Impact of Self-Contacts on Perceived Pose Equivalences

Jean Basset, Badr Ouannas, Ludovic Hoyet, Franck Multon, Stefanie Wuhrer

Impact of Self-Contacts on Perceived Pose Equivalences

Jean Basset [*] jean.basset@inria.fr

Badr Ouannas [*] badr.ouannas@grenoble-inp.org

Ludovic Hoyet [†] ludovic.hoyet@inria.fr

Franck Multon [‡] fmulton@irisa.fr

Stefanie Wuhrer [*] stefanie.wuhrer@inria.fr

Abstract

Defining equivalences between poses of different human characters is an important problem for imitation research, human pose recognition and deformation transfer. However, pose equivalence is a subjective information that depends on context and on the morphology of the characters. A common hypothesis is that interactions between body surfaces, such as self-contacts, are important attributes of human poses, and are therefore consistently included in animation approaches aiming at retargeting human motions. However, some of these self-contacts are only present because of the morphology of the character and are not important to the pose, e.g. contacts between the upper arms and the torso during a standing A-pose. In this paper, we conduct a first study towards the goal of understanding the impact of self-contacts between body surfaces on perceived pose equivalences. More specifically, we focus on contacts between the arms or hands and the upper body, which are frequent in everyday human poses. We conduct a study where we present to observers two models of a character mimicking the pose of a source character, one with the same self-contacts as the source, and one with one self-contact removed, and ask observers to select which model best mimics the source pose. We show that while poses with different self-contacts are considered different by observers in most cases, this effect is stronger for self-contacts involving the hands than for those involving the arms.

1 Introduction

Understanding what defines the meaning of a human pose is an important problem for character animation. More specifically, precisely defining equivalences between character poses is crucial to perform and evaluate deformation transfer, *i.e.* the retargeting of an existing animation to a new character.

Deformation transfer approaches take as input a source character with a given pose and a target character with a given identity, and aim to generate the target character performing a pose that is as similar as possible to the pose of the source character. This implies being able to characterise what makes two poses of different characters equivalent. The classical approach to mimic the pose of a source character is to apply similar geometric deformations to the target model. This is done for example by using the same joint angles for skeletal retargeting, or by applying the same surface deformations [Sumner and Popović(2004)]. In such approaches, the objective is to produce poses of source and target characters that should be as geometrically similar as possible. However, studies on human perception of pose similarity showed that geometrical similarity, measured by a low Euclidean distance between the 3D models, does not necessarily imply perceived pose equivalence for observers [Tang et al.(2008), Durupinar(2021)].

[*]Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP (Institute of Engineering Univ. Grenoble Alpes), LJK, 38000 Grenoble, France

[†]Univ Rennes, Inria, CNRS IRISA, France

[‡]Univ Rennes, Inria, CNRS IRISA, M2S, France

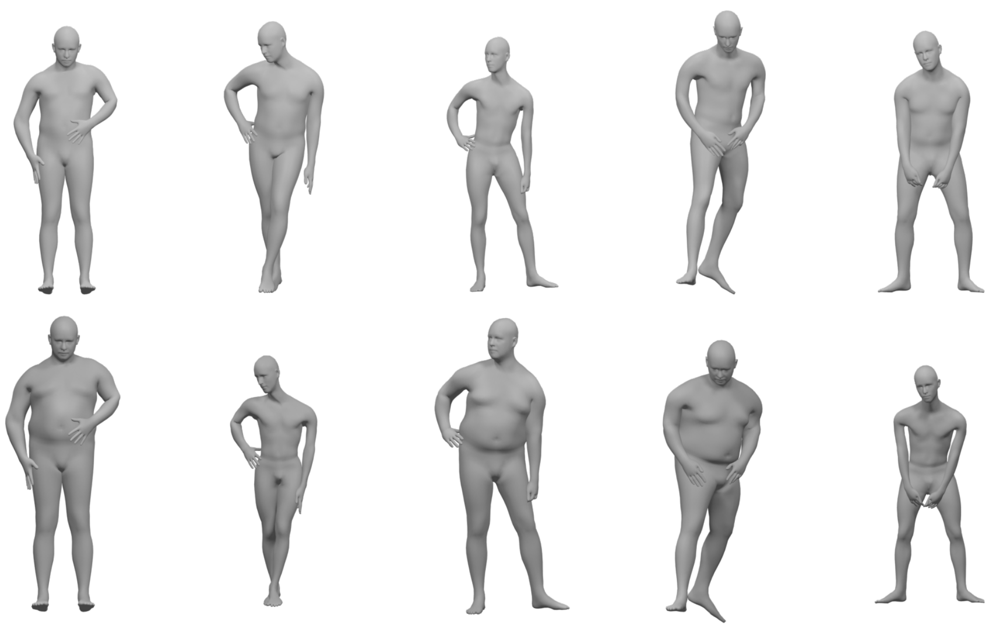

Figure 1: Human characters with varying morphologies performing poses with several self-contacts. Characters with similar poses but different body shapes (columns) can have slightly different spatial relationships between body segments, in particular self-contacts (*e.g.* contact between the left arm and the flank appearing for the middle characters). In this paper we explore which self-contacts are important to the meaning of the pose independently from other parameters such as body shape.

Another popular heuristic explored by methods that aim to adapt poses to the morphology of new characters is to encode and preserve relative positions of body segments [Ho et al.(2010)] or surfaces [Jin et al.(2018), Liu et al.(2018)]. This is efficient to avoid collisions and to preserve important constraints such as contacts. However, some contacts only appear as a consequence of the morphology of the character performing the pose, and do not bring semantic information to the pose. This is illustrated for example in Figure 2 where two characters perform the same pose with different self-contacts: the contact between the arms and the torso of the larger character does not change the meaning of the pose and is simply the result of the larger morphology. Contrarily, the contact between the hands must be preserved between different characters in order to preserve semantic equivalence. Systematically preserving contacts during deformation transfer can thus cause artifacts as unimportant contacts are kept.

