Copying behaviour of expressive motion

Maurizio Mancini[1], Ginevra Castellano[2],
Elisabetta Bevacqua[1], and Christopher Peters[1]

[1] IUT de Montreuil
University of Paris8

[2] InfoMus Lab, DIST
University of Genova

Abstract. In this paper we present an agent that can analyse certain human full-body movements in order to respond in an expressive manner with copying behaviour. Our work focuses on the analysis of human full-body movement for animating a virtual agent, called Greta, able to perceive and interpret users' expressivity and to respond properly. Our system takes in input video data related to a dancer moving in the space. Analysis of video data and automatic extraction of motion cues is done in EyesWeb. We consider the amplitude and speed of movement. Then, to generate the animation for our agent, we need to map the motion cues on the corresponding expressivity parameters of the agent. We also present a behaviour markup language for virtual agents to define the values of expressivity parameters on gestures.

1 Introduction

A critical part of human-computer interaction is the ability for systems to respond affectively to users [Pic97]. That means that systems must be able to detect the user's emotional state and to plan how to give an appropriate feedback.

Virtual agent systems represent a powerful human-computer interface, as they can embody characteristics that a human may identify with and may therefore interact with the user in a more empathic manner [RN96].

In our work we focus on the analysis of human full-body movement and the synthesis of a virtual expressive agent, since our general goal is to create a system capable of affective responses. The automatic extraction of movement characteristics is done by EyesWeb [CMV04], while the synthesis is done by the virtual agent called Greta [PB03].

We present a preliminary approach to the creation such a system: an agent that exhibits copying behaviour. For each gesture performed by the human, the agent will respond with a gesture that exhibits the same quality. We do not aim at copying the user's gesture shape but only the quality: that is, some of the physical characteristics of movement. In fact, we are interested in studying the importance of the quality of motion independently by its shape, in order to understand what kind of information it can convey. So, the novelty of our work is that the agent's animation is not a straight copy of the original: only the global characteristics of the input motion are retained. For

example, if a user moves his/her hand far out to the side fast, the agent may move its hand upward fast. In the near future we also want to extend the set of motion cues by introducing for example the fluidity and continuity of movement.

A possible application of the actual system may be to conduct tests in which subjects are asked to evaluate the emotion perceived in the agent's copying behaviour. Then learning algorithms could be implemented to map user's quality of movement to his emotional state. The system will use such information to influence the response and the behavior of the virtual agent.

The presented work is divided into two main parts. The first part focuses on sensing and analysis of data coming from a real environment. The second part describes how this information is used to generate copying behaviour with a virtual agent.

2 Previous work

In the human-computer interaction field, there is an increasing attention on automated video analysis aiming to extract and describe information related the emotional state of individuals. Several studies focus on the relationships between emotion and movement qualities, and investigate expressive body movements ([SW85],[WS86], [DeM89], [Wal98], [BC98], [Pol04]). Nevertheless, modelling emotional behaviour starting from automatic analysis of visual stimuli is a still poorly explored field. Camurri and colleagues ([CLV03], [Cas06], [CCRV06]) classified expressive gesture in human full-body movement (music and dance performances) and in motor responses of subjects exposed to music stimuli: they identified cues deemed important for emotion recognition and showed how these cues could be tracked by automated recognition techniques. Other studies show that expressive gesture analysis and classification can be obtained by means of automatic image processing ([BRI$^+$05], [DBI$^+$03].

Several systems have been proposed in which virtual agents provide visual feedback/response by analysing some characteristics of the user behaviour. In such systems the input data can be obtained from dedicated hardware (joysticks, hand gloves, etc), audio, movement capture, video source. SenToy [PCP$^+$03] is a doll with sensors in its arms, legs and body. According to how the user manipulates the doll in a virtual game, the system is able to understand the emotional state of the user. Taylor et al. [TTB] developed a system in which the reaction of a virtual character is driven by the way the user plays a music instrument. The user has to try to vary her execution to make virtual character reacts in some desired way. Wachsmuth's group [KSW03] described a virtual agent capable of imitating natural gestures performed by a human using captured data. Imitation is conducted on two levels: when mimicking, the agent extracts and reproduces the essential form features of the stroke which is the most important gesture phase; the second level is a meaning-based imitation level that extracts the semantic content of gestures in order to re-express them with different movements

In two previous works the Greta virtual agent has been used to respond to the user input. The first one, presented by Mancini et al. [MBP05], is a system obtained by connecting emotion recognition in musical execution with Greta. That is, the facial expression and head movements of the agent were automatically changed by the audio in input in real-time. The second work by Bevacqua et al. [BRP$^+$06] aimed at animating

the same virtual agent off-line by providing data coming both from video recognition of the user expression/behaviour and from scripted actions.

