A Proposed Dynamical Analytic Method for Characteristic Gestures in Human Communication

Toshiya Naka[1,2(✉)] and Toru Ishida[1,2]

[1] Kyoto University, Kyoto, Japan
[2] Panasonic Advanced Research Lab, Kyoto, Japan
naka.tosiya@jp.panasonic.com

Abstract. In human communication, nonverbal information such as gestures and facial expressions often plays a greater role than language; and some gesture-driven operations of the latest mobile devices have proved to be easy-to-use and intuitive interfaces. In this paper we propose a method of analyzing gestures that focuses on human communication based on the dynamical kinematic model. We have extended the analysis method of our proposed approach to take into account additional effects, such as those exerted by external forces, and we analyze the effects over the entire body of forces generated by gestures. We found that the degree of exaggeration could be quantified by the value of, and changes in, torque values. Moreover, when calculating them taking into account external forces and the moment of drag that is acting on both feet, it is possible to determine the twisting torque of the main joints with a high degree of precision. We also noted "preparation" or "follow-through" motions just before and after the emphasized motion, and found that each behavior can be quantified by an "undershoot" or "overshoot" value of changes in torque.

Keywords: Nonverbal · Communication · Gesture · Virtual reality · Dynamics

1 Introduction

In face-to-face human communication, nonverbal information, such as gestures and facial expressions, often plays a greater role than language [1]. However, few studies of conventional cognitive psychology have quantitatively analyzed the roles and effects of nonverbal information of this type. We investigated the importance of nonverbal cues in human communication via the following approach, to quantify the function of characteristic gestures [2]. There is a wide range of types of nonverbal information that vary according to the context of a conversation, so we focused on the study of gestures, which are an important channel of human communication. Our initial studies revealed certain gestures that were used in Japanese Kabuki, in dance, and in Disney animation. We analyzed the effects of these special gestures, and examined whether they would fit into our proposed exaggerated-gesture model. By collecting dozens of characteristic gestures and applying them to our model, we found we were able to represent the degree of exaggeration in a quantitative manner. We also discovered our proposed

M. Antona and C. Stephanidis (Eds.): UAHCI 2015, Part II, LNCS 9176, pp. 50–57, 2015.
DOI: 10.1007/978-3-319-20681-3_5

model to be applicable to the speaker's exaggerated/emphasized movements in speeches and presentations that held the audience's attention. We regard these commonalities as being of great interest.

Most people accompany verbal information with exaggerated gestures when wishing to emphatically communicate their message. In our previous research, we defined these basic dynamics using an analytic model of certain characteristic behaviors. Some special gestures were quantified in terms of torque values of elements of a human skeletal model. We tend to apply force to the required portion of the arms and body for the action to be emphasized, so it is possible to quantify those effects in terms of the torque applied to each joint. There is a close correlation between gesture and torque applied. Although the use of dynamics as analytical technique for motion is used in sports kinematics [3, 4] and robotics [5, 6], there has been little research that mathematically analyzes communicative gestures [7]. In our previous studies, we proposed several basic mechanisms for analyzing and representing exaggerated gestures, but some problems remained because it was only an initial approach. In this paper, we attempted to develop our proposed model to resolve these problems. In the section below, we set out an overview of the exaggerated-gesture model suggested so far, and its drawbacks. We follow this with our proposal for a new analytical model designed to solve these problems by taking into account the overall effect on the human body. Finally, we summarize some typical experiments and their results.

2 Definition of Gesture Analytic Model

In previous studies of human gesture, social psychological approaches, which generate many observations and subjective evaluations, have usually been adopted. However, although this approach can give a general picture, it does not allow for a precise numerical evaluation of the role of gesture. We therefore focused on special gestures that are frequently used in human communication, and preceded with research to analyze their effects quantitatively. In general, human gesture can be expressed using a skeletal structure that follows a hierarchical link model, such as that shown in Fig. 1.

