A Perceptual Study of the Relationship Between Posture and Gesture for Virtual Characters

Pengcheng Luo* and Michael Neff

Department of Computer Science
University of California, Davis
Davis, CA, 95616 U.S.A.
{pcluo,mpneff}@ucdavis.edu

Abstract. Adding expressive body motions that synchronizes with gestures appropriately is a key element in creating a lively, intelligent virtual character. However, little is known about the relationship between body motions and gestures or how sensitive humans are to the errors generated by desynchronized gesture and body motion. In this paper, we investigated the motion splicing technique used for aligning body motion and studied people's sensitivity to desynchronized body motions through two experiments. The first experiment is designed to see whether audio will affect people's sensitivity to desynchronization errors and explore the role of Posture-Gesture Mergers in the transferability of body motion. A motion distance metric for measuring the distance between stylistically varied body motions is proposed and evaluated in a second experiment. The experiments revealed that audio does not affect the recognition rate, but the presence of posture gesture mergers in the source motion lowers output quality, and people's sensitivity to motion realism is related to an energy distance metric.

Keywords: gesture and posture, realistic character motion, distance metric

1 Introduction

Virtual characters with communicative abilities have been a significant research focus in both academia and industry. Body motion is an important aspect of representing believable, intelligent virtual characters, and yet, we have limited knowledge of the relationship between body motion and gesture. In particular, we do not understand which aspects of body motion are perceptually relevant for gestures. In interactive character applications, gestures must often be generated or adapted to match the character's selected speech [1–4]. How can body motion be adjusted to match these gestures? Understanding the perceptual connection between posture and gesture provides guidance as to which aspects of motion animation algorithms must maintain and where liberties can be taken.

* The complemental video can be downloaded: http://www.idav.ucdavis.edu/~pcluo/MIG2012/MIG2012.mp4

The relationship between gesture and posture has been studied in movement theory [5] [6] and psyschology [7]. Lamb defined posture as a movement that is consistent throughout the whole body and gesture as a movement of particular body part or parts [5]. In autonomous virtual character animation, gesture is often a movement of the arms and hands. Lamb and his colleagues [5] [6] observed that posture and gesture are integrated into a whole in many daily actions. They named this phenomenon Posture-Gesture Merger (PGM). However, how to quantify this kind of merger is still not clear and little is known about whether people will be sensitive to desynchronization of posture and gesture in animation. Therefore, we wish to understand how desynchronizing body and gesture will affect people's perception of the realism of a virtual character. Questions of particular interest include: Does the presence of voice improve people's ability to distinguish between real and synthetic motion? Is there a metric on motion itself that could be used to predict human sensitivity to changes in realistic motion?

Answering these questions will inform algorithm design for communicative motion, indicating how body motion may be edited or added to gesture. Previous works has designed models to add idle body motion [8], posture turns [9] based on statistical analysis , or add lower body motion using rule-based and statistical models [10]. We are trying to support and advance this research by providing a perceptually basis, ultimately pushing body motion to a more expressive level that includes different kinds of body motion variations.

Motion transplantation is a process whereby the motion of certain body parts are transferred from one clip to another e.g. copying the lower body motion from one clip to another clip that just contains arm gestures. It provides a simple yet powerful solution to add variance to body motions. However, it will break the coordination among body parts and potentially generate unrealistic body motions.

Human motions are characterized by a large number of independent Degree of freedoms (DOF), yet inner coordination exists among joints. We anticipated noticeable errors if we misaligned gesture and body motions; however, people were only sensitive to some of these errors in our pilot study. Therefore we wish to study which factors might affect people's sensitivity to these desynchronization errors and hope to find an economic way to add motion variations while preserving realism. Our first study examined whether the presence of the audio of the character's voice impacted participants judgement of movement realism. Since synchrony with voice is a key part of multimodal communication, it potentially impacts perceived realism. This experiment also explores the role of PGMs in motion transferability. In a pilot study, we also noticed that whether the motion looks realistic or not depends on the similarity of certain high level motion qualities between the body motion that the gestures are extracted from and the source body motion used in the output. Therefore, we parameterize motion distance based on these factors and designed a second perceptual study to determine the most appropriate distance metric.