In addition, preliminary research [CP06] has been presented regarding the calculation of global full-body humanoid motion metrics, in particular quantity and expansiveness of motion, using the EyesWeb expressive gesture processing library [CMV04]. Input can be linked to either a video camera, for sensing a real dancer in the real environment, or a synthetic vision system [Pet05] for sensing a virtual dancer in a virtual environment. Higher level processes, such as behaviour planning, are invariant with respect to the source of the input data, allowing for comparison of behavioural algorithms designed for virtual environments with their real world counterparts.

3 Overview

We present a scenario where an agent senses, interprets and copies a range of full-body movements from a person in the real world. In that scenario, we refer to a system able to acquire input from a video camera, process information related to the expressivity of human movement and generate copying behaviour. Our work focused on full-body motion analysis of a dancer. In our system, the agent responds with copying behaviour accordingly to expressive human motion descriptors like the quantity of motion and the contraction/expansion of movement.

Figure 1 shows an overview of our architecture, based on a *TCP/IP* network connection. The *EyesWeb* block performs the automatic extraction of movement characteristics that are described in the next section. The *Copying* block computes the expressivity parameters (see 6) from the movements cues. Finally the *Animation Computation* and *Visualization* are performed using the virtual agent system Greta.

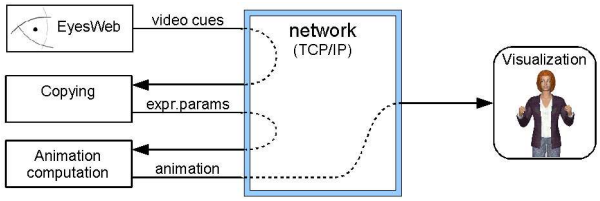


Fig. 1. Overview of the system architecture.

4 Analysis of input data

Our system takes in input video data related to a dancer moving in the space. Video analysis started with the extraction of the silhouette of the dancer from the background. We then automatically extracted motion cues from the full-body movement of the dancer. Analysis of video data and automatic extraction of motion cues were done in EyesWeb

[CCM+04], and particularly by using the EyesWeb Expressive Gesture Processing Library [CMV04]. We analysed global indicators of the movement of the dancer, related to the amount of movement he makes and his use of the space. More specifically, we extracted the following motion cues.

1. **QoM** - *Quantity of motion*
   Quantity of Motion is an approximation of the amount of detected movement, based on Silhouette Motion Images (see Figure 1). A Silhouette Motion Image (SMI) is an image carrying information about variations of the silhouette shape and position in the last few frames.

$$SMI[t] = \{\sum_{i=0}^{n} Silhouette[t-i]\} - Silhouette[t] \quad (1)$$

   The SMI at frame *t* is generated by adding together the silhouettes extracted in the previous n frames and then subtracting the silhouette at frame *t*. The resulting image contains just the variations happened in the previous frames.

Fig. 2. A measure of QoM using SMIs (the shadow along the dancer's body).

   QoM is computed as the area (i.e., number of pixels) of a SMI, normalised in order to obtain a value usually ranging from 0 to 1. That can be considered as an overall measure of the amount of detected motion, involving velocity and force.

$$QoM = Area(SMI[t,n])/Area(Silhouette[t]) \quad (2)$$

2. **CI** - *Contraction Index*
   Contraction Index is a measure, ranging from 0 to 1, of how the dancers body uses the space surrounding it. That can be calculated using a technique related to the bounding region (see Figure 2), i.e., the minimum rectangle surrounding the dancers body: the algorithm compares the area covered by this rectangle with the area currently covered by the silhouette.