2.1 Mathematical Analysis of Gestures

Use of the skeletal hierarchical structure shown in Fig. 1 makes it possible to quantitatively define reproducible gestures in terms of the rotational angles around the x, y, and z axes (local coordinate system) of each joint. This method is termed *kinematics* and has hitherto been used chiefly in the field of robotics. Once we define the structure of a human as in Fig. 1, any gesture can be expressed using the following dynamical Eq. 1. In this equation, θ is each joint's rotational angle as a time-series data set such as [θ_1, θ_2, … θ_n], M is the *inertia matrix*, C is the *Coriolis force*, g is the *gravity* term. Further, θ' and θ'' represent the *angular velocity* and *angular acceleration* of each joint respectively.

$$\tau = M(\theta)\theta'' + C(\theta, \theta')\theta' + g(\theta) \tag{1}$$

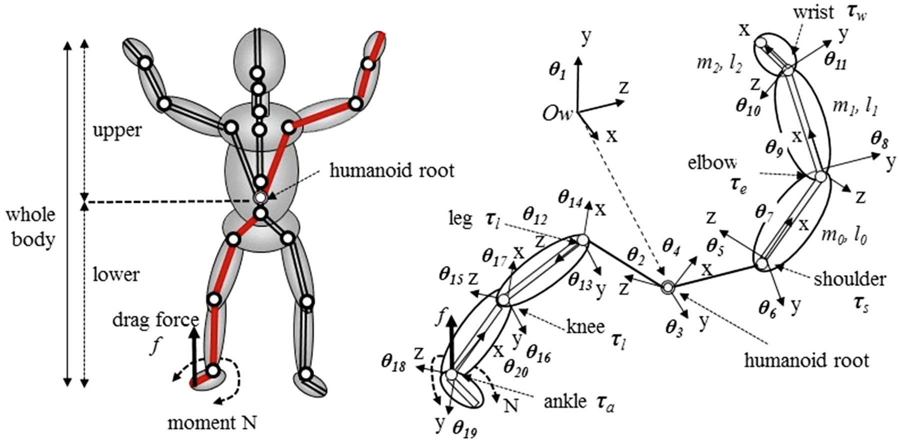

Fig. 1. Hierarchical skeletal model and definition of links

Using this dynamical equation, we can express the relationship between gestures expressed by θ and the torque variation τ of each joint. This equation is a general formula for the dynamical model, and torque τ can be calculated by using the following *Lagrange function* L [2, 8].

$$Q_i = \frac{d}{dt}\left(\frac{\partial L}{\partial \theta_i'}\right) - \frac{\partial L}{\partial \theta_i} \qquad \text{where i } = 0 \sim \text{n} \tag{2}$$

In this equation, Q_i is the equation of motion, specifically the *Lagrange equation*, and for rotational force, Q_i represents the torque value. We can quantitatively analyze the basic mechanism of the enhancement or exaggeration of characteristic gestures using Eq. 2 [2]. In our previous studies, to simplify the problem, we focused chiefly on movement of the upper body while making these special gestures. For the purposes of the generalization of our proposed method, we then extended our analysis to the entire body, including all links from wrist to ankle, via the humanoid root shown in Fig. 1. Generally, humans have to support their upper body and balance their motion using their legs, so it is necessary to consider all the links from foot to hand, shown as solid colored links in Fig. 1, to be able to precisely analyze gestures that are made by the entire body [9]. The feet also need to be regarded as external forces, such as drag and friction forces f and N, from the floor.

$$\tau = J(\theta_i)^T \begin{pmatrix} f \\ N \end{pmatrix} \tag{3}$$

In Eq. 3, f is the drag force acting on the feet, and N represents the moment around the legs. Furthermore, J is the *Jacobian matrix* for θ_i in the general coordinate system. The above equation is the general formula for numerical analysis of extended gestures.

In the following sections, we attempt to parse exaggerated gestures for the whole body using these definitions.

2.2 Target of Our Studies

In this section, we summarize our approach to these problems and the solutions we arrived at. As we have mentioned, in our previous studies it was not possible to analyze gestures while considering their effect on the entire human body and that of other influences such as external forces. The main targets of our studies are as follows.

1. As our newly proposed analysis model of the characteristic gestures, we will extend our original mathematical model to develop a dynamical analysis model that includes the whole body and takes into account all external forces [8].
2. Secondly, we plan to improve the accuracy of identification of the transition point between an ordinary gesture and an exaggerated gesture. This is the most important problem to solve in conventional gesture analysis.

For these two issues and objectives, we devised new strategies and carried out the following verifications.

