Ecological Gestures for HRI : the GEE Corpus

Maxence Girard-Rivier, Romain Magnani, Véronique Aubergé, Yuko Sasa, Liliya Tsvetanova, Frédéric Aman, Clarisse Bayol.

LIG-Lab, University of Grenoble Alps, France

E-mail: name.surname@imag.fr

Abstract

As part of a human-robot interaction project, we are interested by gestural modality as one of many ways to communicate. In order to develop a relevant gesture recognition system associated to a smart home butler robot. Our methodology is based on an IQ game-like Wizard of Oz experiment to collect spontaneous and implicitly produced gestures in an ecological context. During the experiment, the subject has to use non-verbal cues (i.e. gestures) to interact with a robot that is the referee. The subject is unaware that his gestures will be the focus of our study. In the second part of the experiment, we asked the subjects to do the gestures he had produced in the experiment, those are the explicit gestures. The implicit gestures are compared with explicitly produced ones to determine a relevant ontology. This preliminary qualitative analysis will be the base to build a big data corpus in order to optimize acceptance of the gesture dictionary in coherence with the ñsocio-affective glueò dynamics.

Keywords: Human-Robot Interaction, gestures, gesture recognition, ñsocio-affective glueò, Wizard of Oz experiment

1. Introduction

In face-to-face language interactions, the facial expressions, the body gestures and proxemics have been explored but we usually notice them in a wider range of body language studies (Scheflen, 1972; Gallagher, 2005). Gestures are usually considered as a paralinguistic feature but in sign languages, they carry all the complexity of language and interactions (Kendon, 1994). In Human-Robot Interaction (HRI), the human gestuality has been studied mainly as complementary information to speech.

This work is a part of the Interabot (Investissements dôAvenir) Project, held with the robotics Awabot Company developing the Emox robot and the LIRIS Laboratory where an efficient DNN gestures recognition system is developed. One main goal of this presented study is to demonstrate that human gestures produced naturally without focus on the gestural commands are very different from gestures they would have proposed if they would have been explicitely asked to produce gestures for HRI (as for example in ñfocus groupsò as proposed in ergonomy methods). Thus, the scenario implying the human in a quite ecological situation, gives only the gestural modality to communicate with a robot. If the basic gestures of this corpus of Gestural Emoz Expressions (GEE) are possible to be learned by automatic gestures recognition systems, that could be a way to introduce HRI without creating and teaching artifactual gestures to the human user. The corpus collected for this study is set in an ecological micro-world, as we need to observe gestures spontaneously produced by humans in a real HRI situation (Guillaume & al, 2015). As a consequence, to train an HRI system to natural gestures should not impose a gesture language to the human. This could directly optimize the acceptance of technology, and to ensure its durability. These gestures, when validated, can be mimicried in artificial conditions for building enough big data corpus for the LIRIS DNN (Guillaume & al, 2015). Another long term aim is to show that the dynamic of these gestures evolves within the dynamics of the relation. It was shown in a previous work (Aubergé & al, 2014), that (1) some vocal primitives (extracted from human productions) given to the robot can progressively build a socio-affective ñgroomingò relation, named socio-affective ñglueò (2) the speech expressions of the humans change within the gluing process, in particular the voice becomes breathy. Similarly in this present experiment, we want to observe if the gestures become subtle (like breathy for voice within the gluing process.

We present here the corpus methodology and collection for 22 subjects, the defining and labelling of types, ontologies and occurrences of the spontaneous gestures, and the differences with the gestures asked explicitly to be produced by the subjects after the spontaneous experiment for each ontology.

2. GEE corpus

2.1 Experimental setting

To collect GEE, we adapted the wizard of Oz Emoz platform, which is part of the Domus LIG Living lab, and which was developed previously for the Elderly Emoz Expressions (EEE) corpus (Aubergé & al, 2014).

The Domus Smart Home is equipped with 6 ceiling cameras (two for each room), seven microphones placed in the ceiling. In addition to this, a GoPro camera was placed on the forehead of the subject to capture his gestures (in accordance with the pretext task scenario), in order to complete the ceiling cameras. All the experimenters are in the control room outside the apartment, and can control through Emoz the Emox movement and all the domotic perturbations required by the pretext task.

2.2 The Pretext Task Scenario: implicit gestures

The pretext task is based on an IQ-game like scenario. The scenario was built in order to focus the attention of the subjects on a motivating task that gives a very secondary role to Emox, without the subjects could guess that their interactions with the robot are the aim of the experiment.