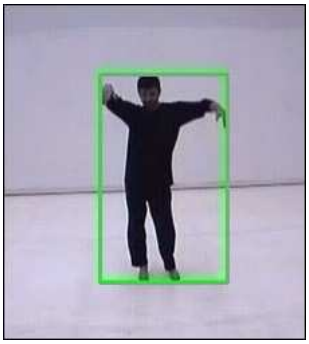


Fig. 3. A measure of CI using the bounding region.

Automatic extraction allows to obtain temporal series of the selected motion cues over time, depending on the video frame rate. As explained in the following sections, the agent motor behaviour generation requires the definition of a specific mapping between the human motion cues and the expressivity parameters of the agent. We segmented the video of the dancer in five main phases, according to the main gestures performed by him. For each phase, we computed the mean of the QoM and the CI of the dancer. These values were mapped onto the expressivity parameters of the agent.

5 Virtual agent expressivity parameters

Several researchers (Wallbott and Scherer [WS86], Gallaher [Gal92], Ball and Breese [BB00], Pollick [Pol04]) have investigated human motion characteristics and encoded them into dimensional categories. Some authors refer to body motion using dual categories such as slow vs fast, small vs expansive, weak vs energetic, small vs large, unpleasant vs pleasant. In particular Wallbott and Sherer have conducted perceptual studies that show that human beings are able to perceive and recognise a set of these dimensions [WS86]. We define *expressivity of behaviour* as the "How" the information is communicated through the execution of some physical behaviour. There are many systems available for synthesising expressive emotional animation of a virtual agent. Badler's research group developed EMOTE (Expressive MOTion Engine [CCZB00]), a parameterised model that procedurally modifies the affective quality of 3D character's gestures and postures motion. From EMOTE the same research group derived FacEMOTE [BB02], a method for facial animation synthesis that altered pre-existing expressions by setting a small set of high level parameters.

The Greta's animation engine is an expressive engine. Starting from the results reported in [WS86], we have defined and implemented expressivity [HMP05] as a set of parameters that affect the gesture (performed by arms or head) quality of execution speed of arms / head, spatial volume taken by the arms / head, energy and fluidity of arms / head movement, number of repetitions of the same gesture. Thus, the same gestures or facial expressions are performed by our agent in a qualitatively different way

depending on these parameters, something with great promise for generating variable character behaviours.

In the present work we will focus on the parameters affecting the spatiality and speed of arm movement, respectively *Spatial extent* and *Temporal extent*:

1. **SPC** - *Spatial extent*
   This parameter basically affects the amplitude of arm movements. The space in front of the agent that is used for gesturing is represented as a set of sectors following McNeill's diagram [McN92]. We expand or condense the size of each sector through scaling. Wrist positions in our gesture language are defined in terms of these sectors (see Figure 4). To find the location of articulation for a gesture, we calculate joint angles needed to reach a some points in the resized sectors.

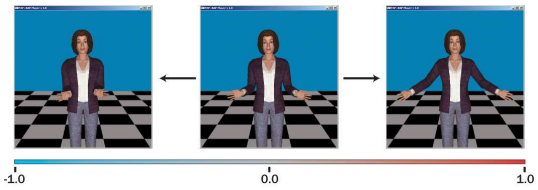


Fig. 4. Spatial Extent - execution of the same gesture can be small or large.

2. **TMP** - *Temporal extent*
   This parameters represents the speed of execution of gestures. We start from a constraint on the position of the end of the gesture stroke because we want to ensure the synchronicity of the gesture end of stroke for example with stressed syllables of speech (in the presented work there is no speech anyway). Then, depending on the temporal parameter we place the preceding and proceeding gesture phases (preparation, stroke start, retraction) in time. Figure 5 illustrates the variation of the stroke start position relative to the stroke end depending on the temporal parameter.

