HRI in an ecological dynamic experiment: the GEE corpus based approach for the Emox robot

Leslie GUILLAUME[1], Véronique AUBERGÉ[1], Romain MAGNANI[1], Frédéric AMAN[1], Cécile COTTIER[1], Yuko SASA[1], Christian WOLF[2], Florian NEBOUT[3], Natalia NEVEROVA[2], Nicolas BONNEFOND[4], Amaury NÈGRE[5], Liliya TSVETANOVA[1], Maxence GIRARD-RIVIER[1]

***Abstract*— As part of a human-robot interaction project, the gestural modality is 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 where the robot is the referee. These gestures are compared with explicitly produced gestures to determine a relevant ontology of gestures. 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 "socio-affective *glue*" dynamics.**

Index Terms: Human-Robot Interaction, gesture recognition, machine learning, "socio-affective *glue*", Wizard of Oz experiment

I. The Interabot Project issues: an automation environment butler robot socio-affectively attached to the user

The common goal of the partners (Awabot Company, LIG CNRS Lab, LIRIS CNRS Lab, Voxler Company) involved in the Interabot project (Investissements d'Avenir BGLE2 of French Industry Ministry) is to develop a butler robot for an automated environment which is socio-affectively bound to the user, and strongly rooted in the ecological reality of this specific Human Robot Interaction (HRI) context. Our approach is consequent to the concept of "companion robot" which is an ill-posed problem: the robot augmenting the human social reality cannot appear directly with a companion social role, except perhaps while it borrows the pet role (the first moment a puppy appears in someone's life, it is affectively *glued*).

The robot Emox used in this project is proposed as a service robot, and because it is a service robot, it is relevant and useful, whereas it is inevitably "another". Therefore, it is perceived as a communicating interactive entity, no matter its given communicative competences. While this robot is introduced as a service robot, we highly expect to see what we call the "socio-affective glue" (that could be explained as the consequence of a grooming engagement) building function to develop companion characteristics. We even noticed with elderly people in need of practical and useful services, that a robot executing smart home automated actions is perceived as a companion, not only as a service robot if it performs a "socio-affective coaching" [1]. Thus, outside the specific pet stance which implies directly a relationship with the owner, all social roles steeped in the social space induce the cause of the relationship's building function.

Once the roles are defined, the relationship is built in a dynamic process which *glues* the two correspondents: roles induce the relationship allowing the *glue* to maintain the roles efficiency which in return modifies the glue modulating the relationship. As a consequence of this dynamic loop, the correspondents tend to be qualified as companions. Indeed the description of companion robot as a primary (and not derived) characteristic does not seem to be so relevant [2][3][4][5]. The animistic process defining robots are constrained by the anthropomorphisation tendency [5][6][7], especially if its appearance is human-like or pet-like.

In order to minimize the biases in this project (induced by all the "uncanny valley" effects in particular), the role given to the robot is absolutely not transferred from a possible human role: it is an automation controller (manipulating computational protocol). Moreover this robot has to be kept away from a human-like or a pet-like appearance to avoid the bias of anthropomorphisation and this justify our choice for the Emox robot developed by our partner Awabot (see Figure1).

Figure 1. The Emox Robot (Awabot Company)

One of the main hypothesis of this work lies on the socio-affective *gluing* process, key tool based on human dialog interaction primitives: the language's nature is dynamically timed with the *glue* level [1]. This timed evolution has also been noticed through the interpersonal synchrony [8]. The hypotheses of the social *gluing* skills given in the Interabot project to the robot, are increasing during the relationship and are based on: (1) no speech, (2) pure prosodic mouth noises supposed to be the *glue's* tools, (3) lexicons with supposed *glue* prosody as interjections and onomatopoeia and (4) subjects commands imitations with supposed *glue* prosody [9]. These increasing *gluing* materials have been confirmed as efficient since it has been observed that the human vocal and

1 University Grenoble Alps, LIG, UMR CNRS 5217, France (e-mail: firstname.surname@imag.fr)
2 INSA-Lyon, LIRIS, UMR CNRS 5205, F69621, France (e-mail: firstname.surname@liris.cnrs.fr)
3 Awabot Company, Villeurbanne, France (e-mail: florian.nebout@awabot.com)
4 AMIQUAL4HOME, INRIA, Grenoble, France (e-mail: nicolas.bonnefond@inria.fr)
5 INRIA, Grenoble, France (e-mail: amaury.negre@imag.fr)

linguistic interactional behaviors are strongly modulated and transformed during the *gluing* process [10]. The communication based on these phenomena is multimodal where gestures are undifferentiated cues.

