Evaluating Perceptual Predictions based on Movement Primitive Models in VR- and Online-Experiments

Benjamin Knopp
Department of Psychology
University of Marburg
benjamin.knopp@uni-marburg.de

Dmytro Velychko
Department of Psychology
University of Marburg
dmytro.velychko@uni-marburg.de

Johannes Dreibrodt
Department of Psychology
University of Marburg
dreibrod@students.uni-marburg.de

Alexander C. Schütz
Department of Psychology
University of Marburg
alexander.schuetz@staff.uni-marburg.de

Dominik Endres
Department of Psychology
University of Marburg
dominik.endres@uni-marburg.de

ABSTRACT

We investigate the role of prediction in biological movement perception by comparing different representations of human movement in a virtual reality (VR) and online experiment. Predicting movement enables quick and appropriate action by both humans and artificial agents in many situations, e.g. when the interception of objects is important. We use different predictive movement primitive (MP) models to probe the visual system for the employed prediction mechanism. We hypothesize that MP-models, originally devised to address the degrees-of-freedom (DOF) problem in motor production, might be used for perception as well.

In our study we consider object passing movements. Our paradigm is a predictive task, where participants need to discriminate movement continuations generated by MP models from the ground truth of the natural continuation. This experiment was conducted first in VR, and later on continued as online experiment. We found that results transfer from the controlled and immersive VR setting with movements rendered as realistic avatars to a simple and COVID-19 safe online setting with movements rendered as stick figures. In the online setting we further investigate the effect of different occlusion timings. We found that contact events during the movement might provide segmentation points that render the lead-in movement independent of the continuation and thereby make perceptual predictions much harder for subjects. We compare different MP-models by their capability to produce perceptually believable movement continuations and their usefulness to predict this perceptual naturalness.

Our research might provide useful insight for application in computer animation, by showing how movements can be continued without violating the expectation of the user. Our results also contribute towards an efficient method of animating avatars by combining simple movements into complex movement sequences.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SAP '20, September 12–13, 2020, Virtual Event, USA

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7618-1/20/09.
https://doi.org/10.1145/3385955.3407940

CCS CONCEPTS

• **Computing methodologies** → **Perception**; *Animation*; *Motion processing*; • **Theory of computation** → *Gaussian processes.*

KEYWORDS

human animation, movement primitives, perception, dynamical systems, psychophysics, Gaussian process dynamical model, dynamical movement primitives

ACM Reference Format:
Benjamin Knopp, Dmytro Velychko, Johannes Dreibrodt, Alexander C. Schütz, and Dominik Endres. 2020. Evaluating Perceptual Predictions based on Movement Primitive Models in VR- and Online-Experiments . In *ACM Symposium on Applied Perception 2020 (SAP '20), September 12–13, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3385955.3407940

1 INTRODUCTION

Predictive coding is one of the hot topics in neuroscience [Friston 2010; Hohwy 2013]. In this framework, the brain is viewed as an engine generating predictions based on previous sensory input. These predictions are then compared to the current sensory input to refine a percept. The investigation of the prediction mechanism is directly relevant for areas of applied perception, such as computer animation: Generating realistic animation could be achieved in the most economical manner possible [Sattler et al. 2005].

Ways of economical movement production have also been proposed to facilitate the motor control problem: movement primitives (MPs) are hypothetical elements used by the central nervous system to build complex movements. Assuming a common code of action and perception [Friston 2010; Prinz 1997], MPs might be used in perception as well. If this would be the case, the MP representation used by the brain should yield the best animation results. Furthermore, we hypothesize that movement perception is Bayes-optimal [Knill and Pouget 2004], i.e. we assume that the complexity of the perceptual representation reflects Bayesian model comparison, which serves as our ideal observer model with MP-Type specific complexity parameters as input (see 3.1). The cross-validatory mean squared error (MSE) as approximate Bayesian model evidence can

then be used to predict the perceived naturalness of movements based on MPs (see 3.3.2).

We use a prediction task (adapted from Graf et al. [2007]) to compare MP representations with different predictive mechanisms. Participants rate movement continuations generated by MP models in a two-alternative forced choice task. One trial consists of two sequences, each with the same lead-in movement followed by a short occlusion, but with one sequence showing the generated movement continuation and the other one showing the actual recorded movement. We implemented this paradigm in VR, as well as a web browser based online experiment.

The movements we study either contained object contact or not. We furthermore manipulate the occlusion timing to control the visibility of the contact event. We can therefore investigate the role of segmentation by a contact event on the perceptual prediction [Zacks and Swallow 2007]: we hypothesize that a contact event breaks the continuity of movement necessary for prediction. A contact event during occlusion should thus widen the expectation of possible continuations.

2 RELATED WORK

This study is inspired by [Knopp et al. 2019], and shares the same assumptions about MPs as possible common representation for action and perception. While this previous work focused on the perception of naturalness of movements, the current study addresses the predictive mechanism inherent in MP models. Similar studies were also conducted for MP models of emotional handshakes [Taubert et al. 2012] and facial expressions [Chiovetto et al. 2018].

In our experiments, we investigate the perceptual extrapolation of a trajectory beyond the actual presented or implied movement of an object, which is termed representational momentum (RM), as a part of the visual prediction process [Bertamini 1993; Freyd and Finke 1984; Thornton and Hayes 2004]. Senior et al. [2000] reviewed functional magnetic resonance imaging (fMRI) results and used transcranial magnetic stimulation (TMS) to identify the middle temporal visual area (V5/MT) as involved in processing RM. Jarraya et al. [2005] found evidence of RM in memory tasks involving movements represented in point-dot figures. Brain areas that process motion, such as V5/MT, respond when motion is implied, for example in pictures, or occluded [Graf et al. 2007]. Kilner et al. [2004] found neural oscillations in the motor cortex without actual motor activity during expectation of a hand movement presentation prior to its onset, presumably due to visual prediction processes. These processes are also found in participants observing imitable actions [Buccino et al. 2004; Wilson and Knoblich 2005] These studies suggest motor activity, or motor simulation [Stadler et al. 2012] to be involved in predicting future percepts of movements in real time, which further supports the functional framework of the mirror neuron system [MNS, Iacoboni and Dapretto 2006; Rizzolatti and Craighero 2004].

