). The mean paths for the MEG datasets steadily increase up to approximately state three. From there, paths bend towards higher states with a larger slope of the curve. During that high slope part, they cross state 4, which is used for duration decoding. After a second bend, back
to a very slight slope, paths ascend slowly towards the final state, which acts as some sort of dump state accounting for periods of picture absence. Conversely, paths for the ECoG datasets make an early jump to state 2 associated with the picture onset. At the end of the image presentation paths abruptly ascend to the highest state (4) where they form a more or less marked plateau. The connecting decrease to state 2 can be interpreted as a return to the starting situation, which is eventually followed by additional increases (‘multi stimulus issue’). The described behavior of MEG and ECoG Viterbi paths is found not only for averaged paths but also consistently on a single trial basis (figure S4). This reliable behavior provides a solid foundation for decoding of duration information. Applying the Viterbi path analysis, presentation durations have been extracted with RMSEs in the range of 123–279 ms. For the best subject this results in discrete duration decoding.

5. Conclusion

In this work, we show—to our knowledge—the first use of the beneficial properties of HMMs to extract additional stimulus information in the context of BCI decoding problems. We present a largely automated routine that shows consistent results across subjects and, with minor adaptations, across acquisition modalities. Our results of duration decoding show strong correlations between Viterbi paths and picture presentation duration. For all four datasets decoding accuracies substantially above chance level were obtained. The best dataset provided classification accuracies of more than 80% for both category and duration decoding with a single classifier trained only on category information. The presented approach to decoding ‘quality’ (category) and ‘quantity’ (duration) information may be transferred to an every-day BCI task, e.g. movement direction and duration. However, for an every-day BCI use online decoding is essential. This can be addressed conveniently with HMMs requiring only slight changes to the routines presented here. More sophisticated algorithms for Viterbi path analysis as well as simulations of online scenarios will be focus of our upcoming work. The work emphasizes the benefit of dynamic classifiers compared to conventional static classifiers. In summary, it provides evidence that the properties of HMMs are efficient for such a BCI utilization.

Acknowledgments

The work of this paper is funded by the Federal Ministry of Education and Research (Germany) within the Forschungscampus STIMULATE under grant number ‘13GW0095A’ and supported by grant 2013070 from the US-Israel binational science foundation to LYD and RTK.

References

[1] Wu T, Yang B and Sun H 2010 Eeg classification based on artificial neural network in brain computer interface Life System Modeling and Intelligent Computing of Communications in Computer and Information Science ed L Kang et al vol 97 (Berlin: Springer) pp 154–162
[2] Hazrati M K and Abbas Erfanian A 2010 An online eeg-based brain–computer interface for controlling hand grasp using an adaptive probabilistic neural network Med. Eng. Phys. 32 730–9
[3] Huan N-J and Palaniappan R 2004 Neural network classification of autoregressive features from electroencephalogram signals for brain–computer interface design J. Neural Eng. 1 142
[4] Akram F, Han S M and Kim T-S 2015 An efficient word typing p300-bci system using a modified {T9} interface and random forest classifier Comput. Biol. Med. 56 30–6

[5] Aydemir O and Kayikcioglu T 2014 Decision tree structure based classification of [EEG] signals recorded during two dimensional cursor movement imagery J. Neurosci. Methods 229 68–75

[6] Kashihara K 2014 A brain–computer interface for potential nonverbal facial communication based on eeg signals related to specific emotions Frontiers Neurosci. 8 00244

[7] Jrad N, Congedo M, Phlypo R, Rousseau S, Flamary R, Yger F and Rakotomamonjy A 2011 sw-svm: sensor weighting support vector machines for eeg-based brain–computer interfaces J. Neural Eng. 8 056004

[8] Shenoy P, Miller K J, Ojemann J G and Rao R P N 2008 Generalized features for electrocorticographic bcs IEEE Trans. Biomed. Eng. 55 273–80

[9] Lotte F, Congedo M, Lécuyer A, Lamarche F and Arnaldi B 2007 A review of classification algorithms for eeg-based brain–computer interfaces J. Neural Eng. 4 R1

[10] Gales M and Young S 2007 The application of hidden Markov models in speech recognition Found. Trends Signal Process 1 195–304

[11] Baker J, Deng L, Glass J, Khudanpur S, Lee C-H, Morgan N and O’Shaughnessy D 2009 Developments and directions in speech recognition and understanding part 1 [dsp education] IEEE Signal Process. Mag. 26 75–80

[12] Rabine L. 1989 A tutorial on hidden Markov models and selected applications in speech recognition Proc. IEEE 77 257–86

[13] Jain A K, Duin R P W and Mao J 2000 Statistical pattern recognition: a review IEEE Trans. Pattern Anal. Mach. Intell. 22 4–37

[14] Wissel T, Pfeiffer T, Frysch T, Knight R T, Chang E F, Hinrichs H, Riegler J W and Rose G 2013 Hidden Markov model and support vector machine based decoding of finger movements using electrocorticography J. Neural Eng. 10 056020

[15] Kübler A, Holz E M, Riccio A, Zickler C, Kaufmann T, Kleih S C, Staiger-Sälzer P, Desideri L, Hoogerwerf E-J and Mattia D 2014 The user-centered design as novel perspective for evaluating the usability of bci-controlled applications PLoS One 9 e112392

[16] Obermaier B, Guger C, Neuper C and Pfurtscheller G 2001 Hidden Markov models for online classification of single trial [EEG] data Pattern Recognit. Lett. 22 1299–309

[17] Chiappa S and Bengio S 2004 Hm and iohmm modeling of eeg rhythms for asynchronous bci systems European Symp. on Artificial Neural Networks ESANN

[18] Lederman D and Tabrikian J 2012 Classification of multichannel eeg patterns using parallel hidden Markov models Med. Biol. Eng. Comput. 50 319–28

[19] Rezaei S, Tavakolklan K, Nasrabadi A M and Setarehdan S K 2006 Different classification techniques considering brain computer interface applications J. Neural Eng. 3 139

[20] Gerber E M, Buchanan K G, Kuperman R A, Auguste K I, Golan T, Parvizi J, Knight R T and Deouell L Y 2015 The neural basis of sustained visual perception Annual Meeting of the Society for Neuroscience, SfN (Chicago,USA)

[21] Pfeiffer T, Heinze N, Gerber E, Deouell L Y, Parvizi J, Knight R T and Rose G 2014 Decoding of picture category and presentation duration—preliminary results of a combined ecog and meg study Proc. 6th Int. Brain–Computer Interface Conference 2014 pages ID 042-1

[22] Murphy K 2005 Hidden Markov model (hmm) toolbox for matlab, University of British Columbia (https://www.cs.ubc.ca/~murphyk/software/HMM/hmm.html)

[23] Chang C-C and Lin C-J 2011 LIBSVM: a library for support vector machines ACM Trans. Intell. Syst. Technol. 2 1–27

[24] Haxby J V, Gobbini M I, Furey M L, Ishai A, Schouten J L and Pietrini P 2001 Distributed and overlapping representations of faces and objects in ventral temporal cortex Science 293 2425–30

[25] Fisch L et al 2009 Neural ignition: Enhanced activation linked to perceptual awareness in human ventral stream visual cortex Neuron 64 562–74

[26] Reddy L, Tsuchiya N and Serre T 2010 Reading the mind’s eye: decoding category information during mental imagery NeuroImage 50 818–25