We proposed, as a fake, to the subject to evaluate their ñglobal IQò supposed to combine IQ relative to emotional abilities. 22 subjects with a high education level in computer science or robotics were selected (see Table 1) between 18 and 45 years old. They are mainly French, but some are from different cultural origins.

	Number of Subjects	
	Men	Women
Number	13	9
French	10	6
Cultures represented (other than French)	Iranian, Japanese, Italian	Japanese, Russian, Colombian

Table 1. Repartition of subjects in the GEE Corpus

The subject is explained that he/she will be left alone in the apartment with the task to solve a reversed rebus, that is they have to find successive objects in each room of the smart home, each object giving points to evaluate their global IQ. He/she is warned that some strange noises, movements of connected objects, and lights perturbations will occur during the game to slow their cognition. 66 rebus objects, were placed in the apartment for the task, 33 of them were in the living room, 21 in the bedroom and 12 in the kitchen. The experimenter explains to the subject that the way to show and validate his/her suggestion of object is to make it to be validated by a robot, with the constraint not to move the object (to keep the setting of the rooms for the next player). Emox is supposed to be efficient in speech understanding. But just before the game starts, the robot is simulated to have a damaged acoustic sensor, and to be available only with the video camera. That is that the experimenter explains to the subjects that the only way to communicate is visual movements. Emox is told to validate each object choice by specific movements: ñyesò with the Emox head, together with a circle movement of the Emox base.

However, the speaker of Emox is told to the subject not to be damaged, and he/she is warned that Emox can emit some sounds. Of course, theses sounds are controlled during the experiment from the control room in the progressive socio-affective gluing protocol (see (Aubergé & al, 2014)).

Once the subject considered to have finished, he/she can show a QR code which is on the main door to call the experimenter back into Domus.

This way, we could observe what types of gestures emerged spontaneously.

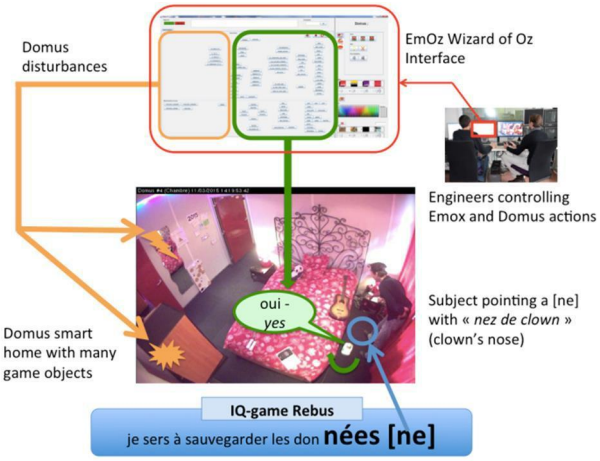

Figure 1. Spontaneous gestures captured with Emoz (pointing exemple)

2.3 Production of Explicit Gestures

Just after the pretext task, we presented them a short questionnaire (asking age, situation etc) to attract their attention and we simulate to need some more information to give them a precise IQ evaluation. Without to explain them that they were tricked, we ask them how they could control the robot with only the robot camera and if they remember which kind of gestures they performed just before. Some ñexplicitò gestures could be recorded, that is the gestures that the subjects conscientiously think they produced during the experiment, and with which meaning.

The spontaneous ï implicit ï gestures can thus be compared to these explicit gestures.

2.4 Auto-Annotation of GEE

In order to label the gestures, the chosen method is not experts annotation, but auto-annotation (see (Aubergé et al. 2006));

Each subject spent 3 hours on average to complete the two experimentôs phases, including the auto-annotation sessions. Auto-annotation sessions last between 36 minutes to 3 hours.

The auto-annotation sessions implied participants several weeks after the experiment. They watch (on ELAN) their recordings and they are asked to segment it into gestural units, and they were asked to comment freely (without any suggestion of the experimenter) everything they remembered about what happened, they felt and so on: even if our goal is to deduce ontologies of gestures, the experimenter must strictly avoid using key words like ñgesturesò, ñmeaningò or ñemotionò when asking the subject to auto-annotate his videos, the experimenter must ask questions like ñso what are you doing there ?ò or ñhow were you feeling at that time ?ò.

In most of the cases, the subjects could use their autobiographical memory to remember what was intended by producing their gestures.