Besides the involvement of the MNS in RM, Graf et al. [2007] also show that visual movement prediction is a real-time process that includes effect estimations of motor commands before the motor action is performed. Visual Movement Prediction also requires prior information [Schröger et al. 2015], such as visual identifications of the percepts, therefore making tasks of visual prediction more difficult compared to sheer tasks of identifying or distinguishing movements, such as in Knopp et al. [2019]. This is consistent with the predictive coding framework, which follows from a Bayesian view of the MNS and also explains how we can infer movement intentions from movement observations [Kilner et al. 2004]. Bayesian model scores would therefore not only serve to identify the model with the best prediction performance, but should also be diagnostic of visual movement prediction performance of humans.

3 MODELS AND METHODS

In this section we shortly review relevant features of the investigated MP model types to make this publication self-contained. We then describe the experimental paradigm and its implementation as VR- and web-browser based online experiment. Finally we describe our methods for data analysis.

3.1 Movement Primitives

MPs refer to building blocks of complex movements, but there is little consensus on an exact definition. Consequently, many different types of MPs have been proposed in literature [Endres et al. 2013]. We focus on dynamical and temporal MPs in this study, as we are interested in finding a higher level representation suitable for modeling perception,

We perceptually validate 3 generative MP Types: Temporal MPs, Dynamical MPs and the coupled Gaussian Process Dynamical Model. Each MP-Type has specific complexity parameters, which should ideally be selected to maximize the Bayesian model evidence. We use the cross-validatory MSE as approximation to the model evidence.

In this section we can only provide a rough overview, just enough to enable readers from different backgrounds to understand parameters of the stimuli for the psychophysical experiment. Please refer to the cited papers for detailed information. Velychko et al. [2018] also provide graphical model representations and summarize the features of the MP models presented in this chapter.

3.1.1 Temporal Movement Primitives [TMP, Clever et al. 2016]. Temporal MPs describe the stereotyped temporal patterns of movement parameters, for example Electromyography (EMG) signals, but also joint trajectories as well as endpoint trajectories. We refer to all signals more generally as Degree-of-freedom (DOF). A possible biological implementation of temporal MPs might be central pattern generators (CPGs) [Ivanenko et al. 2004] combined with cortical top-down control. Temporal MPs incorporate a temporal predictive mechanism: the complete time-course of the movement is determined at its onset. This type of MPs allows for simple concatenation and temporal scaling.

The trajectory $x_k(t)$ of a DOF, e.g. a joint angle, is a weighted sum of Q MPs $y_q(t)$, which are functions of time. $\varepsilon_i(t) \sim \mathcal{N}(0, \sigma_i)$ is Gaussian observation noise:

$$x_k(t) = \sum_{q=1}^{Q} w_{k,q} y_q(t) + \varepsilon_i(t) \tag{1}$$

The posterior distribution of weights and MPs are learned by approximate Bayesian learning via free energy. We use the number of

MPs $Q = 3 \ldots 15$ as ideal observer parameter. In general, more MPs allow for more fine-grained temporal structure of the movement, but might lead to over-fitting.

3.1.2 Dynamic Movement Primitives [DMP, Ijspeert et al. 2013]. While temporal MPs directly model the movement parameters, DMPs describe the stereotyped elements of movement as attractors of a dynamical system, thus enabling the prediction of the next state from the previous ones. Building on the hypothesis of separate brain areas for rhythmic and discrete movements, two kinds of dynamical systems are common: cyclic oscillators and point attractors [Schaal 2006].

More formally: DMP models represent a movement trajectory $x_k(t)$ obeying a differential equation. They rely on a damped spring system which forces $x_k(t)$ to contract to the specified goal g_k, if the dampening factor is high enough. Through the non-linear forcing function f_k (Eq. 2) the trajectories can be modified. This function is modeled as weighted sum of Gaussian basis functions $\Psi_i(\tau)$ (Eq. 4). Time is replaced by τ, which decays exponentially to zero (Eq. 3). DMPs are learned from training data by setting the weights w_i such that the training mean-squared error between predicted and actual movement (MSE) is minimal.

$$\tau \ddot{x}_k = \alpha_z(\beta_z(g_k - x_k) - \dot{x}_k) + f_k(\tau) \tag{2}$$

$$\dot{\tau} \propto -\tau \tag{3}$$

$$f_k(\tau) = \frac{\sum_{i=1}^{N} \Psi_i(\tau) w_{k,i}}{\sum_{i=1}^{N} \Psi_i(\tau)} \tau(g_k - x_k(0)) \tag{4}$$

We investigate $N = 10, 20 \ldots, 100$ basis functions as ideal observer parameters. Basis functions serve a similar role as the number of MPs in the TMP model: more basis functions allow for more complicated forcing functions, which enable richer temporal dynamics.

3.1.3 Coupled Gaussian Process Dynamical Model [CGPDM, Velychko et al. 2014]. CGPDMs compose different dynamical models in a low dimensional latent space for M different body parts. This model is a generalization of the Gaussian Process Dynamical Model [GPDM, Wang et al. 2008]: By setting the whole body as one body part ($M = 1$) the CGPDM becomes the GPDM. If there are more than one body parts, each dynamical system predicts not only the next time-step of their associated body part, but also the temporal evolution of other body parts via coupling functions. This way, flexible coupling between body-parts is possible. The CGPDM can be regarded as a middle ground between DMPs encoding single DOFs, and the monolithic GPDM. The latent dynamical systems can thus be thought of as flexibly coupled CPGs routing commands to the muscles.

