

Corpus Creation and Perceptual Evaluation of Expressive Theatrical Gestures

Pamela Carreno-Medrano, Sylvie Gibet, Caroline Larboulette, Pierre-François Marteau

▶ **To cite this version:**

Pamela Carreno-Medrano, Sylvie Gibet, Caroline Larboulette, Pierre-François Marteau. Corpus Creation and Perceptual Evaluation of Expressive Theatrical Gestures. Intelligent Virtual Agents, Aug 2014, Boston, United States. 10.1007/978-3-319-09767-1_14 . hal-01165575

HAL Id: hal-01165575
https://hal.science/hal-01165575

Submitted on 19 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Corpus Creation and Perceptual Evaluation of Expressive Theatrical Gestures

Pamela Carreno-Medrano, Sylvie Gibet, Caroline Larboulette, and Pierre-François Marteau

Université de Bretagne Sud, IRISA, Bâtiment ENSIBS,
F-56017 Vannes, France
{Pamela.Carreno-Medrano, Sylvie.Gibet, Caroline.Larboulette, Pierre-Francois.Marteau}@irisa.fr

Abstract. While human communication involves rich, complex and expressive gestures, available corpora of captured motions used for the animation of virtual characters contain actions ranging from locomotion to everyday life motions. We aim at creating a novel corpus of expressive and meaningful gestures, and we focus on body movements and gestures involved in theatrical scenarios. In this paper we propose a methodology for building a corpus of full-body theatrical gestures based on a magician show enriched with affective content.
We then validate the constructed corpus of theatrical gestures and sequences through several perceptual studies focusing on the complexity of the produced movements as well as the recognizability of the additional affective content.

1 Introduction

Gestures and movements are increasingly exploited in advanced interactive systems populated with virtual agents. Application domains include entertainment, pedagogy and artistic performance. However, while human-to-human communication involves rich, complex and expressive gestures, often linked to verbal languages [11], available corpora of captured motion used for virtual characters usually comprise actions such as locomotion or everyday life motions, but not a large range of examples of expressive gestures.

Our aim is to study these so-called expressive gestures, i.e., gestures conveying some meaningful information and *expressive content*. Expressive content refers to aspects of motion related to feelings, moods, affect, or intensity of emotional experience. Analysis of expressive content has been conducted for artistic performances [4], everyday actions such as knocking or drinking [19], and interactive embodied conversational agents [17]. More specifically, we focus on theatrical gestures because they have to constantly and deliberately attract attention and use body language to express some meaningful scenarios with an emotional intent [12]. We are also interested in the spatio-temporal properties of those gestures, as we make the assumption that they have a richer and enhanced kinematics compared to everyday life actions.

In this paper we propose a methodology for designing a corpus of theatrical gestures in the context of three magic tricks. Experiments have been defined to carry out perceptual evaluations in order to select and characterize relevant gestures and their expressiveness for further analysis and synthesis studies.

The outline of the paper is as follows: section 2 briefly summarizes the existing expressive and non-expressive motion capture databases. Sec. 3 describes the composition and the capture protocol of the theatrical mocap corpus we propose. Sec. 4 describes the perceptual evaluations used to validate our corpus and protocol. Finally Sec. 5 summarizes our contributions.

2 Related Work

Different motion capture databases have been designed to study human behavior. Among them, we can identify those publicly available and largely used by the academic research community for motion analysis, recognition, or synthesis: the HDM05 database [16] provided by the Max Planck Institute, the CMU database provided by the Carnegie-Mellon University [5], and the UTA database provided by the University of Texas at Arlington [24]. These databases comprise a wide range of mocap data from diverse categories including locomotion, sport activities, and everyday life motions. Even though these databases are useful for analyzing the social and relational behaviors of virtual agents, they lack of expressive data carrying information about the meaning conveyed by body movements [6] and the expressive content.

Recently, an increased interest for expressive variations of body movements has led to the design and construction of affective motion capture databases. Among them, two categories may be considered: *i.) Portrayed emotional gestures*, where expressions are produced by actors upon instructions. This category consists of explicit affective archetype gestures where the subjects are instructed to perform short actions or adopt postures that explicitly represent a given emotion [10], [23]. *ii.) Induced emotional expressions occurring in a controlled setting.* In this category we find databases recorded for emotional dance studies, used either for emotion detection [18] or style synthesis [22], and databases used for emotion recognition [3], [13]. Our research objectives belong to the latter. More specifically, we focus on full-body gestures and their varying forms in theatrical scenarios. The gestures are regarded as actions that manifest deliberate expressiveness induced by the emotional state of the actors.