3 Approach and Experiments

With respect to the first target, listed in Sect. 2.2, we carried out the following experiments to verify our hypothesis. The basic idea is that when reproducing a gesture, we need to take into account all external forces, such as f and N, which will be acting on the links from the hands to the feet. This calculation will be executed using Eq. 3 of the previous section. Regarding the second target, we examined the changes in the joints' torque to be able to precisely determine any enhancement or exaggeration of motion and thus identify the transition point between an ordinary motion and an exaggerated gesture. We named this method the *non-continuous search method*. We will describe the experimental results in more detail in the following sections.

3.1 A Precise Model that Takes into Account the Influence of the Whole Body

In our preliminary experiments, we collected about a hundred characteristic gestures made using the whole body, and classified them into five groups, in a way that resembles the proposal for comparison with our proposed method [2]. Figure 2 shows a typical analysis result for joint torque changes. In this figure, (a) shows the exaggerated gestures in a group of Japanese comedy actions, and (b) the upper graph shows the torque changes, but taking into account only the upper body. The lower graph shows body torque that takes into consideration the entire body, that is, with all external forces calculated using Eq. 3. In this typical example, $\tau 7$ indicates the twisting torque of the elbow joint, and $\tau 9$ shows the twisting torque of the shoulder joint.

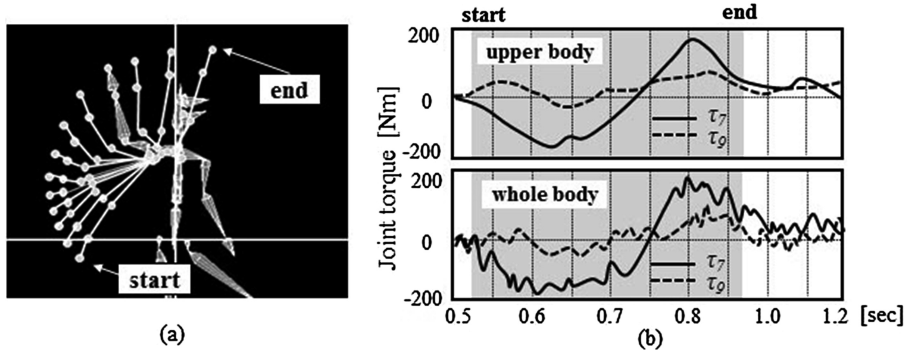

Fig. 2. Comparison of joint torque taking whole body into account (right hand)

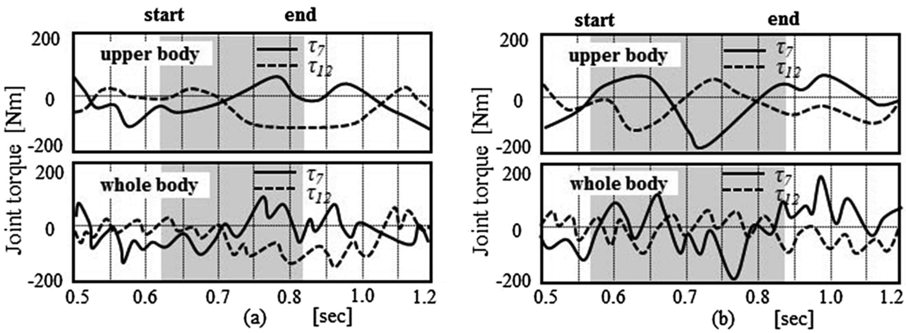

Fig. 3. Comparison of joint torque taking into account the whole body (a) Disney animation and (b) impassioned speech

The entire body needs to be supported with both legs, so we have to take into account the torque due to torsion and drag from the floor. These results reflect the high-frequency component that was applied to the torque due to torsion of the main joints. On the other hand, the changes in the extrapolation of torque curves (low frequency) are maintained.

These patterns were replicated in all the other five gesture groups (fifty-two gestures in all), which we used in our experiments. More analysis results are shown in Fig. 3, in which (a) shows gesture exaggerations which are often used in Disney animations, and (b) shows gestures which are used during impassioned speech. In these figures, the regions marked in gray indicate the torque change of gestures, with τ12 indicating the twisting torque of the right leg joint. From these experimental results, it is possible to quantify the degree of emphasis in terms of the torque change in the twist direction of the main joints when taking into account the whole body, similarly to our conventional approach.