Anyway, the human-to-human gestural communication is a very rich system, involving several studies [11][12][13]...etc. The human gestuality addressed to robots is described both as specific or as directly transplanted from human-to-human interaction [14][15][16]. This present work's final goal is to build a large database of gestures, in order to train the gesture recognition system developed by the LIRIS partner for the Interabot Project.

Gesture recognition is performed automatically based on machine learning. The classifier takes as input very short video sequences captured with a consumer depth camera mounted on the robot, and outputs for each time instant whether a gesture has been performed, and which of the predefined gesture classes has been performed. The targeted gesture recognition system used for this project is based on convolutional deep learning [17][18] in a multimodal setting (RGB color input video, depth input video, audio). Each visual modality captures spatial information at a particular spatial scale, such as motion of the upper body or a hand, and the whole system operates at two temporal scales, namely short dynamical poses and longer periods corresponding to whole gestures.

In a supervised learning setting, a large training set of manually annotated input data samples is created, and a gesture classifier is trained off-line minimizing classification error on this training set. This classical principle, called empirical risk minimization, aims at minimizing classification error on future (unknown) real data, i.e. gestures performed by users in a real setting.

Success depends on the domain shift between the training set and the test set. Classically in machine learning, the difference in the distributions of the training set (the manually annotated training gestures) and the future unknown test set (the real gestures performed in a production setting) must be as small as possible. This requires users to perform the gestures as closely as possible to the way they were performed during the training phase.

Minimizing the domain shift could theoretically be done by forcing users to accept the predefined gestures chosen arbitrarily by the designers of the robot. However, our objective in this project is to optimize acceptance of the gesture dictionary by the target public, and also to maximize the social glue created by this gestures.

Before collecting the training gesture set used for supervised learning of the gesture recognition system, this present study aims at defining which kind of different gestures will be produced for a same command in a real ecological long life Emox-human interaction. In this ecological context, our purpose is to verify if the human gestural behaviors are modulated and transformed through the increasing glue between Emox and the human, similarly to the vocal behaviors modulation and transformation.

To verify this hypothesis, we first collect with a Wizard of Oz protocol, a corpus of spontaneous gestures the subjects use to guide the robot on live, without having any consigns about gestural commands for the robot, and mainly without being aware that their gestural behaviors could be the goal of the experiment. The given goal was set in order to attract their attention on an IQ game-like task which could be related to their gestural behavior. Just after the experiment, subjects are asked to explicitly produce gestures devoted to control the robot.

The main observations are (1) the implicitly vs. explicitly produced gestures are quite systematically different kinds of gestures for the same "command" (2) the implicitly produced gestures clearly evolve during the interaction, alongside the *gluing* process: the gestures become more subtle as the glue level rises and the command ontology is blended with some other kind of information related to the built socio-affective relationship. We observe here how the natural gesture communication of the human is spontaneously adapted to such a robot, without to give any indication to the human of possible restrictions in th gestures processing understanding of robot.

II. Methodology

A. EmOz Wizard of Oz platform to collect an ecological corpus in Domus Living-Lab

The experiment was conducted at the LIG lab's Domus smart home[1], designed to observe the users activities interacting with the environment. This Living-Lab is based on the openHAB[2] home automation open source middleware that centralizes and controls the different actuators and sensors using various protocols (KNX, DMX, UPnP, RFID, MQTT...).

Seven microphones and six video cameras set in the flat ceiling, allowing the supervising experimentation from a control room. This room makes it possible to use a Wizard of Oz method, in which the experimenters are hidden during the experiment to control remotely the automation system.

EmOz, the Wizard of Oz platform [10] was developed to control the Emox robot primitive sounds and moves, as well as the Domus smart home automation. An Xbox wireless controller is also used to drive the robot introduced as the smart home butler, while the subjects consider it completely autonomous because of the EmOz platform and the associated controlled scenario. We used the robot camera view to ease its driving and to increase its behavior credibility during the IQ game-like experiment, in which this robot is also the referee.

With regard to recording and analyzing the subjects' gestures, we mainly used a GoPro camera placed on their forehead. But, sometimes, it could not efficiently capture wider gestures. Wider views were additionally captured with six axis IP cameras (two per room except for the bathroom) and three more D-Link IP Camera in each room to have an extra distant point of view. In total, eleven cameras recorded a subject per session, as well as seven Sennheiser radio microphones set into the ceiling (two per room except one for the bathroom) and a lapel microphone recording

[1] https://hal.archives-ouvertes.fr/hal-00953242/file/puc2010.pdf
[2] www.openhab.org

simultaneously audio channels using the StreamHIS software developed by the LIG lab.