These methods therefore face the important challenge of selecting which contacts are important to the meaning of the pose and which are not. This is a difficult task as such information can be highly subjective, and will depend on *e.g.* context or morphology. However, human observers are able to understand and rate pose similarity, as illustrated by numerous previous studies on human pose perception [Johansson(1973), Hodgins et al.(1998), Harada et al.(2004), Marinoiu et al.(2016)]. Therefore, an interesting direction is to leverage human perception to better understand which contacts are important characteristics of the pose and which are not.

In this paper, we design a perceptual study that is a first step towards understanding the impact of self-contacts on pose equivalences. Human character 3D models are commonly defined by pose parameters that encode the global articulated position of the body parts of the character, and identity parameters that encode its body shape [Anguelov et al.(2005), Loper et al.(2015)]. Our goal is to study which self-contacts on a human model should be preserved when different identity parameters are applied to the same pose parameters, *i.e.* which are important to the meaning of the pose independently of other parameters such as identity. More specifically, we consider self-contacts (*i.e.* contacts between surfaces of different body parts of the same character) between the arms or hands and the upper body of the character, as they are frequent in everyday human poses, such as the ones presented in Figure 1.

We first aim to validate the common assumption that self-contacts are in general important to the meaning of the pose. Our hypothesis is that observers presented with two similar poses, one with a missing self-contact present in the other, would consider the two poses to be visually different in most cases. We also argue that the importance of self-contacts depends on the body parts involved. Indeed, studies on pose similarities [Harada et al.(2004), Marinoiu et al.(2016)] suggest that different body parts have different impacts on the perceived pose meaning. These works also highlight that observers tend to give more attention to the positions of the head, wrists and fingertips when evaluating or imitating a pose. Moreover, we observe that self-contacts involving hands are often the goal of the pose, *e.g.* clapping, touching or the examples in Figure 2. Therefore, we make the hypothesis that self-contacts involving the hands are more important to the meaning of the pose than morphology-dependent contacts, such as when the upper-arms contact with the torso.

To test our hypotheses, we conducted a perceptual study investigating perceived pose equivalences when presenting poses with varying self-contacts to observers. We used a pose dataset of four identities with different body shapes performing 17 different poses containing self-contacts between the arms or hands and the upper body. We presented observers ($N = 81$) with a source pose, a target identity, and two variations of the pose on this target identity, and asked them to select the pose variation which best mimicked the source pose. One of these variations had the same self-contacts as the source pose, and the other had one self-contact present in the source released. These pose variations were generated using the deformation transfer method of Basset et al. basset2020contact with different contact preservation parameters.

The results of our study show that observers tend to consider the pose with the same contacts present in the source pose to be the best imitation of the source pose in most cases. This tendency appears to be more important for contacts involving hands than contacts involving the arms of the characters. Finally, the results show that the release distance of the contacts, *i.e.* the distance between the body surfaces that were originally in contact in the source, has an impact on perceived pose equivalences: the tendency to choose the pose with the same contacts as in the source is stronger when the presented variation has a more important contact release distance. This study thus confirms the hypothesis that self-contacts do not all have the same impact on the pose, and highlights possible parameters to select important contacts to improve animation retargeting approaches.

2 Related Work

In this section, we first review deformation transfer strategies that explored how to define pose equivalence between a source pose and their transfer result. We then present works that explored the perception of pose by human observers.

2.1 Deformation Transfer

Deformation transfer methods take as input a source character performing a specific pose, and a target character with a different identity. They then aim to generate the target character mimicking the pose of the source. An important challenge is thus to define pose equivalences between characters with different body shapes, in order to evaluate if the source character and the result are indeed performing the

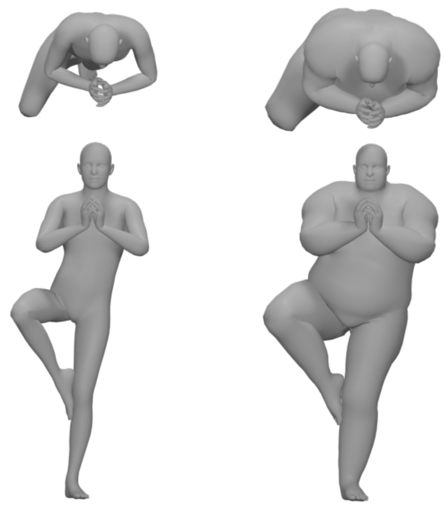

Figure 2: Illustration of dependence of self-contacts on body shape. Characters with different morphologies performing a similar pose with different self-contacts, viewed from front and top. In this pose, one can see the importance of preserving the contacts between the hands, and between foot and knee, while contacts between upper-arm and torso (right) result from the morphology of the character. Meshes generated by an artist, courtesy of [Liu et al.(2018)].

same pose. The common approach in early works is to apply similar geometric deformations than the source to the target character. For skeleton motion retargeting, this is done by applying the same joint angles to the target skeleton, and solving kinematic constraints [Gleicher(1998), Lee and Shin(1999), Popović and Witkin(1999), Choi and Ko(2000)]. For surface-based methods, this is done by encoding the geometrical deformation of the source character's surface, and applying it to the target character [Sumner and Popović(2004), Zhou et al.(2010)]. However, studies have shown that simple geometrical similarity, *e.g.* low Euclidean distance between the characters' 3D positions, is not always correlated with perceived pose equivalence [Tang et al.(2008), Durupinar(2021)].

Other works use spatial relationships between body segments in order to define constraints that give meaning to the pose. They aim to preserve the relative positions of body segments between the source character and the transfer result, using *e.g.* distance constraints [Al-Asqhar et al.(2013), Bernardin et al.(2017)], an egocentric normalization of the body part relative distances [Molla et al.(2018)], or an external graph connecting body segments [Baciu and Iu(2006)]. This approach is well illustrated by the interaction mesh proposed by Ho et al. Ho10spatial. In this representation, all the joints of the skeleton of characters are linked by edges, and their relative positions are preserved by minimizing changes in the Laplacian coordinates of this interaction mesh. This approach has been adapted to 3D meshes, by considering vertices on the surface of the characters instead of skeleton joints [Liu et al.(2018), Jin et al.(2018)]. These approaches allow to better capture pose meaning than simple geometric similarity, however some limitations remain when retargeting between very different morphologies; in this case exactly preserving distances between body parts might result in unrealistic poses, as discussed in Figure 2. Some works have explored adapting the pose to new and very different morphologies, for example by solving physical constraints such as balance [Lyard and Magnenat-Thalmann(2008), Al Borno et al.(2018)], or by simply preserving self-contacts while allowing other inter body part distances to change for the new character [Basset et al.(2020)].

