

# Bayesian approaches for learning of primitive-based compact representations of complex human activities

Dominik Endres<sup>1</sup>, Enrico Chiovetto<sup>2</sup>, Martin A. Giese<sup>2</sup>

1 Theoretical Neuroscience Group, Dept. of Psychology, Philipps-University Marburg, Germany,  
e-mail: dominik.endres@uni-marburg.de

2 Section Computational Sensomotrics, Dept. Cognitive Neurology, University Clinic Tübingen,  
Germany, e-mail:

enrico.chiovetto@uni-tuebingen.de, martin.giese@uni-tuebingen.de

## Abstract

Human full-body activities, such as choreographed dances, are comprised of sequences of individual actions. Research in motor control shows that such individual actions can be approximated by superpositions of simplified elements, called movement primitives. Such primitives can be employed to model complex coordinated movements, as occurring in martial arts or dance. In this chapter, we will briefly outline several biologically-inspired definitions of movement primitives and will discuss a new algorithm that unifies many existing models and which identifies such primitives with higher accuracy than alternative unsupervised learning techniques. We combine this algorithm with methods from Bayesian inference to optimize the complexity of the learned models and to identify automatically the best generative model underlying the identification of such primitives. We also discuss efficient probabilistic methods for the automatic segmentation of action sequences. The developed unsupervised segmentation method is based on Bayesian binning, an algorithm that models a longer data stream by the concatenation of an optimal number of segments, at the same time estimating the optimal temporal boundaries between those segments. Applying this algorithm to motion capture data from a TaeKwonDo form, and comparing the automatically generated segmentation results with human psychophysical data, we found a good agreement between automatically generated segmentations and human performance. Furthermore, the segments agree with the minimum jerk hypothesis about human movement.<sup>32</sup> These results suggest that a similar approach might be useful for the decomposition of dances into primitive-like movement components, providing a new approach for the derivation of compressed descriptions of dances that is based on principles from biological motor control.

## 1 Introduction

Like choreographed dances, complex human full-body activities are comprised of sequences of individual actions. For purposes of teaching and memorization, different often heuristically motivated methods for the abstract notation of such movement sequences have been proposed, e.g. in terms of dance step diagrams or schemes for the execution of forms in martial arts, e.g. katas in Karate, or Hyeongs and Taegueks

in TaeKwonDo. Viewed from a computational perspective, these diagrams are compressed versions of the movements, which have to be decompressed by the actors or dancers during execution. This decompression is possible because the dancers are able to 'fill in the blanks' between subsequent foot positions in a dance step diagram using their own motor repertoire.

A variety of techniques for the modeling of complex human behavior sequences has been proposed in computer science, and a review would exceed the scope of this chapter. Instead we present here several examples from our work where we try to exploit biological concepts to derive mathematical models for such complex human behaviors.

According to a prevalent hypothesis in human motor control, complex coordination patterns within individual movements are organized in terms of movement primitives, i.e. simplified control elements, which can be combined in space and time to whole classes of complex movements. In the biological literature a variety of algorithms have been proposed to estimate such primitives from kinematic or EMG data.<sup>8,31</sup>

In section 2.1, we will review several popular definitions of movement primitives (MP). As an example of a state-of-the art algorithm for the unsupervised extraction of MPs from motion capture data, we describe the Fourier-based anechoic demixing (FADA) algorithm<sup>16</sup> in section 2.2 and show that this algorithm outperforms other learning techniques. In addition, we present in section 2.3 a Bayesian approach for the estimation of the model type and optimal number of primitives.<sup>24</sup> This number is also called the *model order*. We demonstrate that our Bayesian approach results in better model type and order estimates than previously applied schemes.

Most MP extraction methods require a previous segmentation of action streams in individual movements that can be characterized by individually controlled actions (periodic or non-periodic). Therefore, an important question is how to determine such segments from longer action sequences. In section 3, we discuss an efficient probabilistic methods for automatic segmentation. The developed unsupervised segmentation method is based on Bayesian binning (BB), an algorithm that models a longer data stream by the concatenation of an optimal number of segments, at the same time estimating the optimal temporal boundaries between those segments. Applying this algorithm to motion capture data from a TaeKwonDo Taeguek, and comparing the automatically generated segmentation results with human psychophysical data, we found a good agreement between automatically generated segmentations and human performance.<sup>25,26</sup> This was in particular the case when the joint angle trajectories within the segments were modeled by polynomials of order four. This order is consistent with optimal control-based theories of human movements<sup>32</sup> that have been validated in many previous experiments. In particular, these polynomials minimize integrated jerk for given endpoint constraints. Intuitively, this results in movements being as smooth as possible within each segment.

To illustrate that our segmentation approach might be useful for the compressed representation of movements, we create a movement diagram for the *Taeguek Pal-Chang* and compare it to a traditional diagram used in TaeKwonDo teaching (see fig. (4)). Our results suggest that a similar approach could also be useful for the de-

composition of dances into primitive-like movement components, serving as a new form of data-driven method for the derivation of compressed dance descriptions.

## 2 Kinematic movement primitives: an overview

### 2.1 Definitions of movement primitives

A long-standing hypothesis in the neuroscience community is that the central nervous system (CNS) generates complex motor behaviors by combining a small number of stereotyped components, also referred to as muscle synergies or movement primitives.<sup>8,31</sup> Such components consist of movement variables, such as joint trajectories<sup>44,67</sup> or muscle activations<sup>14,18,19,42</sup> that are activated synergistically over time.

Different conceptual definitions of movement primitives have been given in the literature, depending on the mathematical models used to factorize kinematic or electromyographic (EMG) data into different types of temporal, spatial, or spatio-temporal components. One classical definition of movement primitive is based on the idea that groups of degrees of freedom (dofs) might show instantaneous covariations, reflecting a coordinated recruitment of multiple muscles or joints. This implies the assumption that the ratios of the signals characterizing the different dofs remain constant over time. This type of movement primitive has been applied in particular in muscle space, where muscle synergies have been defined as weighted groups of muscle activations.<sup>12,72,73</sup> Such synergies have also been referred to as "synchronous" synergies, since the different muscles are assumed to be activated synchronously without time delays between different muscles. An alternative way to characterize movement primitives is based on the idea that they express invariance across time, so that they can be expressed as basic temporal patterns, defined by functions of time that are combined or superposed in order to reconstruct the movement signals (EMG signals or joint angles). Temporal components based on this definition have been identified in kinematic,<sup>6,17,44</sup> dynamic<sup>71</sup> and EMG signal space.<sup>13,42,43</sup> "Time-varying synergies"<sup>18–20</sup> have been described instead as spatio-temporal patterns of muscle activation, with the EMG output specified by the amplitude and time lag of the recruitment of each synergy. More recently, also models based on the combinations of other definitions have been proposed. For example, Delis and colleagues<sup>21</sup> defined space-by-time EMG organization patterns, where EMG activation patterns are obtained by mixtures of both temporal and synchronous synergies.