3 Building the Corpus of Expressive Theatrical Gestures

When designing our expressive gesture corpus, we started by defining a theatrical context. A motion lexicon was defined as well as sequences of actions following a meaningful body language associated to a narrative scenario. In this section, we present the reasons that form the basis of our work and describe the selected theatrical scenario.

3.1 Motivation

Identifying and producing *expressive gestures*, i.e., body movements carrying expression, meaning and intent, can be a highly subjective and context-dependent process. Gestures that are meaningful for one observer in a given situation can also be considered movements with no expression by a different observer.

When looking for possible sources of widely known *expressive gestures*, the performing arts (theater, dance, magic, mime, etc.) are a good starting point [12] since they aim at expressing emotions and thoughts through different means (e.g. body, voice, objects). As our primary interest is full-body motion as a medium for expressing affect and meaning, we propose a mime theatrical scenario where stories, ideas, and feelings can be solely conveyed by bodily movements [8].

In addition to providing a new source of expressive gestures in a mime theater context, we also aim to provide a new motion capture dataset that will be useful for research in human motion analysis, recognition and synthesis. This goal has strongly influenced many of the decisions that will be presented in this section.

3.2 Selected Scenario and Gestures

We propose a mime theatrical scenario based on a magician performance. The reasons behind this choice of context are threefold. Firstly, by constraining the actors to portray ideas, emotions and meanings through their body only, we expect they will perform gestures for which the main purpose is to be seen and understood by the whole audience. We assume that those gestures will carry more information and thus involve a higher kinematic and dynamic complexity [12].

Secondly, classical mime performances use bodily movements and facial expressions to portray a character and his emotions [2]. As we are solely interested in body, we need a scenario in which everything has to be shown through hand and body motions. Lastly, a magician is an artist of misdirection. He must thus master and use his entire body to mislead the senses of the spectator while performing [9], [14]. Therefore, we consider it an interesting trial case for the kind of scenario we are looking for.

We developed a scenario in which a magician will perform 3 magic tricks: *the disappearing box*, *pulling a rabbit from a hat*, and *appearing scarves in an empty jacket*. Each magic trick involves 3 stages: *i) Introduction:* the magician makes his appearance and introduces him-self to the audience. *ii) Preparation:* the magician shows each object he is going to use in his magic trick to the public. This stage ends when the magician invokes his magical powers. *iii) Conclusion:* the magician shows the result of his trick and makes a bow to the audience. In total the proposed scenario has 3 sequences consisting of 17 isolated gestures.

3.3 Expressive Variations

As stated before, we think that the gestures in our scenario are spatially and temporally more significant, because they must convey meaning, emotion and intent through bodily movements. In order to enhance this kinematic diversity,

it is possible to introduce new sources of variation into the corpus by taking into account the style, personality and emotional state of the actor.

Although we cannot directly influence the style and personality of the actor, we think that it is possible to elicit certain emotional responses that will produce additional spatio-temporal characteristics. Those characteristics may affect the spatiality, the timing, or the fluidity of the performed gestures. Thus, through eliciting different emotions in the performance of the actor we can increase the diversity and expressive richness of the proposed scenario and corpus. A set of 4 emotional states was chosen using the circumplex model of affect proposed by Russell [21]: *happy, sad, stress, and relaxed*. A *neutral* state was added to categorize the motions in which no emotion was intended.

3.4 Experimental Motion Capture Protocol

To produce a new motion capture database that can be used for the analysis, recognition and synthesis of expressive gestures, special care must be given to the number and variety of recorded gestures and sequences.

Technical Setup: the understandability and expressiveness of gestures require accuracy and high definition in the recording of captured motion. A Qualisys motion capture system composed of 8 Oqus400 cameras [20] was used. All full-body actions and hand movements inside a 2.5mx2mx2m volume were recorded. A total of 64 passive markers were placed on the body of the actor including 5 markers on each hand and 2 facial markers. The markers on the hands enabled capturing all the grasping movements involved in a magic performance, and the facial markers gave a more accurate idea of the direction of the head of the actor. We used a 200Hz capture frequency to correctly capture hand motion, since this kind of motion requires a higher accuracy.