2.2 Human Pose Perception

The Computer Graphics literature has widely investigated human perception of virtual characters. Human observers have proved to be able to recognize the identity of actors in stylized or simplified virtual characters. A lot of interest has been given to which factors impacted the recognition of an actor in a deformed face [Tanaka and Farah(1993), Zhao et al.(2003), Olivier et al.(2020)]. Closer to our problem, recognizing the identity or style of an actor from its body pose and motion was also explored. Johansson johansson1973visual represented human motions with a limited number (~ 10) of bright points, and showed that this representation was enough to evoke different kind of motions to observers, such as walking or running, and even to recognize the style of the motion (*e.g.* tired or "wavy"). More recently, Hoyet et al. hoyet2013evaluating animated two realistic characters (one male one female) from actors' captured motion, and explored which motions better allowed to recognize the actor.

Some works have more specifically explored perception of pose or motion similarity. Hodgins et al. hodgins1998perception showed that observers were better at perceiving pose differences in more realistic models than stick figures. Harada et al. harada2004quantitative designed a quantitative pose similarity metric, and optimized it by comparing their results to pose similarity perceived by observers. They highlighted several parameters that impact perceived pose equivalences, such as weights describing the impact of each body part on the perception of the pose. In particular, they showed that the position of the fingertips has a strong impact on the perception of the pose. Chen et al. chen2009perceptual use this approach to design a similarity metric that accounts for relative similarity, *e.g.* which among a set of examples is more similar to a target. Tang et al. tang2008emulating and Pražák et al. pravzak2009perception propose a similar approach applied to human motion similarity, illustrating that simple similarity metrics such as Euclidean distance between joints does not always correlate with human perception of similarity [Tang et al.(2008)]. More recently, Laban Dance Notation [Laban(1928)] has been applied to the study of human poses. Laban Notation serves as a language to evaluate and record human poses and motion based on qualitative parameters. This notation has been applied to motion generation [Durupinar et al.(2016)], and closer to our work perception-based pose similarity metrics [Durupinar(2021)]. The latter further confirms that simple similarity metrics do not correlate with human perception.

Marinoiu et al. marinoiu2016pictorial evaluated

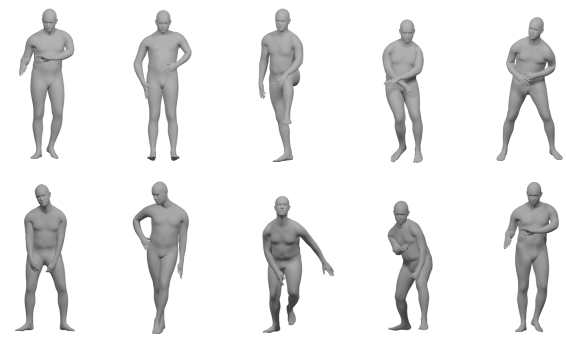

Figure 3: Examples of poses with self-contacts used in the study.

the capacity of human subjects to imitate the pose of a target character. The study showed that subjects (equipped with motion capture markers and an eye tracker) gave more attention to certain joints of the target before imitating the pose, such as the head and the wrists. More recently, Muller et al. muller2021self used a similar approach to annotate a dataset of 2D images of human characters in the wild. They presented 3D models of human characters to subjects tasked to take pictures of themselves imitating the character's pose. Their dataset focuses on poses containing self-contacts, which illustrates that human subjects are able to understand and imitate poses containing self-contacts with acceptable precision.

3 Stimuli Preparation

In this section, we present the data to be used in the perceptual study. This includes how the different body poses were created, including the contact variations and a summary of the resulting stimuli. Our goal is to generate a dataset containing poses with self-contacts on different representative body shapes, and to generate variations where self-contacts are released to evaluate their relative importance on the perception of pose similarity.

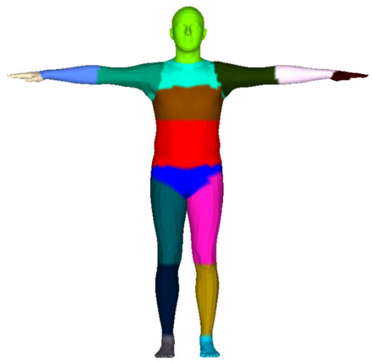

Figure 4: Body part segmentation used to select self-contacts between distinct body parts.

3.1 Selecting Human Poses

To create the data used in this study, we first selected a set of poses with self-contacts. As already mentioned, we focus on self-contacts involving the arms or hands and the upper body of the characters (above the thighs, thighs included), as they are frequent in everyday human poses. We thus aim to obtain a representative set of poses in these configurations. We selected a total of 17 poses from a dataset of 3D human scans [Bogo et al.(2014)] and a dataset of artist-generated animations (Mixamo[1]), and created by hand by modifying pose parameters of a parametric human body model [Loper et al.(2015)]. These poses all present one or two self-contacts between the hands or arms and the upper body, *e.g.* crossing arms, hand resting on the hip, clapping. Figure 3 shows examples of the resulting poses on a fixed body shape.

3.2 Detecting Self-Contacts

Once a set of human poses have been selected, it is then necessary to automatically quantify the location of the different self-contacts, as our goal will be to create poses with and without some of these self-contacts. We only consider self-contacts between distinct body parts of the human character (see Figure 4), as self-contacts within the same rigidly deforming body part rarely bring information on the meaning of the pose. More precisely, two body parts are considered to be in contact if there exist points on the two parts that have Euclidean distance below a threshold. Formally, given a mesh $\mathcal{M}$ representing a character, with vertices $\mathcal{V} \in \mathcal{M}$, we consider two vertices $\{v_i, v_j\} \in \mathcal{V}$ in contact if

$$|v_i - v_j| < t, \quad v_i \in \mathcal{B}_1, v_j \in \mathcal{B}_2, \tag{1}$$

where t is a threshold, and $\mathcal{B}_1$ and $\mathcal{B}_2$ two distinct body parts. When two body parts $\mathcal{B}_1$ and $\mathcal{B}_2$ have one or more points in contact, we call $C_{\mathcal{B}_1,\mathcal{B}_2}$ the set of vertices involved in the self-contact. In practice, we used $t = 1cm$ for estimating the self-contacts of the poses used in this study.

