Perceptual Validation for the Generation of Expressive Movements from End-Effector Trajectories

PAMELA CARRENO-MEDRANO, University of Waterloo
SYLVIE GIBET and PIERRE-FRANÇOIS MARTEAU, Université Bretagne Sud

Endowing animated virtual characters with emotionally expressive behaviors is paramount to improving the quality of the interactions between humans and virtual characters. Full-body motion, in particular, with its subtle kinematic variations, represents an effective way of conveying emotionally expressive content. However, before synthesizing expressive full-body movements, it is necessary to identify and understand what qualities of human motion are salient to the perception of emotions and how these qualities can be exploited to generate novel and equally expressive full-body movements. Based on previous studies, we argue that it is possible to perceive and generate expressive full-body movements from a limited set of joint trajectories, including end-effector trajectories and additional constraints such as pelvis and elbow trajectories. Hence, these selected trajectories define a significant and reduced motion space, which is adequate for the characterization of the expressive qualities of human motion and that is both suitable for the analysis and generation of emotionally expressive full-body movements. The purpose and main contribution of this work is the methodological framework we defined and used to assess the validity and applicability of the selected trajectories for the perception and generation of expressive full-body movements. This framework consists of the creation of a motion capture database of expressive theatrical movements, the development of a motion synthesis system based on trajectories re-played or re-sampled and inverse kinematics, and two perceptual studies.

CCS Concepts: • **Human-centered computing → User studies**; • **Computing methodologies → Perception**; *Animation*;

Additional Key Words and Phrases: Full-body movements, expressive motion analysis and synthesis, trajectories resampling, perceptual validation, emotion perception, motion representation

ACM Reference format:
Pamela Carreno-Medrano, Sylvie Gibet, and Pierre-François Marteau. 2018. Perceptual Validation for the Generation of Expressive Movements from End-Effector Trajectories. *ACM Trans. Interact. Intell. Syst.* 8, 3, Article 24 (June 2018), 26 pages.
https://doi.org/10.1145/3150976

1 INTRODUCTION

One of the main challenges in character animation is to design compelling characters capable of creating a more intuitive, engaging, and entertaining interaction with a user. Whether we play with these animated virtual characters, observe them, or control them, the quality of the interaction

This work was funded by the ANR InGredible Project No. ANR-12-CORD-001 (http://www.ingredible.fr).
The reviewing of this article was managed by associate editor Pelachaud, Catherine.
Q1 Authors' addresses: P. Carreno-Medrano (corresponding author); S. Gibet, P.-F. Marteau.

https://doi.org/10.1145/3150976

and our level of engagement strongly depend on how believable—that is, how well the animated character fits the user's expectations—these characters are perceived. Among the many factors that potentially relate to believable virtual characters, emotion awareness (i.e., the virtual character senses the human user's emotional state) and expression (i.e., the virtual character generates an appropriate and emotionally expressive behavior in response) have received increasing attention during in recent years [16, 35].

Humans perceive and express emotions through the combination of verbal and nonverbal modalities, such as speech, gaze, facial expressions, and bodily movements [39]. Within its movements, a virtual human character expresses emotion mainly through kinematic patterns that are characterized by the subtle variations in the manner movements are performed [18]. Hence, bodily movement represents simultaneously: (i) the source of information from which a virtual character can interpret the emotional state of the human user with whom the virtual character is currently interacting, and (ii) the means through which the virtual character will provide an appropriate and emotionally expressive response to the human user. Furthermore, since our bodies are the medium through which we experience the world around us, they represent a relevant modality to consider when designing engaging and satisfactory interactions within virtual environments and in the presence of animated virtual characters [33, 35].

Animating virtual human characters whose movements are perceived as expressive is difficult for mainly three reasons. First, the human body is a complex and high-dimensional mechanism. Even though animated characters are represented as simplified abstractions of this bio-mechanical system, generating varied, compelling, and plausible motions for those simplified models remains an open challenge. Second, despite extensive research on the important role of human motion in expressing emotions, there is still an incomplete understanding about what aspects of human movement are the most salient to the perception and expression of affect. Last but not least, there is a one-to-many correspondence between emotions and movements. That is, the same emotional state can be communicated through a large choice of kinematically distinct movements. Thus, determining the motion qualities that are common to different bodily expressions of the same emotional state is still an open challenge.

In this article, we adopt a simple, yet powerful, motion representation that we argue is both suitable for: (i) the control and generation of full-body movements for the virtual character, and (ii) the identification and encoding of most—if not all—the information that is salient to the expression of emotions. Furthermore, we expect this representation to be as independent as possible of the actions being performed, which makes possible to analyze its generalization capabilities across diverse motion categories (e.g., cyclic motions, functional actions, performing arts, etc.). And finally, this representation should be appropriate for real-time analysis, thus allowing for synthesizing expressive movements in an interactive process. The idea is not so much to find and use the best low-dimensional control space that allows production of a large variety of expressive full body motion (see, e.g., References [25, 46, 60]), which is in itself a challenging problem in 3D animation, but to investigate to what extent the representation we adopt is able to convey the expressive qualities of the movement. This representation, shared by the analysis and generation processes, consists of a limited set of joint trajectories comprising the extremities of the limbs of a human-like character, i.e., *head*, *hands* and *feet*, the equivalent of the human pelvis (*root joint*), and *elbow* trajectories (so-called target trajectories). We argue that from this relatively low-dimensional space it is possible to: (i) characterize the expressive content laying within their movements; and (ii) synthesize—with the help of additional constraints on joint limits—novel expressive movement for the virtual character. Furthermore, the intuitive nature of the chosen motion representation relies on the following aspects. On the one hand, such a representation can be expressed in a space close to the task space, as it may ease the animators' work in creating expressive gestures by specifying hand movement traces and applying inverse kinematics and/or inverse dynamics. This task space may also facilitate the expression of meaningful gesture components, such as iconic gestures describing a geometric shape or mimicking gestures. On the other hand, from these 3D traces, it is easier to edit and compose movements, comparatively to classical (linear or non linear) dimensional reduction methods, such as Principal Component Analysis (PCA), Principal Geodesic Analysis (PGA), or Gaussian-Mixture-based Methods (GMM), some of them being non intuitive (e.g., PGA and GMM), and producing non-desired artifacts. Third, mocap data can be easily retargeted in this space, so that facilitating the morphological adaptation to various actors and virtual characters.

The aim of the work presented in this article is to develop a methodological framework for perceptually assessing the relevance and usefulness of this representation. Within the proposed framework, our main contributions are: (i) two separate, yet complementary, user studies that evaluate, from a perceptual point of view, how informative is the proposed representation when compared to the full-body and how expressive are the full-body movements obtained from such a representation; and (ii) a synthesis system based on a function that maps low-dimensional trajectories to full-body movements and a re-sampling process suited to the MoCap data that produces non-observed and action-independent low-dimensional trajectories. Both user studies and the proposed synthesis systems were tested in the context of the analysis and synthesis of expressive theatrical bodily movements. We have chosen to work on sequences of theatrical body movements, because each movement within a sequence is deliberately chosen and executed so that the characters portrayed by the actors are perceived as believable.

In the following, we briefly survey the results from different research areas that motivated our hypothesis of end-effector trajectories as a suitable and compact motion representation for both the analysis and generation of emotionally expressive full-body movements. We later review the main approaches on the generation of expressive bodily motions. We then describe and discuss the methodological approach we have adopted to support our hypothesis. Next, we introduce a motion synthesis approach based on a re-sampling of the expressive Mocap Data that exploits this representation. We then discuss the perceptual evaluations that we have conducted to measure the suitability of the proposed representation as well as the expressive quality of the motions generated from it.

2 RELATED WORK

The work presented in this article is based on the idea that end-effector trajectories define a relatively compact and significant motion space that is simultaneously relevant for the analysis and synthesis of expressive full-body movements. In this section, we first review results from three different research areas that motivated our choice of representation. We then survey works on existing approaches for the generation of expressive full-body movements.

2.1 Why End-Effector Trajectories?

Despite the high dimensionality and complexity observed in human movement, results from different research areas suggest that due to the bio-mechanical and functional constraints ruling the human body, most human movements can be represented, with minimal loss of information, in a lower-dimensional space [19]. For instance, studies about the perception of biological motion have shown that human observers seem to effortlessly recognize and extract information about human motion from sparse body representations [7]. In particular, it has been observed that Point-light (PL) displays[1] provide enough information to make possible for humans the determination

[1]Body representation that consists of only a handful of markers attached to the head and main joints of the human body.

of gender [57], affective states [2], and the identity of individuals [58]. Similarly, computing simultaneously body shape and pose from sparse markers provides an efficient way to create lifelike animations [43]. Thus, it seems possible to parameterize human motion and all the nuanced expressive information conveyed by it through a limited set of motion trajectories, i.e, a spatial subset of body joints [20, 49]. In the following, we summarize the main research findings, which suggest that end-effector trajectories might be considered as one of such compact and informative motion parameterizations or representations.

Recent results on human motion and emotion perception have pointed out that although human observers seem to process expressive movements as a whole [50], different body parts convey different amounts of emotion-related motion features [49]. In particular, it has been observed that end-effectors are among the most relevant body parts for emotion perception and recognition. For instance, there is consistent evidence indicating that: (i) head and hand-arm movements are of significant importance for human observers when distinguishing between emotional states [37], and (ii) movement qualities critical for the perception of emotion from gait examples depend only on a small number of joints rather than on the whole body. Moreover, head and arms trajectories have been identified as the most important sources of information for perception of emotion and automatic analysis of expressive bodily motions, even when movements of the entire body are presented [50].

A survey of the most recent literature on automatic recognition of affect reveals that head and hand trajectories are among the bodily cues the most frequently and successfully used [26]. Bernhardt et al. [6] reported 50% recognition rates on upper-body functional movements (knocking motion) depicting neutral, happy, angry and sad emotional states. Bouënard et al. [8] analyzed expressive percussion gestures and found that a reduced dimensional representation consisting of motion features computed from hand trajectories was sufficient for accurately classifying new expressive percussion gestures. Glowinski et al. [24] found that it is possible to define a minimal representation of expressive upper-body movements by analyzing the kinematic qualities of head and hand trajectories. The resulting representation was later used to determine meaningful groups of emotions. Similarly, in our previous work [11], we showed that it is possible to automatically classify emotional movements from kinematic features computed from end-effector trajectories only.

In computer animation, the usefulness of end-effector trajectories for specifying a character's motion has been recurrently highlighted during recent years [13, 53]. From these trajectories, it is possible to generate whole body motions that smoothly follow the specified trajectories and exhibit the correct temporal variations. In the same way, three-dimensional motion sensors located on the two hands and pelvis positions may be used as inputs of a controller to build virtual character poses [38]. End-effector trajectories have also been successfully employed in other animation related areas such as motion indexation and retrieval. For example, Krüger et al. found that the use of richer representations give little or not advantage over the use of end-effector trajectories [41]. In motion compression, Tournier et al. [56] observed that it is possible to recover, with minimal information loss, whole-body poses and motions by specifying only the end-effector and root positions of a skeleton structure as input. However, these techniques have not shown that they can generate expressive movements whose emotional state could be perceptually recognized.