Our contributions are summarized as follows:

- A perceptual study that indicates A) that the presence of voice does not impact people's perception of errors between posture and gesture movement and B) body motion that contains posture-gesture mergers will generate less realistic motions when transplanted to motion with other gestures.
- A comparison of several distance metrics for evaluating mismatches between motions that reveals that the variance in motion energy correlates well with user perception of errors.

In this paper, we begin by first summarizing previous work in Section 2. Section 3 presents our first experiment. Important findings include that voice does not affect the recognition of realistic motion. Body motions without posture-gesture mergers could be aligned with other gestures and generate more realistic animations than those with posture-gesture mergers. Section 4 will describe the experiment that evaluates the distance metrics for body motion. The experiment showed that one of our distance metric is highly correlated with people's rating of realism. A detailed discussion is provided at the end.

2 Literature

Working initially as a student and then colleague of Rudolf Laban, movement theorist Warren Lamb argued that movements can be divided into postural and gestural movements [5]. When moving in a sincere, authentic way, people will often exhibit Posture-Gesture Mergers where the posture and gesture movement is merged into a single, unified, whole body motion [6]. Psychologists have confirmed that "Posture-Gesture Mergers accompany verbal expressions that are truthful, relaxed, sincere or authentic" [7].

Previous researchers have done many perceptual studies trying to discover which aspects of motion are most salient and to develop different perceptual metrics. Harrision et al. [11] conducted several experiments to study the sensitivity to changes in limb length under different amount of attention paid by a subject. Reitsma et al. [12] added errors to human jumping motion and studied user sensitivity to these errors. The most similar work to ours is from Ennis et al. [13]. They investigated two factors - audio mismatches and visual desynchronization - that affect the plausibility of virtual characters in group conversation. We instead focus on the gesture and body synchronization for a single speaking agent.

Some previous work has focused on adding body motion to virtual agents. Egges et. al. [8] provided a statistical framework for adding small posture variation in idle motion for virtual characters. Cassell et al. [9] analyzed the frequency of body motion along with discourse structure and simulated body turns based on these statistical observations. Levine et al. [14] generated body body motion by learning a conditional model from a motion data base and driving motion generation from voice. Luo et al. [10] add lower body motion by combining data-based and rule-based approaches. This paper is focused on developing perceptual metrics that can be used to evaluate acceptable body motion. In particular, we are interested in enabling a wider range of expressive body motion variations.

A considerable amount of previous work focuses on developing computational frameworks for producing gestures for virtual characters. Kopp and Wachsmuth [3] designed a system that generates gesture, facial behaviors and speech from XML-based descriptions and synchronizes them with co-articulation and transition effects. Neff et al. [4] generated gestures for a conversational agent based on learned statistical models from video analysis. Levine et al. [14, 15] animated gestures by learning hidden structure between acoustic features and body language from motion capture data. New motions are created given new voice input. Our work focuses on understanding the relationship between gesture and posture and could be complementary to previous gesture works.

Motion transplantation copies the motion of particular body parts in one clip to another clip. Ikemoto et al. [16] designed several rules for transplanting limbs. Heck et al. [17] studied methods to splice upper body motion and locomotion. Differing from these, we are studying acyclic motions and no clear coordination rules were observed between body parts.

Motion parameterization is one of the topics that many researchers are interested in. Chi et al. [2] presented the EMOTE system that synthesizes gesture based on Effort and Shape qualities derived from Laban Movement Analysis. Zhao et al. [18] designed a neural network that maps from extracted motion features to motion qualities in terms of Laban movement analysis effort factors. Hartman et al. [19] quantify the expressive content of gesture based on a review of the psychology literature. Castellano et al. [20] quantized motion in video based on Laban movement theory. Instead of gestures, we are more focused on quantifying expressive body motions based on perceptual cues.

Based on different parameterization methods, different metrics have been proposed to measure the distance between motions. Muller et al. [21] extracts geometric features from motions and calculate distance using dynamic time warping. Kovar et al. [22] pointed out that numerically similar motion is not logically similar motion and thus they defined a distance metric based on logical criteria. Time correspondence is one of the most important factors they used in their metric. Instead, we are looking at the similarity of body motions from perspective of style and user ratings from a perceptual study are used to evaluate several potential distance metrics. Onuma et al. [23] calculated distance between motion sequences based on the mean of kinetic energy as well as the distribution of energy on different body parts. We utilized the energy concept but computed in a different way.