3.3 Transferring and Releasing Self-Contacts

To evaluate how self-contacts are perceived when transferring poses between different character shapes, we first created a number variations of the poses selected in Section 3.1 by retargeting them on several body shapes (called identity hereafter). We selected four identities for our study, using the identity parameters of the SMPL parametric human body model [Loper et al.(2015)]. This model's identity parameters represent well the variation of human body shapes along its main variation axes, as it was trained on a large dataset of captures of human subjects. We generated four target identities with representative body shapes by sampling models at ± 2 standard deviations of the mean for the two first identity parameters (see Figure 5). To limit the size of our study, we also chose to focus this first study on male characters.

The four target identity shapes were then transferred onto the selected poses using a recent deformation transfer method [Basset et al.(2020)]. Instead of transferring the pose onto the new target character shape, this methods operates by deforming the surface of the source character in the correct pose towards the target identity shape. The method explicitly preserves self-contacts present in the source character through an energy term that can be selectively deactivated for selected body parts. As we are

[1] https://www.mixamo.com/

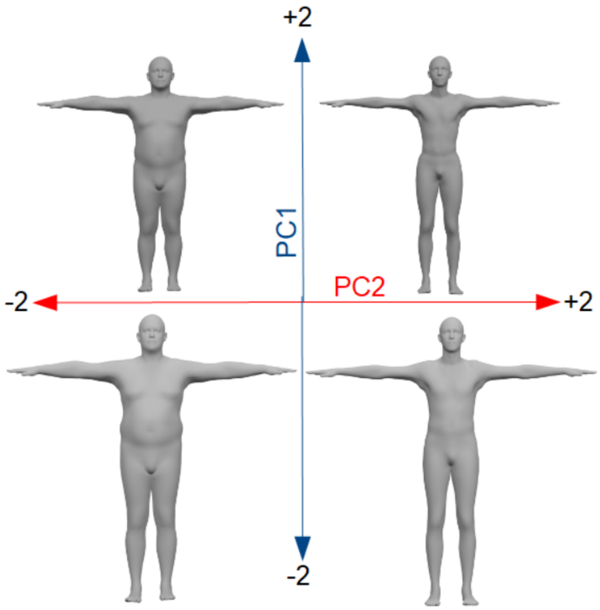

Figure 5: Body shapes used in the study, generated by sampling models at ± 2 standard deviations of the two first principal components (blue and red axes) of a parametric body model's shape space [Loper et al.(2015)].

interested in the impact of the presence or absence of a self-contact on the perceived meaning of the pose, this method was also used to generate pose variations without some of the original self-contacts. More precisely, for each pair of selected pose and identity selected as above, we generated several pose variations by progressively releasing the self-contact present in the source pose. Figure 6 illustrates the different pose variations generated for a specific pose and identity: the pose variation on the new identity with all the self-contacts transferred and with only one self-contact released (left or right hand). We refer to these generated characters as the pose variations.

A self-contact is released when the distance between all pairs of vertices involved exceeds t. The contact release distance is then

$$d_r(\mathcal{B}_1, \mathcal{B}_2) = \frac{1}{n} \sum_{\{v_i, v_j\} \in C_{\mathcal{B}_1,\mathcal{B}_2}} |v'_i - v'_j| - |v_i - v_j| \quad (2)$$

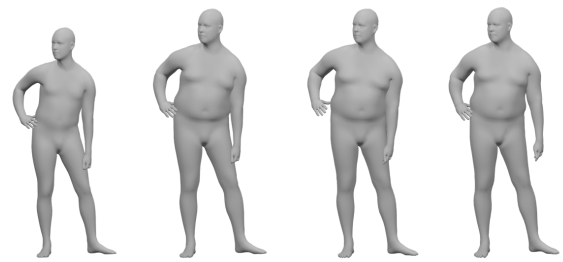

Figure 6: Example of generated pose variations. Left to right: body pose of interest, reference solution for one body shape, pose variation with self-contact involving right hand removed, pose variation with self-contact involving left hand removed.

with v_i, v_j (respectively v'_i, v'_j) the positions of vertices at indexes i and j of the mesh of the character with (respectively without) self-contact, $v_i, v'_i \in \mathcal{B}_1$, $v_j, v'_j \in \mathcal{B}_2$, and n the number of couples of vertices involved in the self-contact $C_{\mathcal{B}_1,\mathcal{B}_2}$.

This results in a dataset containing four body shapes in 17 different poses. For each pose, a reference solution (with all the self-contacts transferred) is available, as well as several variations with each individual self-contact involving the hand, arm or upper-arm released. Our dataset contains 68 reference solutions, and one pose variation for each self-contact present in our poses. Among our 17 poses, 4 (respectively 13) presented one (respectively two) self-contacts involving the arms or the hands and the upper body of the character. This resulted in 30 pose variations with a self-contact released, and thus 120 combinations of body shape and pose variation. We compare these combinations to our reference solutions with no released self-contacts in the study presented in the next section.

3.4 Summary of stimuli

This section summarizes the stimuli shown to the users during the perceptual study. We consider two categories of contacts, *arm* and *hand*, depending on whether the self-contact involves parts of a hand or parts of an arm. Each pose variation is labeled w.r.t the category of the released self-contact. Our dataset

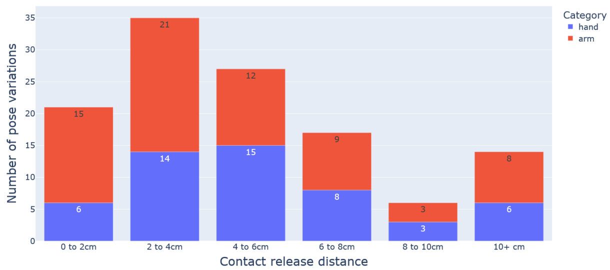

Figure 7: Histogram of the contact release distance labeled by category.

contains 68 pose variations in the category *arm* and 52 in the category *hand*. We also measured the contact release distance (see Equation 2) for each pose variation generated with our method. Our pose variations had a mean release distance of 4.98cm and a standard deviation of 3.42cm. For the *hand* and *arm* category, the release distances had a mean of 5.23cm and 4.80cm, respectively, and standard deviation of 2.99cm and 3.70cm, respectively. Figure 7 shows a histogram of all the contact release distances on our dataset, labeled by category.