2.2 Generation of Expressive Body Movements for Virtual Characters

One of the constraints we considered when selecting a low-dimensional representation suitable for our context of application was the suitability of such representation for the control and generation of expressive movement for virtual characters. In the following, we review the two main existing approaches for animating a virtual human character with expressive movements: *rule-based methods*, generally coupled to procedural synthesis methods, and *example-based methods*.

Rule-based methods have their foundations on psychology studies and suppose that movement qualities important for the generation of expressive motions can be determined by analyzing how humans perceive emotions. The intuition is that by using perceptually validated motion qualities to guide the generation of the character's movements, the expressive qualities of the new motions will be ensured. New motions are synthesized through an ensemble of motion generation and editing rules that specify how the identified motion qualities can be mapped to low-level motion parameters. Some examples of approaches belonging to this category are: the EMOTE model for effort and shape proposed by Reference [14], the expressive gesture synthesis architecture for embodied conversational agents presented in Reference [27], and the supporting software inspired by the study of artistic performance literature built by Reference [48].

Example-based methods apply data-driven techniques based on machine learning methods to automatically determine and learn what movement aspects characterize the expressiveness in the captured data. This kind of methods can be further categorized into four main groups: **Motion blending**, in which new motions are generated through weighted interpolation of structurally similar (i.e., motions that depict the same action with different styles) but distinct motion examples (see, e.g., References [51, 55]). **Component models**, in which human motion is seen as the combination of many different, sometimes mutually independent, motion components. New motions are generated by exchanging, merging and/or interpolating one or several components according to some high-level constraints or control parameters (see, e.g., References [29, 52]). **Style translation**, in which an input motion is transformed into a new style while preserving its original content (see, e.g., References[34, 64]). The relevant transformations are estimated through the analysis of the differences between the emotionally expressive and non-expressive realizations of the same kinematic action, e.g., a neutral and a sad walk. **Stochastic generative models**, in which statistical models such as hidden Markov models (HMM) [9], stochastic adaptations of motions graphs [45], among others, are used to implicitly capture and model the content and the expressiveness of motions examples. In particular neural networks [31, 32, 54] have proven their high scalability and efficiency for controlling motion, but they need a large amount of data, offline computation, and they fail to deal with precise and expressive movements. The appeal behind these stochastic models is their capability to generate motions that are significantly different from the training data.

Each approach has its own merits and drawbacks. On the one hand, rule-based methods offer better control, which results in a higher flexibility and variability in the produced motions. However, these motions are often described as stiff and less visually appealing. Furthermore, the definition of the relations and rules that map control parameters to motion features and motion features to emotional states is a complex task [37]. On the other hand, example-based methods generate novel movements with a high level of details and a great realism. A MoCap database containing examples of perceptually validated expressive motions ensures that the new generated movements will be equally expressive. Nevertheless, the flexibility and variability of example-based methods entirely depends on the richness and vastness of the database. Furthermore, example-based methods often provide a very limited control over the possible output motions and styles, which is a limitation when working with virtual animated characters.

To benefit from the strengths and advantages of both approaches, while potentially addressing their shortcomings, we propose a synthesis approach that combines their main principles. First, we propose to use end-effector trajectories as control signals. They are intuitive, easy to specify, and can be directly mapped to whole-body motions. They also define a relatively low-dimensional motion representation that facilitates the construction of higher level ruled-based and editing approaches. Furthermore, since these trajectories will be extracted or generated from a labeled and perceptually validated dataset, using a statistical re-sampling method to synthesize new end-effector trajectories, we expect to preserve most of the visual appeal of example-based methods.

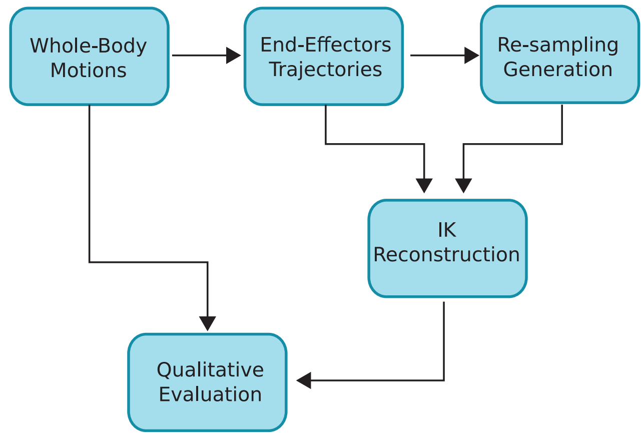


Fig. 1. Methodology used to evaluate end-effector trajectories.

Second, we propose to use inverse kinematics as the function that maps end-effector trajectories to whole-body motions. By doing so, we retain the control and flexibility inherent to rule-based approaches and decrease the dependency of the example-based methods on the motion database.

3 METHODOLOGICAL FRAMEWORK

The validation of the proposed hypothesis is assessed for the perception and synthesis of expressive bodily motions through the methodology illustrated in Figure 1. Two main elements are considered:

- **An expressive synthesis system from end-effector trajectories:** We need to define a function that maps the end-effector trajectories to the high-dimensional full-body space. This mapping should preserve all motion cues indicative of expressed emotions that are present in the end-effector trajectories. Furthermore, to assess whether the proposed representation generalizes beyond the movement categories in the MoCap database, it is necessary to evaluate full body motions generated from novel end-effector trajectories. To do so, we have proposed and implemented:
  - —An inverse kinematics (IK) reconstruction model. Since the geometry and configuration of an anthropomorphic limb is quite dependent on the limb's extremity position [41], we generate full-body motions by defining an inverse kinematic controller for each limb within the character's body. To overcome the redundancy of the system, further joint limits and elbow constraints were also added.
  - —A motion re-sampling scheme that can generate novel end-effector and pelvis trajectories while preserving the underlying emotional content. The main principle behind the proposed re-sampling scheme is the generation of random and semantically void trajectories whose content significantly differs from the original trajectories.
- **Perceptual evaluations:** Two perceptual studies have been conducted:
  - —First, a user study that provides a qualitative measure of how expressive end-effector trajectories are in comparison to whole-body motions. This study measures how observers' perception of emotions changed according to the type of motion representation we presented to them.
  - —Second, we perceptually evaluate the synthesized motions. More precisely, we measure the impact of the movement generation source (original MoCap database, movements

reconstructed from end-effector trajectories issued from the MoCap database, or movements generated using re-sampled trajectories) on the perception of emotion.

4 A MOTION CAPTURE DATABASE OF EXPRESSIVE THEATRICAL MOVEMENTS

In this section, we review the expressive motion capture database used to evaluate the methodological framework introduced in Section 3 and depicted in Figure 1. This database was initially reported in Reference [10] and further extended for the work presented here.

The proposed database was designed for the analysis and synthesis of expressive full-body movements. It contains several subjects, different types of movements in which all body limbs are employed, several emotional states and various repetitions for each possible combination of type of movement and emotional state. Furthermore, it was designed in the context of a form of theater—known as *physical theater*—that privileges the use of the human body and its wide variety of movements and postures to communicate emotions and interact with others.

Specifically, this database contains sequences from a mime-magician scenario in which skilled actors were asked to convey meaning and express emotions only through their body movements. In this scenario, each actor embodied a magician during performance. Three magician tricks: *the disappearing box*, *pulling a rabbit from a hat*, and *taking scarves from an empty jacket*, were to be performed under one of the following emotional states: *happiness*, *sadness*, *stress*, *relaxedness*, and *neutral*. A combined mood induction procedure (story-based and imagination-based MIP) was used to facilitate the enacting of the selected emotional states. It is important to notice that although the magician sequences include individual actions only observable during a magician performance, e.g, a bow toward the public or the use of a magical wand, other actions also present in these sequences are analogous to actions we execute on a daily basis, e.g, pick and place an object, clean a flat surface, among others.[2]

For the work presented here, two more motion sequences were added to the proposed MoCap database. These new sequences are not related to the mime-magician scenario and consist of: a walk example and an improvisation sketch that was freely chosen by each actor. With these additional motion sequences, we wished to extend the movement categories available in our database as well as increase the range of expressive movements to be considered in our evaluation of the proposed motion representation. In total, our database contains: movements from five actors (two females and three males, ranging in age from 38 to 54), five emotional states, examples of at least four different movement categories (i.e., magician actions, everyday life actions, locomotion, and theatrical improvisations), three repetitions of each magician trick for each emotional state by actor, and one repetition of the walk example and the improvisation sketch for each emotional state by actor. All motions were recorded at a rate of 200fps with a Qualysis motion capture system consisting of nine cameras. In total, we have approximately 206min of high-quality recorded full-body motion.

5 SYNTHESIS OF EXPRESSIVE WHOLE BODY MOTIONS

Below, we describe the components of the proposed synthesis system as well as the synthesis tasks we designed to evaluate this system.

5.1 Inverse Kinematics Controllers

When mapping low-dimensional control signal to high-dimensional whole-body movements, two types of approaches are usually privileged: data-driven reconstruction methods [13, 53] and *inverse*

[2]We refer the reader to the accompanying videos for some examples of the different sequences in our database.

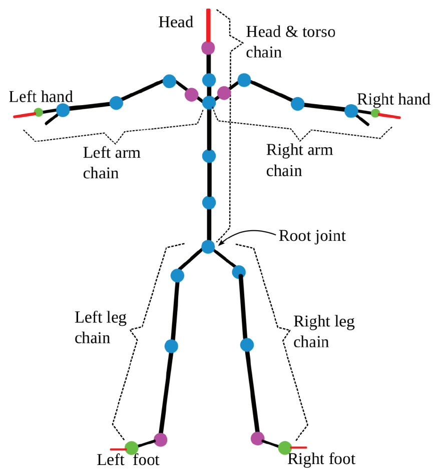


Fig. 2. Articulated chains and DOF controlled through IK. Red segments indicate end-effector associated to each articulated chain. Degrees of freedom of each joint are indicated as follows: green dots = 1 DOF, pink dots = 2 DOF, and light-blue dots = 3 DOF.

kinematics (IK) solvers. In the former, examples closely approximating the trajectories described by the control signals are retrieved from a large MoCap database. Those examples are later used to build local models capable of producing continuous and smooth whole body motions. In the latter approach, an iterative optimization solver computes the motion, posture by posture, for which the end effectors follow as smoothly and accurately as possible the control signals. Other than the knowledge about the hierarchical representation of the character's body and its degrees of freedom (DOF), an IK solver has no prior information about the expressive characteristics of the motions to be produced. Thus, if the resulting bodily motion exhibits variations related to a particular emotional state, it is because the relevant expressive cues were encoded in the control signals. For this reason, we have adopted an IK-based solver within our synthesis system.