In contrast to DMPs, CGPDMs learn a full dynamical model of latent variables Y in discrete time, which are mapped onto the observed DOFs X. Both the dynamics mapping $f^{i,j}()$ (Eq. 5) from the latent space of body part j to body part i ($i, j = 1 \ldots M$), as well as the mapping from latent to observed space $g()$ (Eq. 6) are drawn from Gaussian process priors. dt denotes the time discretization step-size:

$$Y^i(t) = f^{j,i}(Y^j(t - dt)) + \varepsilon^i_{Y,t} \tag{5}$$

$$X^i(t) = g^i(Y^i(t)) + \varepsilon^i_{X,t} \tag{6}$$

The model can be trained in two ways: by maximum-a-posteriori inference (MAP), or by free energy minimization using variational approximations (Variational (Coupled) Gaussian Process Dynamical Model [v(C)GPDM, Velychko et al. 2018]). In our study we use $M = 1, 3$ body parts and use both training methods: GPDM ($M = 1$, MAP), CGPDM ($M = 3$, MAP), vGPDM ($M = 1$, variational), vCGPDM ($M = 3$, variational).

Without variational approximations, due to the non-parametric GPs prior, the movements *are* the movement representation, which is not compact. Therefore, MAP-trained (C)GPDMs, do not provide a complexity parameter.

The representation can be compressed by introducing sparse variational approximations. Now, each v(C)GPDM is parameterized by a small set of inducing points (IPs) and associated inducing values (IVs). The initial choice of IPs/IVs is the only remaining source of stochasticity in the training process. It may have measurable effects as we will show below.

We use IPs for both mappings, serving as ideal observer model parameters: "dynamics" IPs for the dynamical model mapping, and "pose" IPs for the latent-to-observed variable mapping. More dynamics IPs allow for richer dynamics (similar to the parameters of DMP and TMP), while more pose IPs will allow for more (spatial) variability of poses. An IP/IV pair might be thought of as a prototypical example for the mappings drawn from their associated Gaussian process. They thus provide some abstraction from the observed movement and might be implemented by small neuronal populations.

3.2 Experiments

This study includes two experiments: first, we conducted a highly controlled and ecologically valid VR-Experiment. Then, we decided to specifically study effects of contact events on perceptual predictions using the same paradigm with additional occlusion timing conditions. After we made this decision, the COVID-19 pandemic forced us to close our VR-Lab. This triggered us to port the VR-Experiment to an online setting. As benefit we could collect more data with less effort, but we as drawback we we could not control the viewing conditions under which participants performed the experiment. The VR experiment was implemented using Vizard 5 [WorldViz 2019] and the online experiment was implemented using the javascript library jsPsych [De Leeuw 2015] and webGL. A test version of the online experiment can be tried online[1].

In general, the methods of this work first comprise learning the recorded movements via extraction of MPs from mocap data, resulting in 3D joint locations and trajectories. The joint locations of both model-extracted and natural movement data are then connected (rigged) to a digital avatar (VR experiment) or a skeleton stick figure (online experiment). For the VR experiment, the rigged avatar, containing both natural and model-generated movements is then imported in a VR environment. For the online experiment, the movements are rendered in webGL.

[1] http://vhrz1092.hrz.uni-marburg.de/javascriptbvh/experiment.html?subject=xyz.

We use this stimulus material for a psychophysical experiment in the form of a Graphics Turing Test [McGuigan 2006] on human movement prediction performance. In both experiments, the participants execute forced-choice trials, deciding which movement continuation fits best to a given beginning. The experiments' data comprises the relative frequency of a MP model successfully confusing participants to prefer its generated movement to a natural movement continuation. We call this frequency 'confusion rate'.

3.2.1 Movements. All presented movement consists of putting a bottle from one side of a table in front of the torso, where the bottle is passed to the other hand, to the other side of the table while sitting on a chair. Four kinds of movements are used: Passing the bottle from the left side to the right side (pass-bottle-movements), and vice versa (return-bottle- movements), and from the left to the right side without a pause (pass-bottle-hold-movements) but instead passing the bottle directly to the right hand, and vice versa (return-bottle-hold). Motor expertise/experience [Graf et al. 2007; Stadler et al. 2012] and visual familiarity of the movements to one's own movements [Loula et al. 2005] influences prediction performances. Simple movements of passing a bottle are actions with a low demand of motor expertise. Therefore, participants are not expected to strongly differ in their prediction performance due to expertise or familiarity.

3.2.2 Stimulus Generation. We recorded movements from one actor for the experiment with a PhaseSpace Impulse X2 System and 44 active LED markers. We inferred skeleton and joint angles from the recorded C3D-files, which contain marker positions in the recorded time frames using our own skeleton estimation software. These are used by computational implementations of the MP models to learn from five different bottle-passing movements for each movement type. The models then generate Biovision bvh-files containing joint locations and their trajectories from 5 different starting positions.

For the VR experiment, the bvh-files are then imported into the Autodesk MotionBuilder environment, where the bvh-joints are manually rigged onto a custom skeleton of a gray avatar polygon mesh. The rigging is adopted for all other bvh files with a custom script. The rigged avatar is then imported to the Autodesk 3dsMax environment, where the avatar and the movements were converted into a cfg-file, containing avatar mesh, skeleton and animation files, which was then importable for the Vizard 5 software, with which the experiment was designed.

For the online experiment, we used a simple stick figure to display the movement 2: the bvh-files produced by the MP-models are converted into pairs of 3D positions, where each pair is start- and end-point of a segment specified by the skeleton. Each pair is then rendered using the `GL_LINES` OpenGL-primitive. As we have no control over the setting and state of the subject when she is running the experiment, we added attention check trials. For this, we used movements generated by DMP models which obviously failed, such as avatars floating up from the chair. We excluded experimental runs where participants failed to correctly identify the floating movement in more than 40% of attention checks. In the VR experiment no such attention checks were needed, and we fixed the avatar's pelvis to the chair for these movements.

3.2.3 Stimulus Presentation. Elements of the trial structure were adopted from Sparenberg et al. [2012], who implemented experiments on internal simulation of movement and Graf et al. [2007], who tested various occlusion times in a psychophysics experiment of movement prediction performance, where participants were instructed to identify 1 of 2 action continuations as the most fitting to the beginning of the action before the occlusion. The structure of a trial can be viewed in Figure 1. Textures and objects for the experiment environment were provided by WorldViz and the website www.sketchfab.com. Presenting the two stimuli sequentially instead of simultaneously has the advantage of participants not having to distribute their fixations across the stimuli and instead could focus each stimulus separately.