3.2 Improving the Accuracy of Gesture Analysis

Next, we examined the potential for improving the accuracy of identification when there was a transition between ordinary motion and the exaggerated part of characteristic gestures. In the previous section, we showed that exaggerated gestures could be quantitatively evaluated by their torque value, as could changes that were generated at the main joints like the gray regions in Figs. 2 and 3. Moreover, we thought that the start and end points of exaggerated gestures could be accurately determined using the following method. The key point noted in our proposed model was that a "preparation" or "follow-through" motion was added just before or after the main part of the gesture [2]. When qualitatively described, these motions correspond to the concept of an "undershoot" or "overshoot," just before or after the main motion, in the variation of the torque curve.

The challenge here is to identify a method of accurately defining the start and end positions of the exaggerated motion. Our response to this problem was to focus on the synchronicity of verbal information, represented by voice intensity, with the exaggerated gestures. We named this method the *non-continuous search method*. The results of our preliminary experiments showed some human motions to be closely and synchronously correlated with voice intensity. The experimental results of monitoring the relationship between major torque variations and the voice variation of typical exaggerated gestures are shown in Fig. 4. The upper graph shows the changes in torque of the elbow, shoulder, and wrist joints respectively, and the lower graph shows the variation in sound data with which they are mutually synchronized. These data show the results for exaggerated gestures during an impassioned address by a famous public speaker. In the following section, we would like to show an example of a new mobile application environment that takes advantage of these methods.

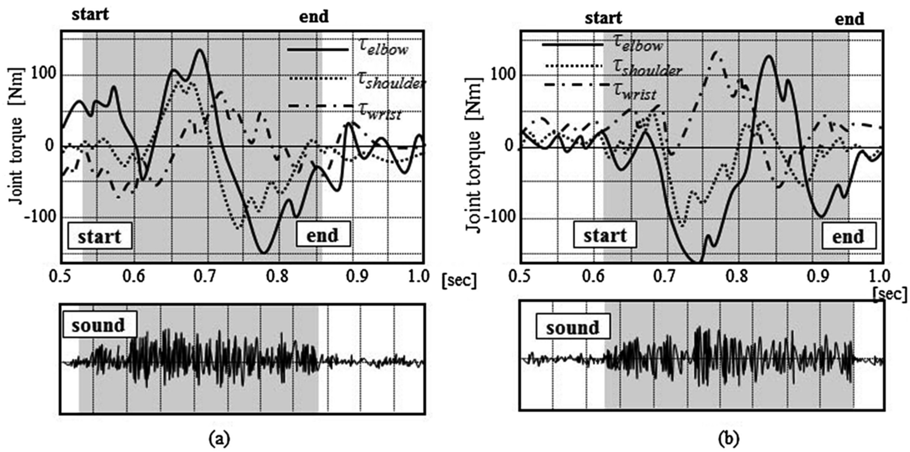

Fig. 4. Typical examples of exaggerated gesture torque (upper) and speech (lower)

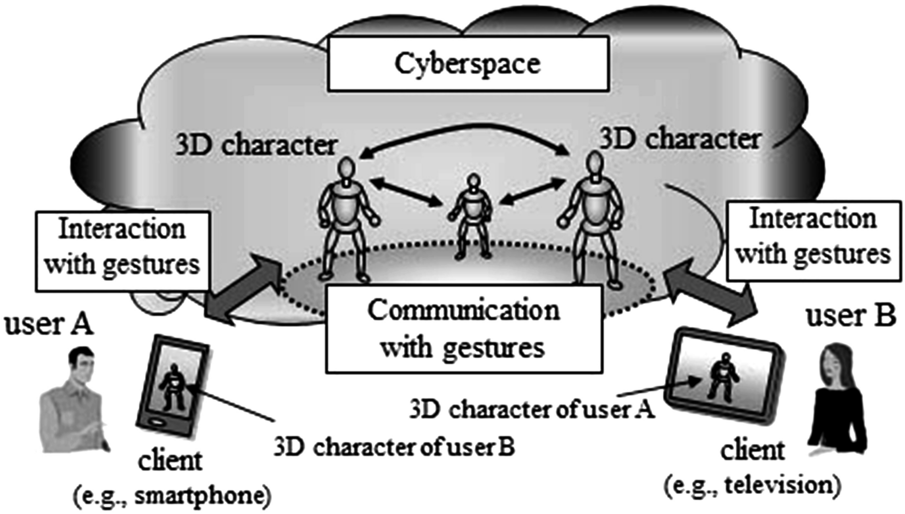

Fig. 5. Overview of mobile communication using 3D characters with realistic gestures