Previously introduced in Reference [12], our inverse kinematics implementation consists of a set of independent IK controllers that map end-effector trajectories to whole-body motions. We have associated a controller to each one of the 5 articulated chains in the character's body. Figure 2 shows the five selected chains and their corresponding end-effectors (highlighted in red) as well as the number of DOFs associated to all joints within each chain. We observe that both leg chains have 12 DOFs each, arm chains have 15 DOFs each (including the clamp of the hand), and the head-torso chain has 17 DOFs. As the inverse kinematics is achieved simultaneously on five articulated chains, this gives this possibility of generating uncoordinated and unsynchronized movements and consequently offers great variability in the synthesized movements. Conversely, this makes the tasks more difficult to achieve. To constrain the space of possible IK solutions to more natural and humanly plausible postures, a combination of joint limits and multiple tasks for the arm chains (i.e., targets for both elbows and hands) were used during reconstruction.

5.2 Re-Sampling Scheme: Local Grid Bootstrap

To further evaluate the suitability of the proposed low-dimensional motion representation and its applicability to the generation of expressive whole-body motions, we need to produce and analyze new and sufficiently different expressive end-effector trajectories. To that end, we have selected and implemented a re-sampling procedure.

The main idea behind the use of a re-sampling method in our framework is to produce, from our limited set of capture movement data, new movement trajectories that share most of the statistical properties of the captured ones. We hypothesize that, as far as we select a convenient size for the time window from which the temporal dependence of the data is modeled, we are able to filter out most of the semantic dependency (i.e., the dependencies to the action that is performed) while keeping the expressiveness that characterizes the captured data.

Re-sampling methods share many similarities with Monte Carlo simulations. The main difference is that, in re-sampling methods, the simulated samples are drawn from the available data. In physics-related problems, Monte Carlo methods are known to be useful for simulating systems with many coupled degrees of freedom, such as fluids movement or strongly coupled solids [42]; this makes it particularly suitable for synthesizing articulated chain movements.

In this way, the *Local Grid Bootstrap* (LGB) bootstrap procedure proposed by Monbet et al. [47] and adopted in this work is a Monte Carlo method (MCM), since it shares a common objective: given some probability measure μ defined on some state space $\mathcal{S}$ (for us the space in which the movement data is embedded), generate as many random sequences from μ as required. The solution to achieve this objective is to construct a Markov chain with state space $\mathcal{S}$ keeping μ invariant. In other words, LGB, similarly to MCM, constructs a transition probability between states to create a stochastic time evolution for the system that has produced the observation data. Our choice to select LGB was motivated by its capability to:

- capture the temporal dependence of both MoCap data and end-effector trajectories,
- generate sequences whose length may be chosen independently from the length of the observed sequences, and
- produce unobserved states within the new sequences.

5.2.1 Definition and LGB procedure. If $X \in \mathbb{R}^d$ represents a complete posture or any given part of it (e.g., a specific articulated chain), then the LGB procedure constructs a Markov chain of order p (which is a parameter of the method) with state space $\mathcal{S}$, according to

$$P(X_{t+1} \in A|X_j,\ j \le t) = P(X_t \in A|X_j,\ t-p+1 \le j \le t), \tag{1}$$

where A is any compact subset[3] of $\mathbb{R}^d$. Hence, for all $t \in \mathbb{T}$, the state of the process $\{\mathcal{X}_t\}_{t\in\mathbb{T}}$ at time t is supposed to depend only on the p previous samples. Furthermore, the state space $\mathcal{S}$, on which this Markov chain is defined, is a subset of $\mathbb{R}^{dp}$ that contains all the subsequences (vector $Y = (X_t, X_{t-1}, \ldots, X_{t-p+1})$) in the captured data that will be re-sampled.

Then, the LGB re-sampling algorithm generates a new sequence $(\hat{X}_i)_{i\in\{1,\ldots,N\}}$, N being independent from the length T of the observed time series. Each new synthesized sample $\hat{X}_{t+1}$ is obtained by assigning probabilities to a finite subset of convenient states and sampling this subset according to a discrete probability distribution evaluated from a probability measure μ, which is defined using a local density kernel estimator (K_{dp}).

Let $\hat{Y}_t = (\hat{X}_t, \hat{X}_{t-1}, \ldots, \hat{X}_{t-p+1}) \in \mathbb{R}^{dp}$ be the state of the generated sequence at time t and let $\mathcal{S}_Y = \{Y \in \mathcal{S} | \exists i \text{ s.t. } Y_i = (X_i, X_{i-1}, \ldots, X_{i-p+1})\}$ be the set of all possible states that can be extracted from the observed sequence. Hence, the Markov chain constructed by LGB is a statistical model of the underlying dynamical process Φ responsible for the production of the observed sequence of movement:

$$X_{t+1} = \Phi(Y_t). \tag{2}$$

[3]That is, a closed and bounded subset.

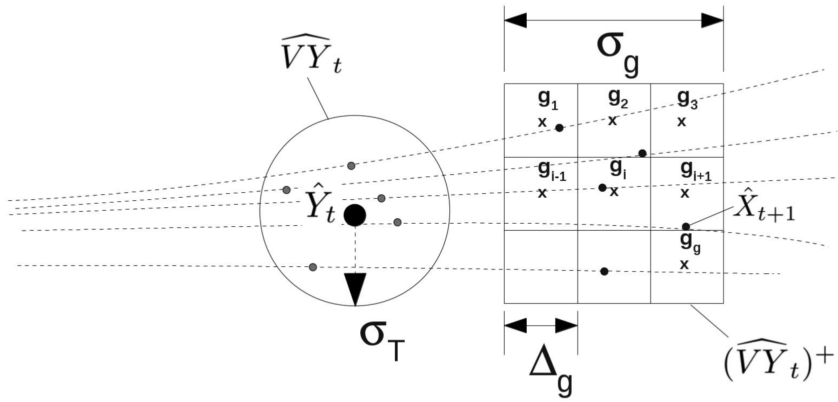


Fig. 3. Local grid bootstrap procedure.

The neighborhood $\widehat{VY}_t$ of $\hat{Y}_t$ is defined by the set of observed states that fall into the hyper-sphere centered on $\hat{Y}_t$ whose diameter is σ_T. $Y_l \in \{Y \in \mathcal{S}_Y | d(Y, \hat{Y}_t) \leq \sigma_T/2\}$.

Let $(\widehat{VY}_t)^+ = \Phi(\widehat{VY}_t)$ be the image of this neighborhood through the dynamic operator Φ. We define hG as the hyper-grid (i.e., the discretized hyper-cube) centered on the barycenter of $(\widehat{VY}_t)^+ \subset \mathbb{R}^d$ whose edge length is σ_g and discretization step Δ_g.

The LGB principle is schematized in Figure 3 and proceeds as follows:

(1) **Initialization step:** We select an initial state $\hat{Y}_1 \in \mathcal{S}_Y$, the width σ_T of the neighborhood of a given state and the two hyper-grid parameters σ_g and Δ_g.
(2) **Step $t + 1$:**
- Let us suppose that the state $\hat{Y}_t$ is already sampled. The neighborhood $\widehat{VY}_t$ of $\hat{Y}_t$ is extracted, depicted by the circle in the left side of Figure 3 and its image through ϕ, $(\widehat{VY}_t)^+$, evaluated.
- A hyper-grid hG_t is then constructed that embeds $(\widehat{VY}_t)^+$, as depicted by the square grid in the left part of figure 3, and a local density K_d is used to affect a probability of presence for each element of the hyper-grid. Basically, if numerous observed trajectories going through the neighborhood $\widehat{VY}_t$ fall inside element k of the hyper-grid, then the affected probability P_k to this grid element will be high. Conversely, if none or few trajectories going through the neighborhood $\widehat{VY}_t$ fall into grid element k, then its probability P_k will be small.
- The next generated state, $\hat{Y}_{t+1}$, is constructed as follows. First a grid element is randomly drawn according to the previous probability distribution. Then, $\hat{X}_{t+1}$ is defined as the barycenter of the state vector of the observed trajectories falling into the selected grid element, and this gives $\hat{Y}_{t+1} = (\hat{X}_{t+1}, \hat{X}_t, \ldots, \hat{X}_{t-p+1})$.

5.3 Generation of End-Effector Trajectories

To enhance LGB transition probabilities, closed and continuous end-effector trajectories are necessary. Such trajectories can be produced by smoothly transitioning between the beginning and end postures of each pair of motions in our database. Namely, given the motion sequences from which end-effector trajectories are extracted, for each targeted emotional state, we generated transition motions using the last and first $L = 250$ frames, respectively, of any two motion sequences. This

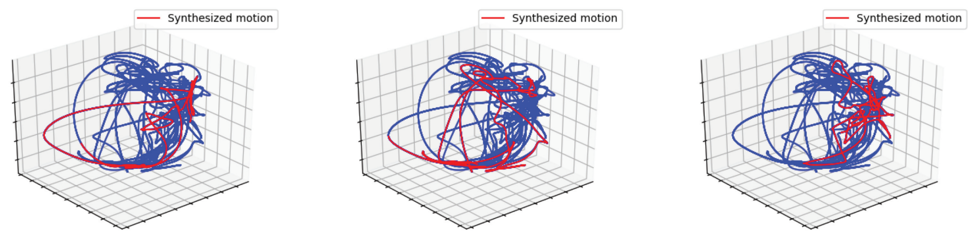


Fig. 4. Three original trajectories resampled from a same 3D manifold corresponding to the "right-arm" end effector.

value was selected such as to ensure that all individual actions within any magician sequence are completed before and after transitioning from one sequence to another. Once all transitions have been generated, the continuous end-effector (including elbow positions) and pelvis trajectories necessary for LGB re-sampling are computed for each articulated chain.

Each new time series generated by LGB re-sampling scheme consists of four processes characterized respectively by: the 3D head, the 6D right elbow-hand, the 6D left elbow-hand, the 6D+3D lower body trajectories. Each of these processes is independently re-sampled according to the LGB procedure for which X_t corresponds to the 3D or 6D trajectories. In the case of the leg chains, we have sampled them together to guarantee the character's stability during motion. Additionally, since the motion of the character's body root joint highly depends on the leg chains displacement and vice versa, every time a new state is sampled, the corresponding root's position is taken from the observed sequence $X = (X_i)_{i \in \{1,\ldots,T\}}$ associated to that sampled state.

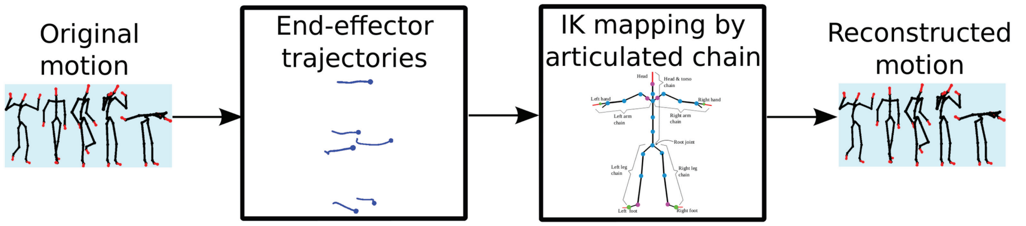
The order of the Markov process representing each observed sequence has been empirically adjusted to $p = 1$ for the 6D elbow-hand and 6D+3D lower-body trajectories, and to $p = 3$ for the 3D head trajectories. This choice was made such that the density of the area explored by each articulated chain can be sufficiently well estimated given the available data and so that the synthesized trajectories were reliable and smooth enough after considering the number of points and their dimensionality by emotional state.