3.2.4 Catch Trials. Instead of predicting the correct movement continuation, participants might instead use the unintended strategy of only distinguishing the first and second movement continuation as more or less natural-looking, ignoring the movement onset presented before the occlusion. Participants also might be less attentive to the experiment, resulting in higher confusion rates on average. To measure these variables, the experiment includes so-called "catch-trials", of which 24 were implemented for each participant in the VR experiment, and 2 for each experimental run in the online experiment. A catch-trial has the structure of a standard trial, but replaces the model-generated movement continuation with the same natural movement continuation as in the other movement sequence. This catch-continuation sequence will be time-incoherent to the movement-offset before the occlusion: catch-movements of the VR experiment start either 400 ms, 700 ms or 1000 ms (8 trials per participant) before the natural movements and therefore make the natural movement look as if they skipped movement frames during the occlusion. Catch-trials measure the rate of erroneously choosing the time-incoherent action continuation as a natural continuation. Time delays of the catch-trials were inspired from Graf et al. [2007]. We adapted the skip-timings for the online experiment slightly to 375 ms, 667 ms and 1000 ms to increase the range of investigated shifts.

3.2.5 VR Experiment.

Participants: N = 34 participants (23 female, 18-39 years old, mean age = 22,7 years, SE = 3,3 years) were recruited. As recruitment criteria, participants had to be 18 years or older and had to have no impaired vision. They also should not suffer from a disease of the musculoskeletal system, in order to handle HTC Vive controllers for the experiment. Participant recruitment was organized and promoted with the Sona Systems ® participant management software. They received financial compensation (8€/h) or course credits for participation. An ethics application for the experiment had been approved by the local ethics commission (Ethikantrag 2015-19K). Participants received written information about the experiment in the participant management software and on the participant information sheet as well as in an instructional text in the VR environment. Participants gave their informed consent to participate.

Experimental Procedure: Participants were asked to sit on the experiment chair and were instructed to wear the head-mounted display (HMD) and HTC Vive controllers. As soon as participants

felt comfortable wearing the HMD, the experiment environment was loaded and the experiment started with an instructional text on the trial structure followed by nine familiarization trials where participants received feedback on their performance after each trial. After the familiarization trials, participants were additionally instructed to keep their gaze mainly focused on the avatar and their arms rested on their laps. Participants then started with the first of 269 trials. The trial number is derived from 24 catch-trials plus 5 repetitions of all 49 MP models. The trials were separated into four blocks of each 67 to 68 trials. After each block participants could take off the HMD and take a break of up to 5 minutes. Both catch- and normal trials were distributed randomly through all trials. Whether a trial presents a pass-hold-, pass- or return-movement was also randomized but is selected for both natural and model-generated movements presented in it. Each experiment run took about 70 to 85 minutes including all aforementioned procedures. Nine participants reported fatigue and one participant reported eye fatigue. One participant reported a headache after finishing a few trials of the first experiment block and aborted the experiment.

3.2.6 *Online Experiment.*

Participants: We collected data from 220 experiment runs of N = 98 participants using the university's participant management system (SONA System). The only metadata collected was a participant ID assigned by SONA. Participants were psychology students. They received course credits for participation.

Experimental Procedure: The experimental procedure is similar that of the VR experiment. We skipped the familiarization trials. Each experiment had 55 normal trials and two catch trials. This results in a length of approximately 15 minutes. Trials were sampled randomly from all possible trials for each experiment. Therefore each participant was allowed an arbitrary number of repetitions of the experiment. Each participant has a fixed anonymous ID assigned by the SONA participant management system. In the advertisement of the experiment we recommended 6 repetitions, but we did not control this number.

3.3 Data Analysis

3.3.1 Confusion Rate. Participants were forced to choose one of the two sequences in each trial. Therefore, in trial i the participant's response is $r_i = 0$ if she guessed the wrong sequence and $r_i = 1$ if the participant chose the correct sequence. We pooled across participants to achieve sufficient statistical power.

The confusion rate (p) is defined as the number of times a participant erroneously chooses the sequence containing a model-generated movement continuation as more fitting to the movement onset divided by the total number of trials N.

$$p = \frac{N - \sum_i r_i}{N} \tag{7}$$

We assume that p approaches 0.5 if the model perfectly matches the observers perceptual predictions. The confusion rate measures the model success while $1 - p_i$ measures human discrimination ability. We chose to report the confusion rate, as we are interested in comparing the models.

Each trial is specified by:

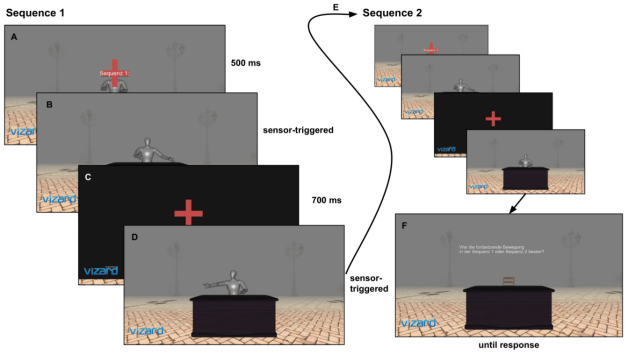

Figure 1: Trial structure of the psychophysics experiment. Each trial consists of two sequences, each beginning with (A) a red fixation cross appearing for 500ms in front of the desk for fixating the gaze towards the avatar, followed by (B) the onset of a natural movement randomly chosen from the set of 6 pass-hold-, 10 pass- and 9 return-movements, but based on the model-generated movement type. As soon as the hand returns to the front of the avatar, (C) an occlusion is triggered that lasts for 700ms. During the occlusion the movement is continued. After the occlusion, (D) the movement is continued by either the avatar performing the natural movement, with which the sequence has started, or an avatar performing the model-generated movement. The occurrence of the natural movement continuation in the first or second sequence is randomized. The end of the movement triggers either (E) starting sequence 2 or (F) making the visible avatar disappear and asking the participant for choosing the sequence with the correct movement continuation: "Which sequence did you perceive as more natural?". The second sensor is activated 300 ms after the hand of the avatar enters it. This ends the movement sequence as soon as the hand is about to return to a position in front of the avatar. After choosing a sequence (by pressing the trigger-button on either the left HTC Vive controller for sequence 1, or the button on the right HTC Vive controller for sequence 2) the next trial starts.