## 2.2 Unsupervised learning techniques for the identification of movement primitives

In the literature, a variety of unsupervised learning methods have been used for the identification of movement primitives from experimental data sets. This includes well-known classical unsupervised learning techniques based on instantaneous mixture models, such as principal component analysis (PCA) and independent component analysis (ICA),<sup>14,22</sup> and also more advanced techniques that include, for instance, the estimation of temporal delays of the relevant mixture components. An example is the work by D’Avella and colleagues,<sup>18,19</sup> who extended the classic non-negative matrix factorization (NMF) algorithm introduced by Lee and Seung<sup>50</sup> to identify spatiotemporal EMG synergies. Omlor and Giese<sup>58–60</sup> developed a new algorithm based on the Wigner-Ville Transform for the extraction of time-shifted temporal components that is based on an anechoic mixture model (1), as used in acoustics for the modeling of acoustic mixtures in reverberation-free rooms.<sup>9,23,79</sup> This model assumes that a set of  $N_s$  recorded acoustic signals  $x_i, i = 1, 2, \dots, N_s$ , is caused by the superposition of  $N$  acoustic source functions (signals)  $s_j(t)$ , where time-shifted versions of these source functions are linearly superposed with the mixing weights  $a_{ij}$ . The time shifts are given by the time delays  $\tau_{ij}$ , and in the acoustical model are determined by the traveling times of the signals. The model has the following mathematical form:

$$x_i(t) = \sum_{j=1}^N a_{ij} s_j(t - \tau_{ij}) \quad (1)$$

For the special case that  $\tau_{ij} = 0$  for all pairs  $(i, j)$ , this model (1) coincides with the classical linear combination models underlying PCA and ICA. To analyze kinematic data associated with a pointing task accomplished during locomotion, and inspired by the previous work by Omlor and Giese,<sup>58,59</sup> we<sup>16</sup> developed recently a new algorithm (Fourier-based Anechoic Demixing Algorithm, FADA), that is based on the same generative model (1), but includes additional smoothness priors for the identified functions. The introduction of such priors is justified because EMG or kinematic data from motor tasks usually have limited band-width, and it substantially improves the robustness of the estimation method. Band limited source functions in (1) can be approximated by a truncated Fourier series of the form:

$$x_i(t) \cong \sum_{k=-M}^M c_{ik} e^{ikt} \quad (2)$$

and

$$s_j(t - \tau_{ij}) \cong \sum_{k=-M}^M v_{jk} e^{-ik\tau_{ij}} e^{ikt} \quad (3)$$

$M$  being a positive integer which is determined by Shannon's theorem according to the limit frequency of the signals. The symbol  $\iota$  signifies the imaginary unit and  $c_{ik}$  and  $v_{jk}$  complex numbers ( $c_{ik} = |c_{ik}| e^{\iota\phi_{c_{ik}}}$  and  $v_{jk} = |v_{jk}| e^{\iota\phi_{v_{jk}}}$ ). Substituting (2) and (3) in (1), and assuming uncorrelatedness of the sources  $s_j(t)$  as it was in other previous works,<sup>58,59</sup> the following iterative algorithm can be derived for the identification of the unknown parameters in model (1):

After a random initialization of the estimated parameters, the following steps are carried out until convergence:

1. Compute the absolute values of the coefficients  $c_{ik}$  and solve the following positive demixing problem using positive ICA or non-negative matrix factorization:

$$|c_{ik}|^2 = \sum_{j=1}^N |a_{ij}|^2 |v_{jk}|^2 \quad (4)$$

with  $i = 0, 1, \dots, N_s$  and  $k = 0, 1, \dots, M$ .  $N$  is the number of sources. Since the signals are real the Fourier coefficients equations (2) and (3) for positive and negative indices  $k$  are complex conjugates of each other. For this reason it is sufficient to solve the demixing problem by considering only the coefficients with indices  $k \geq 0$ . For the shown implementation we used non-negative independent component analysis<sup>40</sup> for solving the underlying demixing problem with non-negative components.

2. Initialize  $\phi_{v_{jk}} = 0$  for all pairs  $(j, k)$  and iterate the following steps:
  - a. Update the phases of the Fourier coefficients of the sources, which are defined by the identity  $\phi_{v_{jk}} = \text{angle}(v_{jk}) = \arctan(\text{Im}(v_{jk})/\text{Re}(v_{jk}))$  by solving the following non-linear least square problem

$$\min_{\Phi} \|\mathbf{C} - \mathbf{V}\|_F^2 \quad (5)$$

where  $(\mathbf{C})_{ik} = c_{ik}$ ,  $(\mathbf{V})_{ik} = \sum_{j=1}^N a_{ij} e^{-\iota k \tau_{ij}} |v_{jk}| e^{\iota \phi_{v_{jk}}}$  and  $\Phi_{jk} = \phi_{v_{jk}}$ .  $\|\cdot\|_F$  indicates the Frobenius norm.

- b. Assuming that the source functions  $s_j(t)$ , defined by the parameters  $v_{jk}$  are known, the mixing weights  $a_{ij}$  and the delays  $\tau_{ij}$  can be optimized for each signal  $x_i$  by minimization of the following cost function:

$$\left[ \hat{\mathbf{a}}, \hat{\mathbf{t}} \right] = \arg \min_{\mathbf{a}, \mathbf{t}} \|x_i(t) - \mathbf{s}(t, \mathbf{t})' \mathbf{a}\|_F^2 \quad (6)$$

Optimization with respect to  $\mathbf{a}$  and  $\mathbf{t}$  is feasible, assuming uncorrelatedness of the sources and independence of the time delays.<sup>70</sup> The column vector  $\mathbf{a}$  concatenates all weights associated with dof  $i$ , i.e.  $\mathbf{a} = [a_{i1}, \dots, a_{iN}]'$ . The vector function  $\mathbf{s}(t, \mathbf{t}) = [s_1(t - \tau_{i1}), \dots, s_N(t - \tau_{iN})]'$  concatenates source functions associated with dof  $i$ , shifted by the associated time delays.

By the approximation of the signals by a truncated Fourier series, compared to more general algorithms, the FADA algorithm has a substantially smaller number of

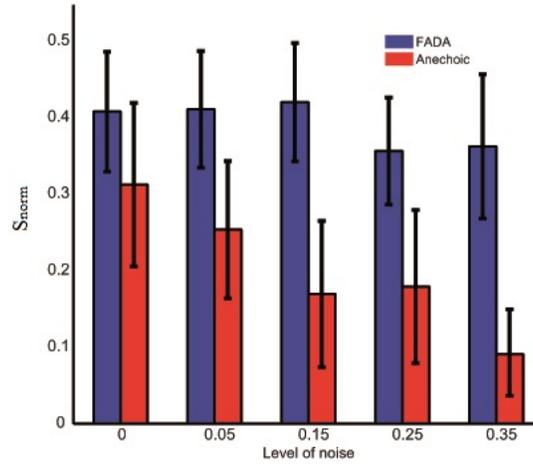
free parameters that have to be estimated. This makes the algorithm for cases where the underlying assumption about the frequency spectrum of the relevant signals are fulfilled more efficient and robust than more general algorithms (such as general anechoic demixing algorithms<sup>58,59</sup> or classic PCA and ICA applied to temporal data<sup>22,42</sup>). Since the underlying optimization problems have fewer local minima, convergence of the algorithm is faster and it is less prone to be trapped in irrelevant local minima. We have confirmed these properties in extensive simulations<sup>15</sup> using synthetic ground-truth data that was derived from known generative models.

For example, the FADA algorithm performed better than other methods when identifying anechoic primitives (fig. 1). In order to evaluate the performance of different algorithms, artificial kinematic data sets were simulated, using the generative model defined by equation (1). The source functions were generated by filtering white noise with a Butterworth filter with a cut off frequency that mimicked the spectrum of real kinematic data. The values of the mixing weights and the time delays were also drawn from uniform distributions over fixed intervals. For the simulations the number of sources was set to four, and the number of simulated signals was 250 for each data set. The data sets were also corrupted with signal-dependent noise drawn from a Gaussian distribution of variance  $\sigma = |\alpha x_i(t)|$ .<sup>37</sup> The scaling parameter  $\alpha$  was adjusted in order to adjust the correlation between the unperturbed data and the perturbed data, realizing the values  $1 - R^2$  equal to 0.05, 0.15, 0.25 and 0.35. The similarity (S) between original and identified primitives (source functions) was quantified by computing the maximum of the scalar products between original and recovered primitive (over all possible time delays).