For each process the grid parameter σ_g was locally adjusted to maintain between 50 and 100 observations inside the neighborhood $\widehat{VY}_t$. The parameter h_T was set up to $\sigma_g/3$, which seems to nicely fit with the implementation of a grid that has 3 subdivisions along each of the dimensions that are considered. A Kd-tree [5] is used to index $p.d$ dimensional samples collected along the observed end-effector trajectories. The search for the neighborhood of the current $\hat{Y}_t$, which conditions the algorithmic complexity of the method, is near logarithmic with the size of the data as far as $p.d$ is sufficiently small, basically below 20.

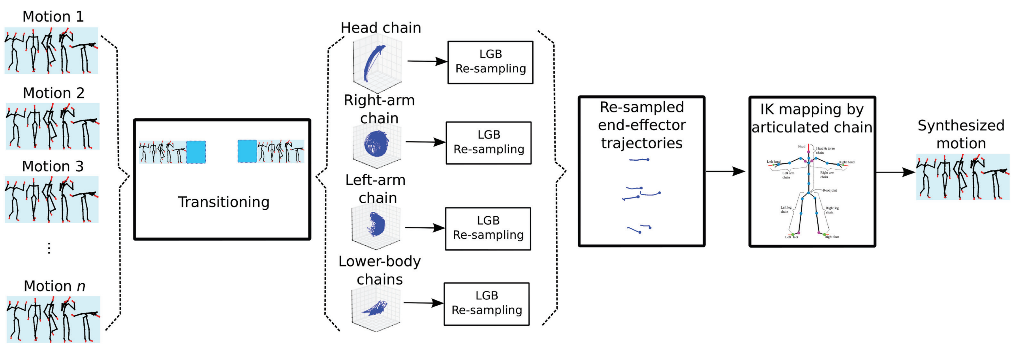
Figure 4 presents three runs of the resampling procedure. Each run corresponds to a random walk on the manifold and provides as a result a new trajectory (i.e., never observed in the Mocap data) shown in red color on the figure.

6 SYNTHESIS TASKS

We have proposed two distinct yet complementary synthesis tasks: *motion reconstruction* and *motions from re-sampled trajectories*. In the first task, we seek to evaluate whether the proposed IK implementation generates motions that are similar to those from which the end-effector trajectories guiding the reconstruction were extracted. In the second task, we want to assess whether the full body movements obtained from the randomly sampled trajectories show the same expressive patterns than the motions in our database. Both tasks are considered for each one of the five emotional states in the database, i.e., *happiness, neutral, relaxedness, sadness,* and *stress.*



(a) Motion reconstruction task



(b) Motions from re-sampled trajectories task

Fig. 5. Synthesis tasks proposed to evaluate suitability of end-effector trajectories for the generation of expressive full-body movements.

6.1 Motion Reconstruction

This task aims to evaluate the quality of the whole-body motions reconstructed through the combination of observed end-effector trajectories and IK-based motion controllers. We proceed as shown in Figure 5(a). For each original motion, we extract the trajectories of the end-effectors associated to each articulated chain (see Figure 2) as well as the trajectory of the root joint. Additionally, we also consider the elbow trajectories for both arm chains to further constraint the space of possible solutions. Each trajectory is expressed with respect to the coordinate frame system associated to its respective articulated chain. Furthermore, all trajectories are invariant to the position of the character's body in world space. Once all target trajectories have been defined, the reconstructed motion is generated, posture by posture, using the proposed IK-based mapping.

6.2 Motions from Re-Sampled Trajectories

This task aims to fully assess whether expressive end-effector trajectories are suitable control signals for generating new expressive body movements. New expressive end-effector trajectories are generated for each target emotional state using LGB re-sampling scheme. Once we have synthesized target trajectories for each one of the articulated chains in the character's body, we proceed to generate a new motion, posture by posture, using the proposed IK-based mapping. The main difference between our two synthesis tasks lies in the origin of the end-effector trajectories guiding the reconstruction process. All the steps involved in this synthesis task are illustrated in Figure 5(b).

7 PERCEPTUAL EVALUATIONS

Since human perception of emotionally expressive body movements is highly subjective, we used perceptual studies to evaluate: (i) the expressive information conveyed by end-effector trajectories compared to full-body representations, and (ii) the quality and expressiveness of the synthesized motions compared to the motions directly observed in our MoCap database.

7.1 End-Effector Trajectories vs Full-Body Motions

This study analyzed the impact of the body representation—full-body vs. end-effector trajectories—on the perception of emotion. This evaluation was conducted using a subset of the motion capture database presented in Section 4.

7.1.1 Study Design. To avoid any possible carry-over effect between representations and guarantee that participants remained naive to the purpose of the study, participants were randomly assigned to either the full-body or the end-effector trajectories condition. Similarly, to avoid the effects of boredom and fatigue, we reduced the number of sequences to evaluate from 250 to 125. Only one repetition for each of the 5 sequence examples in the database was considered by actor and emotion (5 sequences $\times$ 5 emotions $\times$ 5 actors = 125 sequences to rate). The selected sequences were later randomly assigned to one of 5 groups in such a way that all emotions were depicted in each group, and each group included at least two different sequences and actors. The assignment was done so that from the initially selected 125 sequences a single participant rated no more than 25 sequences.

Video clips at 30fps were created for each selected sequences, one for each representation. Hence, a total of 10 groups were constituted. For each video clip, the character representation was placed at the center of a 3D virtual space and facing the virtual camera at approximately 45 degrees. For the end-effector trajectories, we displayed the 3D position of the selected joints along with the trace of their respective trajectories. For the full-body condition, stick figures were used instead.

Participants were asked to answer the following questions for each clip:

(1) *"Which of the 5 listed emotions do you think is conveyed through the motion?"* A forced-choice question was used, since we were only interested in assessing whether the sequences in the database effectively conveyed the target emotions.
(2) *"How do you qualify the emotion conveyed in the video?"* Participants had to rate the arousal and valence components of the emotion perceived for each movement on a scale from 1 to 7.

A total of 200 participants, 98 women and 102 men, ranging in age from 21 and 75, were recruited through Amazon Mechanical Turk (MTuk) service. They were randomly assigned to one of the 10 groups. The same participant could not be part of more than one group. We had 20 participants by group.

7.1.2 Detection of Outlier Participants. One of the main concerns when using crowd-sourcing services such as Amazon Mechanical Turk is how to ensure the quality of the answers submitted by the participants. Since the participants are not in a controlled environment and within the reach of the experimenter, it becomes harder to ensure that the participants understand the task they are asked to accomplish and that they do it with the care, diligence and seriousness expected by the experimenter [63].

Among the strategies commonly used to evaluate the quality of the answers submitted by MTurk workers, we have adopted an approach similar to the one suggested by Feng et al. [21]. With this approach, it is possible to improve the quality of the answers we later use in our analysis without

Table 1. F-statistics, p-values, and Effect Size (η^2) Results from Two-way Mixed Repeated Measures ANOVAs for Main Effect of Intended Emotion and Representation

	Ratings(i):	*Accuracy*	*Valence*	*Arousal*
	F(4, 752) =	34.918	71.452	193.713
$H_0(1, i)$	p-value =	**<0.001**	**<0.001**	**<0.001**
	η^2 =	0.406	0.502	0.713
	F(1, 188)=	54.308	2.477	29.732
$H_0(2, i)$	p-value =	**<0.001**	0.117	**<0.001**
	η^2 =	0.224	0.013	0.136
	F(4, 752) =	2.951	14.16	12.981
$H_0(3, i)$	p-value =	**0.019**	**<0.001**	**<0.001**
	η^2 =	0.061	0.175	0.136

directly influencing participants answers or changing the manner in which they approach the task as it is the case with catch trial questions[4] [28].

Specifically, we use inter-rater agreement measures to automatically detect outlier participants and improve the general quality of the collected answers. We start by computing the agreement coefficients of all participants within the same group, we then identify which participants are outliers with respect to the group they belong to by using the Tukey's method [59]. That is, all participants whose agreement scores lie outside a determined interval are considered as outliers. We have used the interval defined by

$$I = [Q_1 - k \times (Q_3 - Q_1),\ Q_3 + k \times (Q_3 - Q_1)]\ with\ k = 1.5, \tag{3}$$

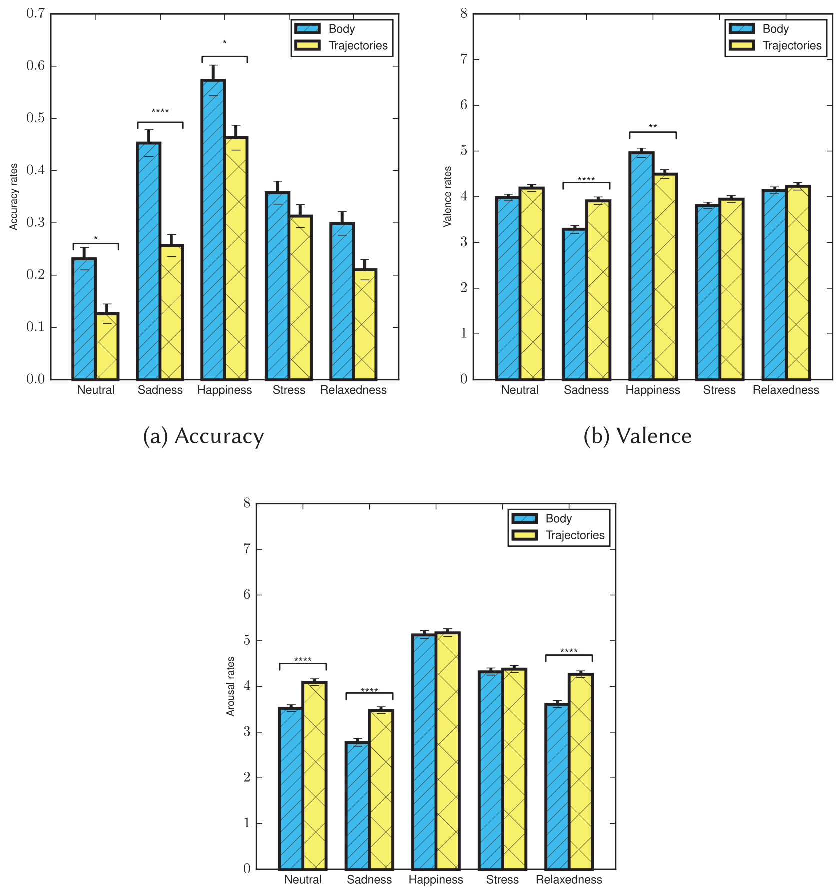
where Q_1, Q_3 correspond to the first and third quantiles of all agreement coefficients computed within the same group. Using this procedure, among 200 participants who took part of the study, 190 were finally retained (10 were estimated as outliers, one for each group).