3 Experiment 1: The Impact of Voice and Posture-Gesture Mergers

3.1 Data Collection

We recruited a professional female dancer to generate a set of motion samples for our experiment. We asked her to describe getting her first car, using two different types of body motion in two separate sequences. The first used normal,

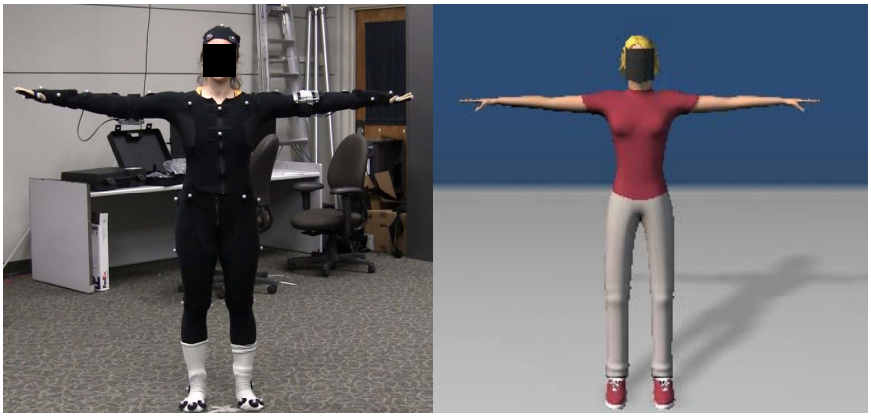

Fig. 1. Left: actress is performing action in our mocap lab; Right: the character we used for retargetting.

"random" body motion(i.e. natural movements such as weight shifts or small steps, but without an explicit effort to coordinate the motion throughout the body - her default style of movement). For the second type, she made an effort to make more full body motions, encouraging greater posture gesture mergers. We post-reviewed the videos to ensure that the captured motion sequences satisfy the requirements.

Each sequence lasted about 2 minutes. We attached 53 makers to performer's body and the motion was recorded with 12 Vicon motion capture cameras. A regular video camera was used to record video and audio. The performer faced towards the camera most of the time. The finger motion was recorded with a separate pair of data gloves and integrated into the body motion using third-party software. The recorded motions were retargeted to a virtual character and voice was manually aligned. As head and facial motion is not studied in our experiment, we used a mask to cover the character's face in the resulting animations. The motions were rendered in Maya (Fig. 1).

3.2 Stimuli and Procedure

motion name	motion parameters
gesture	finger motion, arm motion, head/neck motion
body motion	torso motion, lower body motion

Table 1. Body parts included in "gesture" and "posture" components of motion.

For the purpose of studying the desynchronization of gesture and body motions, we picked two small samples, 15 seconds in length, from each motion

sequence. This provides four motion samples, two of them with strong posture-gesture mergers (G_1B_{1-M} and G_2B_{2-M}), and the other two without posture-gesture mergers(G_3B_{3-NM} and G_4B_{4-NM}). Each clip can be separated into a "posture" and "gesture" component. The body parts considered for gestures versus posture motions are defined in Table 1. The motion samples are selected with random start times and there is no overlap between any two motion samples. Each motion is represented in the parameterization developed in Neff and Kim [24], which facilitates editing. Each body motion was combined with all the gesture clips. Therefore we have 16 motion clips in total including 4 original body motion, 6 clips that mismatched body motion with posture-gesture merger(G_iB_{j-M}) as well as 6 clips that mismatched body motion without posture-gesture merger(G_iB_{j-NM})($i \neq j$). As the face was covered with a mask, no facial motion was shown to participants.