Number of Actors and Repetitions: for the analysis and recognition of human motion, numerous repetitions of a set of actions performed by several subjects are needed. Each magic trick was recorded twice per emotional state. In addition, the most representative gestures (8 in total) were selected among the initial 17. For each selected gesture, 2 sequences of 5 repetitions per emotional state were recorded. Currently, our database contains the motions of 2 skilled amateur actors (1 man and 1 woman). For each actor, 110 motion capture files were produced. We intend to further record 8 additional actors.

Emotion Elicitation: another challenge concerns how the instructions for performing the scenario are given to the actor and how the emotional state is induced. First, a video of each magic trick was presented to the actors the day before the capture. This made possible for the actors to learn the gestures and perform more fluently. Second, on the day of the capture, the actors were asked to perform each magic trick several times before we started to record. By doing so, we could correct all possible doubts about how each gesture should be performed. Last, an emotional state was randomly chosen and the emotion elicitation was done using an imagination mood induction procedure. During the elicitation process, each actor was instructed to remember an emotional event in their lives that corresponded to the selected emotion. After performing the whole

scenario, i.e., the 3 sequences plus the 8 isolated gestures in a given emotional state, a debriefing was done to re-establish the initial emotional state of the actor

4 Experiments and Results

Three perceptual experiments were performed to validate the suitability of the chosen scenario, and the effectiveness and efficiency of our experimental motion capture protocol. The experimental setup also enabled evaluating the usability of the produced mocap data for tasks such as motion analysis, motion recognition and motion synthesis. We were aiming to answer the following questions:

1. Do people perceive theatrical gestures as being more kinematically and dynamically significant than daily actions? Do people perceive theatrical gestures as motions conveying more information?

2. Can observers associate the spatio-temporal variations introduced through the elicitation of emotional states to one of the five selected emotions? If they can do so, how expressive do they find the theatrical gestures?

4.1 Stimuli Creation

For the theatrical gestures, 1 realization for 8 theatrical gestures and for 2 magic tricks were chosen per emotion and per actor. Additionally, the actors were asked to perform 8 daily actions that we consider are the most frequently found in available mocap databases (cf. Table 1 for a list of the chosen stimuli).

Table 1: Stimuli used for the perceptual evaluations

Daily gestures	**Theatrical gestures**	**Sequences**
Lifting	Show empty jacket	The disappearing box
Waving	Take scarves out of jacket	Scarves appear in a jacket
Kicking	Invoke magic with wand	
Hand shake	Show box disappeared	
Walking	Cover box	
Knocking	Invoke magic with hand	
Throwing	Introduction bow	
Punching	Final bow	

All stimuli were played on a of point-light like character (cf. Figure 1). We chose this kind of representation as we did not want to convey any additional information about the avatar's gender and appearance that might influence the categorization of the selected emotions. Additionally, previous studies have shown that this type of representation does not stop observers from perceiving any emotional state at any intensity [1], [15].

For the theatrical gestures and the daily actions, individual video clips of the same duration (10s) were created at 25Hz. For the magic trick sequences videos of 42s were also produced. The character was displayed in the center of the screen, facing forward at the beginning of each clip. Video clips were presented at 1280x1024 resolution and 116 videos were generated in total.

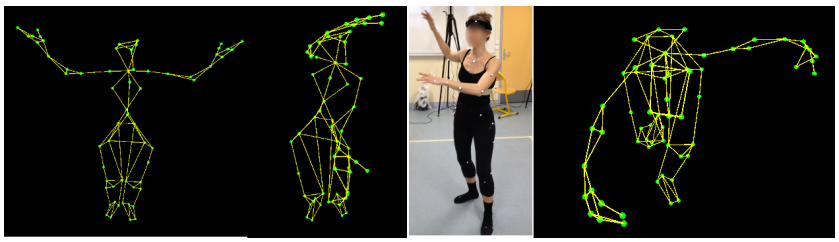


Fig. 1: Marker set and posture examples

4.2 Participants and Duration of Each Study

Twenty participants took part in the studies we will detail in the following section, a total of 100 different individuals contributed to our experiments. Participants came from various educational backgrounds and were all naive to the purpose of the experiment. They only knew they would watch some avatar videos and answer a few questions about what they perceived from those videos. Detailed informations about the gender and age distribution of each group of participants and the duration of each study are presented in Table 2.