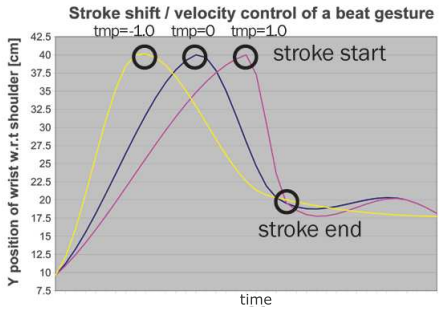


Fig. 5. Temporal Extent - execution of stroke is faster or slower.

5.1 Expressivity specification

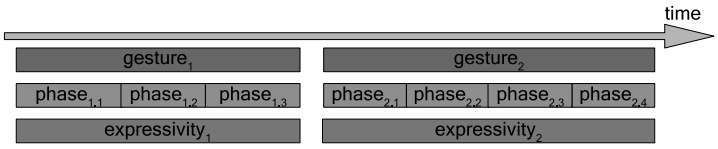
In the Greta system we specify expressivity parameters values either on a sequence of consecutive gestures (that is all the gestures have the same) or on single gestures. In Greta, gestures are stored in a pre-defined gesture library. At runtime they are instantiated following the given timing and expressivity. This information can be passed to Greta by writing a text file in a *ad-hoc* behaviour markup language.

Figure 5.1(a) shows an example of such file. In the example we want Greta to produce two gestures (*gesture1* and *gesture2*) and we added the information about the expressivity of each gesture (see tags *expressivityspc* and *expressivitytmp*). The content of the file can be seen as the temporal diagram in Figure 5.1(b).

```
<gesture id='gesture1' type='BEAT=UPRIGHT' start='0.32' end='1.14'/>
<gesture id='gesture1' stroke='0.500'/>
<gesture id='gesture1' expressivityspc='0'/>
<gesture id='gesture1' expressivitytmp='1.0'/>

<gesture id='gesture2' type='ADJECT=LARGE' start='2.06' end='4.86'/>
<gesture id='gesture2' stroke='0.5'/>
<gesture id='gesture2' expressivityspc='1.0'/>
<gesture id='gesture2' expressivitytmp='0'/>
```

(a)



(b)

Fig. 6. Expressivity parameters values are specified separately for each gesture: the behaviour specified in the example (a) can be seen as a temporal diagram of gestures with different expressivity values (b).

6 Copying of expressive behaviour

As explained in section 4, starting from the real video, the dancer's performance is manually divided in 5 sequences of movements with almost the same shape. In fact, since the dancer moves continuously, without pauses, we have to find a way to identify gestures whose quality we can copy. We select sequences of similar movements as a single gesture with repetitions. For each gesture performed by the dancer the analysis system computes its quality calculating the value of the motion cues CI and QoM.

Then, to generate the animation for our agent, we need to map the motion cues on the corresponding expressivity parameters. In this work we take into account Spatial Extent and Temporal Extent. In fact, the Contraction Index can be mapped to the Spatial Extent since, as we have seen before, the first one describes how the dancer's body uses the space surrounding it and the second one describes the amplitude of movements. Instead, the Quantity of Motion can be mapped to the Temporal Extent since both of them are linked to the velocity of movements.

The analysis system calculates the motion cues for each frame of the video, while the synthesis system needs a single value for each expressivity parameters to define the quality of a whole gesture. For such a reason the analysis system must calculate the mean value of QoM and CI for every gesture performed by the dancer, in this way we obtain a single value that we can map to the corresponding expressivity parameter.

Now, since the motion cues and the expressivity parameters varies in different ranges, a scaling procedure is needed. Regarding the CI, we reversed the sign of Spatial Extent. In fact, if a gesture has a high value of CI, it means that during its performance the arms remain near the body (as explained in 4). So, a high value of CI corresponds a low value of SPC. First of all we define the boundaries mapping the minimum and the maximum values of CI computed during the whole dancer's performance respectively to the maximum and the minimum values of the spatial parameter. Then we can map the CI of each gesture to a meaningful value of SPC using the formula:

$$SPC = (((QoM_{ges} - QoM_{min})/(QoM_{max} - QoM_{min})) * 2) - 1;^{3} \qquad (3)$$

Instead, for the QoM, we simply map its minimum and the maximum values computed during the whole dancer's performance to the minimum and the maximum values of the Temporal Extent to determine the boundaries of the interval. Then we can map the QoM of each gesture to a meaningful value of TMP. Figure 7 shows an example of the mapping between QoM and TMP.

