Gesture vs. Gesticulation: A Test Protocol

Francesco Carrino[1,2], Antonio Ridi[1,2], Rolf Ingold[2], Omar Abou Khaled[1], and Elena Mugellini[1]

[1] University of Applied Sciences Western Switzerland, Fribourg, Switzerland
[2] University of Fribourg, Fribourg, Switzerland
{francesco.carrino,antonio.ridi,omar.aboukhaled, elena.mugellini}@hefr.ch, rolf.ingold@unifr.ch

Abstract. In the last years, gesture recognition has gained increased attention in Human-Computer Interaction community. However, gesture segmentation, which is one of the most challenging tasks in gesture recognition applications, is still an open issue. Gesture segmentation has two main objectives: first, detecting when a gesture begins and ends; second, recognizing whether a gesture is meant to be meaningful for the machine or is a non-command gesture (such as gesticulation). This paper proposes a novel test protocol for the evaluation of different techniques separating command gestures from non-command gestures. Finally, we show how we adapted adopted our test protocol to design a touchless, always available interaction system, in which the user communicates directly with the computer through a wearable and "intimate" interface based on electromyographic signals.

Keywords: Gesture segmentation, gesture interaction, test protocol, muscle-computer interface, system evaluation and interaction.

1 Introduction

For human beings, the gesture is a natural mean for effectively interacting with objects, tools and for communicating with other people. In 1980, the system developed by R. A. Bolt, known as "Put that there" [1] makes his entrance in the Human-Computer Interaction (HCI) field as the first system combining vocal and gesture interaction. After about thirty years, Microsoft launches Kinect going beyond existing devices such as Nintendo Wii or PlayStation Move, since it offers hands-free interaction. The success of this type of devices, also in research contexts, highlights the interest of exploiting gestures looking for novel interfaces and more natural ways of interaction with computers.

Previous works have explored input techniques based on a variety of input modalities: computer vision, special gloves, muscular activity, etc. However, independently from the approach used, the *gesture segmentation* remains an important problem to be addressed.

Within "gesture segmentation" we include two main aspects: detecting when a gesture is performed (beginning and ending) and therefore separate it from the previous

M. Kurosu (Ed.): Human-Computer Interaction, Part IV, HCII 2013, LNCS 8007, pp. 157–166, 2013.

and the subsequent gestures; recognizing whether the subject is performing a command or a non-command gesture, and consequently changing the machine behavior. The typical example is the normal gesticulation that can be erroneously interpreted by the system as a valid command.

While detecting the beginning and the conclusion of a gesture is a problem often handled in literature, discriminating command and non-command gestures is an aspect rarely discussed. The problem is often avoided restraining the interaction to precise moments (synchronous systems) or using a limited set of unnatural gestures. Aiming at a more natural and always-available interaction experience, what is currently missing is a method that can aid evaluating different segmentation approaches, taking into account also the gesticulation.

In this paper we propose a novel test protocol focused on the evaluation of gesture segmentation approaches. In particular, the proposed method allows the differentiation between command and non-command gestures.

Finally, we show a first evaluation of our protocol in a test on the field. In particular, we adopted our test protocol during the design of an electromyography (EMG) based gesture interaction/segmentation approach. The main goal of the interaction system was to provide to the user an always-available human-computer interface based on subtle and motionless gestures. Triceps contractions were adopted to segment the gestures (i.e., performed gestures are recognized by the system as command-gestures only if they are performed during a contraction of the triceps).

This paper is structured as follow: the next chapter illustrates the works related to the domain of gesture segmentation and offers a survey of the different approaches used in recent studies; in section 3, we will show in detail the developed test protocol and how it has been adapted to the design of an EMG based interaction and segmentation system; finally, we will discuss the results and the possible test protocol improvements.

2 Background and Related Work

In literature, the segmentation problem has rarely been addressed independently, but rather within the gesture recognition context. Command gestures were often distinguished from gesticulation, setting aside very particular gestures for the interaction choosing command gesture too much wide and tiring and so too restrictive for an interaction claimed as “natural”.