7.1.3 Results. We evaluated the main and interaction effects of motion representation (i.e., full-body or end-effector trajectories) and intended emotion on the perception of expressive body movements using two-way mixed repeated ANOVA measures. We defined *representation* as a between-subject factor and *intended emotion* as a within-subject factor. With $i = \{$*accuracy, valence, arousal*$\}$, below we list the null hypotheses tested in this study:

- $H_0(1, i)$: The means of the participants' ratings of i for the different intended emotions are equal.
- $H_0(2, i)$: The means of the participants' ratings of i for full-body and end-effector trajectories stimuli are equal.
- $H_0(3, i)$: Representation type and intended emotion are independent factors and no interaction between the two is present on the participants' ratings of i.

The resulting F-statistics (with Greenhouse-Geisser correction if necessary), p-values, and effect sizes (η^2) are listed in Table 1. Effects and interactions were evaluated at a significant level of $\alpha = 0.05$ (highlighted p-values). We present in Figure 6 the average accuracy rates, valence, and arousal ratings across motion representations for all intended emotional states. Significant pair-wise differences (post-hoc paired Tukey-HSD tests with Bonferroni corrections) are also indicated.

[4]Questions with obvious answers are presented at specific points during the perceptual study.



(a) Accuracy

(b) Valence

(c) Arousal

Fig. 6. Mixed two-way ANOVA accuracy and other ratings for end-effector trajectories vs full-body movements study. Significant differences were labeled using the following convention: ****: p < 0.0001; ***: p < 0.001; **: p < 0.01; *: p < 0.05.

7.1.4 Analysis 1: Effect of Intended Emotions. Intended emotion was found to have a significant effect on the participants' perception of expressive movements for all ratings i. Hence, we reject $H_0(1, i)$ for all dependent variables. Similarly, the effect of intended emotion is large ($\eta^2 > 0.16$) in all cases, indicating that the conveyed emotion was a critical factor on participants' answers. Follow-up post-hoc analyses show that accuracy for all emotional states is above chance level (20%) for full-body representation. In contrast, only four of the five emotional states reported recognition rates superior to chance for the end-effector trajectories condition. Significant pairwise differences between participants' accuracy were found at a significant level of $p < 0.01$ for most of the intended emotions. Specifically, we observe that for both representations, *happiness* registered the highest accuracy rates followed, in no particular order, by *sadness* and *stress*. However, no significant difference (p < 0.05) between the accuracy rates of *relaxedness* and *neutral* were found for both representations.

Regarding valence ratings, we found that for full-body representations, both *happy* and *sad* movements have valence ratings significantly different (p < 0.001) from the other emotions. As both *happiness* and *sadness* were more accurately recognized than other emotional states, assessing their valence was possibly a much easier task. No significant differences were found for the valence rating of *neutral*, *relaxedness* and *stress* ($p > 0.05$). For end-effector trajectories only pairwise differences between *happiness* and emotional states with a negative valence, i.e., *sadness* and *stress*, were found statistically significant ($p < 0.001$).

Post-hoc analysis of arousal ratings showed that for the full-body representation there are significant differences ($p < 0.0001$) for almost all pairwise comparisons. No notable difference between the arousal ratings of *neutral* and *relaxedness* emotional states was observed. This result suggests that the examples of *neutral* and *relaxed* motions in our database might be kinematically very similar, thus making it harder to perceive and rate the subtle differences of activation between them. In the case of end-effector trajectories, the pairwise differences on arousal ratings follow the same patterns observed with participants' accuracy rates. Namely, arousal ratings for *happiness* and *sadness* were found to be significantly different ($p < 0.0001$). For the other emotional states, participants do not seem to have perceived any difference on arousal levels.

7.1.5 Analysis 2: Effect of Representation and Interactions. Representation was found to have a significant effect only on accuracy and arousal ratings. Hence, we reject the null hypotheses $H_0(2,\ accuracy)$ and $H_0(2,\ arousal)$. From Table 1, we observe that the main effect of representation is large on participant's accuracy ($\eta^2 > 0.16$) but medium ($0.06 < \eta^2 < 0.16$) on arousal ratings. On the one hand, the large effect on accuracy indicates that the body representation presented to the observers had a significant impact on their perception of expressive movements and emotions. On the other hand, the medium effect on arousal ratings suggests that representation alone can not account for all the variability observed on the perception of the activation and kinematic patterns of our expressive motions.

Although ANOVA tests reported that representation only had a significant effect on accuracy and arousal ratings, we found that interactions between this factor and intended emotion were tested as statistically significant at $p < 0.05$ for all dependent variables. Thus, we reject $H_0(3, i)$ for all dependent variables. These interactions registered, in average, a medium effect size ($\eta^2 \approx 0.124$), indicating that the interplay between the type of representation and the intended emotion is responsible for approximately 12% of the variability observed on participants' perception of expressive bodily emotions.

Follow-up post-hoc tests showed that the main differences in accuracy are observed on the perception of motions depicting *sadness, happiness,* and the *neutral* state. As shown in Figure 6(a), we found that the recognition rates of these three emotional states consistently decreased with the change of representation. This effect is much more significant for *sad* movements and for the *neutral* state. The former showed a loss of accuracy of approximately 20%, while the latter was no longer recognized above chance level for the end-effector trajectories condition. *Happiness* accuracy also decreased, but it remained the best recognized emotional state among participants.

Regarding arousal, we observed that motions depicting *neutral, sadness* and *relaxedness* emotions obtained higher ratings when the end-effector trajectories representation was used. Arousal ratings for *happiness* and *stress* remained the same across the two types of representations, indicating that emotional states characterized by low-activation seem to be perceived as more energetic when representations with reduced body information are employed.

For the perception of valence, we observed statistically significant differences ($p < 0.01$) between the average valence ratings of *sadness* and *happiness* (see Figure 6(b)). While valence ratings of motions depicting *happiness* registered a significant decrease between both representations, *sad* motions were rated with a higher valence level when depicted through end-effector trajectories. No significant differences between the valence ratings of *neutral*, *stress* and *relaxedness* emotional states for both representation were found (see Figure 6(b)).

7.1.6 Findings Summary. Overall the change from a richer, i.e., full-body, to a sparser representation, i.e., end-effector and pelvis trajectories, was found statistically significant for two of the three dependent variables we analyzed. In summary, we observed that:

- All intended emotional states were recognized above chance level (20%) for both representations. Only the recognition rates of motions depicting the *neutral* state fell below chance level with the change of representation.
- For both representations, *happiness*, *stress*, and *sadness* were among the best recognized emotional states. Similarly, the recognition rates obtained for *neutral*, *happiness*, and *sadness* decreased the most for the end-effector trajectories representation.
- The arousal ratings of emotional states associated with high activation levels such as *happiness* and *stress* did not significantly change across representations. However, for emotional states characterized with low activation levels, i.e., *sadness* and *relaxedness*, the arousal ratings obtained for the end-effector trajectories representation were significantly higher than those obtained for the full-body representation.
- The valence ratings participants gave to motions depicting *stress*, *relaxedness* or the *neutral* state did not significantly changed across both representations. However, the valence ratings of *happy* and *sad* motions respectively decreased and increased for the end-effector trajectories representation.
- Independently of the representation being used, the similarity between the average valence and arousal ratings that participants associated to opposed emotional states, e.g., *relaxedness* and *stress* seem to suggest that some of our motion sequences and actors failed to successfully convey the intended emotional states.

7.2 Perceptual Evaluation of Synthesis Tasks

In this user study participants rated the emotional content and expressiveness of the motions obtained from three sources. The first source, hereinafter referred to *MoCap*, corresponds to the original motions performed by an actor. The other two sources correspond to the synthesis tasks described in Section 6 and are hereinafter respectively referred to as *IK reconstruction* and *re-sampled trajectories*. The results obtained for the actor's motions provide a base-line of perceived expressiveness against which we can compare the perception rates obtained from the synthetically generated motions.

7.2.1 Study Design. For this study, we used the MoCap recordings of a single actor. This choice was motivated by the significant differences on emotion perception we observed across actors in our previous study. The data we used consists of 24 motion sequences from the magician scenario: 6 sequences depicting happiness, 6 realizations for the neutral state, 4 sequences for relaxedness, 5 sequences for sadness, and 3 examples for stress. From these sequences, we obtained the full-body animations used as base-line as well as the end-effector, elbow and pelvis trajectories necessary to control the synthesis tasks previously described in Section 6. Figure 7 shows an example of the end-effector trajectories used as control signals and the postures generated from them.

A total of 10 video clips at 30fps were created for each movement generation source. For the *MoCap* source, two realizations by emotional state were randomly selected among the 24 available sequences. In the case of the stimuli belonging to the *IK reconstruction* source, video clips were generated by applying the already described IK controllers on the end-effector, elbow and pelvis trajectories extracted from each one of the 10 realizations representing the *MoCap* source. The stimuli associated to the third source, i.e., *re-sampled trajectories*, were obtained through a three-step process. First, for each emotional state, 30,000 *frames* long synthesized full-body motions were generated using the procedure introduced in Section 5.2. Second, a group of short motion candidates was

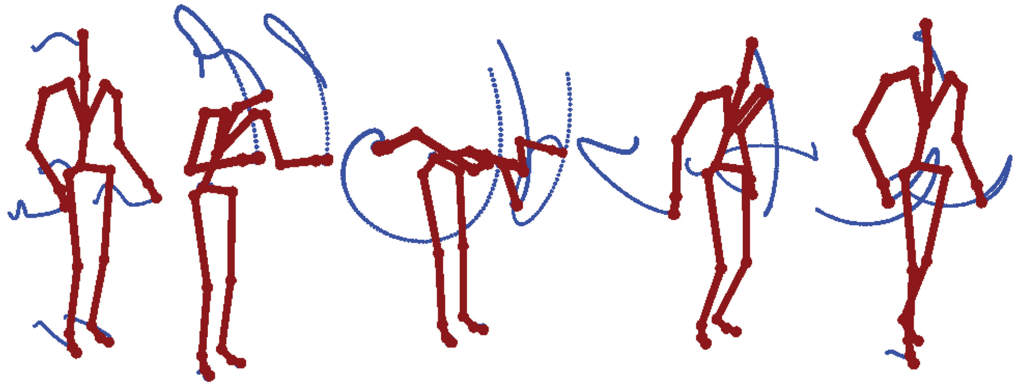


Fig. 7. Example of end-effector and root joint trajectories used as control signal and the whole-body postures generated from them.

generated by selecting several motion segments from the 30,000 *frames* long synthesized motions. These candidates were chosen such as to limit the number of rest poses within the synthesized motions and to reduce strong visual artifacts. Third, two annotators were asked to choose two most dynamic and variable motion segments by emotional state. Videos for which there was total agreement between the annotators were automatically selected. If there was no agreement, then we randomly selected one or two videos from the lists provided by the annotators for each emotional state.

Participants were randomly assigned to one of the movement generation sources being evaluated; the same participant could not be appointed to more than one source. Participants were presented with each video twice. Video clips were presented in random order; however, we made sure that all video clips to be evaluated were rated once before showing any video a second time. Participants answered the same questions used in our first study.

A total of 72 participants (24 by motion generation source) took part in this study. In total, we had 35 women and 37 men, ranging in age from 22 and 69 years old. We used MTuk as an intermediate platform for recruiting participants and conducting our study. Since each video clip was rated twice, each participant was presented with 20 videos in total. Participants took in average approximately 25min to answer all questions. Participants whose agreement score within their group was outside a determined interval were marked as outliers. Among the initial 72 participants, only two outliers were detected.