In Figure 8 three sequences of copying behaviour are shown. The corresponding video can be seen at:

`http://www.iut.univ-paris8.fr/greta/clips/copying.avi`

7 Future work

We are working towards the creation of a real-time system that recognises emotions of users from human movement and an expressive agent that shows empathy to them. In the near future we want to extend the set of motion cues by introducing for example the fluidity and continuity of movement. Next steps of our work will be to map movement expressivity of the user to his emotional state, so that the agent can plan affective response accordingly to that. In fact, if the agent can understand the user's emotional state analysing the quality of his gesture, it can decide which facial expression to show.

[3] the SPC paramater can vary in the interval [-1,1]; the value of 2 represents the amplitude of such interval.

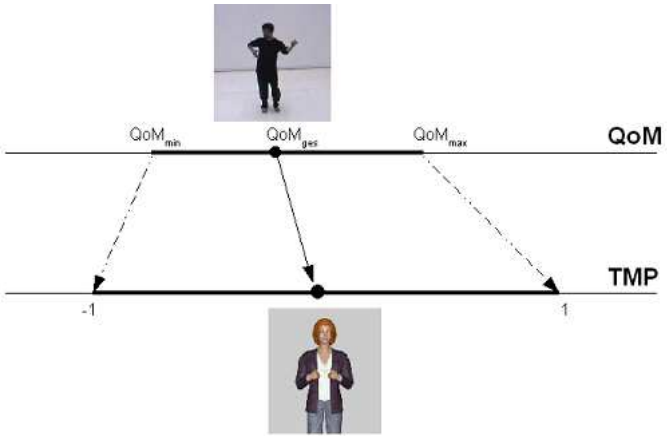


Fig. 7. Example of caling from QoM to Temporal Extent.

We are particularly interested in responding to threat and showing empathy with expressions of happiness or sadness. Here are some examples about how motion cues can be perceived by the agent and how it could respond to them:

- high QoM, low CI and low fluidity (that is large, fast and rigid movement) could be perceived as a threat: in this case Greta could show fear.
- high QoM, low CI and high fluidity (large, fast and fluid movement) could be seen as happiness or enthusiasm: Greta could show happiness.
- low QoM and medium CI (slow and small movements) can be associated to sadness: Greta could respond with sadness.

In the last two examples we want to link the motion cues to empathy.

Further, the definition of a mapping between movement and emotions can contribute to validate algorithms for movement expressivity analysis and synthesis. Perceptive tests with users aimed to associate emotional labels to affective - real and virtual - videos are planned.

8 Acknowledgements

This work has been partially funded by the Network of Excellence Humaine (Human-Machine Interaction Network on Emotion) IST-2002-2.3.1.6 / Contract no. 507422 (http://emotion-research.net/).

References

[BB00] Gene Ball and Jack Breese, *Emotion and personality in a conversational agent*, Embodied conversational agents, MIT Press, Cambridge, MA, USA, 2000, pp. 189–219.

[BB02] Meeran Byun and Norman Badler, *Facemote: Qualitative parametric modifiers for facial animations*, Symposium on Computer Animation (San Antonio, TX), July 2002.

[BC98] R. Thomas Boone and Joseph G. Cunningham, *Childrens decoding of emotion in expressive body movement: the development of cue attunement*, Developmental Psychology **34** (1998), 10071016.

[BRI+05] Themis Balomenos, Amaryllis Raouzaiou, Spiros Ioannou, Athanasios Drosopoulos, Kostas Karpouzis, and Stefanos Kollias, *Emotion analysis in man-machine interaction systems*, Machine Learning for Multimodal Interaction (Hervé Bourlard Samy Bengio, ed.), Lecture Notes in Computer Science, vol. 3361, Springer Verlag, 2005, pp. 318–328.

[BRP+06] Elisabetta Bevacqua, Amaryllis Raouzaiou, Christopher Peters, Georges Caridakis, Kostas Karpouzis, Catherine Pelachaud, and Maurizio Mancini, *Multimodal sensing, interpretation and copying of movements by a virtual agent*, PIT06:Perception and Interactive Technologies (Germany), June 19-20 2006.