When the gesticulation was considered in the context of interaction, this was often done aiming to deduct subject’s characteristics [2] or as a complement in voice-based interaction systems [3]. Alan Wexelblat [4] analyzed the gesticulation in order to deduct the most natural gestures and thus designing an interactive system to communicate naturally with virtual environments; the gestures were intended as command when used along with vocal orders.

The problem of separating continuous gestures has been addressed in particular contexts such as dance [5], theatre [6] and recognition of American Sign Language [7]. Kahol et al. [5] proposed a model called "Hierarchical Activity Segmentation to Represent the Human Anatomy" and they used low-level parameters to characterize motion in the various segments of the human body. Kelly et al. [8] presented a framework for continuous multimodal sign language recognition: they proposed a solution taking into account the epenthesis (i.e. the movement of the hand between the end point of the previous gesture and the start point of the next gesture).

However, these works did not considered the case in which the subject can move, interact with objects, and gesticulate, etc. between a gesture and the subsequent one.

Studies on always-available interfaces explored the problem of interacting with computer asynchronously during daily activities. Saponas et al. [9], in their EMG based interaction system, indicated the command gestures to the machine by clenching a fist: only when the fist was closed, the gestures performed with the other hand should be taken into account by the recognition system. However, in the following analysis they did not consider the influence of the segmentation system on the global system performance. Referring to similar experiences, Costanza et al. [10] tested their system in a particular scenario, providing a reasonably realistic environment (the subjects were walking through a predetermined path) maintaining enough experimental control to measure performance.

Actually, each work focusing on gesture recognition needs to develop its own approach to segment gestures. What currently lacks is an approach to evaluate the performances of the method used for the segmentation independently from the recognition and that could take into account non-command gestures.

In this paper we propose a test protocol that can be adopted to measure the performances of gesture segmentation approaches with regards to the gesticulation.

3 Test Protocol

In the first part of this section we will present an overview on the test protocol structure; subsequently, we will detail each step using as example our gesture recognition/segmentation approach for a real recognition/segmentation study presented in [11].

3.1 Protocol Overview

The test protocol (see Figure 1) is divided into four **phases** composed of eight different *steps*, concluded with one additional stage for the usability evaluation.

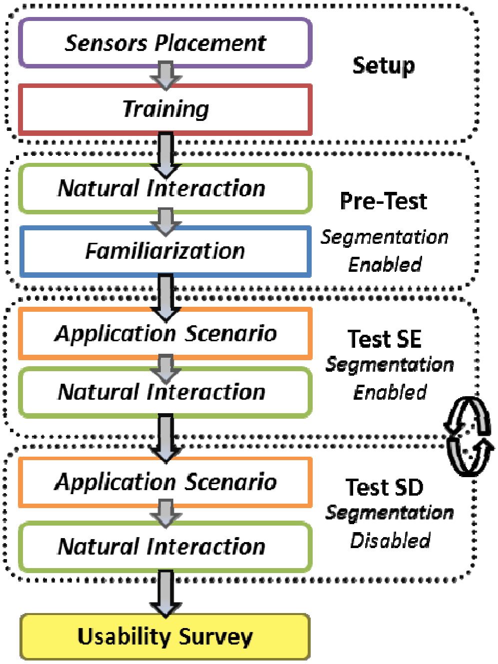

Fig. 1. Schematic representation of the test protocol structure

The first phase is the **setup** of the system. This part is required for two main reasons: the correct *sensors placement* (in the environment and/or on the subject) and *training* the system to adapt the recognition to the participants. This phase depends strictly on the approach and sensors used. For instance, cameras have different requirements than electrodes, and also the training is related with the classification strategy adopted. Once the setup is over, the segmentation system is activated and can detect when a command gesture is performed.

The **pre-test** session begins with the subjects performing the first *Natural Interaction* step. It consists in accomplishing predetermined tasks such as walking, manipulating as natural as possible objects with different characteristics, etc. Tasks and objects are closely related with the features used to distinguish the gestures. For example, during an approach based on muscular activity, "objects weight" and "handgrip type" are relevant features. On the opposite, during experiments involving computer vision techniques, the choice of objects according to features such as "color" or "shape" would be more suitable. This step precedes the *Familiarization* stage in which the subjects could finally understand and try the functioning of the system with the segmentation paradigm enabled. We chose to precede the *Familiarization* with a first *Natural Interaction* phase in order to evaluate whether or not knowing the interaction/segmentation approach could change the method of interaction.