To evaluate the effect of voice, we designed two versions of the stimuli, with and without voice. In the first part of the experiment, we showed the 16 videos without voice to participants. Video order was randomized to avoid any ordering effects. In the second part of the experiment, subjects viewed the clips with voice, also in random order. After watching each video, the participants answered the prompt "Please indicate whether this video seems real or synthetic " with one of the three options: "real", "synthetic" or "not sure". The reason for using three options is that some participants felt it hard to rate videos with only two options to be either real or synthetic, therefore we provided the third, neutral option. Some participants also felt it hard to rate videos accurately when given many options, so we restricted the choices to three. Before the survey, we only told participants that "real" is playing back the real captured motion while "synthetic" clips are altered in some way, as was done with the experiment design in Ennis et al. [13]. In designing this experiment, our main hypothesis was that voice would help subjects recognize the synthetic body motion.

3.3 Result and Analysis

We recruited 18 people(4 female and 14 males with ages from 21 to 35) to participate in our study. For analysis of the experiment, we conducted a repeated measures analysis of variance (ANOVA) with the factor "voice"(on, off). The results showed that there is no significant difference between the average rating for synthetic motion with and without voice ($F(1,34) = 0.114$, $p = 0.204$). The average rating of synthetic motion without voice is -0.0185 (.079 SE) and 0.0138 with voice (.085 SE).

We also examined at the factor body motion with and without posture-gesture mergers by comparing the average rating of G_iB_{j-M} and $G_iB_{j-M}(i \neq j)$. ANOVA test shows that this factor makes a significant impact to people's sensitivity to realistic motions ($F(1,22)=11.87$, $p=0.002$). From Fig. 2, we can see clips that mismatched body motion with posture-gesture merger will generate motions that look synthetic while those that mismatched body motion without PGM will look more realistic on average. This means body motions with posture-gesture mergers have a stronger connection with specific gestures,

thus require more careful consideration when transplanting them to other gestures. An ANOVA test confirmed that voice does not affect the ratings in both cases(Body motion with PGM: $F(1,10)=0.006$, $p=0.94$) ; Body motion without PGM: $F(1,10)=0.04$, $p=0.83$)). One thing worth noting is that ratings of clips that align with PGM body motions with voice have a larger variance compared to other cases(0.55 compared with 0.25, 0.25, 0.18). This might mean that voice has the largest effect on PGM body motions.

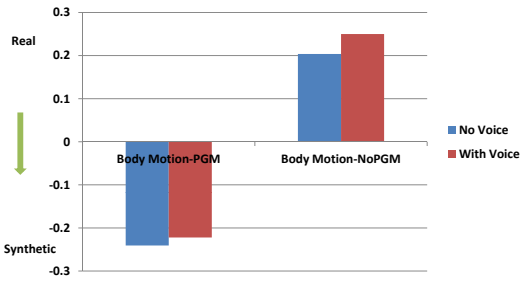

Fig. 2. Left: average rating for mismatched Body Motion of PGM; Right: average rating for mismatched Body Motion Without PGM.

4 Experiment 2: Evaluation of Distance Metrics

4.1 Data Collection

To evaluate people's sensitivity to various motion differences, we asked the same performer to perform multiple motions with variations of either topic or emotion. The prompts we used included:

- Please tell a story about purchasing your first car.
- Pretend you are arguing with your boss over unfair treatment at work.
- Please tell us a story about catching a fish.
- Please tell the story about going fishing using the emotions: happy/tired/angry/calm (each emotion was performed separately).

For each of the first three prompts, the performer did one variation in which she tried to perform full body motion in a unified way and one in which she tried to make no particular connection between posture and gesture. The reason for having two different types of body motions is to have more variance in the motion samples. Using different emotions in the scenario provides an easy and effective way to elicit different motion samples. The capture session lasted about 2 hours. All the motions were retargetted to the virtual character mentioned before.

4.2 Stimuli and Procedure

For the purpose of examining people's sensitivity to motion realism, we picked 17 small samples from captured motion sequences. The motion samples covered all the captured motion sequence and there was no overlap between them. Each example lasts 15 seconds. We picked one motion sample as the base and used the gestures and audio from it, adding all the other body motions to it. This resulted in 17 motion stimuli, all with the same gesture and voice, but different body motions. These 17 motions stimuli were shown to participants in random order. After reviewing each motion clip, participants answered the following prompt "Please rate whether this video seems real or synthetic on the following 5-point scale from 1 (Clearly synthetic) to 5 (Clearly real)". They could view each video as often as they liked.