### 2.3 Model selection criteria

Different established approaches for the extraction of movement primitives from trajectory and EMG data differ, on the one hand, by the type of generative model that is used (e.g. instantaneous mixtures vs. models containing time delays). On the other hand they also can differ in terms of the number of model parameters, e.g. the number of primitives or source functions in the mixture model. The number of primitives is also called the *model order*. To our knowledge only very few motor control studies have so far addressed the problem of model selection in a principled way, see e.g. Delis and colleagues<sup>21</sup> and Hart and Giszter<sup>38</sup> for notable exceptions. The existing generative models for the extraction of movement primitives have indeed been demonstrated to provide a low-dimensional decomposition of the experimental data, but no clear criterion has been developed to objectively determine which model is best suited for describing the statistical properties of the data under investigation.

In the field of machine learning various methods for the optimization of model complexity have been developed, either using heuristic approaches or methods derived from Bayesian inference. The well-known Akaike Information criterion (AIC) and Bayesian Information Criterion (BIC) have the advantage of being easy to use when a likelihood function for a given model is available. Hence, they are often the



**Fig. 1** The figure shows the average level of the similarity between actual and identified anechoic primitives for different levels of signal dependent noise. The primitives were estimated from artificial ground-truth data sets with the anechoic demixing algorithm by Omlor and Giese<sup>58,59</sup> and the new algorithm FADA.<sup>16</sup> The similarity values are normalized according to the formula  $S_{\text{norm}} = (S - S_b) / (1 - S_b)$ , where  $S_b$  indicates the baseline value that was obtained by assessing the average similarity between the randomly generated source functions. The value  $S_{\text{norm}} = 0$  corresponds to chance level similarity, and the maximum value of the normalized similarity is one.

first choice for model order estimation, but not necessarily the best one. In the work by Tu and Xu<sup>74</sup> several criteria for probabilistic PCA (or factor analysis) models were evaluated, including AIC, BIC, MIBS<sup>54</sup> (Minka’s Bayesian model selection) and Bayesian Ying-Yang.<sup>78</sup> The authors found that MIBS and Bayesian Ying-Yang work best. AIC and BIC criterion have also been used to estimate the number of independent components in fMRI data. This was done for instance by Li and colleagues<sup>52</sup> that, however, found AIC and BIC estimation performance to be adversely dependent on temporal correlations between signals. Other heuristic methods have been used on the literature for model order selection. Such approaches typically utilize some features of the reconstruction error (or conversely, of the variance-accounted-for (VAF)) as a function of the model order. For instance, the usual procedure is to search for a ”knee” in that function, a procedure which is inspired by the scree test for factor analysis.<sup>11</sup> For example, multiple authors<sup>12,18,42</sup> used them to determine the number of EMG synergies underlying different human behaviors.

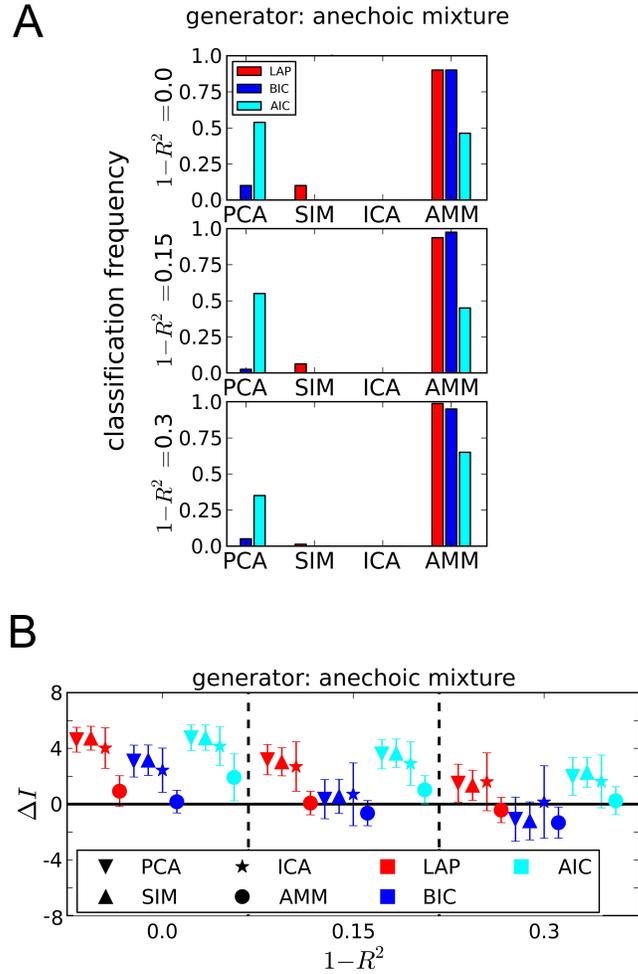
To improve the accuracy provided by standard Bayesian estimator, we developed a new objective criterion (which we called *LAP*) in the framework of Bayesian generative model comparison<sup>7</sup> for model-order selection that extends the other classical ones based on information-theoretic and statistical approaches. The criterion is based on a Laplace approximation of the posterior distribution of the parameters of a given blind source separation method, re-formulated in the framework of Bayesian generative model comparison.<sup>7</sup> Exploiting the Laplace approximation allowed us to approximate some intractable integrals appearing in the computation of the marginal

likelihood of the models that, after the approximation, assumed the following form:

$$\begin{aligned}
 p(D|\Phi, M) \approx & \underbrace{\log(p(D|\Theta^*, \Phi, M))}_{\text{log-likelihood}} + \underbrace{\log(p(\Theta^*|\Phi, M))}_{\text{log-prior}} \\
 & + \underbrace{\frac{\dim(\Theta)}{2} \log(2\pi) - \frac{1}{2} \log(|\mathbf{H}|)}_{\text{log-posterior-volume}} \quad (7)
 \end{aligned}$$

where  $D$  indicates the observable data,  $\Theta_M$  is a tuple of model parameters for a model indexed by  $M$  (the 'model index') and  $\Phi$  indicates a tuple of hyperparameters. In addition,  $\Theta^*$  is a tuple of model parameters that maximize the log-likelihood subject to the regularization provided by the parameter prior and  $\mathbf{H}$  is the Hessian matrix (second derivatives of the log-posterior=log-likelihood + log-prior at  $\Theta^*$ )<sup>24</sup>. Equation (7) comprises three parts, which can be interpreted. The first term the log-likelihood measures the goodness of fit, similar to explained variance or VAF. The second term is the logarithm of the prior, which corresponds to a regularization term for dealing with under-constrained solutions for  $\Theta$  when the data set is small. Finally, the third part measures the volume of the parameter posterior, since  $\mathbf{H}$  is the posterior precision matrix (inverse covariance) of the parameters in the vicinity of  $\Theta^*$ , i.e. it indicates how well the data constrain the parameters. Given different models, to discriminate the most suitable one to describe the available data the criterion requires to compute the values of the model evidence (7) associated with each model and to choose the one which maximizes this evidence as the most appropriate model. With a similar procedure, it is also possible to identify, given a specific model, the right model order. This can be done by computing the model evidences associated with different model orders and by taking the order associated with the highest value as the best one to represent the data.

We showed in our previous work<sup>24</sup> how the *LAP* is more reliable than other already existing classical criteria in selecting the generative model underlying a given data set, as well as in determining the best model order. The criterion performance was evaluated on synthesized data and compared to the performance provided by AIC and BIC. Some results adapted from the work of Endres and colleagues<sup>24</sup> are reported in fig. (2). Panel A shows that the generating model, here an anechoic mixture, is correctly identified by *LAP* with near certainty for all three testes noise levels. In panel B, we plotted the difference  $\Delta I$  between estimated and true model order. *LAP* provides the best estimates across noise levels, if the correct model type (AMM) is used for the estimation procedure.