An important point to take into consideration is that all of our participants were not from the same culture, and a lot of gestural semantic behaviors are culturally determined. Typically ñemblemsò (Ekman & Friesen, 1969) fall in the

category of culturally motivated gestures, but the other types of gestures listed may also be dependent of culture and or context.
The auto-annotation process well-formedness has a direct consequence on the relevance and the boundary of the gestures that will be analyzed.

3. GEE analysis

In order to determine which subjectsô motions correspond to a significant gesture and what information/label is associated to a defined gesture, the data is first auto-annotated following a specific methodology. The label will then be used to determine if the detected motion is a variation of the same gesture or a different gesture, by analyzing their occurrences and distributions in the corpus. Finally, the chosen gesture can be used as a model to produce prototypes of the gesture ontology in order to compose a wide data corpus for a gesture recognition system based on machine learning (Guillaume & al, 2015).

. Another important point to take into consideration is that all of our participants were not from the same culture, and a lot of gestural semantic behaviors are culturally determined. Typically ñemblemsò (Ekman & Friesen, 1969) fall in the category of culturally motivated gestures, but the other types of gestures listed may also be dependent of culture and or context.

3.1 Ontology of Implicit gestures

The gesture data includes 1350 labelled gestures. According to the auto-annotation labels, we selected 453 prototypical that are clear and representative variants among all the labelled gestures. They were regrouped, on the base on auto-annotated labels, into 37 variant ontologies. These ontologies could be classified, in term of meaning, in three main categories of ontologies: ñmovement indicationsò, ñobject pointingò and ñDraw attentionò. They contain respectively 229, 187 and 37 prototypical gestures. The considerable number of gestures of the two first categories is linked to the experimental protocol: to solve the rebus, the participants needed to show to Emox different objects disposed in various places. The results for the third category seem to be related to the peopleôs need of feedbacks. This category emerged from the data. Generally, people express their need of a feedback from the robot when they are not ordering him around and it just stays idle waiting for a command.

Movement indications	Show direction	47
	Follow me	21
	Go Forward	50
	Go in this direction	3
	Come near me	4
	Turn the robot	59
	Go backward	6
	Come	26
	Stop the robot	13
Object Pointing	Pointing	148
	Pointing with insistence	28
	Rapid pointing	7
	Pointing oneself	1
	Show a zone	3
Draw Attention	Draw attention	37

Table 2. Repartition and name of gestures observed

As seen in Table 2, the repartition of gestures is fairly disproportioned as some gestures are very frequent (i.e. ñpointingò) while others appear only once. This is mainly a direct consequence of the ñguidance & pointingò orientation of the pretext task. Yet it did allow for the ñattentionò label to emerge. Some labels may seem very similar, i.e. ñcomeò and ñcome near meò (the similarity might seem stronger in French, as the experiment was made in French) but according to the auto-annotation, those were not similar for the subject. Moreover these two gestures are articulatory different. The same explanation prevails for ñshow directionò and ñgo in that directionò.

3.2 The gluing effect

In some cases, as for ñpointingò and ñpointing with insistenceò the explanation is different: by further analysis and decomposition of gestures (McNeill, 2012) a modification of the gestural dynamics was noticed. On one hand, a subject had a tendency to do repeated pointing gestures (ñpointing with insistenceò) in the beginning of the experiment but as time passed he tends to stop (ñpointingò). On the other hand, there was one subject that started touching the objects while ñpointingò in the late part of the experiment.

Thus, when extracting the ontology from the auto-annotation we could observe that the participants described several subtle differences of the same gesture by using the same description.
We noticed that the gesture dynamics becomes quite systematically more subtle, as it was observed in ECA (Dibris & al, 2015). That is the gesture evolves from an ñhyper articulationò at the beginning of the experiment, to an ñundershootò gesture for some of the same command gestures in the ecological context.
It could be interpreted as first level as a least effort tendency when the subjects can evaluate that they have well understood, but in acoustic signals, in similar gluing procedure, a systematically increasing breathy voice and

lax global structures (in terms of rhythm and morpho-syntax) could be shown as characterizing a positive attachment (Sasa & al, 2014), in coherence with the care cues of breathy voice often observed.
It has been partially confirmed by the free comments of the subjects which express some positive attachment when they watch their subtle gestures (for example: ñI kindly show it/himò).
Consequently, we propose the subtle gestures (with lax movements) as characterizing the glue effect as well as the increasing confidence in robot understanding.