[Cas06] Ginevra Castellano, *Human full-body movement and gesture analysis for emotion recognition: a dynamic approach*, Paper presented at HUMAINE Crosscurrents meeting, Athens, June 2006.

[CCM+04] Antonio Camurri, Paolo Coletta, Alberto Massari, Barbara Mazzarino, Massimiliano Peri, Matteo Ricchetti, Andrea Ricci, and Gualtiero Volpe, *Toward real-time multimodal processing: Eyesweb 4.0*, in Proc. AISB 2004 Convention: Motion, Emotion and Cognition, 2004.

[CCRV06] Antonio Camurri, Ginevra Castellano, Matteo Ricchetti, and Gualtiero Volpe, *Subject interfaces: measuring bodily activation during an emotional experience of music*, Gesture in Human-Computer Interaction and Simulation (J.F. Kamp S. Gibet, N. Courty, ed.), vol. 3881, Springer Verlag, 2006, pp. 268–279.

[CCZB00] Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler, *The emote model for effort and shape*, ACM SIGGRAPH '00 (New Orleans, LA), July 2000, pp. 173–182.

[CLV03] Antonio Camurri, Ingrid Lagerlöf, and Gualtiero Volpe, *Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques*, International Journal of Human-Computer Studies, Elsevier Science **59** (2003), 213–225.

[CMV04] Antonio Camurri, Barbara Mazzarino, and Gualtiero Volpe, *Analysis of expressive gesture: The eyesweb expressive gesture processing library*, Gesture-based Communication in Human-Computer Interaction (G.Volpe A. Camurri, ed.), LNAI 2915, Springer Verlag, 2004.

[CP06] Ginevra Castellano and Christopher Peters, *Full-body analysis of real and virtual human motion for animating expressive agents*, Paper presented at the HUMAINE Crosscurrents Meeting, Athens, June 2006.

[DBI+03] Athanasios Drosopoulos, Themis Balomenos, Spiros Ioannou, Kostas Karpouzis, and Stefanos Kollias, *Emotionally-rich man-machine interaction based on gesture analysis*, Human-Computer Interaction International, vol. 4, june 2003, p. 1372 1376.

[DeM89] Marco DeMeijer, *The contribution of general features of body movement to the attribution of emotions*, Journal of Nonverbal Behavior **28** (1989), 247 – 268.

[Gal92] Peggy E. Gallaher, *Individual differences in nonverbal behavior: Dimensions of style*, Journal of Personality and Social Psychology **63** (1992), no. 1, 133–145.

[HMP05] Bjoern Hartmann, Maurizio Mancini, and Catherine Pelachaud, *Towards affective agent action: Modelling expressive ECA gestures*, Proceedings of the IUI Workshop on Affective Interaction (San Diego, CA), January 2005.

[KSW03] Stefan Kopp, Timo Sowa, and Ipke Wachsmuth, *Imitation games with an artificial agent: From mimicking to understanding shape-related iconic gestures*, Gesture Workshop, 2003, pp. 436–447.

[MBP05] Maurizio Mancini, Roberto Bresin, and Catherine Pelachaud, *From acoustic cues to an expressive agent*, Gesture Workshop, 2005, pp. 280–291.

[McN92] David McNeill, *Hand and mind - what gestures reveal about thought*, The University of Chicago Press, Chicago, IL, 1992.

[PB03] Catherine Pelachaud and Massimo Bilvi, *Computational model of believable conversational agents*, Communication in Multiagent Systems (Marc-Philippe Huget, ed.), Lecture Notes in Computer Science, vol. 2650, Springer-Verlag, 2003, pp. 300–317.

[PCP+03] Ana Paiva, Ricardo Chaves, Moisés Piedade, Adrian Bullock, Gerd Andersson, and Kristina Höök, *Sentoy: a tangible interface to control the emotions of a synthetic character*, AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems (New York, NY, USA), ACM Press, 2003, pp. 1088–1089.