**Fig. 2** Panel A shows the model classification performance provided by different model selection criteria (AIC, BIC and LAP) when applied to artificial data sets (based on model (1) and generated as explained in the previous subsection). The data is corrupted by three different levels of noise. Four different models were tested for each data set, with PCA indicating the probabilistic Principal Component Analysis linear model, Smooth Instantaneous Model (SIM), Independent Component Analysis (ICA) model and Anechoic Mixture Model (AMM). The model selection criteria BIC and LAP both performed well. The AIC criterion frequently confuses PCA and AMM. In panel B the estimated number of sources provided by the criteria are displayed.  $\Delta I$  indicates the difference between the estimated and actual number of primitives. Symbol shapes stand for analysis algorithm, colors indicate the selection criteria. If AMM was used for analysis, BIC and LAP provided similar estimations for the number of primitives. AIC tended to overestimate the number of primitives. For incorrect analysis model, all criteria provided a higher number of sources.

### 3 Generating compressed movement descriptions with Bayesian binning

#### 3.1 Objective and related approaches

We now turn to the problem of extracting individual actions from longer streams of activities. This problem is particularly important in dance, where a dance is typically choreographed by concatenating a sequence of individual dance movements. In addition, the approaches described in the last section typically assume previous segmentation into elementary actions, which then can be modeled by superposition of a set of control primitives. This raises the problem of an automatic segmentation of natural action streams, which has been a central problem in movement analysis and modeling for several decades.

One of the earliest approaches achieved automatic segmentation into piecewise linear functions,<sup>5</sup> employing dynamic programming for the solution of the resulting least-squares trajectory fitting problem. There, it is also mentioned that the approach might be extended to arbitrary polynomial orders, but exactly how is not shown. An online version of this algorithm for the piecewise linear model was later developed,<sup>45</sup> it was subsequently extended to work with polynomials of higher order.<sup>51</sup> The intended application domain is data mining, where online-capable algorithms are required, since the databases are typically too large to fit into working memory. Recently, real-time versions of these algorithms were derived which achieve the necessary speed-up by using orthogonal polynomials as basis functions for segment contents.<sup>33,34</sup>

Most of these approaches use heuristics for the determination of the polynomial order (or leave it up to the user to choose one). We developed an exact Bayesian approach addressing the model complexity issue.<sup>25</sup> One might wonder why polynomials are popular segment models, as opposed to the other movement primitive models described above. The reason for this is twofold: first, polynomials are mathematically convenient and well understood. Second, trajectories that minimize integrated jerk are polynomials of (at most) order five,<sup>32</sup> and human movement production seems to implement this optimality principle in many instances. Thus, B-splines have found applications in humanoid robotics, for example for human-robot transfer<sup>75</sup> and goal-directed movement synthesis.<sup>76</sup>

Important non-polynomial action segmentation methods are based on hidden Markov models (HMM)<sup>36,48</sup> (see also Liu et al.'s contribution<sup>53</sup> in this book), change-of-contact events,<sup>77</sup> Gaussian mixtures,<sup>3</sup> and probabilistic principal component analysis (pPCA) approaches.<sup>3</sup> Because pPCA encodes a full covariance model, it has the potential advantage of handling correlated sensor noise better than models with a polynomial time dependence of the mean and an isotropic noise assumption. We showed<sup>25</sup> how these two approaches can be combined.

However, movement primitive extraction is only one application where action segmentation is interesting: many methods for movement synthesis in computer graphics generate individual segments by (parametric) interpolation between exam-

ple trajectories, e.g. by PCA<sup>41,66</sup> or the classical verb-adverbs approach,<sup>64</sup> which uses B-splines. For these methods to work, accurately pre-segmented data are required. Similarly, motion graphs<sup>46,65</sup> or motion database sequencing methods<sup>2</sup> give most convincing results when segment boundaries are consistent with human perception.

Finally, in Psychology and Neuroscience, models for human segmentation performance are expected to provide information about the structure of action representation in the brain, since human brain activity appears to be time-locked to perceived event boundaries.<sup>80</sup> Thus, the segmentation of sequences of piecewise linear movements in the two-dimensional plane was studied.<sup>1,69</sup> Polyakov et. al<sup>61</sup> fitted monkey scribbling trajectories with parabolic segments, which minimize jerk<sup>32</sup> and comply with the two-thirds power law.<sup>49</sup> They determined that neural signal changes correlate with the segment boundaries thus computed. We compared 3D action segmentation by human observers to polynomial segments computed with Bayesian Binning (BB),<sup>25</sup> finding a good correspondence between 4th-order polynomials (which minimize jerk) and human observers.

In the next section, we describe the basic idea behind BB and apply it to the segmentation of a TaeKwonDo Taeguek. A Taeguek is a stylized martial arts action sequence, in which the performer fights against a number of imaginary opponents. For teaching and memorization purposes, these Taeguek can be described by diagrams of key poses<sup>i</sup> similar to the dance steps diagrams used e.g. in ballroom dancing<sup>10</sup>. This kind of description can be 'decompressed' by a martial artist if he has learned the individual techniques that correspond to the key poses. Our goal is to automatically extract a set of key poses that succinctly describe the Taeguek with BB.

## 3.2 Bayesian binning

Bayesian Binning (BB) is a method for modeling data by piecewise defined functions. It can be applied if the data have a totally ordered structure, such as a time series. Since it is an (exact) Bayesian approach, it allows for an automatic control of model complexity. In this context, this means that the number of segments (bins), their length and the model for segment contents (the movement primitives) are determined with minimal user intervention. Originally, BB was developed for probability density estimation of neural recordings and their information theoretic evaluation.<sup>27</sup> Later, it was extended for regression of piecewise constant functions<sup>39</sup> and further applications in neuroscience.<sup>28,29</sup> A similar Bayesian formalism for dealing with multiple change point problems was concurrently developed.<sup>30</sup>

In the following subsection, we describe BB in terms of a probabilistic graphical model. For a mathematical treatment of the polynomial movement primitive model, we refer the reader to a previous publication<sup>25</sup>. We also forgo developing the algo-

---

<sup>i</sup> The diagrams for all Taegueks can e.g. be viewed on [www.taekwodo.de](http://www.taekwodo.de)

rhythm in detail, since we did that elsewhere.<sup>27</sup> BB operates in discrete time, hence the time axis (see fig. (3), B) is discretized into intervals of duration  $\Delta$ .  $\Delta$  has to be small enough so that all relevant features of the data do not change noticeably within such an interval. A natural lower bound on  $\Delta$  is given by the time resolution of the data recording equipment. The VICON motion capture system which was used for the TaeKwonDo recordings had a frame-rate of 120 Hz, hence  $\Delta \geq \frac{1}{120\text{Hz}} \approx 8.3\text{ms}$ . The intervals are labeled by a discrete time index  $t$ , which runs from 0 to  $T$ . BB concatenates these intervals into contiguous, non-overlapping segments. In the example shown in Fig. 3, C, there are two segments: the first one extends from  $t = 0$  to  $t = i$  (inclusive), the second one comprises time steps  $t = i + 1$  to  $t = T - 1$ . In Fig. 3, B, the segment boundaries are indicated by the orange lines (there is an implicit boundary at the end,  $t = T$ ).