4.3 Motion Parameters

In the pilot study, we found that whether motion looks realistic or not is related to how different the style of the body motion source clip was from the style of the body motion clips in gesture source clip. Therefore we proposed several distance metrics and evaluated them based on people's ratings of motion realism. Interesting parameters studied in our experiment includes space and "energy". Space describes the volume a character occupies during speech based on the observation that a person, having a large stance width/large step size, appears different from the one with small stance width/small step size. In our work, space is calculated as the volume of the minimum cuboid that covers a person's body, excluding the arm gestures. "Energy" measures the amount of energy and tension invested in movement and is calculated as the summation of the square of velocity:

$$\sum {v_i}^2$$

where v_i is the velocity of joint i. We looked at these joints: *Left Foot, Right Foot, Left Knee, Right Knee, Left Thigh, Right Thigh, Left Shoulder, Right Shoulder, Chest, Abodomen*. The intuition of using "Energy" parameter is that a person, moving fast looks different from the one that moves slow. Examples of Energy and Space parameters for different motions can be found in fig. 3.

Each body motion b_i can be characterized as $(p_{1i}, p_{2i}, \cdots, p_{ni})$, where n is the number of parameters or "descriptors" calculating particular properties of the motion ($n \geqslant 1$). Interesting statistical descriptors used in our experiment include *the mean of space*, *the variance of space*, *mean of energy* as well as *the energy variance*. We were not sure which descriptor or descriptors are most related to perceptually noticeable differences in motion, therefore we proposed two distance metrics to evaluate different combinations of descriptors: Manhattan distance and Mahalanobis distance. The distance is calculated between the body motion of the base clip where the gestures come from and the other body motion used to align with the gestures. For a single descriptor, the distance is calculated using the Manhattan distance:

$$d_{ij} =\mid b_i - b_j \mid \tag{1}$$

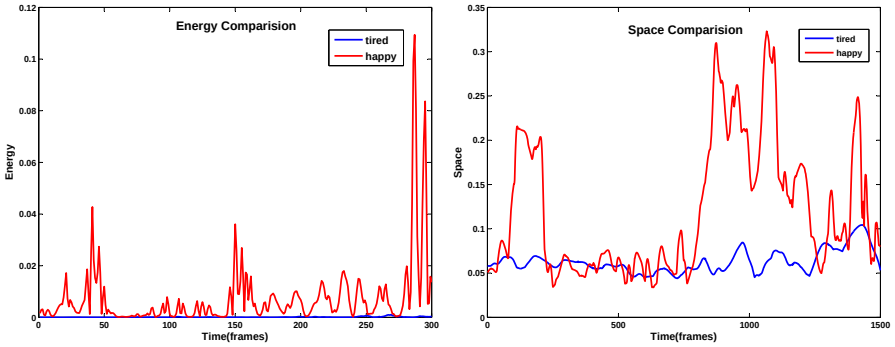

Fig. 3. Left: energy comparison between tired body motion and happy body motion along time; Right: space comparison between tired and happy body motion along time

For multiple descriptors (multiple dimensions), the Mahalanobis distance is used in our experiment:

$$d_{ij} = \sqrt{(b_i - b_j)^T S^{-1} (b_i - b_j)} \tag{2}$$

where S is covariance matrix.

4.4 Result Analysis

18 people were recruited to watch the videos including 14 males and 4 females with ages ranging from 21 to 35. They were paid with a gift card after the experiment. The Pearson's correlation coefficients between different distances and user's ratings (clearly synthetic to clearly real) are listed in Table 2.

	distance on space variance	distance on space mean	distance on energy variance	distance on energy mean	distance on combination of all four measures
Correlation Coefficient	-0.16	0.32	-0.76	-0.36	-0.56

Table 2. The correlation measure between different distance metrics and participant ratings on motion realism.