- MP type with parameters:
  - TMP: Number of MPs Q.
  - DMP: Number of basis function N.
  - v(C)GPDM: Number of dynamical and pose IPs.
  - MAP-GPDM: No parameters.
- Movement: With or without table contact.
- Direction: From left to right, or vice versa.
- Training data set.
- Model scores after training.

In the online experiment there are furthermore three occlusion conditions:

- Occlusion timing: before, during, or after passing the center of the table

We assume that confusion rate p depends on a subset of these parameters. It might also be participant-specific.

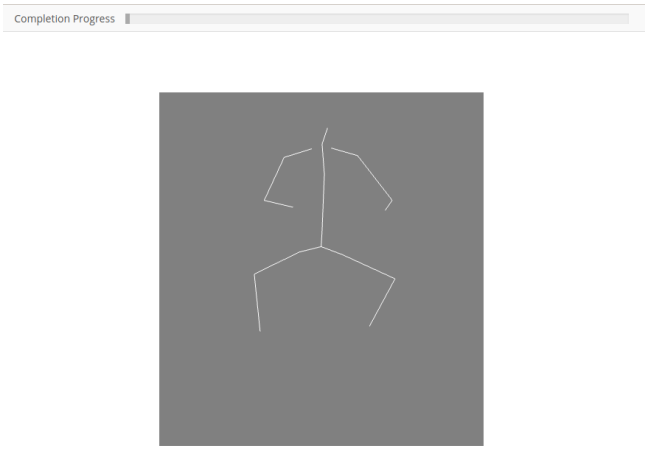

Figure 2: Screenshot of online experiment. The procedure was the same as described in Fig. 1, but participants respond by clicking buttons instead of using controllers.

3.3.2 Logistic Regression. We assume that variable r_i is Bernoulli distributed and investigate the effect of cross-validatory test set mean squared error (MSE), which is a proxy for the Bayesian model evidence. We obtain this MSE by training the MP models on 4 out of 5 movements, and then compute the mean squared residual between the reconstruction and actual observation of the 5th movement.

We use the *centered MSE* $= \text{MSE} - \mathbb{E}[\text{MSE}]$ as predictor for the participants' responses using a Bayesian logistic regression model:

$$r_i \sim \text{Bernoulli}(p_i) \tag{8}$$

$$p_i = \frac{1}{1 + \exp(-(\alpha + \beta \cdot \text{MSE}_i))} \tag{9}$$

$$\alpha, \beta \sim \mathcal{N}(0, 10) \tag{10}$$

The participants' responses are Bernoulli distributed, with parameter p_i being the output of the sigmoid model with parameters α and β. We set a wide Gaussian prior on these parameters and compute their posterior using Markov chain Monte Carlo[2].

4 RESULTS

First, we compare the results of the VR- and the online experiments and contrast these with previous findings in a naturalness perception experiment [Knopp et al. 2019]. We demonstrate that our paradigm works as intended by presenting the catch trial results. We then show the predictions of logistic regression for different MP types and finally present results demonstrating the effect of contact events in our experiments.

4.1 Comparison of Experiments

Figure 3 compares the mean confusion rates over complexity parameters of MP-Types of a naturalness perception experiment [Knopp et al. 2019] with the two experiments described in this study. The previous experiment measured confusion rate in a task where participants had to choose the more natural one of two walking movements. One of the movements in each trial of that experiment was MP generated, the other one was a replay of a natural movement recording.

Considering the differences regarding movement (walking vs. object-passing), experimental paradigm (prediction vs identification), setting (desktop vs. VR vs. web-based), and representation (full avatar vs. stick-figure), the confusion rates are remarkably similar.

TMP models consistently perform best. DMP models perform well in all settings, too. The prediction and identification paradigms differ very much regarding the training mode of the (C)GPDMs: MAP training failed to fool subjects to mistake the generated walking movements in the identification task, but performed on par with the variationally trained models in the pass-object prediction task.

We used attention checks in the online experiment (highly unrealistic floating movements were shown), to filter experiment runs with inattentive subjects. Still, there is a slight tendency of slightly higher confusion rates in the online experiment compared to the VR setting.

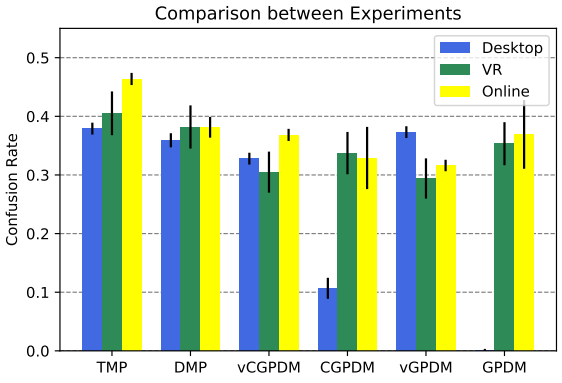

Figure 3: Confusion rate for MP models in three different experiments. 1. Previously published desktop experiment, 2. VR experiment and 3. online experiment from this study. Error bars depict beta-distributed standard error.

4.2 Catch Trial Results

We recorded participants' performance of falsely identifying the discontinuous movement continuation as the one most fitting the movement onset before the occlusion in 809 catch trials in the VR experiment, and in 318 catch trials (up to now) in the online experiment. Figure 4 shows the resulting confusion rates. The smallest shift of 375/400 ms is not detected, as the confusion rate is close to 0.5. The rate decreases for the conditions with larger shifts. The decrease is more pronounced for the VR data compared to data collected by the online experiment.

[2]We use the No-U-Turn Sampler implemented in Python library PyMC3 [Salvatier et al. 2016].