4 Study

The goal of our perceptual study is to better understand in which conditions a self-contact will be important to the meaning of the pose. More specifically, we explore the impact of the involved body parts on the importance of self-contacts. We first aim to validate the assumption that self-contacts are generally an important constraint defining the pose. Based on observations and state-of-the-art studies [Harada et al.(2004), Marinoiu et al.(2016)], we make the hypothesis that self-contacts involving the hands are more important to the meaning of the pose than others. The hypotheses explored in this study are therefore:

H1 Pose variations with released self-contacts will be perceived as further away from the source pose than the variation with similar self-contacts.

H2 A pose variation with a released self-contact of the category *hand* will be perceived as different from the source pose more often than for the category *arm*.

4.1 Population

The study was designed to be shared online to subjects. We shared the study to laboratory staff, and using academic mailing lists. The study was validated by an ethics evaluation committee. Participants gave written and informed consent before starting the study. We ensured participants had correctly performed the study by removing those who answered less than 10 questions, as well as those who answered questions in less than 5 seconds on average (time of rotation of the presented videos). After this selection, a total of eighty-one (52 male, 26 female, and 3 who chose not to specify gender) subjects were included in our study, with ages ranging from 19 to 61 years old, with an mean of 31 years old. Forty-four participants reported having prior experience with 3D animation. Participants did the study on computer screens, with an mean and standard deviation screen size of 21.5 ± 5.8 inches.

4.2 Protocol

Before the study, participants were presented with instructions illustrated with an example of question. They were asked to sign an informed consent form before proceeding to the study. They then filled a short demographic questionnaire to gather the information summarized above.

3D models from the pose dataset presented in Section 3.4 were used in this study. We randomly split our variation dataset in 4 subsets of 30 questions. Each participant was presented with only one of these subsets. For each question, observers were tasked to select which of two characters best imitated the pose of a source character displayed in the upper left of the screen. The choices were deformation transfer results of a new identity (displayed in the screen lower left) to the source pose. One displayed similar self-contacts as the source, and the other had one self-contact released, presented randomly on the

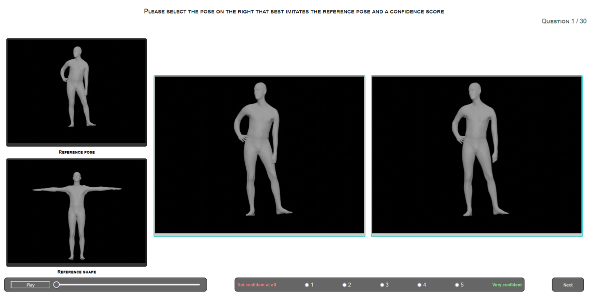

Figure 8: Interface presented to users during the study.

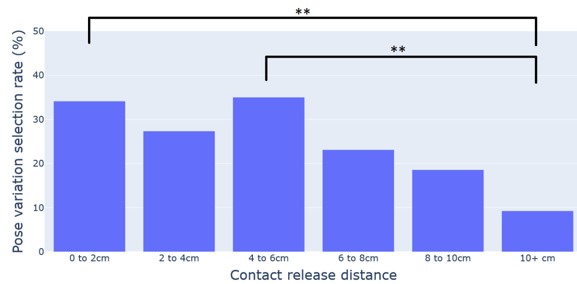

Figure 9: Rate at which observers selected the pose variation with a contact present in the source released, depending on the contact release distance. 50% corresponds to the chance level.

left or the right. In the remainder of this paper, we call these choices the "original contacts pose" and the "pose variation" respectively. The display is illustrated in Figure 8. Characters were displayed with no surrounding environment, in a short video (5 seconds) rendered using Blender, rotating around the 3D model in order to present different angles of the pose to the observers. The video could be controlled with a slider and paused, in order to observe a chosen angle. We measured response time for each question but no time limit was imposed to the subjects during the study. For each question, we gathered between 14 and 30 answers (mean 20) from different observers.

For each question, observers were asked to report their confidence in their response on a scale from 1 (not confident at all) to 5 (very confident).

4.3 Results

In this section, we present descriptive statistics on the results of this study. We also test for statistical significance of the effects of our different stimuli. We assessed normality of the data using the Shapiro-Wilk test. As our data was found to be non-normally distributed, we used the non-parametric Kruskal-Wallis one-way analysis of variance test to study potential main effects of a specific factor on our results. When an effect is found for a factor, we explore it further using Tukey post-hoc test to compare pair-wise means. We chose a statistical significance threshold of 5%, meaning that we consider an effect significant if the probability that the difference between means is due to chance is under 5%.

Pose Variation Selection Rate. We first study the rate of pose variation selection, *i.e.* the proportion of observers choosing the pose variation with a contact present in the source pose released. As a reminder, if participants always preferred the pose that retains all self-contacts this rate would be 0%, while if participants had no preference for maintaining self-contacts or not the rate would be close to 50% (chance level). Observers chose the pose variation in 27.16% of answers. We averaged for each observer the number of cases they chose the pose variation, obtaining an independent pose variation selection rate per observer. We conducted a one-sample t-test in order to determine if the observed selection rate was significantly lower than a random decision that would give a 50% selection rate. The result of this test showed that this difference was indeed significant ($p = 3.159 \times 10^{-29}$).