After the *Familiarization* the **first test session** of tests begins. In this phase the segmentation is enabled and the subject performs one *Application Scenario* and one

Natural Interaction task. The *Application Scenario* is conceived to evaluate the accuracy of the system while the subjects interact with the system using gestures. The scenario has to deal with the challenging task to induce gesticulation in the user. The work presented in [4] gives some guidelines to follow in order to build a scenario enough involving to produce gesticulation.

In order to evaluate how the segmentation paradigm influences the recognition accuracy, the cognitive load and the system usability, a **second test session** is performed without segmentation.

The tests sessions (with and without segmentation) can be repeated several times in order to collect the desired amount of information.

Finally, the subjects conclude the experiment filling a *Usability Survey* based on System Usability Scale (SUS) [12].

3.2 EMG Test Protocol

We tested the protocol presented in the previous section in a real case [11]: evaluating the segmentation of two simple gestures of the wrist, segmented and recognized using EMG signals. Our test involved 11 participants, 8 men and 3 women, all of them were volunteers. They ranged in age between 23 and 30. Ten of the participants were right-handed and one was left-handed. The gestures were all performed with the preferred hand and none of the subjects had any known neuromuscular disease.

In order to use only not tiring gestures for the interaction, we adopted the wrist flexion and the wrist extension (see Figure 2). Keeping the palm of the hand in the sagittal plane, for the right-handed the wrist flexion corresponds to a left movement and the wrist extension to a right movement. For the left-handed the association is reversed.

Fig. 2. Adopted gestures and electrodes placement: A and B detect wrist flexion and wrist extension; S is used for the segmentation; R is the reference electrode

The muscle used for the segmentation was the triceps. Contracting the triceps meant that the gesture performed at the same time had to be considered as a command gesture. All other movements, gestures and gesticulation had to be considered non-command gestures. We choose the triceps contraction aiming at an almost-invisible, not tiring segmentation approach. These kinds of gestures have been defined “motionless gestures” [10].

Setup Phase

Sensors Placement. In our scenario, the system had to detect three gestures: the wrist flexion, the wrist extension, the triceps contraction. Therefore, the muscles directly involved in these movements were selected: for the wrist flexion the muscle designated was the *palmaris longus* and for the wrist extension the *extensor carpi ulnaris*. For the triceps contraction we monitored the *triceps brachii* activity. The skin was cleaned with alcohol according to SENIAM-recommendations [13], in order to assure a good skin-electrode contact. We did not shave or scratch the skin in order to evaluate a more interesting scenario for the HCI.

Training. A training phase was used in order to adapt the parameters of the used classifiers (Linear Discriminant Analysis) to each participant. We asked the subjects to perform the gestures as naturally as possible. A visual input indicated to the subject when to execute a specific gesture. A left arrow was used to indicate a left movement of the wrist, a right arrow for a right movement. The triceps contraction was associated with an up arrow. The training session lasted about one minute for each subject.

Pre-Test Phase

Natural Interaction. During the *Natural Interaction* step, the subjects were required to execute a series of tasks commonly done in the everyday life in order to quantify the false positives rate during these activities. The subjects were asked to:

- “Manipulate” five objects.
  - Displace the objects from a desk to another.
  - Putting the objects on the floor and then lifting them on the desk.
- Write and draw on a blackboard.

During the manipulation activities, the participants had to handle five different objects, with different weights. Also the grip changed but no indication on how to manipulate them was provided. The objects used were:

- Phone. The cell phone weighed about 150 grams. Given that it is a small object, many pretensions are possible.
- Bottle. The bottle was filled since the weight of 1 kg was reached. The material of the bottle was glass.
- Book. A thick book of 2 kg was used.
- A filled computer bag. It weighed 3 kg.
- A chair of about 7 kg.