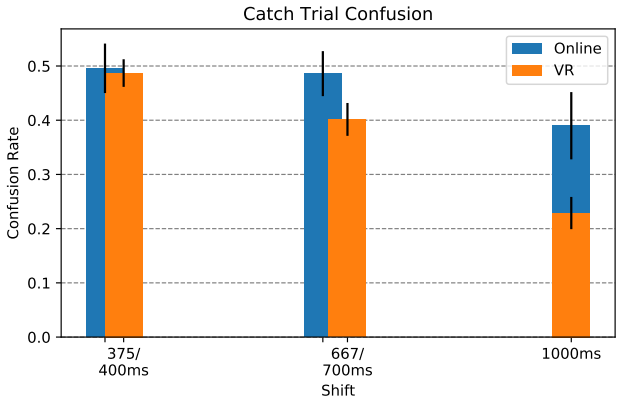

Figure 4: Confusion rate for catch trials with different time shifts for the two experimental conditions. The bars of the two experiments are shifted relative to each other, because we changed the shift timings slightly for the online experiment.

Table 1: Mean and standard deviation of posterior samples of parameters α and β.

	VR	Online
α		
DMP	-0.48 ± 0.05	-0.50 ± 0.08
TMP	-0.41 ± 0.04	-0.15 ± 0.04
vCGPDM	-0.83 ± 0.05	-0.55 ± 0.05
vGPDM	-0.89 ± 0.05	-0.78 ± 0.05
β		
DMP	0.09 ± 0.14	-0.65 ± 0.18
TMP	-1.29 ± 0.16	-0.09 ± 0.21
vCGPDM	0.67 ± 0.54	-3.54 ± 0.95
vGPDM	-1.49 ± 0.27	-3.09 ± 0.72

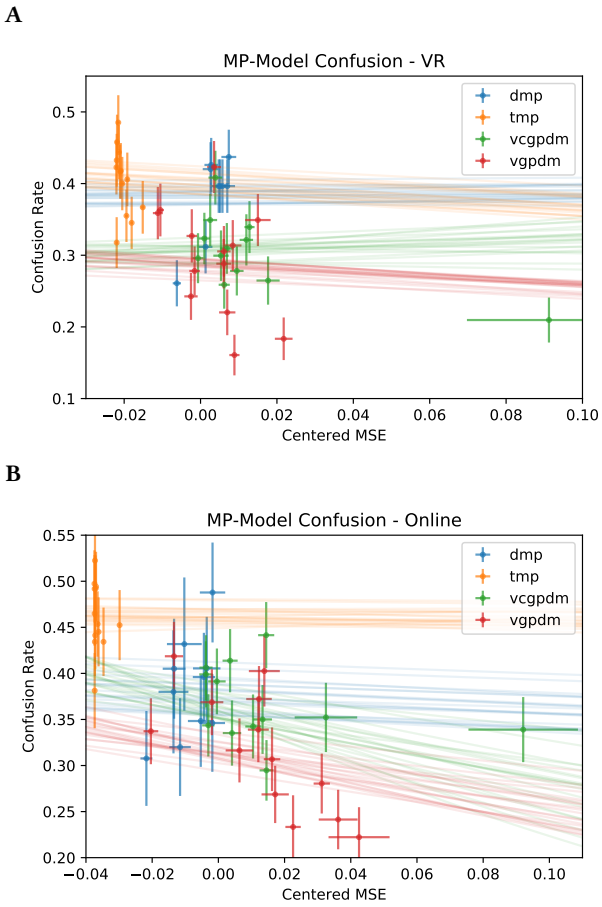

Figure 5: Confusion rate and logistic regression for different MP model types for data collected in (A) VR- and (B) online experiment. Each point shows the mean confusion rate for a MP model with specific parameters against the centered MSE (which is the MSE with subtracted mean). Error bars show the beta distributed standard deviation and the standard error of the MSE. Lines are predictions of the logistic regression model with 20 samples of parameters α and β.

4.3 Predicting Perceptual Predictions from Centered MSE

We predict the confusion rate, which is our measure for the different MP types' ability to generate movements in line with human perceptual predictions, from centered MSE using logistic regression (3.3.2). Figure 5 shows confusion rates of MP models over mean MSE. In general, lower MSE corresponds to higher confusion rate (negative slope β). We do not observe this relationship for DMP, vCGPDM models tested in the VR experiment. TMP models tested in the online experiment have a near-zero negative slope β. TMP models of the VR experiment on the other hand show the strongest dependence of the confusion rate on the MSE, together with vGPDM models. We summarize the posterior of parameters in Table 1.

4.4 Effect of Contact Event on Perceptual Predictions

In our online experiment we collected data for movements where a bottle is passed from one hand to the other either with or without touching the table. We varied the occlusion timing to investigate the effect of the table contact on perceptual prediction performance of MSE: The movement was occluded before, during, or after bottle was passed. We compare the influence of MSE on the confusion rates of trials with occlusion during table contact with the rest of the trials. For this we use logistic regression [3.3]. Here, the slope β is a measure of influence of MSE on the confusion rate. Given the posteriors of β we can compute the probability of $|\beta_{nc}|$ for trials without occluded contact being greater than $|\beta_c|$ for trials with occluded contact: $p(|\beta_{nc}| > |\beta_c|) = 0.998$. We are thus fairly

certain that MSE loses predictive capability if object-table contact is occluded.

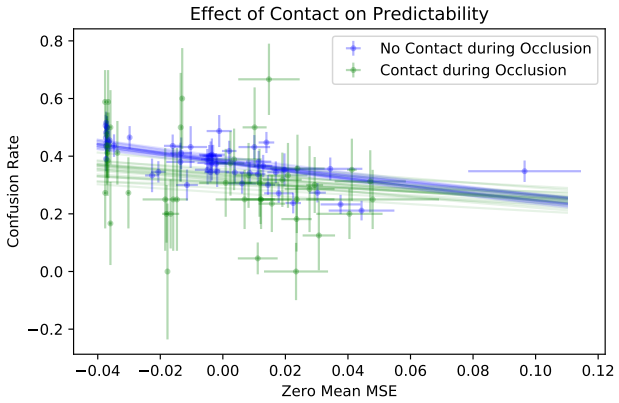

Figure 6: Predictions by logistic regression model for trials either with or without contact during occlusion, and confusion rate of models plotted against the centered MSE (same as fig. 5). In case of contact during occlusion, the slope of the fit is smaller, making MSE less useful predictor.