Suppose we measured a time series of data points  $D_t$  consisting of three-dimensional joint angles for  $Q$  joints, i.e.  $D_t = (x_t^0, \dots, x_t^{3Q})$ . BB makes two central modeling assumptions about such data: (1) within a segment, the parameters  $\Theta_t$  of the data generating model do not change, and (2) these models are independent across segments. Hence, the joint probability (density) of the data  $(D_i, \dots, D_j) = D_{i:j}$ , that is  $P(D_{0:T-1} | \Theta_{0:T-1})$  factorizes across segments, in our example:

$$P(D_{0:T-1} | \Theta_{0:T-1}) = P(D_{0:i} | \Theta_0) P(D_{i+1:T-1} | \Theta_{i+1}). \quad (8)$$

This factorization property, combined with the total order of time points, facilitates efficient evaluation of expectations of segmentation point locations and segment parameters. To understand why, consider the graphical model of BB in Fig. 3, A, where we use standard graphical modeling notation:<sup>7</sup> circles are random variables, lines with arrows denote conditional dependencies, and dashed rectangles are gates,<sup>55</sup> each of which being controlled by a binary gating variable  $b_t$  with a Bernoulli prior<sup>ii</sup>. Depending on the value of this variable, either one or the other alternative part of the model is instantiated. Here, if  $b_t = 0$ , then the parameters  $\Theta_t$  for time step  $t$  are simply copied from the previous time-step. In our example, this is the case for all  $t = 1, \dots, i$ . On the other hand, if  $b_t = 1$ , then the corresponding  $\Theta_t$  is drawn from a suitably chosen prior distribution (e.g. at  $t = i + 1$ ). This algorithm for determining the parameters effectively concatenates time steps into contiguous, non-overlapping segments. Note that the graphical model is *singly connected*: there is at most one path between any two circles, if one travels along the lines. Hence, the sum-product algorithm<sup>47</sup> can be applied for the efficient and exact evaluation of expectations of interest, if conjugate priors are chosen for the (continuous) parameters<sup>iii</sup>. Readers with a machine learning background may observe the similarity of Fig. 3, A to a HMM,<sup>4,62</sup> which is probably the most well-known example of a singly-connected

<sup>ii</sup> The a-priori independent gating variables and their Bernoulli priors induce a Binomial prior on the number of segments, which is a special case of the general priors on segment number boundaries which we developed previously.<sup>27</sup> The latter need a dependency model between the gating variables, which we do not consider here for the sake of simplicity.

<sup>iii</sup> Strictly speaking, any priors that allow for an evaluation of posterior expectations in closed form are suitable, but conjugate priors are particularly convenient

graphical model and allows for efficient computation of marginal expectations for the same reason.

The observation model for each time step, given by eqns. (9)-(11), is a  $Q$ -dimensional multivariate Gaussian. It is defined by the parameters  $\Theta_i = (z_i, \theta_i^{0,0}, \dots, \theta^{3Q,S})$ , that encode a  $S$ -th order polynomial time dependence of the mean and a  $(3Q \times 3Q)$  covariance matrix  $\Sigma_i$ . The variable  $z_i$  specifies the initial time step of the current segment:

$$\mu_i^q = \sum_{s=0}^S \theta_i^{q,s} (i - z_i)^s \quad (9)$$

$$\mu_i = (\mu_i^0, \dots, \mu_i^{3Q}) \quad (10)$$

$$P(D_i | \Theta_i) = \mathcal{N}(\mu_i, \Sigma_i) \quad (11)$$

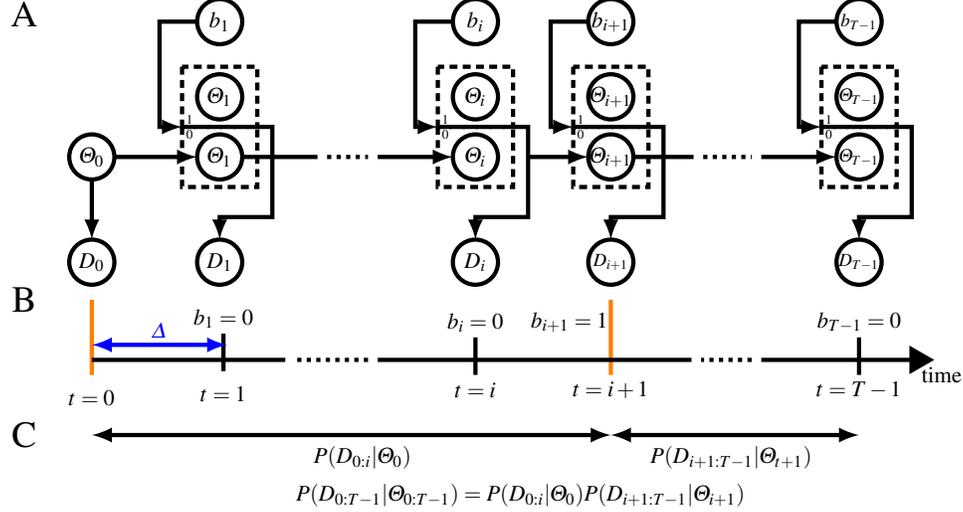
We showed<sup>25</sup> how to construct a conjugate Gauss-Wishart prior to this observation model (eqns. 9-11), enabling sum-product message passing.

### 3.3 TaeKwonDo Taeguek segmentation

#### 3.3.1 Data recording

The action streams we studied are TaeKwonDo Taegueks carried out by internationally successful martial artists. Each artist performed the same fixed sequence of 27 kicks and punches of the Taeguek Pal-Chang. The complete form has a full length of about 40 seconds. We obtained kinematic data by motion capture using a VICON 612 system with 11 cameras. This setup yields 3D positions of 41 passively reflecting markers attached to the performers' joints and limbs with a 3D reconstruction error below 1 mm and at a sampling frequency of 120 Hz.

We used the kinematic data for two purposes. First, we computed joint angles trajectories from a hierarchical kinematic body model (skeleton) which we fitted to the original 3D marker positions. This yielded Euler angles describing the rotations between adjacent skeleton segments in terms of flexion, abduction and rotations about the connecting joint (e.g.<sup>57,63</sup>). Second, from the derived joint angle trajectories we created movie clips showing computer animations of the TaeKwonDo movements. Those videos served as stimuli in a psychophysical experiment to obtain segmentation boundaries according to human perception. For this experiment, we split the motion capture data of each Taeguek manually into sub-sequences of comparable length each containing between three and eight separable TaeKwonDo moves. We presented the videos of these sub-sequences to naïve human observers and asked them to segment the videos by pressing a key. We computed a segmentation density from the key-press timings pooled across all observers, and identified peaks in this density. We refer to the peak locations as 'population-averaged human segment boundaries'. For details of the experiment and evaluation procedures, see.<sup>25,26</sup>



**Fig. 3** **A**: the graphical model of Bayesian Binning. We follow standard graphical modeling terminology<sup>7</sup>: circles represent random variables, arrows denote conditional dependencies, and dashed boxes are gates<sup>55</sup>. The observable joint angle data  $D_i$  are modeled by a (latent) movement primitive model with parameters  $\Theta_i$ . Subscripts denote discrete time steps. Presence or absence of a segment boundary at a time step  $i$  is indicated by a binary gating variable  $b_i$ : if  $b_i = 1$ , then the corresponding gate below it is active, which means that a new set of parameters  $\Theta_i$  is instantiated at this time step. Otherwise, if  $b_i = 0$ ,  $\Theta_i$  is a copy of the parameters of the previous time step. The graph is singly connected, hence marginal expectations can be computed efficiently with sum-product message passing<sup>47</sup>. For details, see text. **B**: the time axis is discretized into non-overlapping, contiguous intervals of duration  $\Delta$  small enough so that all relevant features of the data are preserved. These intervals are labeled with an integer time index  $t$ . There are  $T$  such intervals, hence  $t \in \{0, \dots, T-1\}$ . Segment boundaries are effectively controlled by the gating variables  $b_t$ : in this example, the interval from 0 to  $T\Delta$  is subdivided into two segments, as indicated by the orange segment boundaries at  $t =$  and  $t = i + 1$ . **C**: BB models the time series by contiguous, non-overlapping segments, which are assumed to be independent. With the gating variable setting shown in panel **A** (i.e. only  $b_{i+1} = 1$ , all others are 0), the joint probability of the data given the parameters  $P(D_{0:T-1}|\Theta_{0:T-1})$  factorizes into two contributions:  $P(D_{0:i}|\Theta_0)$  for time steps  $0, \dots, i$  and  $P(D_{i+1:T-1}|\Theta_{i+1})$  for time steps  $i + t, \dots, T - 1$ .