7.2.2 Results. A two-way mixed repeated measures ANOVA was used to evaluate the main and interaction effects of intended emotion and movement generation source on the perception of emotionally expressive body movements. In this analysis, emotion and generation source where modeled as within-subject and between-subject factors respectively. Three ANOVAs tests were performed on the average accuracy, valence, and arousal ratings across participants. With $i = \{accuracy, valence, arousal\}$, we list below the null hypotheses tested in this study:

- $H_0(1, i)$: The means of the participants' ratings of i for the different intended emotions are equal.
- $H_0(2, i)$: The means of the participants' ratings of i for the different movement generation sources are equal.
- $H_0(3, i)$: Movement generation source and intended emotion are independent factors and no interaction between the two is present on the participants' ratings of i.

Table 2. F-statistics, p-values, and Effect Size (η^2) Results from Two-way Mixed Repeated Measures ANOVAs for Main Effect of Intended Emotion and Generation Source

	Ratings(i):	*Accuracy*	*Valence*	*Arousal*
	F(4, 268) =	40.158	22.484	146.168
$H_0(1, i)$	p-value =	**<0.0001**	**<0.0001**	**<0.0001**
	η^2 =	0.775	0.586	0.874
	F(2, 67) =	0.627	4.424	0.319
$H_0(2, i)$	p-value =	0.538	**0.015**	0.728
	η^2 =	0.018	0.117	0.009
	F(8, 268) =	1.146	1.749	0.663
$H_0(3, i)$	p-value =	0.332	0.099	0.684
	η^2 =	0.054	0.098	0.044

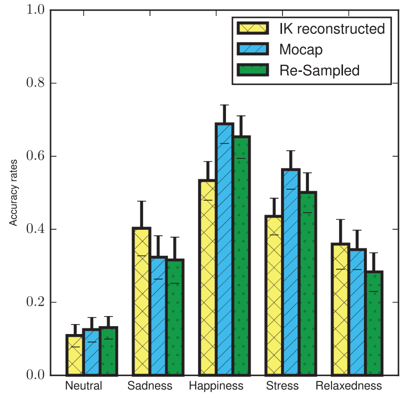
In the same manner, the resulting F-statistics (with Greenhouse-Geisser correction if necessary), p-values and effect sizes (η^2) are listed in Table 2. Effects and interactions were evaluated at a significant level of $\alpha = 0.05$ (highlighted p-values). Furthermore, we present in Figure 8 the average accuracy rates, and valence and arousal ratings across motion generation sources for all intended emotional states. Significant pair-wise differences are also indicated.

7.2.3 Analysis. Intended emotion was found to have a significant effect in all dependent variables ($p < 0.0001$); hence $H_0(1, i)$ for $i \in \{accuracy, valence, arousal\}$ are rejected. This effect is large in size ($\eta^2 > 0.16$), which suggests that the variance observed on participants' ratings is mainly due to the emotion being conveyed by the character's motion. This detected main effect suggests that the motions generated by our synthesis tasks retained almost, if not all, the subtle cues necessary to perceive and differentiate all intended emotional states.

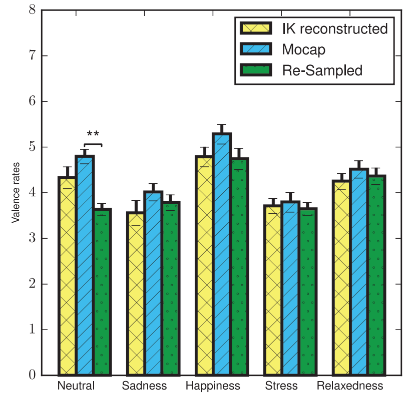
Similarly, we observed that movement generation source and the paired differences in the participants' ratings of accuracy and arousal (see Figures 8(a) and 8(c)) were not found significant ($p > 0.05$) . Therefore, we retain $H_0(2, i)$ for $i \in \{accuracy, arousal\}$, which states that the average ratings of movements from different generation sources are equal. However, since a significant effect of generation source on valence ratings was found ($p < 0.05$), we reject $H_0(2, valence)$ in favor of the alternative hypothesis.

A further analysis of the effects of movement generation source suggests that although this factor had little impact on the variability observed in participants' perception of the intended emotions and their associated activation levels, it still produced different valence ratings for the same intended emotions. Follow-up post-hoc tests showed that the most significant effect of movement generation source on valence ratings ($p < 0.01$) was obtained for the motions conveying a *neutral* emotional state and that were generated using *re-sampled trajectories* (see Figure 8(b)). However, it is important to notice that these perceived differences had little impact on the perception and recognition of emotionally expressive bodily motions belonging to this emotional state for all sources (each source registered 13% average recognition rate as shown in Figure 8(a)).

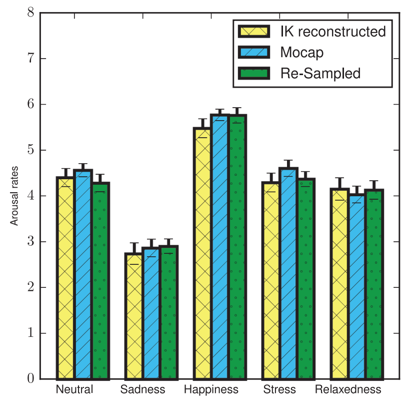
Finally, regarding our third set of null hypotheses $H_0(3, i)$ on the interaction of movement generation source and intended emotion, we found no significant effect ($p > 0.05$) on participants' accuracy, valence and arousal ratings. Hence, we retain $H_0(3, i)$ for $i \in \{accuracy, valence, arousal\}$, which states that intended emotion and movement generation source factors are independent and that when combined they have no effect on the mean participants' ratings.



(a) Accuracy



(b) Valence



(c) Arousal

Fig. 8. Mixed two-way ANOVA accuracy and other ratings for all movement generation sources. Significant differences were labeled using the following convention: ****: $p < 0.0001$, ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$.

7.2.4 Findings Summary. Overall, when comparing participants' perception of the expressive cues conveyed by full-body movements produced by three distinct movement generation sources, we observed that:

- The accuracy rates obtained for motions generated by either *IK reconstruction* or *re-sampled trajectories* sources were found to be not significantly different from the rates obtained for the original *Mocap* sequences. This suggests that the proposed synthesis framework seems to encode most of the motions cues that are necessary for the perception of the different intended emotions.
- The arousal and valence ratings participants associated to the almost all intented emotional states did not significantly change across the different movement sources, indicating that the movements generated with the proposed synthesis tasks are perceived very similarly to the original motions executed by the human actor.
- Once again, four of the five target emotional states were recognized above chance level (20%). Moreover, independently of the movement generation source, the neutral state was

perceived as conveying some particular expressive content even though it was initially intended to convey no expressive content at all.

8 GENERAL DISCUSSION

In this article, we stated and challenged the hypothesis that end-effector trajectories form a compact and informative motion representation that is both suitable for the analysis and synthesis of emotionally expressive full-body movements. The necessity of a motion representation with this dual analysis-synthesis functionality was motivated by the context of application of the *Ingredible* project. To prove our claim, we conducted two separate, yet complementary, user studies.

In the first study, described in Section 7.1, we evaluated how participants' perception of the expressive content conveyed through full-body movements changed according to the type of representation, i.e., end-effector trajectories or full-body stick figures, that we showed them. This study provided us with a qualitative measure of the loss of information we might have incurred with the change of motion representation. The results obtained for this study showed that participants presented with end-effector trajectories—28% average recognition rate—were in average 10% less accurate than the participants who rated motions depicted by full-body stick figures—38% average recognition rate. In spite of this loss of accuracy, it is important to notice that although observers were less accurate in their judgments of the emotional states conveyed through end-effector trajectories, four of the five intended emotional states were still recognized above chance level. This observation is in accordance with results obtained in previous studies about the contribution of form and motion information to the recognition of emotions from bodily movements. In References [3, 17], it was observed that in the light of significantly disturbances of form-related motion cues, kinematic information alone can help to distinguish basic emotions above chance level. Thus, although the use of end-effector and pelvis trajectories alone may considerable weaken form and shape cues previously identified as relevant for emotion perception (e.g., head orientation, elbow bend, curvature of spine, etc. [15]), the motion qualities of these selected trajectories (e.g., quantity and duration of the movement; the observed velocity, acceleration and jerkiness of a movement; among others [39]) still convey relevant motion cues for the perception of the intended emotional states above chance level. Overall, from this first study, we can conclude that end-effector and pelvis trajectories seem not to be as perceptually and visually rich as full-body representations. However, the impact of this observation is attenuated by the results we obtained from the evaluation of movements partially synthesized from end-effector trajectories.

In our second and last study, we aimed at evaluating the suitability of end-effector trajectories for the synthesis of expressive full-body movements. We hypothesized that although, perceptually speaking, end-effector and pelvis trajectories seem not to be as informative as full-body representations, numerically they still encode enough expressive content to successfully control the generation of full-body expressive movements. For this purpose, we measured whether observers' perception of expressive content would be influenced by the source from which full-body movements were obtained. We considered three possible sources: MoCap data, motions reconstructed using end-effector and pelvis trajectories extracted from the MoCap data, and motions generated from re-sampled control trajectories. The results from this study showed that bodily movements generated either from observed or synthesized end-effector and pelvis trajectories are perceived as being as expressive as the movements produced by a human actor. This indicates that although emotion perception was impaired when only end-effector and pelvis trajectories were displayed, this impairment was not longer present when rating full-body motions generated using these trajectories and our synthesis system. Hence, end-effector and pelvis trajectories seem to encode most of motions cues that are necessary for both the numerical analysis and generation of expressive full-body motions.

The results obtained from this second study are arguable from the point that our synthesis tasks did not only receive end-effector and pelvis trajectories as input and control signals. As previously indicated in Section 5.1, due to the redundancy and increased number of DOFs in the kinematic chains we defined as the characters' arms, we decided to additionally provide elbow trajectories. In this case, we employed elbow trajectories as an IK solver constraint analogous to the commonly used end-effector's orientation, the main purpose of which is to limit the space of possible IK solutions to more natural and humanly plausible postures. However, it is likely that by providing both hand and elbow trajectories as control signals, our IK solver was implicitly capable of replicating some of the form and shape cues that were missing in our first study and that from end-effector trajectories alone it might not be possible to obtain. Thus, our claim that end-effector trajectories are a sufficient representation for the generation of expressive full-body motions should be revisited. Thus, given our results and the addition of elbow trajectories to our synthesis framework, we can conclude that as long as expressive end-effector and elbow trajectories guide the synthesis process, emotions are still successfully conveyed and perceived in the resulting synthesized motions.