[Pet05] Christopher Peters, *Direction of attention perception for conversation initiation in virtual environments*, International Working Conference on Intelligent Virtual Agents (Kos, Greece), September 2005, pp. 215–228.

[Pic97] Rosalind Picard, *Affective computing*, Boston, MA: MIT Press, 1997.

[Pol04] Franck E. Pollick, *The features people use to recognize human movement style*, Gesture-based Communication in Human- Computer Interaction (G. Volpe A. Camurri, ed.), LNAI 2915, Springer Verlag, 2004, pp. 20–39.

[RN96] Byron Reeves and Clifford Nass, *The media equation: How people treat computers, television and new media like real people and places*, CSLI Publications, Stanford, CA, 1996.

[SW85] Klaus R. Scherer and Harald G. Wallbott, *Analysis of nonverbal behavior*, HANDBOOK OF DISCOURSE: ANALYSIS, vol. 2, Academic Press London, 1985.

[TTB] Robyn Taylor, Daniel Torres, and Pierre Boulanger, *Using music to interact with a virtual character*, The 2005 International Conference on New Interfaces for Musical Expression.

[Wal98] Harald G. Wallbott, *Bodily expression of emotion*, European Journal of Social Psychology **13** (1998), 879–896.

[WS86] Harald G. Wallbott and Klaus R. Scherer, *Cues and channels in emotion recognition*, Journal of Personality and Social Psychology **51** (1986), no. 4, 690–699.

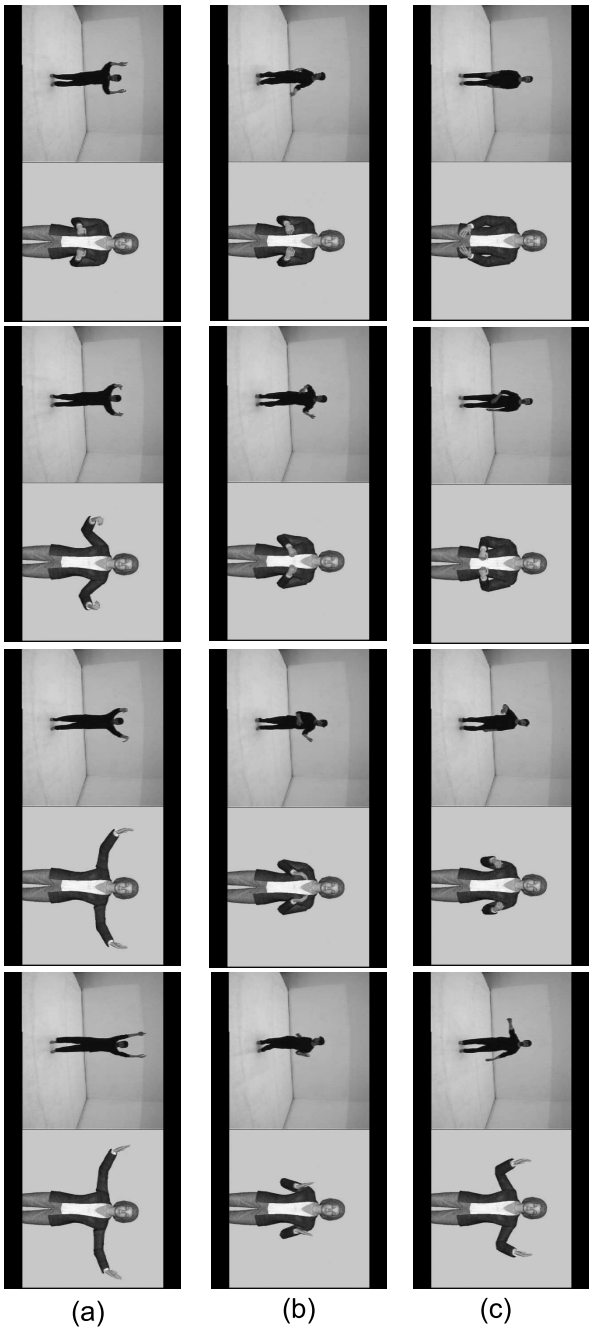


Fig. 8. Three sequences showing copying behaviour with different level of CI: (a) large, (b) small, (c) medium.