From the table, we can see that people's sensitivity to realistic character motion is highly correlated with the variance of motion energy. Other parameters (space mean, space variance, energy mean) are not strong enough to describe the difference between body motion which matches people's perception of the realism; the combination of four parameters also decreases this ability compared with energy variance alone. Therefore, we produced a regression based on least

square estimation to represent this relationship. The formula is listed as follows:

$$y = -253.07x + 3.66$$

By running the analysis of variance(ANOVA) test, we found the relationship is significant($F(1,15)=8.68$, $p = 0.0004$) (Fig. 4). Therefore the variance of motion energy is a good descriptor to differentiate motions and we can use this descriptor to predict which body motion could be selected to align with gesture in order not to produce noticeable artifacts.

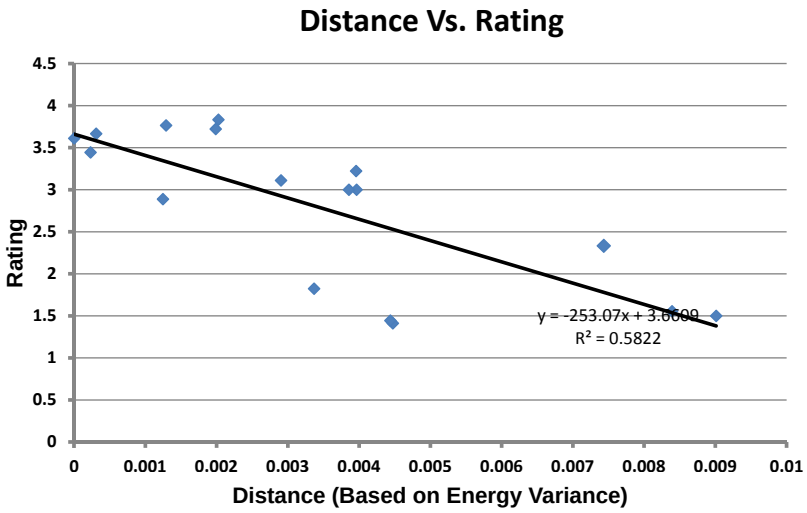

Fig. 4. Regression Analysis between distance and rating

5 Discussion and Conclusion

This paper presented a pair of perceptual studies that evaluated several parameters that might affect people's perception of gesture and body motion synchronization. Through these experiments, we are trying to understand the fundamental of movement and how the unified body is working together to make movements. The studies produced three main results. Result 1 showed that the presence of voice did not affect people's ability to perceive synthetic motion. Result 2 showed that, compared with body motions that have posture-gesture mergers, body motions without posture-gesture mergers could be aligned with other gestures and generate more realistic animations. Result 3 showed that people's sensitivity to errors in body motion is highly correlated with the variance of energy. It is therefore possible to predict whether a motion will look realistic by comparing this motion feature.

Our findings revealed that in some cases, combing different body motions with gesture could also generate realistic motions and how sensitive viewers are to this kind of error can be measured using the proposed motion metric (energy variance). This will support the design of algorithms that allows the reuse of body motion for gesturing characters while maintaining acceptable levels of realism.

This paper examines how mismatching body and gesture could create unrealistic motions. However, realism is one of many metrics that could be used to judge a motion. People could have more complex interpretations of mismatches between body and gesture. For example, mismatching body and gesture might change the personality of a virtual character, which warrants future investigation.

Many other parameters that might affect people's perception of realistic motion have not been studied in this paper. For example, in cyclic body motion, the timing correspondence between gesture and posture might be very important. In this paper we only try to characterize body motion difference using energy and space without considering correspondence of timing, which was considered in previous papers [22, 21]. Although most of the body motions that occur along with gestures are noncyclic motions, it is worth investigating some of the cyclic examples in the future. Besides energy and space, more parameters could be proposed and are worth further evaluation.

The ideal duration for of each clip in the experiment also needs further study. [13] used 10 second clips while we used 15 seconds. Fatigue would be another potential factor that affects people's sensitivity, especially after watching many videos. A rigorous experiment is planned to evaluate this in the future.

Acknowledgments. We would like to thank all the performers in the motion capture and the participants in our perceptual study. Financial support for this work was provided in part through NSF grant 0845529. The models were built by Jonathan Graham.