### 3.3.2 Segmentation analysis

We analyzed the joint angle trajectories with BB, as described above. To determine the polynomial order which best matches human perception, we conducted a hit rate analysis:<sup>25</sup> to determine 'hits', we counted how often a BB segment boundary was placed in close temporal proximity ( $\pm 250$ ms) to a population-averaged human segment boundary. All BB segment boundaries that could not be matched to a human one were counted as 'false positives'. The best compromise between hits and false positives is reached at order 4, which agrees with the minimum jerk hypothesis for human movement generation<sup>32</sup>. Put differently, our results indicate that minimum jerk is not only a movement production principle, but also perceptually relevant.

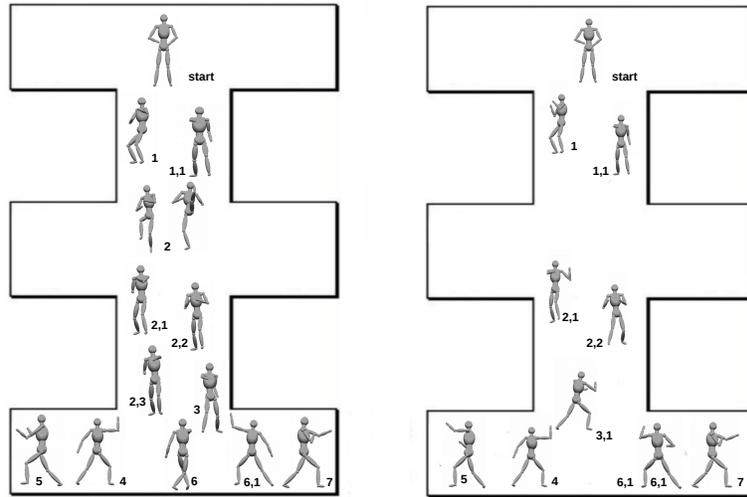
Furthermore, the hit rate analysis also revealed that naïve observers' behavior is best explained by the segment boundaries computed from shoulder and elbow joints taken together. If we were to repeat this analysis with skilled TaeKwonDo martial artists as observers, we would expect to find that the leg joints need to be included in the analysis as well, since kicks are techniques of significant importance in this martial art. This fact, however, was not picked up by the naïve observers in our experiment.

To generate a compressed representation of the Taeguek, we compute the performer's pose at each segment boundary. The results, for the first part of the Taeguek, are shown in fig. (4). For visual comparison with TaeKwonDo teaching diagram (fig. (4), left), we arranged the poses in the same fashion on the right of this figure. Poses match, except for the kicks (pose 2 on the left), pose 2,3 on the left and the transition (pose 3,1) on the right. This indicates that the compressed description generated by BB may be useful for semi-automatic movement description. We attribute the missing kicks to the fact that we segmented using only arm joints. The naïve observers we tested did not segment at the kicks, too.

## 4 Discussion

In this chapter we have summarized some of our recent work that applies biologically-inspired learning-based models for the modeling of complex human activities. At the level of individual actions, we proposed a new efficient algorithm for anechoic demixing that outperforms other approaches for the learning of movement primitives from trajectory and EMG data. In addition we presented how Bayesian inference can be exploited for the selection of the correct generative model and specifically the order of the model. We have shown elsewhere that movement primitives based on anechoic demixing are also suitable for the online synthesis of movements and the embedding in control architectures. For this purpose the learned source function are mapped onto dynamic primitives that are formulated as dynamical systems.<sup>35,56</sup> At the level of sequences of actions within complex activities, we proposed a new probabilistic method for segmentation that is based on Bayesian Binning, and which applied exact Bayesian inference for determining an optimal segmentation of the action stream into subsequent component actions.

We described an anechoic demixing algorithm in section 2.1. It subsumes as special cases multiple other existing methods for the estimation of movement primitives and can be combined with further constraints on weights and sources, e.g. positivity. In this way, the new model provides a unifying framework for the study of the relationship between different mathematical models for the extraction of movement primitives from data.<sup>15</sup> The investigation of movement primitives, extracted as components of kinematic (joint angle) data seems interesting for dance, on the one hand to characterize the full-body coordination patterns, e.g. between locomotion and upper body movements in dance. On the other hand, such primitives might



**Fig. 4** **Left:** stylized representation of the first part of the movements that comprise the *Taeguek Pal-Chang*, a solo TaeKwonDo form. Only key poses are shown, a trained martial artist can fill in the movements in between these poses using their own motor repertoire. The bounding box represents the space within which the martial artist is supposed to move during the Taeguek. Pose numbering corresponds to the numbering found in the diagrams on [www.taekwondo.de](http://www.taekwondo.de). **Right:** key poses determined with Bayesian binning employing a 4th order polynomial observation model. This observation model minimizes jerk, and provides the best match to segmentation points determined by naïve human observers, if elbow and shoulder joints are used for the segmentation. Numbering of poses is the same as in the left panel. Pose 2,3 is missing here, as are the kicks (number 2 on the left), which could not be decoded from arm movements. Pose 3,1, which is a transition between poses 3 and 4, does not appear in the Taeguek image. All other poses match.

also be helpful to understand 'co-articulations', that is overlaps between subsequent coordination patterns that are important for the smooth flow of motion in dance.

With respect to the method for the segmentation of action streams exploiting Bayesian binning in section 3, our work shows that the obtained segmentation is close to the one provided by humans. However, for other types of movements likely additional features beyond the ones characterizing the arm movements would have to be included. Determining optimal sets of such features from data could also be realized by Bayesian model comparison.

It would also be interesting to investigate if the actor's poses at the segmentation boundaries are a sufficiently rich description of the movement so that human observers can 'fill in blanks' between two subsequent boundaries using their own motor repertoire. In that case, Bayesian binning could be used to derive compressed movement representations similar to dance step diagrams.

As a next step, it would be interesting to investigate in how far the obtained compressed representation is suitable for movement production with modular movement primitives. For this purpose, the general polynomial models within the individual segments would have to be replaced by models that are based on primitives of the type discussed in the first part of this chapter.

The modeling assumptions made by BB presuppose that human actions can be viewed as a sequence of clearly separable primitives in time, see eqn.(8) . While the match between 4th order polynomial BB and naïve human perception indicate that this presupposition is at least approximately fulfilled for a TaeKwonDo Taeguek, it seems likely that this assumption is violated for other types of human action, such as dancing. There, the transitions from one movement to the next are typically more continuous than in TaeKwonDo. Between-segment dependencies would also be relevant in sign language production, where co-articulation between neighboring signs has been observed.<sup>68</sup> The BB approach can be extended to include dependencies between segments, as long as the graphical model (see fig. 3, A) remains singly connected. Exact inference will then still be possible, albeit with a higher computational effort. Another way of extending our approach would be to include task or context information into the segmentation process,<sup>77</sup> to supplement the purely kinematic information which we currently use. Humans use such context information for segmentation when available,<sup>81</sup> and rely increasingly on kinematics when context is reduced. Here, again the advantage of the proposed probabilistic approach is that it can be connected easily to probabilistic representations of context and even to semantic representations, as long as they can be expressed in the language of graphical models.