5 DISCUSSION

We compared 3 different types of MP-models using a predictive paradigm in two settings: VR and web-browser based. The representation of the movement was different as well: 3D avatars in the VR-, and stick figures in the online experiment. We also compared these results to published data in [Knopp et al. 2019] , which used different movements, and a non-predictive paradigm. Our results indicate that measured confusion rates generalize across movements, paradigm, and rendering specifics. A notable exception is the dramatic performance increase of MAP-trained (C)GPDMs. We suspect that the initialization of this model in the previous experiment might have been unfavorable.

Participants of the online experiment have shown slightly worse prediction performance. We expect this is due to attentional and motivational shortcomings of a non-lab environment. Still, considering the substantially lower effort of running the experiment and highly increased reach for recruiting participants, this drawback is more than compensated. Reaching out to many participants is very important, as our experimental design, even though simple and elegant, is collecting very little information per trial (1 bit). Still, a problem remaining are potential inter-individual differences. Because participants are exhausted very quickly, we can only test a small subset of all models and conditions. Pooling across participants while still accounting for inter-individual differences might be useful and we will explore this in the next study.

Catch trials show decreasing confusion rate for increasing timeshift of natural movements, which indicates that participants actually predict the movement, instead of rating the naturalness of the movement continuation without regard to the lead-in movement. This decrease is less pronounced in the data of the online experiment.

In our experiments, we found that TMP-models produce the most realistic movement. This is in line with previous findings [Knopp et al. 2019]. Therefore, TMPs might be used by the visual system for perceptual predictions. Dynamical models might still be involved in movement production because of their ability to handle perturbations. The shared representation between perception and production may therefore be more abstract: one dynamics model paired with a corresponding TMP model that encodes typical (unperturbed) solutions of the dynamics model, for fast perceptual predictions [Giese and Poggio 2000].

We use the centered MSE to predict perceived naturalness by using a logistic regression model for the confusion rate. The prediction worked well for TMP and vGPDM models of the VR experiment, and for vCGPDM and vGPDM of the online experiment. The online experiment might be the decisive bit harder for subjects, such that many TMP models come close enough to indistinguishability, impeding prediction. The vCGPDM has increased number of IP sets (one set for each body part) compared to the monolithic vGPDM. This introduces more stochasticity during training, resulting in large variation of the MSE. This might explain the different prediction results of the vCGPDM for the different experiments. Compared to previous findings [Knopp et al. 2019], the predictions are less reliable. This is because our experimental design is more complex, adding different movements and switching to a predictive sequential task. As previously discussed, more data is required to disentangle effects of different MP types, movements, and occlusion conditions.

Contact events are a common heuristic for the task of segmenting movements. Yet, there is little psychophysical investigation measuring the effect of models on segmentation, but see [Endres et al. 2011]. In our online experiment, we manipulated the occlusion timing to investigate the existence of perceptual segmentations induced by contact events. We found a higher expected increase of confusion rate for increasing MSE in trials where table contact was not occluded, which we interpret as follows: participants, who expect a contact event based on the previous trajectory, but can not see it, will have a less precise expectation about the continuation of the movement, making them less susceptible to higher deviations from their expectations. This is not the case for participants who see the contact, and can use frames after contact to build a more precise expectation.

The current work is the new and unexplored implementation of a Graphics Turing Test of movement prediction performances. Even though the structure of the prediction task was mostly adapted from other works [Graf et al. 2007; Knopp et al. 2019], conducting this task in the context of a Graphics Turing Test in a VR and web-based environment is novel and has yet to be established more firmly in psychophysical research.

6 CONCLUSIONS

The present work created a psychophysical task for visual prediction performances in a VR and web-based environment and implemented it to gather psychophysical data on six different representations of motor actions based on MPs. MP models can be used to generate natural-appearing novel movements on virtual avatars, which is important for neuroscientists searching for a common code

of action and perception and might be applied to build realistic computer animation with less effort in the future. In future studies we want to validate the assumptions that the influence of different movement representations (stick-figure vs. 3D avatar) is small and compare different movements, to investigate the generalizability of our results.

ACKNOWLEDGMENTS

This work was funded by DFG, IRTG1901 - The brain in action, and SFB-TRR 135 - Cardinal mechanisms of perception. We thank Olaf Haag for help with rendering of the stimuli and collecting data.

REFERENCES

Marco Bertamini. 1993. Memory for position and and dynamic representations. *Memory & Cognition* 21, 4 (1993), 449–457.

Giovanni Buccino, Ferdinand Binkofski, and Lucia Riggio. 2004. The mirror neuron system and action recognition. *Brain and language* 89, 2 (2004), 370–376.

Enrico Chiovetto, Cristóbal Curio, Dominik Endres, and Martin A. Giese. 2018. Perceptual integration of kinematic components in the recognition of emotional facial expressions. *Journal of Vision* 18, 4 (April 2018), 13–13. https://doi.org/10.1167/18.4.13

Debora Clever, Monika Harant, Henning Koch, Katja Mombaur, and Dominik Endres. 2016. A novel approach for the generation of complex humanoid walking sequences based on a combination of optimal control and learning of movement primitives. *Robotics and Autonomous Systems* 83 (Sept. 2016), 287–298. https://doi.org/10.1016/j.robot.2016.06.001

Joshua R De Leeuw. 2015. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods* 47, 1 (2015), 1–12.

Dominik Endres, Enrico Chiovetto, and Martin A. Giese. 2013. Model selection for the extraction of movement primitives. *Frontiers in Computational Neuroscience* 7 (2013), 185. https://doi.org/10.3389/fncom.2013.00185

Dominik Endres, Andrea Christensen, Lars Omlor, and Martin A. Giese. 2011. Emulating human observers with Bayesian binning: segmentation of action streams. *ACM Transactions on Applied Perception (TAP)* 8, 3 (2011), 16:1–12.

Jennifer J Freyd and Ronald A Finke. 1984. Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10, 1 (1984), 126.