Table 2: Information about each study's participants and duration in minutes

Study	Gender	Mean age	Duration
Daily actions vs theatrical gestures	11M, 9F	24.0+10.0	15
Isolated gestures (emotions male actor)	10M, 10F	23.5+6.0	40
Isolated gestures (emotions female actor)	15M, 5F	23+7.0	40
Gestures sequences (emotions male actor)	13M, 7F	21.6+7.5	15
Gestures sequences (emotions female actor)	13M, 7F	25.0+13.0	15

4.3 First Experiment: Everyday Life Movements vs. Skilled Theatrical Movements

In our first experiment we wished to determine whether observers perceived theatrical movements as more kinematically and dynamically significant than everyday life movements. Additionally, we wished to investigate whether participants regarded theatrical gestures as motions conveying more information compared to everyday life actions.

For this study, we presented participants with 32 video clips of 10s duration, depicting 8 daily actions and 8 theatrical movements for each actor. Participants viewed each video clip in a random order, played it as many times as they wished, and after each clip were asked to rate on a scale of 1-7 whether the performed action was considered as current, spontaneous and habitual (1 on the scale) or as skilled, meaningful and elaborated (7 on the scale).

Since the answers of the participants were nominal variables, we did not think the data fits the assumptions of an ANOVA. Results for this study were analyzed using Kruskal-Wallis one-way of variance and paired T-Tests for all post-hoc

analyses. We found that the gender of the participants and actors had no effect on the ratings of daily actions and theatrical gestures. A significant difference ($H = 158.5377, 1d.f, p < 0.001$) between the mean rank scores of the two types of gestures was found. As we confirmed a significant divergence between the two categories of gestures, we were then interested in identifying which particular motions were considered more kinematically significant and conveying more information. The results of the Kruskal-Wallis test ($H = 270.15, 15d.f, p < 0.001$) were significant; the mean ranks scores of 7 of our 8 theatrical gestures were significantly different among the 16 different movements presented to the participants. For daily gestures, we found that *kicking* and *punching* gestures were perceived as the most kinematically significant actions among the everyday motions. A possible reason for this could be that both actions are considered more sportive actions than everyday motions, thus a higher kinematic variance can be attributed to them. Mean rank scores for both categories and for the 16 gestures are shown in Figure 2.

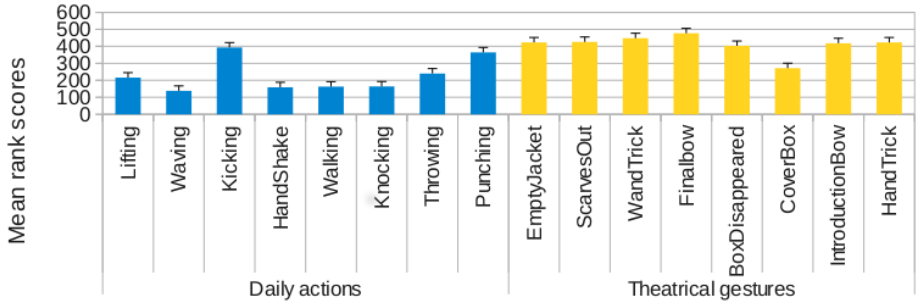


Fig. 2: Mean ranks scores for each category and each one of the sixteen presented actions

4.4 Second and Third Experiment: Perception of Emotion in Isolated Gestures and Sequences

In this study, we take into account the fact that emotional states are expressed differently depending on the subjects and that such states might be more easily recognized over longer stimuli. For this reason, we used 4 separated groups. Two groups rated the emotions of our male character and female character over isolated gestures, and the 2 other groups did the same over the sequences.

We wished to determine whether the 5 emotions portrayed in our theatrical gestures could be recognized among a 6 non-forced-choice list of emotions (the 5 emotions already listed plus the *other* option). Additionally, we wished to investigate the intensity with which each emotion was perceived.

For the isolated gestures, we presented participants with 8 video clips of 10s, representing our 8 theatrical gestures (cf. Table 1 for a detailed list) in each one of the 5 emotional states, where each gesture was presented twice. Participants

viewed each video clip in a random order as many times as they wished and were asked to choose an emotion among the 6 possible options (also randomly presented). They were also asked to rate the intensity of the selected emotion on a scale from 1 (not intense) to 7 (very intense)

For the magic trick sequences, we followed the same methodology applied in the evaluation of isolated gestures. However, instead of using short videos of a unique gesture, we presented participants with a whole realization of a magic trick. For this study only the recordings of *the disappearing box* and *appearing scarves in an empty jacke*t were considered.