Interesting for online synthesis is also the question in how far individual movements within individual segments can be derived from optimal control problems with boundary conditions derived from the task, refining the rather unspecific polynomial model used for the signal time-courses in the presented model. Ongoing work focuses on this question, aiming at finding optimized segmentations for specific pre-defined control architectures.

**Acknowledgements** The research leading to these results has received funding from the European Union under grant agreements Koroibot FP7-ICT-2013-10/ 611909, AMARSi- EC FP7-ICT-248311; FP7-PEOPLE-2011-ITN(Marie Curie): ABC PITN-GA-011-290011, 604102 (HBP), CogIMon H2020 ICT-23-2014 /644727, and from the DFG under grants GI 305/4-1, DFG GZ: KA 1258/15-1, and from BMBF grant FKZ: 01GQ1002A. DE has received support from the DFG under grant IRTG-GRK 1901 "The Brain in Action".

## References

1. Y. Agam and R. Sekuler. Geometric structure and chunking in reproduction of motion sequences. *Journal of Vision*, 8(1):1–12, 2008.
2. O. Arikan and D. A. Forsyth. Interactive motion generation from examples. *ACM Trans. Graph.*, 21:483–490, July 2002.

3. J. Barbič, A. Safonova, J.-Y. Pan, C. Faloutsos, J. K. Hodgins, and N. S. Pollard. Segmenting motion capture data into distinct behaviors. In *Proceedings of Graphics Interface 2004*, GI '04, pages 185–194, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2004. Canadian Human-Computer Communications Society.
4. L. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171.
5. R. Bellman. On the approximation of curves by line segments using dynamic programming. *Communications of the ACM*, 4(6):284–, 1961.
6. B. Berret, F. Bonnetblanc, C. Papaxanthis, and T. Pozzo. Modular control of pointing beyond arm’s length. *J Neurosci*, 29(1):191–205, Jan 2009.
7. C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2007.
8. E. Bizzi, V. C. K. Cheung, A. d’Avella, P. Saltiel, and M. Tresch. Combining modules for movement. *Brain Research Reviews*, 57(1):125–133, 2008.
9. P. Bofill. Underdetermined blind separation of delayed sound sources in the frequency domain. *Neurocomputing*, 55(34):627 – 641, 2003. Evolving Solution with Neural Networks.
10. P. Bottomer. Anness Publishing, London, UK, 2012.
11. R. B. Cattell. The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2):245–276, 1966.
12. V. C. K. Cheung, A. d’Avella, M. C. Tresch, and E. Bizzi. Central and sensory contributions to the activation and organization of muscle synergies during natural motor behaviors. *Journal of Neuroscience*, 25(27):6419–6434, 2005.
13. E. Chiovetto, B. Berret, I. Delis, S. Panzeri, and T. Pozzo. Investigating reduction of dimensionality during single-joint elbow movements: a case study on muscle synergies. *Frontiers in Computational Neuroscience*, 7:11, 2013.
14. E. Chiovetto, B. Berret, and T. Pozzo. Tri-dimensional and triphasic muscle organization of whole-body pointing movements. *Neuroscience*, 170(4):1223–1238, 2010.
15. E. Chiovetto, A. d’Avella, and M. A. Giese. A unifying algorithm for the identification of kinematic and electromyographic motor primitives. April 2013. not reviewed.
16. E. Chiovetto and M. A. Giese. Kinematics of the coordination of pointing during locomotion. *PLoS One*, 8(11):e79555, 2013.
17. E. Chiovetto, L. Patan, and T. Pozzo. Variant and invariant features characterizing natural and reverse whole-body pointing movements. *Experimental Brain Research*, 218(3):419–431, 2012.
18. A. d’Avella, A. Portone, L. Fernandez, and F. Lacquaniti. Control of fast-reaching movements by muscle synergy combinations. *Journal of Neuroscience*, 26(30):7791–7810, 2006.
19. A. d’Avella, P. Saltiel, and E. Bizzi. Combinations of muscle synergies in the construction of a natural motor behavior. *Nature Neuroscience*, 6(3):300–308, 2003.
20. A. d’Avella and M. C. Tresch. Modularity in the motor system: decomposition of muscle patterns as combinations of time-varying synergies. In S. A. S. Michael I. Jordan, Michael J. Kearns, editor, *Advances in Neural Information Processing Systems 14*, pages 141–148. MIT Press, Cambridge, MA, 2002.
21. I. Delis, S. Panzeri, T. Pozzo, and B. Berret. A unifying model of concurrent spatial and temporal modularity in muscle activity. *J Neurophysiol*, 111(3):675–693, Feb 2014.
22. N. Dominici, Y. P. Ivanenko, G. Cappellini, A. d’Avella, V. Mond, M. Cicchese, A. Fabiano, T. Silei, A. Di Paolo, C. Giannini, R. E. Poppele, and F. Lacquaniti. Locomotor primitives in newborn babies and their development. *Science*, 334(6058):997–999, 2011.
23. B. Emile and P. Common. Estimation of time delays between unknown colored signals. *SIGNAL PROCESSING*, 68(1):93–100, 1998.
24. D. Endres, E. Chiovetto, and M. Giese. Model selection for the extraction of movement primitives. *Frontiers in Computational Neuroscience*, 7:185, 2013.
25. D. Endres, A. Christensen, L. Omlor, and M. A. Giese. Emulating human observers with Bayesian binning: segmentation of action streams. *ACM Transactions on Applied Perception (TAP)*, 8(3):16:1–12, 2011.

26. D. Endres, A. Christensen, L. Omlor, and M. A. Giese. Segmentation of action streams: human observers vs. Bayesian binning. In S. Edelkamp and J. Bach, editors, *KI 2011, LNAI 7006*, pages 75–86. Springer, 2011.
27. D. Endres and P. Földiák. Bayesian bin distribution inference and mutual information. *IEEE Transactions on Information Theory*, 51(11):3766 – 3779, 2005.
28. D. Endres and M. Oram. Feature extraction from spike trains with bayesian binning: latency is where the signal starts. *Journal of Computational Neuroscience*, 29:149–169, 2010.
29. D. Endres, M. Oram, J. Schindelin, and P. Földiák. Bayesian binning beats approximate alternatives: estimating peri-stimulus time histograms. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*. MIT Press, Cambridge, MA, 2008.
30. P. Fearnhead. Exact and efficient bayesian inference for multiple changepoint problems. *Statistics and Computing*, 16(2):203–213, 2006.
31. T. Flash and B. Hochner. Motor primitives in vertebrates and invertebrates. *Current Opinion in Neurobiology*, 15(6):660–666, 2005.
32. T. Flash and N. Hogan. The coordination of arm movements: an experimentally confirmed mathematical model. *J. Neurosci.*, (5):1688–1703, 1985.
33. E. Fuchs, T. Gruber, J. Nitschke, and B. Sick. Online segmentation of time series based on polynomial least-squares approximations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(12):2232–2245, 2010.
34. A. Gensler, T. Gruber, and B. Sick. Blazing fast time series segmentation based on update techniques for polynomial approximations. In *13th IEEE International Conference on Data Mining Workshops, ICDM Workshops, TX, USA, December 7-10, 2013*, pages 1002–1011, 2013.
35. M. Giese, A. Mukovskiy, A. Park, L. Omlor, and J. Slotine. Real-time synthesis of body movements based on learned primitives. In D. Cremers, B. Rosenhahn, and A. L. Yuille, editors, *Statistical and Geometrical Approaches to Visual Motion Analysis*, page 107127. Springer Verlag, Heidelberg, 2009.
36. R. D. Green. Spatial and temporal segmentation of continuous human motion from monocular video images. In *Proceedings of Image and Vision Computing*, pages 163–169, New Zealand, 2003.
37. C. M. Harris and D. M. Wolpert. Signal-dependent noise determines motor planning. *Nature*, 394(6695):780–784, 1998.
38. C. B. Hart and S. Giszter. Distinguishing synchronous and time varying synergies using point process interval statistics: Motor primitives in frog and rat. *Frontiers in Computational Neuroscience*, 7:52, 2013.
39. M. Hutter. Exact bayesian regression of piecewise constant functions. *Journal of Bayesian Analysis*, 2(4):635–664, 2007.
40. P. A. Hjen-Srensen, O. Winther, and L. K. Hansen. Mean field approaches to independent component analysis, 2001.
41. W. Ilg, G. Bakir, J. Mezger, and M. Giese. On the representation, learning and transfer of spatio-temporal movement characteristics. *International Journal of Humanoid Robotics*, 1(4):613–636, 2004.
42. Y. P. Ivanenko, G. Cappellini, N. Dominici, R. E. Poppele, and F. Lacquaniti. Coordination of locomotion with voluntary movements in humans. *Journal of Neuroscience*, 25(31):7238–7253, 2005.
43. Y. P. Ivanenko, R. E. Poppele, and F. Lacquaniti. Five basic muscle activation patterns account for muscle activity during human locomotion. *Journal of Physiology*, 556(Pt 1):267–282, 2004.
44. T. R. Kaminski. The coupling between upper and lower extremity synergies during whole body reaching. *Gait and Posture*, 26(2):256–262, 2007.
45. E. Keogh, S. Chu, D. Hart, and M. Pazzani. An online algorithm for segmenting time series. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 289–296, 2001.