We are also aware that the significance of the conclusions presented in this article is limited by the relatively low accuracy rates we obtained for both user studies. Although our participants mostly showed perception rates above chance level, we still observed that the motion sequences within or generated using our database were less accurately recognized than movements considered in other existing databases (see, e.g., References [40] and [36]). We have identified several possible reasons for these rates. First, our choice of emotional states. Our database includes examples of two emotional states, i.e., *relaxedness* and *stress*, that are not often considered when investigating the perception of emotions from motion and/or when designing expressive MoCap corpora. It is probable that additional contextual factors (e.g., who the character is, where she/he is, what her/his current task is) and visual cues such as facial expressions, gaze direction, and so on, are needed for a more accurate identification of these two emotional states [26]. Second, our choice of scenario and motion embodiment. Gunes et al. [26] pointed out that the understanding of the action being performed is critical for emotion perception from human body motion. It might be possible that the motion sequences and individual actions contained in our database were not fully understood by the observers that took part of our studies, which in turn penalized the accuracy rates we obtained from them. Furthermore, although several studies have shown that humans are capable of perceiving emotions from less detailed representations and/or embodiments [44], it is still possible that by presenting participants with stick-figures displays, some of the subtle cues that make the discrimination between the intended emotions much easier were not longer present in our video-clips [30]. Finally, our choice of emotion elicitation over emotion portrayal might be critical. During our MoCap recordings, actors were given a short scenario that helped them to better contextualize, interpret, and enact each of the intended emotions in the database. This kind of induction mood procedures privileges more natural and spontaneous expressive motions over more easily recognized movements [4]. As a result, accuracy rates for this kind of expressive motion (e.g., Volkova et al. obtained accuracy rates between 20% and 50% [62] while Fourati and colleagues obtained an average accuracy rate of 40% for their *Emilya* database [23]) are often lower—and similar to the rates we have obtained in our user studies—than the ones reported for portrayed and acted motion databases [2, 36, 49].

9 FUTURE WORK

Future work will be oriented to overcoming the limitations of this study and testing the proposed motion representation and synthesis approach on different types of motions and emotional states. For instance, we are interested in applying the perceptual evaluations we presented in this article to other MoCap databases. It will be indeed interesting to determine whether end-effector trajectories generalize without problem to other types of movements such as daily actions [22] or dance sequences [1], as well as to other emotional states not initially considered in our database. Futhermore, by considering databases in which more exaggerated movements were employed to convey emotional content (see, e.g., References [36, 49]), it will be possible to obtain additional insights about the nature of the motion data to be used when synthesizing expressive bodily motions. That is, whether much more exaggerated motion examples should be employed over more naturalistically obtained motion (e.g., elicited emotions) when animating believable virtual characters.

Another aspect to consider for future work will be the motion retargeting issue, i.e., the usability of end-effector trajectories for the generation of expressive movements when controlling animated characters or employing examples from actors with different body proportions and sizes. Indeed, since we wanted to reduce as much as possible any confounding effect (e.g., differences in acting style or expressiveness) that might have influenced the results of our second perceptual study, we decided to test our synthesis system on the data recorded from a single actor. Because of this, it is unknown whether the adaptation of end-effector trajectories through existing motion retargeting will result in artifacts during the full-body reconstruction step and changes in the perception of emotions.

Similarly, we are also interested in improving the function that maps end-effector trajectories to whole-body motions, since it is one of the key elements of the motion synthesis approach proposed in this article. Although the inverse kinematics controllers we used to approximate this mapping provide an efficient and flexible control over the resulting motions, additional constraints (joint limits and elbow trajectories) were needed to enhance the solutions provided by this mapping. Nonetheless, we observed that the generated motions might still suffer from visual artifacts inherent to both the redundancy of the articulated chains being controlled and the use of purely procedural synthesis techniques. We believe that deep-learning techniques represent an interesting direction for the improvement of this mapping function. Specifically, convolutional networks have shown to successfully encode within the hidden units the bio-mechanical constraints governing human motion [31, 32]. By combining this type of network and a generation scheme as the one recently proposed by the deep generative network WaveNet [61], we believe it is possible to generate biologically correct and visually appealing whole-body motions that closely follow the control signal and exhibit the intended emotional content.

Overall, the results from both qualitative evaluations show that end-effector trajectories are an interesting choice of motion representation and have the potential to be sufficient for the generation of expressive bodily motions. However, when mapping functions with no prior knowledge about the biomechanical constraints of the human body are used, such as IK kinematics, additional control signal, such as the elbow trajectories we used, might be necessary to obtain plausible and rich expressive full-body movements.

REFERENCES

[1] Andreas Aristidou, Panayiotis Charalambous, and Yiorgos Chrysanthou. 2015. Emotion analysis and classification: Understanding the performers' emotions using the LMA entities. *Comput. Graph. Forum* 34, 6 (2015), 262–276. DOI: http://dx.doi.org/10.1111/cgf.12598

[2] Anthony P. Atkinson, Winand H. Dittrich, Andrew J. Gemmell, and Andrew W. Young. 2004. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception* 33 (2004), 717–746. Retrieved from https://doi.org/10.1068/p5096.

[3] Anthony P. Atkinson, Mary L. Tunstall, and Winand H. Dittrich. 2007. Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition* 104, 1 (2007), 59–72. DOI: http://dx.doi.org/10.1016/j.cognition.2006.05.005

[4] Tanja Bänziger, Marcello Mortillaro, and Klaus R. Scherer. 2012. Introducing the Geneva multimodal expression corpus for experimental research on emotion perception.*Emotion* 12, 5 (2012), 1161. DOI: http://dx.doi.org/10.1037/a0025827

[5] Jon Louis Bentley. 1975. Multidimensional binary search trees used for associative searching. *Commun. ACM* 18, 9 (Sept. 1975), 509–517. DOI: http://dx.doi.org/10.1145/361002.361007

[6] Daniel Bernhardt and Peter Robinson. 2007. Detecting affect from non-stylised body motions. In *Affective Computing and Intelligent Interaction*. Springer, Berlin, 59–70.

[7] Randolph Blake and Maggie Shiffrar. 2007. Perception of human motion. *Annu. Rev. Psychol.* 58 (2007), 47–73. DOI: http://dx.doi.org/10.1146/annurev.psych.57.102904.190152 PMID: 16903802.

[8] Alexandre Bouënard, Marcelo MM Wanderley, and Sylvie Gibet. 2010. Gesture control of sound synthesis: Analysis and classification of percussion gestures. *Acta Acust. United Acust.* 96, 4 (2010), 668–677. DOI: http://dx.doi.org/doi:10.3813/AAA.918321

[9] Matthew Brand and Aaron Hertzmann. 2000. Style machines. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'00)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, 183–192. DOI: http://dx.doi.org/10.1145/344779.344865

[10] Pamela Carreno-Medrano, Sylvie Gibet, Caroline Larboulette, and Pierre-François Marteau. 2014. Corpus creation and perceptual evaluation of expressive theatrical gestures. In *Proceedings of the 14th International Conference on Intelligent Virtual Agents (IVA'14)*. 109–119. DOI: http://dx.doi.org/10.1007/978-3-319-09767-1_14

[11] Pamela Carreno-Medrano, Sylvie Gibet, and Pierre-François Marteau. 2015. End-effectors trajectories: An efficient low-dimensional characterization of affective-expressive body motions. In *Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII'15)*. 435–441. DOI: http://dx.doi.org/10.1109/ACII.2015.7344607

[12] Pamela Carreno-Medrano, Sylvie Gibet, and Pierre-François Marteau. 2016. From expressive end-effector trajectories to expressive bodily motions. In *Proceedings of the 29th International Conference on Computer Animation and Social Agents (CASA'16)*. ACM, New York, NY, 157–163. DOI: http://dx.doi.org/10.1145/2915926.2915941

[13] Jinxiang Chai and Jessica K. Hodgins. 2005. Performance animation from low-dimensional control signals. *ACM Trans. Graph.* 24, 3 (July 2005), 686–696. DOI: http://dx.doi.org/10.1145/1073204.1073248

[14] Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler. 2000. The EMOTE model for effort and shape. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'00)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, 173–182. DOI: http://dx.doi.org/10.1145/344779.352172

[15] Mark Coulson. 2004. Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *J. Nonverb. Behav.* 28, 2 (2004), 117–139. DOI: http://dx.doi.org/10.1023/B:JONB.0000023655.25550.be

[16] Virginie Demeure, Rados Niewiadomski, and Catherine Pelachaud. 2011. How is believability of a virtual agent related to warmth, competence, personification, and embodiment?*Presence: Teleoper. Virtual Environ.* 20, 5 (Oct. 2011), 431–448. DOI: http://dx.doi.org/10.1162/PRES_a_00065

[17] Winand H. Dittrich and Anthony P. Atkinson. *The Perception of Bodily Expressions of Emotion and the Implications for Computing*. I-Tech education and publishing, Chapter 9, 157–184.

[18] Funda Durupinar, Mubbasir Kapadia, Susan Deutsch, Michael Neff, and Norman I. Badler. 2016. PERFORM: Perceptual approach for adding OCEAN personality to human motion using laban movement analysis. *ACM Trans. Graph.* 36, 1, Article 6 (Oct. 2016), 16 pages. DOI: http://dx.doi.org/10.1145/2983620

[19] Ahmed Elgammal and Chan-Su Lee. 2008. *The Role of Manifold Learning in Human Motion Analysis*. Springer Netherlands, Dordrecht, 25–56. DOI: http://dx.doi.org/10.1007/978-1-4020-6693-1_2

[20] S. Ali Etemad, Ali Arya, and Avi Parush. 2014. Additivity in perception of affect from limb motion. *Neurosci. Lett.* 558 (2014), 132–136. DOI: http://dx.doi.org/10.1016/j.neulet.2013.11.010

[21] Donghui Feng, Sveva Besana, and Remi Zajac. 2009. Acquiring high quality non-expert knowledge from on-demand workforce. In *Proceedings of the 2009 Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources (People'sWeb'09)*. Association for Computational Linguistics, Stroudsburg, PA, 51–56. http://dl.acm.org/citation.cfm?id=1699765.1699773.

[22] Nesrine Fourati and Catherine Pelachaud. 2014. Emilya: Emotional body expression in daily actions database. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association (ELRA), 26–31.

[23] N. Fourati and C. Pelachaud. 2016. Perception of emotions and body movement in the emilya database. *IEEE Trans. Affec. Comput.* PP, 99 (2016), 1–1. DOI: http://dx.doi.org/10.1109/TAFFC.2016.2591039

[24] Donald Glowinski, Nele Dael, Antonio Camurri, Gualtiero Volpe, Marcello Mortillaro, and Klaus Scherer. 2011. Toward a minimal representation of affective gestures. *IEEE Trans. Affect. Comput.* 2, 2 (April 2011), 106–118. DOI: http://dx.doi.org/10.1109/T-AFFC.2011.7

[25] Keith Grochow, Steven L. Martin, Aaron Hertzmann, and Zoran Popović. 2004. Style-based inverse kinematics. *ACM Trans. Graph.* 23, 3 (Aug. 2004), 522–531. DOI: http://dx.doi.org/10.1145/1015706.1015755

[26] Hatice Gunes, Caifeng Shan, Shizhi Chen, and YingLi Tian. 2015. Bodily expression for automatic affect recognition. (2015), 343–377. DOI: http://dx.doi.org/10.1002/9781118910566.ch14

[27] Björn Hartmann, Maurizio Mancini, and Catherine Pelachaud. 2006. *Implementing Expressive Gesture Synthesis for Embodied Conversational Agents*. Springer, Berlin, 188–199.