The chair was chosen as border case to test the limits of our segmentation approach.

During the writing task, the subjects were asked to write a specific sentence on a blackboard using a chalk. The phrase was: "*It's a rather rude gesture, but at least it's clear what you mean*" (Katharine Hepburn).

The *Natural Interaction* phase ended drawing standard geometrical figures.

Familiarization. During this phase, the subjects received instructions about the functioning of the system and how to manage the *Application Scenario* in the following phases. In particular we explained how to use the triceps contraction to segment the other gestures.

The participants tested the system until they felt confident with the interaction. In our experiment this phase lasted in average 5 minutes.

Test Segmentation Enabled (SE) Phase

Application Scenario. In order to involve the user in the task and attempt to produce "natural" gesticulation, we asked each subject to interact with a slideshow presentation. In each slide there were the indication of the current slide and the indication to the slide they should reach.

To complete one tour, a total of 16 gestures, 8 *left* and 8 *right* were needed. In order to complete the *Application Scenario* two tours were required for a total of 32 gestures.

Each tour, during two slides, the subjects were asked to describe the pictures shown to them, evaluating if the images are good looking or not and arguing their reasons. This step was added to discern if the system detects false positives when the subject is gesticulating in a presentation context.

The participants were free to choose the duration of the interval between two interactions.

The *Application Scenario* was alternated with a *Natural Interaction* step.

Test Segmentation Disabled (SD) Phase

The whole first test phase was repeated a second time with the segmentation disabled (i.e., the input from the triceps were ignored) allowing to evaluate the impact of the segmentation system on usability, cognitive load and overall performances.

System Usability Scale

The SUS [12] was used in order to evaluate the system usability in terms of perceived complexity and difficulties, consistency and other aspects. The SUS consists of ten statements for which each participant specified the level of agreement on a 5 points scale.

Additional open questions were added to cover aspects missing in the SUS survey, such as muscular fatigue, and user opinion about strengths and weaknesses of the segmentation and interaction.

4 Discussion and Conclusion

Using the proposed test protocol to design an EMG-based interaction system [11], we could effectively collect different kinds of information (Figure 3) related to the selected segmentation approach: the *Natural Interaction* steps permitted to perform an accurate false positives analysis; the *Application Scenario* provided information about false negatives, accuracy rate and cognitive load; the *Usability survey* provided an important feedback about the system cumbersomeness and ergonomics. Finally, anticipating the first *Natural Interaction* phase before to the *Familiarization* phase allowed evaluating conscious or subconscious biases introduced when the subject knows the functioning of the system.

In our experimentation, the length of the experiment resulted in a good trade-off between completeness and fatigue, aspects particularly important when the muscular activity is monitored.

(a) False Positives Analysis

Objects (*weight*)	False Positives (with segmentation)	False Positives (without segmentation)
Phone (*0.15 kg*)	0	9
Glass Bottle (*~1 kg*)	4.5	18
Book (*~2 kg*)	9	27
Computer Bag (*~3 kg*)	24	33
Chair (*7 kg*)	58.5	90
Writing tasks	15.5	27

(b) Accuracy Analysis

	Step I	Step II	No Seg. Step
Average	92.32	95.88	96.87
SD	4.27	3.29	2.79
Paired t-test (SI-SII)	N = 11, α = 0.05 t(10) = 2.88, P(T<t) two tails = 0.0163		
Paired t-test (SII-No Seg.)	N = 11 α = 0.05 t(10) = 1.64, P(T<t) two tails = 0.1314		

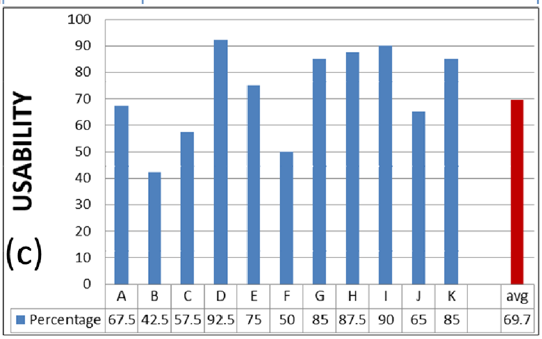

Fig. 3. (a) Results of false positives analysis, (b) accuracy analysis, and (c) usability analysis [11]