For each participant, we then computed a pose variation selection rate, which corresponds to how often the participant selected the pose variation over the original contacts pose during the study. When testing for a specific factor, we computed the participant's pose variation selection rate for each group of this factor (*e.g.* selection rates for *hand* and *arm* categories were computed separately when testing the category of self-contact). We found a significant main effect of category of self-contacts (*i.e.*

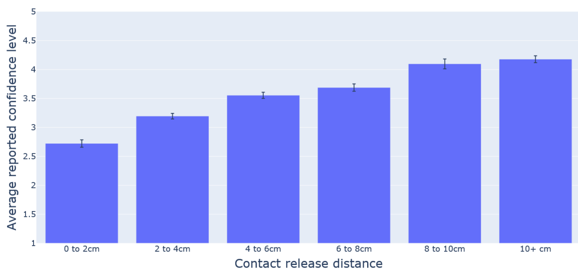

Figure 10: Mean and standard error of confidence level reported by observers, depending on the contact release distance.

hand or *arm*) on variation selection ($H(2) = 72.3$, $p = 1.854 \times 10^{-17}$). This effect was confirmed by the post-hoc test ($p = 0.0001$). These tests show that observers chose the pose variation significantly less often for the category *hand* (18.00% of answers) than for the category *arm* (34.05% of answers). We also found a main effect of contact release distance (measured with Equation 2), when grouping the questions by intervals of 2cm of release distance as shown in Figure 9 ($H(2) = 74.4$, $p = 1.267 \times 10^{-14}$). Post-hoc Tukey test showed that this effect was only significant between release distances inferior to 2cm and superior to 10cm, and between distances in the 4 to 6cm interval and distances superior to 10cm. We found a significant effect of the target identity in the pose variation selection rate ($H(2) = 10.4$, $p = 0.015$). Post-hoc test showed that participants chose less often the pose variation with different contacts with the top right body shape of Figure 5 than of the two left shapes.

Confidence Level. For all questions, the mean confidence score reported is 3.42 ± 1.29. Similarly to the previous paragraph, we compute a mean confidence score per participant for each studied factor. We found a significant main effect of category on the average confidence score ($H(2) = 20.7$, $p = 5.298 \times 10^{-6}$). For the category *hand*, observers reported a significantly higher confidence in their answers (3.66 ± 1.29) than for category *arm* (3.24 ± 1.26). We also found a significant main effect of release distance on confidence scores ($H(2) = 169.3$, $p = 1.046 \times 10^{-34}$). The results of the Tukey test showed that the confidence score significantly increased between groups of release distance separated by at least 2cm, except between distances in the 4 to 6cm interval and the 8 to 10cm interval, and between distances in the 6 to 8cm interval and distances superior to 10cm (see Figure 10). We found significant effect of the target identity ($H(2) = 13.1$, $p = 0.004$). Post-hoc tests showed that participants were less confident in their answer for the lower left body shape of Figure 5 than for the two right ones.

Table 1: Number of observers reporting a given confidence score after choosing the original contacts pose (same self-contacts as the source pose) or the pose variation (one self-contact released).

Confidence score / Choice	1	2	3	4	5
Pose Variation	105	132	161	186	56
Original Contacts Pose	165	185	303	574	489

We also observed that participants reported different confidence levels depending on their choice. We observe an average confidence score of 2.93 ± 1.22 when the observer chose the pose variation, and 3.60 ± 1.27 when the observer chose the original contacts pose with similar contacts than the source pose. We show in details reported confidence levels of observers depending on their choice in Table 1. We observe that observers tend to report higher confidence levels when choosing the original contacts pose over the pose variation. This observation is shown to be significant by a chi-square independence test between the confidence scores and the choice of the observers ($p = 8.95 \times 10^{-31}$). For questions in the category *hand*, when choosing the pose variation (respectively the original contacts pose), observers reported an average confidence of 2.81 ± 1.30 (respectively 3.84 ± 1.21). For the category *arm*, observers reported when choosing the pose variation (respectively the original contacts pose) a confidence level of 2.97 ± 1.19 (respectively 3.38 ± 1.27). In both categories, these higher confidence scores when choos-

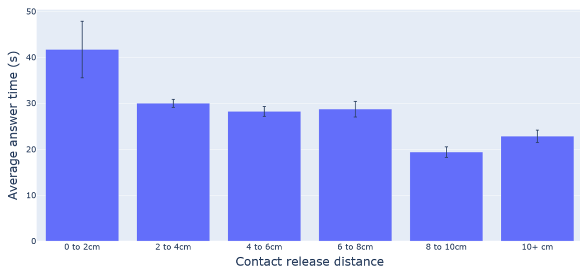

Figure 11: Mean answer time and standard error, depending on the contact release distance.

ing the original contacts pose are significant (chi-square for category *hand*: $p = 4.58 \times 10^{-20}$, *arm*: $p = 7.38 \times 10^{-9}$). However, we observed that the increase in reported confidence is significantly higher ($p = 0.0002$) for the category *hand* (1.03) than for the category *arm* (0.4).

Time Spent. Participants required on average 15 minutes to finish the study. For each question, the mean answer time was 30.16 seconds. As above, we computed the average answer time for each participant for each studied factor. We found a main effect of the category on answer time ($H(2) = 26.5$, $p = 2.578 \times 10^{-7}$), confirmed by the post-hoc Tukey test ($p = 0.0158$). Observers thus answered significantly faster for questions in the category *hand* (26.51s) than for questions in the category *arm* (32.92s). We also found a main effect of contact release distance on answer time ($H(2) = 55.7$, $p = 9.279 \times 10^{-11}$). Tukey post-hoc test showed that observers took longer to answers for questions with release distance between 0 and 2cm than for any of the other intervals but the 6 to 8cm interval (see Figure 11). We found no effect of the target identity on answer time ($H(2) = 3.2$, $p = 0.355$).

Demographic Data. We found no significant effect of the demographic data collected in the study on the choice of transfer result, or on the confidence level reported.

5 Discussion

In this study, we investigated the impact of self-contacts on the perception of pose similarity. We presented observers with characters with a source pose and a target identity, and proposed two new characters with the target identity performing a pose similar to the source pose. One of these transfer results had similar self-contacts than the source pose, while the other had one self-contact released. Observers were tasked to choose which transfer result best imitated the source pose, and to report their confidence in their response on a 1 to 5 scale.

When presented with two possible transfer results, observers chose the variation with a self-contact present in the source pose released in only 27.16% of cases. This selection rate was shown to be significantly lower than a random 50% choice. This tends to confirm our hypothesis **H1**: Pose variations with released self-contacts were perceived as further away from the source pose's meaning. Moreover, observers were significantly more confident in their response when choosing the original pose with the same self-contacts than the source pose (average confidence level: 3.60) than when choosing the pose variation (average confidence level: 2.93). Choosing the pose with the same contacts was thus perceived as the right choice with more confidence, further validating our hypothesis.