Results for these studies were analyzed using standard analysis of variance (ANOVA) and paired T-Tests for all post-hoc analysis. As done in [7], we calculated and analyzed the accuracy rate for emotion, i.e., how many emotions were correctly recognized for each participant. We found no effect of participants and actors gender on the accuracy of emotion identification.

For the isolated gestures experiment we found a main effect of emotion ($F = 18.68, 4d.f, p < 0.001$). Post-hoc tests showed that the 5 emotional states were recognized with means ranging from 29% to 64%. The most accurately identified emotions were *stress* and *sadness*. No main effect of actor gender and type of action were found. However, an interaction between these 2 factors was shown as significant ($F = 2.81, 7d.f, p < 0.007$). This interaction might be due to both actors having different acting qualities for each type of emotion and action.

For the sequences experiment we also found a main effect of emotion ($F = 6.04, 4d.f, p < 0.001$). Post-hoc tests showed that the 5 emotional states were recognized with means ranging from 40% to 70%. Contrary to the isolated gestures experiment, in this study participants were more accurate in emotion categorization. This could be explained by the length of the stimuli presented to participants. We found that the most accurately identified emotional states were *stress* and *sadness*, followed by *relaxed* and *neutral*. As for the isolated gestures, no main effect for actor gender and action type were found. However, an interaction between the emotion and actor factors was again observed ($F = 4.75, 4d.f, p < 0.001$). We believe this interaction might be due to the acting qualities of our two actors.

To have a better insight of where the miscategorizations happened, the confusion matrices of both studies is shown in Table 3. For both studies, the accuracy rate of the participants is above chance (16.6%) and the results obtained for *neutral* and *sad* emotional states are consistent with previous works [7], [25]. Additionally, *stress*, an emotional state that is not frequently used, has the highest recognition rate. However, the *happy* and *relaxed* emotional states were frequently misclassified between them or with the *neutral* state. We think possible reasons for this could be the proximity of these 3 emotional states in the circumplex model [21], the utilization of gestures whose sole purpose is not to convey emotional cues, and the recording of non-professional actors. Furthermore, as we are using gestures that are already spatially and temporally rich, we think it is also possible that the variations added by those 3 emotional states were perceived by the participants as indistinct.

Table 3: Confusion matrices

Isolated gestures						
Correct answer	**Relaxed**	**Happy**	**Neutral**	**Sad**	**Stress**	**Other**
Relaxed	**29.22%**	13.44%	28.13%	14.06%	5%	10.16%
Happy	16.41%	**35.31%**	16.72%	2.03%	16.09%	13.44%
Neutral	17.03%	18.44%	**38.91%**	5%	10.16%	10.47%
Sad	8.44%	1.56%	15.63%	**49.84%**	8.91%	15.63%
Stress	2.81%	6.72%	5.47%	2.34%	**64.69%**	17.97%
Sequences						
Correct answer	**Relaxed**	**Happy**	**Neutral**	**Sad**	**Stress**	**Other**
Relaxed	**50.63%**	10%	13.13%	11.88%	4.38%	10%
Happy	11.88%	**52.50%**	13.75%	1.25%	13.75%	6.88%
Neutral	20%	19.38%	**40%**	3.13%	7.5%	10%
Sad	11.88%	1.88%	13.75%	**53.13%**	1.88%	17.50%
Stress	8.13%	7.50%	6.88%	2.50%	**70.63%**	4.38%

To identify possible significant differences in the intensity of the emotional states, Kruskal-Willis tests were applied. We found no difference between the emotional intensities portrayed for both actors and for each action or sequence. For both studies, the emotions rated as the most intense are those the most accurately categorized ($H = 30.82, 4d.f, p < 0.001$ for the isolated gestures and $H = 14.09, 4d.f, p < 0.001$ for the sequences).

5 Conclusions

In this paper, we have proposed a new motion capture corpus composed of 17 theatrical gestures, in the context of a magic performance, into which emotional variations were added. Three perceptual studies were conducted in order to validate the suitability and usability of our mocap data as well as the relevance of the capture protocol. We found that theatrical gestures can be globally considered more skilled, meaningful and elaborated gestures compared to everyday life actions. We also established that for the selected theatrical scenario, emotional states can be successfully elicited in the laboratory setting, and most importantly that our recognition results are very close to those of previous studies in which archetypal affective actions and more complete visual clues were used [15].