3.3 Explicit Gestures

The implicit vs. explicit gestures are completely different for the equivalent labels. For instance, for a same subject, the ñfollow meò label can differ from a hand wave in the implicit gesture to a snapping gesture in an explicit gesture.
We also observed a strong inter-subjects variation as observed for implicit gestures. The subjects can use different gestures for the same command

Only a few subjects shared the same explicit gestures: ñfollow meò (repeated ï horizontal- hand wave) ñcome closerò (repeated ï vertical- hand wave) ñgo over thereò (dynamic pointing) which are all path indication.
The subjects tend to have different implicit gestures to steer the robot. Though they can be grouped easily into variants of a same ontology thanks to the auto-annotation made by the subjects. To show an object to the robot, they often used a pointing gesture (fingers closed with index in direction to the object of interest), nonetheless this pointing gesture varies a lot between subjects.

4. Conclusion

This study is aimed to show that the communicative behavior in HRI cannot be expressively conduced with human, but need ecological methods to make emerge the spontaneous adaptation of the human communication abilities to HRI. GEE is a spontaneous and quite ecological corpus relative to a micro-world, which is later auto-annotated to avoid interpretation biases.
In relation of our pretext task, only three groups of ontologies emerged. But these only three groups are related to 37 variant ontologies that are expressed by more than 453 prototypical morphologies of gestures.
The pretext task is very specific and too reduced to be generalized to larger ontologies: the gestures mainly address path, localization and deixis. But the auto-annotations reveal for GEE some socio-affective cues (for example some gestures were labeled to express by themselves irritation of the subject). These socio-affective dimension need to be deeper explored in GEE together with the ñprosodyò of the gestures.

5. Perspectives

In parallel to extend the micro-world to large ontologies, some cross-perception experiments are ongoing between deaf and hearing subjects, since the hearing subjects are handicapped in our conditions whereas deaf can express the whole dimension of language in the GEE visual only conditions.

6. Acknowledgements

This work was partially funded by French grants Interabot parts of BGLE no2 Investissements dôAvenir and it has been partially supported by the LabEx PERSYVAL-Lab (ANR- 11-LABX-0025-01).

7. References

Aubergé V., Audibert N. & Rilliard A (2006). *Auto-annotation: an alternative method to label expressive corpora*. Proceedings of the International Workshop on Emotion: Corpora for research on emotion and affect, Genoa, Italy.

Aubergé, V; Sasa, Y; Bonnefond, N; Meillon, B; Robert, T; Rey-Gorrez, J; Schwartz, A; Batista Antunes, L; De Biasi, G; Caffiau, S; and Nebout, F (2014).ñ*The EEE corpus: socio-affective áglueô cues in elderly-robot interactions in a Smart Home with the Emoz platform*ò presented at the 5th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data, Reykjavik, Iceland.

Dibris, RN and Pelachaud, C (2015). The Effect of Wrinkles, Presentation Mode, and Intensity on the Perception of Facial Actions and Full-Face Expressions of Laughter, *ACM Transactions on Applied Perception*, Vol. 12, No. 1, Article 2.

Ekman, P and Friesen, W.F (1969). The repertoire of nonverbal behavioral categories ï origins, usage, and coding. *Semiotica*, Vol. 1 (1969) 49-98

Gallagher S (2005). *How the body shapes the mind*. Cambridge Univ Press.

Guillaume, L; Aubergé, V; Magnani, R; Aman, F; Cottier, C; Sasa, Y; Wolf, C; Nebout, F; Neverova, N; Bonnefond, N; Nègre, A; Tsvetanova, L and Girard-Rivier, M (2015). ñ*HRI in an ecological dynamic experiment: the GEE corpus based approach for the Emox robot*ò IEEE Int. Workshop Adv. Robot. Its Soc. Impacts, July 1st ï July 3rd.

Kendon, A (1994). ñDo Gestures Communicate? A Review,ò *Res. Lang. Soc. Interact.*, vol. 27, no. 3, pp. 175ï 200, Jul.

Loyau, F and Aubergé, V (2006). ñ*Expressions outside the talk turn: ethograms of the feeling of thinking*ò in 5th LREC, pp. 47ï 50.

McNeill, D (2012). *How language began ï Gesture and speech in human evolution.*

Sasa, Y and Aubergé, V (2014). ñ*Socio-affective interactions between a companion robot and elderly in a Smart Home context: prosody as the main vector of the ósocio-affective glueôò* presented at the Speech Prosody 7, Dublin, Ireland.

Scheflen A.E (1972). ñ*Body Language and the Social Order; Communication as Behavioral Control*ò.