The category of self-contact of the questions, *i.e.* whether the self-contact released in the pose variation involved the *hand* or the *arm*, had a clear effect on responses. Observers chose significantly less often the pose variation in the category *hand* (18.00%) than in the category *arm* (34.05%). Observers were also more confident in their answer in the category *hand* (average confidence level 3.66) than in the category *arm* (3.24). This suggests that observers were more confident in their choice, and perceived more easily when there was a difference in the hand contacts than when there was a difference in the arm contacts. Moreover, while observers are in general more confident when choosing the original pose over the pose variation, this confidence increase is significantly higher for the category *hand* (1.03) than for the category *arm* (0.4). This suggests that observers felt the pose with similar contacts was the right choice with more confidence when the contact involved the *hand* over the *arm*. Observers were also significantly faster in their response when the question was in the category *hand* (26.5s) than for the category *arm* (32.92s). This suggests that observers had less difficulties answering when there was a difference in hand contacts. These results confirm our hypothesis **H2**; pose variations of the category *hand* were perceived as different from the source pose more often than for the category *arm*.

We also found an effect on our results of the contact release distance, *i.e.* the distance separating body part surfaces of the pose variation that were in contact in the source pose, measured with Equation 2. Observers tended to choose the pose variation less often when the release distance was very high. They reported increasing confidence levels when the release distance augmented by several centimeters, and took significantly longer to answer when the release distance was very low. All these results suggest that observers considered a high release distance as a different pose, and had trouble choosing which proposition best imitated the source pose when the release distance was low. An interesting question would thus be to determine a threshold above which self-contacts are perceived as released by the observers. However as we did not control exactly the release distance in our generated examples, our data is not suitable for this exploration. An interesting future direction would be to duplicate our pose variations by applying increasing release distances to each. By presenting observers with similar poses with a self-contact released at 1, 2, 3, 4, etc. centimeters, we should be able to determine the threshold above which the self-contact is perceived as released, depending on the body parts involved in the contact.

We found an effect of the body shape of the characters on the participants' answers. This first exploration showed that participants tended to be more confident in their answer and to choose the pose variation with different contacts than the source less often when the target character was thinner. As we focused in this study on the effect of the body parts involved in the contact, our study tested 4 target identities, and explored transfers from a source pose with the average identity from the chosen parametric human body model. Another study design focusing on the impact of identity would be necessary to understand which self-contacts are important for a given morphology. As discussed earlier and illustrated in Figure 2, we argue that some contacts are only present to adapt to the morphology of the characters. To test this hypothesis, an interesting future direction would be to design a similar study where observers are presented with transfer results from an extreme morphology to another, and not only from an average to an extreme morphology. Presented pose variations should also contain new self-contacts absent in the pose in addition to released self-contacts, *e.g.* new contacts between the elbows and the torso for a character with a larger torso.

6 Conclusion

In this paper, we explored the importance of self-contacts to the meaning of the pose. We designed a perceptual study to measure the perception of pose similarity between a source pose and two poses with either similar self-contacts than the source or one self-contact present in the source released. We found that two poses with different self-contacts were perceived as less similar than poses with the same self-contacts in the majority of cases. Moreover, we found that the body parts involved in the self-contact had an impact on its importance. Poses with a difference in contacts involving the hands were perceived as different more often than poses with a difference in contacts involving the arms. In deformation transfer applications preserving inter body part relationships, such as [Jin et al.(2018), Liu et al.(2018), Basset et al.(2020)], preserving all self-contacts involving the hands while allowing contacts involving the arms to change should thus give results that better preserve the meaning of the poses, while allowing changes due to other parameters such as identity.

Our study is a first step towards a perception-based automatic constraint selection for deformation transfer algorithms. Our results confirm that self-contacts are a meaningful property of the pose, and highlight parameters that impact the importance of a specific self-contact. However, further exploration must be conducted to be able to robustly choose which contacts to preserve in a deformation transfer application. When removing a self-contact from a pose, we necessarily change the general pose of the character, *e.g.* slightly different angles between body segments. While our study is focused on the impact of the presence or absence of self-contacts, further exploration could be done in order to understand the impact of this change on the perception of the pose. Another interesting direction for future work is to study the importance of the morphology of characters in self-contact preservation, as our results showed that body shape had an impact on pose perception. Moreover, in this study we chose to focus on male characters to limit the size of the study, in particular the number of required participants. A follow-up study on morphology should explore morphologies of all genders to better understand the impact of body shape on the importance of self-contact. Another interesting direction that was not explored is the second body part involved in the contact; while we compared self-contacts involving the hand or the arm and another body part, we did not compare the relative importance of contacts involving *e.g.* the torso, the head or the thighs of the characters. We also did not explore other categories of contacts, such as contacts with another character or the environment. In these scenarios, contacts involving the arms could have a different importance for the meaning of the pose, *e.g.* when leaning against a wall or holding a pole under and arm. Finally, while we found that higher contact release distances were more likely to change the perceived meaning of the pose, further study must be performed to determine a precise threshold above which self-contacts are perceived as released by most observers.

As shown in related studies on perception of pose equivalences, simple similarity metrics often do not correlate with human perception [Tang et al.(2008), Durupinar(2021)]. Moreover, very few studies explored pose equivalences between characters with very different morphologies. Our results combined with further studies could provide tools to build a pose similarity metric between characters with varying morphologies. This metric would greatly help evaluating or designing deformation transfer approaches, in particular to automatically select constraints to be preserved.

Acknowledgments

This project was funded by the Inria IPL AVATAR project.

References

[Al-Asqhar et al.(2013)] Rami Ali Al-Asqhar, Taku Komura, and Myung Geol Choi. 2013. Relationship Descriptors for Interactive Motion Adaptation. In *ACM SIGGRAPH/SCA (SCA '13)*. 45–53.

[Al Borno et al.(2018)] Mazen Al Borno, Ludovic Righetti, Michael J Black, Scott L Delp, Eugene Fiume, and Javier Romero. 2018. Robust Physics-based Motion Retargeting with Realistic Body Shapes. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 81–92.