We will improve our set of emotional states and our emotional elicitation procedure. We plan on recording more actors in order to provide a better insight about the influence of acting skills, gender and personal style in the perception of the expressiveness of theatrical gestures.

Acknowledgements This work is funded by the French National Reasearch Agency under the project INGREDIBLE.

References

1. Atkinson, A.P., Dittrich, W.H., Gemmell, A.J., Young, A.W.: Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception 33, 717–746 (2004)
2. Aubert, C.: The art of pantomime. Dover publications, inc. (2003)
3. Bernhardt, D., Robinson, P.: Detecting emotions from connected action sequences. In: Proceedings of the the International Visual Informatics Conference (2009)
4. Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., Volpe, G.: Multimodal Analysis of Expressive Gesture in Music and Dance Performances. In: Gesture-Based Communication in Human-Computer Interaction, pp. 357–358. LNCS (2004)
5. Carnegie Mellon University: Motion capture database (2003), `http://mocap.cs.cmu.edu/`
6. Cowie, R., Douglas-Cowie, E., Cox, C.: Beyond emotion archetypes: Databases for emotion modelling using neural networks. Neural networks 18(4), 371–388 (2005)
7. Ennis, C., Hoyet, L., Egges, A., McDonnell, R.: Emotion capture: Emotionally expressive characters for games. In: Proceedings of Motion on Games. pp. 31:53–31:60. MIG '13 (2013)
8. Examinations, L.: Mime matters. Tech. rep., London Academy of Music & Dramatic Art (2012)
9. Jones, G.M.: Trade of tricks: inside the magician's craft. University of California Press (2011)
10. Kapur, A., Kapur, A., Virji-Babul, N., Tzanetakis, G., Driessen, P.F.: Gesture-based affective computing on motion capture data. In: ACII'05. pp. 1–7 (2005)
11. Kendon, A.: Gesture: visible action as utterance. Cambridge University Press (2004)
12. Lecoq, J.: Theater of movement and gesture. Taylor & Francis (2006)
13. Ma, Y., Paterson, H.M., Pollick, F.E.: A motion capture library for the study of identity, gender, and emotion perception from biological motion. Behavior research methods 38(1), 134–141 (2006)
14. Macknik, S.L., King, M., Randi, J., Robbins, A., Thompson, J., Martinez-Conde, S., et al.: Attention and awareness in stage magic: turning tricks into research. Nature Reviews Neuroscience 9(11), 871–879 (2008)
15. McDonnell, R., Jörg, S., McHugh, J., Newell, F.N., O'Sullivan, C.: Investigating the role of body shape on the perception of emotion. ACM TAP 6(3), 14 (2009)
16. Müller, M., Röder, T., Clausen, M., Eberhardt, B., Krüger, B., Weber, A.: Documentation mocap database hdm05. Tech. rep., Universität Bonn (2007)
17. Niewiadomski, R., Bevacqua, E., Mancini, M., Pelachaud, C.: Greta: An interactive expressive eca system. In: AAMAS '09 - Volume 2. pp. 1399–1400 (2009)
18. Piana, S., Staglianò, A., Odone, F., Verri, A., Camurri, A.: Real-time automatic emotion recognition from body gestures. CoRR (2014)
19. Pollick, F., Paterson, H., Bruderlin, .A., Sanford, A.: Perceiving affect from arm movement. Cognition 82(2), B51–61 (2001)
20. Qualisys AB: Qualisys motion capture systems, `http://www.qualisys.com/`
21. Russell, J.A.: A circumplex model of affect. Journal of personality and social psychology 39(6), 1161 (1980)
22. Torresani, L., Hackney, P., Bregler, C.: Learning motion style synthesis from perceptual observations. In: NIPS. pp. 1393–1400 (2006)
23. University College London: UCLIC Affective body posture and motion database, `http://web4.cs.ucl.ac.uk/uclic/people/n.berthouze/AffectME.html`

24. University of Texas at Arlington: Human motion database (2011), http://smile.uta.edu/hmd/
25. Zibrek, K., Hoyet, L., Ruhland, K., McDonnell, R.: Evaluating the effect of emotion on gender recognition in virtual humans. In: SAP '13. pp. 45–49 (2013)