Karl Friston. 2010. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience* 11, 2 (Feb. 2010), 127–138. https://doi.org/10.1038/nrn2787

Martin A. Giese and Tomaso Poggio. 2000. Morphable Models for the Analysis and Synthesis of Complex Motion Patterns. *International Journal of Computer Vision* 38 (June 2000), 59–73. https://doi.org/10.1023/A:1008118801668

Markus Graf, Bianca Reitzner, Caroline Corves, Antonino Casile, Martin Giese, and Wolfgang Prinz. 2007. Predicting point-light actions in real-time. 36 (2007), T22–T32. https://doi.org/10.1016/j.neuroimage.2007.03.017

Jakob Hohwy. 2013. *The predictive mind.* Oxford University Press.

Marco Iacoboni and Mirella Dapretto. 2006. The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience* 7, 12 (2006), 942.

Auke Jan Ijspeert, Jun Nakanishi, Heiko Hoffmann, Peter Pastor, and Stefan Schaal. 2013. Dynamical Movement Primitives: Learning Attractor Models for Motor Behaviors. *Neural Computation* 25, 2 (Feb. 2013), 328–373. https://doi.org/10.1162/NECO_a_00393

Yuri P. Ivanenko, Richard E. Poppele, and Francesco Lacquaniti. 2004. Five basic muscle activation patterns account for muscle activity during human locomotion: Basic muscle activation patterns. *The Journal of Physiology* 556, 1 (April 2004), 267–282. https://doi.org/10.1113/jphysiol.2003.057174

Mohamed Jarraya, Michel-Ange Amorim, and Benoît G Bardy. 2005. Optical flow and viewpoint change modulate the perception and memorization of complex motion. *Perception & psychophysics* 67, 6 (2005), 951–961.

James M Kilner, Claudia Vargas, Sylvie Duval, Sarah-Jayne Blakemore, and Angela Sirigu. 2004. Motor activation prior to observation of a predicted movement. *Nature neuroscience* 7, 12 (2004), 1299–1301.

David C. Knill and Alexandre Pouget. 2004. The bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neuroscience* 27 (2004).

Benjamin Knopp, Dmytro Velychko, Johannes Dreibrodt, and Dominik Endres. 2019. Predicting Perceived Naturalness of Human Animations Based on Generative Movement Primitive Models. *ACM Trans. Appl. Percept.* 16, 3 (Sept. 2019), 15:1–15:18. https://doi.org/10.1145/3355401

Fani Loula, Sapna Prasad, Kent Harber, and Maggie Shiffrar. 2005. Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance* 31, 1 (2005), 210.

Michael D. McGuigan. 2006. Graphics Turing Test. *CoRR* abs/cs/0603132 (2006).

Wolfgang Prinz. 1997. Perception and Action Planning. *European Journal of Cognitive Psychology* 9, 2 (June 1997), 129–154. https://doi.org/10.1080/713752551

Giacomo Rizzolatti and Laila Craighero. 2004. The mirror-neuron system. *Annu. Rev. Neurosci.* 27 (2004), 169–192.

J Salvatier, TV Wiecki, and C Fonnesbeck. 2016. Probabilistic programming in python using pymc3. PeerJ Computer Science, 2, e55.

Mirko Sattler, Ralf Sarlette, and Reinhard Klein. 2005. Simple and efficient compression of animation sequences. (2005), 209–217. https://doi.org/10.1145/1073368.1073398

Stefan Schaal. 2006. Dynamic Movement Primitives -A Framework for Motor Control in Humans and Humanoid Robotics. In *Adaptive Motion of Animals and Machines*, Hiroshi Kimura, Kazuo Tsuchiya, Akio Ishiguro, and Hartmut Witte (Eds.). Springer-Verlag, Tokyo, 261–280. https://doi.org/10.1007/4-431-31381-8_23

Erich Schröger, Anna Marzecová, and Iria SanMiguel. 2015. Attention and prediction in human audition: a lesson from cognitive psychophysiology. *European Journal of Neuroscience* 41, 5 (2015), 641–664.

Carl Senior, J Barnes, V Giampietroc, A Simmons, ET Bullmore, M Brammer, and AS David. 2000. The functional neuroanatomy of implicit-motion perception or 'representational momentum'. *Current Biology* 10, 1 (2000), 16–22.

Peggy Sparenberg, Anne Springer, and Wolfgang Prinz. 2012. Predicting others' actions: Evidence for a constant time delay in action simulation. *Psychological Research* 76, 1 (2012), 41–49.

Waltraud Stadler, Anne Springer, Jim Parkinson, and Wolfgang Prinz. 2012. Movement kinematics affect action prediction: comparing human to non-human point-light actions. *Psychological research* 76, 4 (2012), 395–406.

Nick Taubert, Andrea Christensen, Dominik Endres, and Martin A. Giese. 2012. Online Simulation of Emotional Interactive Behaviors with Hierarchical Gaussian Process Dynamical Models. *Proceedings of the ACM Symposium on Applied Perception (ACM-SAP 2012)* (2012), 25–32. https://doi.org/10.1145/2338676.2338682

Ian Thornton and Amy Hayes. 2004. Anticipating action in complex scenes. *Visual Cognition* 11, 2-3 (2004), 341–370.

Dmytro Velychko, Dominik Endres, Nick Taubert, and Martin A. Giese. 2014. Coupling Gaussian Process Dynamical Models with Product-of-Experts Kernels. In *Proceedings of the 24th International Conference on Artificial Neural Networks, LNCS 8681*. Springer, 603–610.

Dmytro Velychko, Benjamin Knopp, and Dominik Endres. 2018. Making the Coupled Gaussian Process Dynamical Model Modular and Scalable with Variational Approximations. *Entropy* 20, 10 (Sept. 2018), 724. https://doi.org/10.3390/e20100724

Jack Meng-Chieh Wang, David J. Fleet, and Aaron Hertzmann. 2008. Gaussian Process Dynamical Models for Human Motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 2 (Feb. 2008), 283–298. https://doi.org/10.1109/TPAMI.2007.1167

Margaret Wilson and Günther Knoblich. 2005. The case for motor involvement in perceiving conspecifics. *Psychological bulletin* 131, 3 (2005), 460.

WorldViz. 2019. Vizard 6.

Jeffrey M. Zacks and Khena M. Swallow. 2007. Event Segmentation. 16, 2 (2007), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x