References

1. Hartmann, B., Mancini, M., Pelachaud, C.: Formational parameters and adaptive prototype installation for MPEG-4 compliant gesture synthesis. In: Proc. Computer Animation 2002. (2002) 111–119
2. Chi, D.M., Costa, M., Zhao, L., Badler, N.I.: The EMOTE model for effort and shape. In: Proc. SIGGRAPH 2000. (2000) 173–182
3. Kopp, S., Wachsmuth, I.: Synthesizing multimodal utterances for conversational agents. Computer Animation and Virtual Worlds **15** (2004) 39–52
4. Neff, M., Kipp, M., Albrecht, I., Seidel, H.P.: Gesture modeling and animation based on a probabilistic re-creation of speaker style. ACM Transactions on Graphics **27**(1) (March 2008) 5:1–5:24
5. Lamb, W.: Posture and gesture: an introduction to the study of physical behavior. Duckworth, London (1965)
6. Lamb, W., Watson, E.: Body code: The meaning in movement. Londres

7. Nann Winter, D., Widell, C., Truitt, G., George-Falvy, J.: Empirical studies of posture-gesture mergers. Journal of Nonverbal Behavior **13**(4) (1989) 207–223
8. Egges, A., Molet, T., Magnenat-Thalmann, N.: Personalised real-time idle motion synthesis. In: 12th Pacific Conference on Computer Graphics and Applications. (October 2004) 121–130
9. Cassell, J., Nakano, Y., Bickmore, T., Sidner, C., Rich, C.: Annotating and generating posture from discourse structure in embodied conversational agents. In: Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents, Autonomous Agents 2001 Conference. (2001)
10. Luo, P., Kipp, M., Neff, M.: Augmenting gesture animation with motion capture data to provide full-body engagement. In: Intelligent Virtual Agents, Springer (2009) 405–417
11. Harrison, J., Rensink, R.A., van de Panne, M.: Obscuring length changes during animated motion. ACM Transactions on Graphics **23**(3) (August 2004) 569–573
12. Reitsma, P., Pollard, N.: Perceptual metrics for character animation: sensitivity to errors in ballistic motion. In: ACM Transactions on Graphics (TOG). Volume 22., ACM (2003) 537–542
13. Ennis, C., McDonnell, R., O'Sullivan, C.: Seeing is believing: body motion dominates in multisensory conversations. ACM Trans. Graph. **29**(4) (July 2010) 91:1–91:9
14. Levine, S., Theobalt, C., Koltun, V.: Real-time prosody-driven synthesis of body language. ACM Transactions on Graphics (TOG) **28**(5) (2009) 1–10
15. Levine, S., Krähenbühl, P., Thrun, S., Koltun, V.: Gesture controllers. ACM Transactions on Graphics (TOG) **29**(4) (2010) 124
16. Ikemoto, L., Forsyth, D.: Enriching a motion collection by transplanting limbs. In: Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation, Eurographics Association (2004) 99–108
17. Heck, R., Kovar, L., Gleicher, M.: Splicing upper-body actions with locomotion. In: Computer Graphics Forum. Volume 25., Wiley Online Library (2006) 459–466
18. Zhao, L., Badler, N.: Acquiring and validating motion qualities from live limb gestures. Graphical Models **67**(1) (2005) 1–16
19. Hartmann, B., Mancini, M., Pelachaud, C.: Implementing expressive gesture synthesis for embodied conversational agents. In: Proc. Gesture Workshop 2005. Volume 3881 of LNAI., Berlin; Heidelberg, Springer (2006) 45–55
20. Castellano, G., Camurri, A., Mazzarino, B., Volpe, G.: A mathematical model to analyse the dynamics of gesture expressivity. In: Proc. of AISB. (2007)
21. Müller, M., Röder, T., Clausen, M.: Efficient content-based retrieval of motion capture data. In: ACM Transactions on Graphics (TOG). Volume 24., ACM (2005) 677–685
22. Kovar, L., Gleicher, M.: Automated extraction and parameterization of motions in large data sets. ACM Transactions on Graphics **23**(3) (August 2004) 559–568
23. Onuma, K., Faloutsos, C., Hodgins, J.: Fmdistance: A fast and effective distance function for motion capture data. Short Papers Proceedings of EUROGRAPHICS **2** (2008)
24. Neff, M., Kim, Y.: Interactive editing of motion style using drives and correlations. In: Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, ACM (2009) 103–112