46. L. Kovar, M. Gleicher, and F. Pighin. Motion graphs. *ACM Trans. Graph.*, 21:473–482, July 2002.
47. F. Kschischang, B. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001.
48. D. Kulic, W. Takano, and Y. Nakamura. Online segmentation and clustering from continuous observation of whole body motions. *IEEE Transactions on Robotics*, 25(5):1158–1166, 2009.
49. F. Lacquaniti, C. Terzuolo, and P. Viviani. The law relating kinematic and figural aspects of drawing movements. *Acta Psychologica*, 54:115–130, 1983.
50. D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *In NIPS*, pages 556–562. MIT Press, 2000.
51. D. Lemire. A better alternative to piecewise linear time series segmentation. *CoRR Arxiv*, abs/cs/0605103, 2006.
52. Y. Li, T. Adahi, and V. Calhoun. Estimating the number of independent components for functional magnetic resonance imaging data. *Human brain mapping*, 28(11):1251–1266, 2007.
53. J. Liu, Y. C. Nakamura, and N. S. Pollard. Annotating everyday grasps in action. In *Dance Notations and Robot Motion*, chapter ZZ, pages XX–YY. Springer, 2015.
54. T. Minka. Automatic choice of dimensionality for pca. Technical report, M.I.T. Media Laboratory Perceptual Computing Section, 2000.
55. T. Minka and J. Winn. Gates: A graphical notation for mixture models. In *Proceedings of NIPS*, 2008.
56. A. Mukovskiy, S. J.J.E., and M. A. Giese. Dynamically stable control of articulated crowds. *Journal of Computational Science*, 4(4):304–310, 2013.
57. L. Omlor. *New methods for anechoic demixing with application to shift invariant feature extraction*. PhD in informatics, Universität Ulm. Fakultät für Ingenieurwissenschaften und Informatik, 2010. urn:nbn:de:bsz:289-vts-72431.
58. L. Omlor and M. Giese. Blind source separation for over-determined delayed mixtures. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 1049–1056. MIT Press, Cambridge, MA, 2007.
59. L. Omlor and M. A. Giese. Extraction of spatio-temporal primitives of emotional body expressions. *Neurocomputing*, 70(10-12):1938–1942, 01 2007. reviewed.
60. L. Omlor and M. A. Giese. Anechoic blind source separation using wigner marginals. *Journal of Machine Learning Research*, 12:1111–1148, 2011.
61. F. Polyakov, E. Stark, R. Drori, M. Abeles, and T. Flash. Parabolic movement primitives and cortical states: merging optimality with geometric invariance. *Biol. Cybern.*, 100(2):159–184, 2009.
62. L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
63. C. L. Roether, L. Omlor, A. Christensen, and M. A. Giese. Critical features for the perception of emotion from gait. *Journal of Vision*, 9(6):1–32, 2009.
64. C. Rose, B. Bodenheimer, and M. F. Cohen. Verbs and adverbs: Multidimensional motion interpolation using radial basis functions. *IEEE Computer Graphics and Applications*, 18:32–40, 1998.
65. A. Safonova and J. K. Hodgins. Construction and optimal search of interpolated motion graphs. In *ACM SIGGRAPH 2007 Papers*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.
66. A. Safonova, J. K. Hodgins, and N. S. Pollard. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM Trans. Graph.*, 23:514–521, 2004.
67. M. Santello, M. Flanders, and J. F. Soechting. Postural hand synergies for tool use. *J Neurosci*, 18(23):10105–10115, Dec 1998.
68. J. Segouat and A. Braffort. Toward the study of sign language coarticulation: Methodology proposal. In *Proceedings of the 2009 Second International Conferences on Advances in Computer-Human Interactions*, ACHI '09, pages 369–374, Washington, DC, USA, 2009. IEEE Computer Society.
69. T. F. Shipley, M. J. Maguire, and J. Brumberg. Segmentation of event paths. *Journal of Vision*, 4(8), 2004.

70. A. Swindlehurst. Time delay and spatial signature estimation using known asynchronous signals. *IEEE Trans. Signal Processing*, 46:449–462, 1997.
71. J. S. Thomas, D. M. Corcos, and Z. Hasan. Kinematic and kinetic constraints on arm, trunk, and leg segments in target-reaching movements. *J Neurophysiol*, 93(1):352–364, Jan 2005.
72. L. H. Ting and J. M. Macpherson. A limited set of muscle synergies for force control during a postural task. *Journal of Neurophysiology*, 93(1):609–613, 2005.
73. G. Torres-Oviedo, J. M. Macpherson, and L. H. Ting. Muscle synergy organization is robust across a variety of postural perturbations. *Journal of Neurophysiology*, 96(3):1530–1546, 2006.
74. S. Tu and L. Xu. An investigation of several typical model selection criteria for detecting the number of signals. *Frontiers of Electrical and Electronic Engineering in China*, 6:245–255, 2011.
75. A. Ude, C. Atkeson, and M. Riley. Programming full-body movements for humanoid robots by observation. *Robotics and Autonomous Systems*, 47:93–108, 2004.
76. A. Ude, M. Riley, B. Nemeč, T. Asfour, and G. Cheng. Synthesizing goal-directed actions from a library of example movements. In *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, pages 0–0, 2007.
77. M. Wächter, S. Schulz, T. Asfour, E. Aksoy, F. Wörgötter, and R. Dillmann. Action sequence reproduction based on automatic segmentation and object-action complexes. In *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, pages 189–195, 2013.
78. L. Xu. Bayesian ying yang learning. *Scholarpedia*, 2(3):1809, 2007.
79. Ö. Yilmaz and S. Rickard. Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions on Signal Processing*, 52:1830–1847, 2004.
80. J. M. Zacks, T. S. Braver, M. A. Sheridan, D. I. Donaldson, A. Z. Snyder, J. M. Ollinger, R. L. Buckner, and M. E. Raichle. Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, 4(6):651 – 655, 2001.
81. J. M. Zacks, S. Kumar, R. A. Abrams, and R. Mehta. Using movement and intentions to understand human activity. *Cognition*, 112(2):201 – 216, 2009.