[Anguelov et al.(2005)] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. 2005. Scape: shape completion and animation of people. In *ACM SIGGRAPH 2005 Papers*. 408–416.

[Baciu and Iu(2006)] George Baciu and Bartholomew K. C. Iu. 2006. Motion retargeting in the presence of topological variations. *Computer Animation and Virtual Worlds* 17, 1 (2006), 41–57.

[Basset et al.(2020)] Jean Basset, Stefanie Wuhrer, Edmond Boyer, and Franck Multon. 2020. Contact preserving shape transfer: Retargeting motion from one shape to another. *Computers & Graphics* 89 (2020), 11–23.

[Bernardin et al.(2017)] Antonin Bernardin, Ludovic Hoyet, Antonio Mucherino, Douglas Gonçalves, and Franck Multon. 2017. Normalized Euclidean distance matrices for human motion retargeting. In *Proceedings of the Tenth International Conference on Motion in Games*. 1–6.

[Bogo et al.(2014)] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. 2014. FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA.

[Choi and Ko(2000)] Kwang-Jin Choi and Hyeong-Seok Ko. 2000. Online motion retargetting. *The Journal of Visualization and Computer Animation* 11, 5 (2000), 223–235.

[Durupinar(2021)] Funda Durupinar. 2021. Perception of Human Motion Similarity Based on Laban Movement Analysis. In *ACM Symposium on Applied Perception 2021*. 1–7.

[Durupinar et al.(2016)] 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 Transactions on Graphics* 36, 1 (2016), 1–16.

[Gleicher(1998)] Michael Gleicher. 1998. Retargetting motion to new characters. In *Proceedings of the 25th annual conference on Computer Graphics and Interactive Techniques*. ACM, 33–42.

[Harada et al.(2004)] Tatsuya Harada, Sou Taoka, Taketoshi Mori, and Tomomasa Sato. 2004. Quantitative evaluation method for pose and motion similarity based on human perception. In *4th IEEE/RAS International Conference on Humanoid Robots, 2004.*, Vol. 1. IEEE, 494–512.

[Ho et al.(2010)] Edmond S. L. Ho, Taku Komura, and Chiew-Lan Tai. 2010. Spatial Relationship Preserving Character Motion Adaptation. *ACM Transactions on Graphics* 29, 4 (July 2010), 33:1–33:8.

[Hodgins et al.(1998)] Jessica K Hodgins, James F O'Brien, and Jack Tumblin. 1998. Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics* 4, 4 (1998), 307–316.

[Jin et al.(2018)] Taeil Jin, Meekyoung Kim, and Sung-Hee Lee. 2018. Aura Mesh: Motion Retargeting to Preserve the Spatial Relationships between Skinned Characters. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 311–320.

[Johansson(1973)] Gunnar Johansson. 1973. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics* 14, 2 (1973), 201–211.

[Laban(1928)] Rudolf Laban. 1928. Tanzkomposition und Schrifttanz.

[Lee and Shin(1999)] Jehee Lee and Sung Yong Shin. 1999. A hierarchical approach to interactive motion editing for human-like figures. In *Proceedings of the 26th annual conference on Computer Graphics and Interactive Techniques*. 39–48.

[Liu et al.(2018)] Zhiguang Liu, Antonio Mucherino, Ludovic Hoyet, and Franck Multon. 2018. Surface Based Motion Retargeting by Preserving Spatial Relationship. In *Motion in Games (MIG '18)*. ACM, Article 7, 7:1–7:11 pages.

[Loper et al.(2015)] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. 2015. SMPL: A skinned multi-person linear model. *ACM transactions on graphics* 34, 6 (2015), 1–16.

[Lyard and Magnenat-Thalmann(2008)] Etienne Lyard and Nadia Magnenat-Thalmann. 2008. Motion adaptation based on character shape. *Computer Animation and Virtual Worlds* 19, 3-4 (2008), 189–198.

[Marinoiu et al.(2016)] Elisabeta Marinoiu, Dragos Papava, and Cristian Sminchisescu. 2016. Pictorial human spaces: A computational study on the human perception of 3D articulated poses. *International Journal of Computer Vision* 119, 2 (2016), 194–215.

[Molla et al.(2018)] E. Molla, H. G. Debarba, and R. Boulic. 2018. Egocentric Mapping of Body Surface Constraints. *IEEE*

Transactions on Visualization and Computer Graphics 24, 7 (July 2018), 2089–2102. https://doi.org/10.1109/TVCG.2017.2708083

[Olivier et al.(2020)] Nicolas Olivier, Ludovic Hoyet, Fabien Danieau, Ferran Argelaguet, Quentin Avril, Anatole Lecuyer, Philippe Guillotel, and Franck Multon. 2020. The impact of stylization on face recognition. In *ACM Symposium on Applied Perception 2020*. 1–9.

[Popović and Witkin(1999)] Zoran Popović and Andrew Witkin. 1999. Physically based motion transformation. In *Proceedings of the 26th annual conference on Computer Graphics and Interactive Techniques*. 11–20.

[Sumner and Popović(2004)] Robert W Sumner and Jovan Popović. 2004. Deformation transfer for triangle meshes. In *ACM Transactions on Graphics*, Vol. 23. ACM, 399–405.

[Tanaka and Farah(1993)] James W Tanaka and Martha J Farah. 1993. Parts and wholes in face recognition. *The Quarterly journal of experimental psychology* 46, 2 (1993), 225–245.

[Tang et al.(2008)] Jeff KT Tang, Howard Leung, Taku Komura, and Hubert PH Shum. 2008. Emulating human perception of motion similarity. *Computer Animation and Virtual Worlds* 19, 3-4 (2008), 211–221.

[Zhao et al.(2003)] Wenyi Zhao, Rama Chellappa, P Jonathon Phillips, and Azriel Rosenfeld. 2003. Face recognition: A literature survey. *ACM computing surveys (CSUR)* 35, 4 (2003), 399–458.

[Zhou et al.(2010)] Kun Zhou, Weiwei Xu, Yiying Tong, and Mathieu Desbrun. 2010. Deformation transfer to multi-component objects. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 319–325.