During the experiment, the *Natural Interaction* phase resulted in a uniform behavior among the subjects, with small variations in terms of seizing the objects and the energy/dexterity used in the write/draw tasks. On the contrary, the *Application Scenario* was visibly influenced by subjects' cultural or behavioral aspects. Subjects not naturally inclined to gesticulate produced some movements only when strictly required by the task. Different and maybe more stimulating scenarios, such as presented in [4], in which subjects were asked to describe in detail specific movies scenes, could improve the protocol.

From our experience, we learned that, in order to obtain significant results from the *Natural Interaction* phase, the choice of the features (e.g. weights and handgrip) and the consequent limit case (the chair) are of crucial importance.

In conclusion, in this paper we presented a novel test protocol for gesture recognition and segmentation methods. In particular our approach allows to easily separate command gesture from non-command gesture, such as gesticulation, providing an important tool to design new interaction systems. We showed a concrete example involving subtle gestures, for the interaction, and a motionless gesture, for the segmentation, demonstrating how the test protocol can be adopted to design a touchless interaction system.

Finally, in order to gain more relevance, our test protocol needs to be employed in similar researches comparing the results to estimate the need to include other overlooked gesture segmentation aspects.

References

1. Bolt, R.A.: Put-that-there: Voice and gesture at the graphics interface. In: Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques, pp. 262–270. ACM (1980)
2. McNeill, D.: Gesture and thought. Continuum (2005)
3. Kettebekov, S.: Exploiting prosodic structuring of coverbal gesticulation on Multimodal interfaces. In: Proceedings of the 6th International Conference on Multimodal Interfaces, pp. 105–112 (2004)
4. Wexelblat, A.: An approach to natural gesture in virtual environments. ACM Transactions on Computer-Human Interaction (TOCHI) 2(3), 179–200 (1995)
5. Kahol, K., Rikakis, T.: Gesture segmentation in complex motion sequences. In: Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), vol. 3, pp. II-105–II-108 (2003)
6. Billon, R., Nédélec, A., Tisseau, J.: Gesture recognition in flow based on PCA analysis using multiagent system. In: Proceedings of the 2008 International Conference in Advances on Computer Entertainment Technology - ACE 2008, vol. 139 (2008)
7. Li, Y., Chen, X., Tian, J., Zhang, X., Wang, K., Yang, J.: Automatic recognition of sign language subwords based on portable accelerometer and EMG sensors. In: International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multi-modal Interaction, vol. 17. ACM (2010)
8. Kelly, D., Reilly Delannoy, J., Mc Donald, J., Markham, C.: A framework for continuous multimodal sign language recognition. In: Proceedings of the 2009 International Conference on Multimodal Interfaces - ICMI-MLMI 2009, vol. 351 (2009)
9. Saponas, T.S., Tan, D.S., Morris, D., Balakrishnan, R., Turner, J., Landay, J.A.: Enabling always-available input with muscle-computer interfaces. In: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology - UIST 2009, vol. 167 (2009)

10. Costanza, E., Inverso, S.A., Allen, R.: Toward subtle intimate interfaces for mobile devices using an EMG controller. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI 2005, vol. 481 (2005)
11. Carrino, F., Ridi, A., Mugellini, E., Khaled, O.A., Ingold, R.: Gesture Segmentation and Recognition with an EMG-Based Intimate Approach - An Accuracy and Usability Study. In: 2012 Sixth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS), 4-6 July, pp. 544–551 (2012)
12. Brooke, J.: SUS-A quick and dirty usability scale, pp. 189–194 (1996)
13. Hermens, H.J., Freriks, B., Disselhorst-Klug, C., Rau, G.: Development of recommendations for SEMG sensors and sensor placement procedures. Journal of Electromyography and Kinesiology: Official Journal of the International Society of Electrophysiological Kinesiology 10(5), 361–374 (2000)