4 Application Using Gestures in Cyberspace

A new mobile communication system which uses 3D characters who make realistic gestures is shown in Fig. 5. Assisted by the 3D character of her or his agent, we are able to engage in natural interactions with multiple other users in 3D over cyberspace. To improve the quality of this communication via 3D characters, it is necessary to reproduce the reality of the appearance and expression of actual gestures. Some user gestures can be captured in real time using cameras on mobile devices. However, simply playing back captured movements does not give the characters a convincingly realistic quality. By using the gesture analysis method described in Sect. 3, however, it is possible to emphasize the torque that is applied to the main joints. As a result, the realistic nature of movement is enhanced, markedly improving the quality of communication. Our proposed model has the potential to play an important role in this kind of application when designing user-gesture interfaces.

5 Conclusions

We took note of the special gestures that are used in face-to-face conversations and studied how to quantitatively analyze their mechanism. In our investigation, we extended our conventional approach to analyze ordinary gestures, taking the effects of external forces into consideration. As a result, we were able to analyze whole-body motions with precision. After an investigation of dozens of characteristic gestures, we obtained the following conclusions.

1. The degree of exaggeration of human motion can be quantified by the total value and the extent of changes in torque value that are acting on the major joints of the

skeletal structure. By taking into consideration, during the calculation, the external forces and moments acting on both feet, it is possible to determine the torque of the main twisting components of joints with a high degree of precision. Generally, the more a speaker wants to emphasize his or her message, the greater the additional changes in rate of torque and torque value applied to the joints.

2. Just before or after the exaggerated or emphasized motion, there is a pattern of addition of "preparation" or "follow-through" motions. Each gesture can be quantified as an "undershoot" or "overshoot" value of torque. Furthermore, there is a high correlation of 0.87 between an exaggerated gesture and simultaneous volume of speech, making it possible to accurately identify and measure the start and end of exaggeration of a main motion.

Using these results makes it possible to quantitatively analyze gesture interactive mechanisms, making this technique widely applicable to processes such as the design of new interactive user interfaces using anthropomorphic agents.

References

1. Mehrabian, A.: Silent Messages: Implicit Communication of Emotions and Attitudes, 2nd edn. Wadsworth, Belmont (1981)
2. Naka, T., Ishida, T.: Proposal of the effective method of generating characteristic gestures in nonverbal communication. HCI **2**, 102–112 (2014)
3. Uno, Y., Kawato, M., Suzuki, R.: Formation and control of optimal trajectory in human multijoint arm movement-minimum torque-change model. Biol. Cybern. **61**, 89–101 (1989)
4. Putnam, C.: Sequential motions of body segments in striking and throwing skills: descriptions and explanations. J. Biomech. **26**(1), 125–135 (1993)
5. Badler, N., Allbeck, J., Zhao, L., Byun, M.: Representing and parameterizing agent behavior. In: Proceedings of Computer Animation, pp. 133–143. IEEE Computer Society (2002)
6. Kanda, T., Miyashita, T., Osada, T., Haikawa, Y., Ishiguro, H.: Analysis of humanoid appearances in human-robot interaction. IEEE Trans. Rob. **24**(3), 725–735 (2008)
7. Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., Stone, M.: Animated conversation: rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In: ACM SIGGRAPH (1994)
8. Mochizuki, Y., Inokuchi, S., Omura, K.: Generating artificially mastered motions for an upper limb in baseball pitching from several objective functions. IEEE Trans. Syst. Man Cybern. B Cybern. **30**(3), 373–382 (2000)
9. Kudoh, S., Komura, T., Ikeuchi, K.: Stepping motion for a human-like character to maintain balance against large perturbations. In: IEEE International Conference on Robotics and Automation (2006)