[28] David J. Hauser and Norbert Schwarz. 2015. It's a trap! Instructional manipulation checks prompt systematic thinking on tricky tasks. *SAGE Open* 5, 2 (2015), 2158244015584617. DOI: http://dx.doi.org/10.1177/2158244015584617

[29] Zhiying He, Xiaohui Liang, Jian Wang, Qinping Zhao, and Chengyu Guo. 2014. Flexible editing of human motion by three-way decomposition. *Comput. Animation Virtual Worlds* 25, 1 (2014), 57–68. DOI: http://dx.doi.org/10.1002/cav.1534

[30] Jessica K. Hodgins, James F. O'Brien, and Jack Tumblin. 1998. Perception of human motion with different geometric models. *IEEE Trans. Visual. Comput. Graph.* 4, 4 (Oct. 1998), 307–316. DOI: http://dx.doi.org/10.1109/2945.765325

[31] Daniel Holden, Taku Komura, and Jun Saito. 2017. Phase-functioned neural networks for character control. *ACM Trans. Graph.* 36, 4, Article 42 (July 2017), 13 pages. DOI: http://dx.doi.org/10.1145/3072959.3073663

[32] Daniel Holden, Jun Saito, and Taku Komura. 2016. A deep learning framework for character motion synthesis and editing. *ACM Trans. Graph.* 35, 4, Article 138 (July 2016), 11 pages. DOI: http://dx.doi.org/10.1145/2897824.2925975

[33] Kristina Höök. 2009. Affective loop experiences: Designing for interactional embodiment. *Philos. Trans. Roy. Soc. London B: Biol. Sci.* 364, 1535 (2009), 3585–3595. DOI: http://dx.doi.org/10.1098/rstb.2009.0202

[34] Eugene Hsu, Kari Pulli, and Jovan Popović. 2005. Style translation for human motion. *ACM Trans. Graph.* 24, 3 (July 2005), 1082–1089. DOI: http://dx.doi.org/10.1145/1073204.1073315

[35] Eva Hudlicka. 2003. To feel or not to feel: The role of affect in human-computer interaction. *Int. J. Human-Comput. Studies* 59, 1–2 (July 2003), 1–32. DOI: http://dx.doi.org/10.1016/S1071-5819(03)00047-8

[36] Asha Kapur, Ajay Kapur, Naznin Virji-Babul, George Tzanetakis, and Peter F. Driessen. 2005. Gesture-based affective computing on motion capture data. In *Proceedings of the 1st International Conference on Affective Computing and Intelligent Interaction (ACII'05)*. Springer-Verlag, Berlin, 1–7. DOI: http://dx.doi.org/10.1007/11573548_1

[37] Michelle Karg, Ali-Akbar Samadani, Rob Gorbet, Kolja Kühnlenz, Jesse Hoey, and Dana Kulić. 2013. Body movements for affective expression: A survey of automatic recognition and generation. *IEEE Trans. Affect. Comput.* 4, 4 (2013), 341–359.

[38] Jongmin Kim, Yeongho Seol, and Jehee Lee. 2013. Human motion reconstruction from sparse 3D motion sensors using kernel CCA-based regression. *Comput. Animation Virtual Worlds* 24, 6 (2013), 565–576. DOI: http://dx.doi.org/10.1002/cav.1557

[39] Andrea Kleinsmith and Nadia Bianchi-Berthouze. 2013. Affective body expression perception and recognition: A survey. *IEEE Trans. Affect. Comput.* 4, 1 (Jan 2013), 15–33. DOI: http://dx.doi.org/10.1109/T-AFFC.2012.16

[40] Andrea Kleinsmith, Nadia Bianchi-Berthouze, and Anthony Steed. 2011. Automatic recognition of non-acted affective postures. *IEEE Trans. Syst., Man, Cybernet., Part B (Cybernetics)* 41, 4 (Aug. 2011), 1027–1038. DOI: http://dx.doi.org/10.1109/TSMCB.2010.2103557

[41] Björn Krüger, Jochen Tautges, Andreas Weber, and Arno Zinke. 2010. Fast local and global similarity searches in large motion capture databases. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA'10)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 1–10. Retrieved from http://dl.acm.org/citation.cfm?id=1921427.1921429.

[42] David Landau and Kurt Binder. 2005. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, New York, NY.

[43] Matthew Loper, Naureen Mahmood, and Michael J. Black. 2014. MoSh: Motion and shape capture from sparse markers. *ACM Trans. Graph.* 33, 6, Article 220 (Nov. 2014), 13 pages. DOI: http://dx.doi.org/10.1145/2661229.2661273

[44] Rachel McDonnell, Sophie Jörg, Joanna McHugh, Fiona N. Newell, and Carol O'Sullivan. 2009. Investigating the role of body shape on the perception of emotion. *ACM Trans. Appl. Percept.* 6, 3, Article 14 (Sept. 2009), 11 pages. DOI: http://dx.doi.org/10.1145/1577755.1577757

[45] Jianyuan Min and Jinxiang Chai. 2012. Motion graphs++: A compact generative model for semantic motion analysis and synthesis. *ACM Trans. Graph.* 31, 6, Article 153 (Nov. 2012), 12 pages. DOI: http://dx.doi.org/10.1145/2366145.2366172

[46] Jianyuan Min, Huajun Liu, and Jinxiang Chai. 2010. Synthesis and editing of personalized stylistic human motion. In *Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D'10)*. ACM, New York, NY, 39–46. DOI: http://dx.doi.org/10.1145/1730804.1730811

[47] Valérie Monbet and Pierre-François Marteau. 2006. Non parametric resampling for stationary Markov processes: The local grid bootstrap approach. *J. Stat. Plan. Infer.* 136, 10 (2006), 3319–3338. DOI: http://dx.doi.org/10.1016/j.jspi.2004.11.014

[48] Michael Neff and Eugene Fiume. 2005. AER: Aesthetic exploration and refinement for expressive character animation. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA'05)*. ACM, New York, NY, 161–170. DOI: http://dx.doi.org/10.1145/1073368.1073391

[49] Aline Normoyle, Fannie Liu, Mubbasir Kapadia, Norman I. Badler, and Sophie Jörg. 2013. The effect of posture and dynamics on the perception of emotion. In *Proceedings of the ACM Symposium on Applied Perception (SAP'13)*. ACM, New York, NY, 91–98. DOI: http://dx.doi.org/10.1145/2492494.2492500

[50] Claire L. Roether, Lars Omlor, Andrea Christensen, and Martin A. Giese. 2009. Critical features for the perception of emotion from gait. *J. Vision* 9, 6 (2009), 15. DOI: http://dx.doi.org/10.1167/9.6.15 arXiv:/data/journals/jov/933553/jov-9-6-15.pdf

[51] Charles Rose, Michael F. Cohen, and Bobby Bodenheimer. 1998. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Comput. Graph. Appl.* 18, 5 (1998), 32–40. DOI: http://dx.doi.org/10.1109/38.708559

[52] Ari Shapiro, Yong Cao, and Petros Faloutsos. 2006. Style components. In *Proceedings of the Conference on Graphics Interface (GI'06)*. Canadian Information Processing Society, Toronto, Ont., Canada, 33–39. http://dl.acm.org/citation.cfm?id=1143079.1143086.

[53] Jochen Tautges, Arno Zinke, Björn Krüger, Jan Baumann, Andreas Weber, Thomas Helten, Meinard Müller, Hans-Peter Seidel, and Bernd Eberhardt. 2011. Motion reconstruction using sparse accelerometer data. *ACM Trans. Graph.* 30, 3, Article 18 (May 2011), 12 pages. DOI: http://dx.doi.org/10.1145/1966394.1966397

[54] Graham W. Taylor and Geoffrey E. Hinton. 2009. Factored conditional restricted boltzmann machines for modeling motion style. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML'09)*. ACM, 1025–1032. DOI: http://dx.doi.org/10.1145/1553374.1553505

[55] Lorenzo Torresani, Peggy Hackney, and Christoph Bregler. 2007. In *Advances in Neural Information Processing Systems*, vol. 19, B. Schölkopf, J. C. Platt, and T. Hoffman (Eds.). MIT Press, 1393–1400. Retrieved from http://papers.nips.cc/paper/3109-learning-motion-style-synthesis-from-perceptual-observations.pdf.

[56] Maxime Tournier, Xiaomao Wu, Nicolas Courty, Elise Arnaud, and Lionel Reveret. 2009. Motion compression using principal geodesics analysis. In *Computer Graphics Forum*, Vol. 28. Wiley Online Library, Blackwell Publishing Ltd., 355–364. DOI: http://dx.doi.org/10.1111/j.1467-8659.2009.01375.x

[57] Nikolaus F. Troje. 2002. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *J. Vis.* 2, 5 (2002), 371–387. DOI: http://dx.doi.org/10.1167/2.5.2

[58] Nikolaus F. Troje, Cord Westhoff, and Mikhail Lavrov. 2005. Person identification from biological motion: Effects of structural and kinematic cues. *Percept. Psychophys.* 67, 4 (2005), 667–675. DOI: http://dx.doi.org/10.3758/BF03193523

[59] John W. Tukey. *Exploratory Data Analysis.* Addison-Wesley Series in Behavioral Science: Quantitative Methods.

[60] Raquel Urtasun, Pascal Glardon, Ronan Boulic, Daniel Thalmann, and Pascal Fua. 2004. Style-based motion synthesis. In *Computer Graphics Forum*, Vol. 23. Wiley Online Library, Blackwell Publishing Ltd., 799–812. DOI: http://dx.doi.org/10.1111/j.1467-8659.2004.00809.x

[61] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew W. Senior, and Koray Kavukcuoglu. 2016. WaveNet: A generative model for raw audio. *CoRR* abs/1609.03499 (2016). arxiv:1609.03499. Retrieved from http://arxiv.org/abs/1609.03499.

[62] Ekaterina Volkova, Stephan de la Rosa, Heinrich H. Bülthoff, and Betty Mohler. 2014. The MPI emotional body expressions database for narrative scenarios. *PLOS ONE* 9, 12 (Dec. 2014), 1–28. DOI: http://dx.doi.org/10.1371/journal.pone.0113647

[63] Andy T. Woods, Carlos Velasco, Carmel A. Levitan, Xiaoang Wan, and Charles Spence. 2015. Conducting perception research over the internet: A tutorial review. *Peer J.* 3 (July 2015), e1058. DOI: http://dx.doi.org/10.7717/peerj.1058

[64] Shihong Xia, Congyi Wang, Jinxiang Chai, and Jessica Hodgins. 2015. Realtime style transfer for unlabeled heterogeneous human motion. *ACM Trans. Graph.* 34, 4, Article 119 (July 2015). DOI: http://dx.doi.org/10.1145/2766999

Received July 2017; revised February 2018; accepted February 2018

Author Queries

Q1: AU: Please provide authors' complete mailing and email addresses.

Q2: AU: You present your p values in mixed styles, as text (roman, no spacing) and also as math (italic, with spacing). Is this as meant?