B. A pretext task: an IQ game-like experiment

In order to collect a spontaneous gesture corpus in an HRI ecological situation, we present to the subjects a false experimental goal in which we evaluate their intelligence and their ability to keep focused on a task while their environment is disturbing them. Where in fact we want to capture and analyze their gestural behaviors in their interactions with the robot.

The scenario was based on a reversed rebus where the subjects had to find a maximum of objects illustrating a phonological part of a given sentence. These objects were scattered all around Domus, and the task was to gesturally guide Emox, the game referee, in front of each objects to be validated. In fact, we pretexted a malfunction of the robot voice recognition system (developed by the LIG lab) due to its microphone, so we highly recommended using the gesture recognition system (the future LIRIS partner system). We told that this robot had learned the valid objects previously with a pattern recognition software. The subjects were not allowed to move anything in the smart home during their task. We never explained how to guide the robot, as we wanted to observe what types of gestures would emerge naturally. To maintain the participants attention on the pretext task, we activated punctual and varied automated actions in the smart home (shutters closing, lights dropping, LED blushing, thematic music, TV blinking, robot dancing, strange noises like the hair dryer... and so on, all set in four different scripts) as we pretended to measure the subjects resistance to environmental disturbances. The IQ game proposed also to calculate a score regarding the number of objects the subjects found out, because each part of the rebus is associated to several objects. That brought the subjects to forget about the gestures they made. The sentence given to the subjects was: “I am used to save data”[3]. The subjects were alone with the robot to resolve this rebus during the experiment which was not limited in time. Once the subjects decided they were done with the game, they had to show a pictogram to the robot that ended the game and sent a call to the experimenter who came back in Domus.

C. Ecological situation: spontaneous gestures implicitly produced to interact with the robot

During the ecological part of the experiment, the subjects conducted Emox near the objects they wanted to validate. Therefore, after each suggestion, the robot did a positive or a negative gesture associated to a vocal feedback.

At first, the Wizard of Oz engineers, as well as the robot were not used to the subjects gestures, so the reaction of the robot was slower and not as fluent as it could be later in the experiment. Then it became more and more aware of the subjects gestures to produce smoother reactions. Nevertheless, as Wizards of Oz, we also wanted to get other kind of gestures from the subjects so we induced them, using Emox “impotence” to some reactions that could lead to other form of gestures (e.g. Emox continues a trajectory after a sign like “follow me” to see if the subjects ask for the robot to stop or not, and how they will do that). The figure 2 illustrates the EmOz scheme to collect these spontaneous gestures.

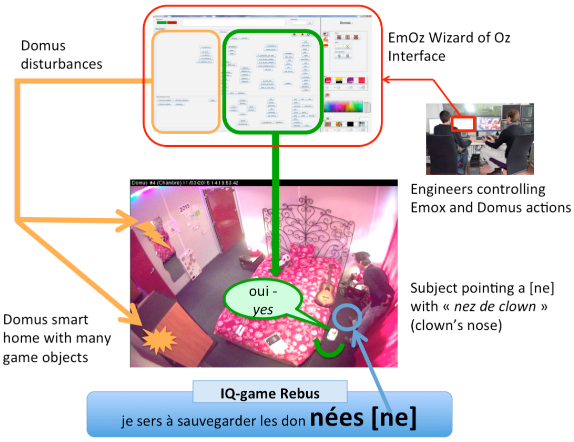

Figure 2. Spontaneous gestures capture with EmOz (pointing example)

D. Explicitly produced gestures

Once the IQ game was finished, during a debriefing that explained the experiment’s real goal, the participants were explicitly asked to do the same kind of gestures they did implicitly during the experiment in order to compare them. The explicit commands to address to Emox were: “go to a specific location”, “follow me”, “come close to me”, “confirm that you have understood the instructions”, “I disagree with your proposal”, “I agree with your choice”, “retreat”, “pass in front of me”, “wait”, “you made a mistake”, “go recharge your batteries”, “turn left”, “turn right”, “speed up and slow down”, “repeat”. The participant was placed in front of the robot, set with a lapel microphone to be recorded. The experimenter was behind the robot, holding a GoPro camera to film the subject. The subject was asked to do each command twice, the first only gesturally, the second time vocally. We noticed that sometimes it is difficult for the subject to do a vocal command without adding the gestural feature, which is also the case for some participants during the ecological context. Mostly, the gesture associated to a vocal cue is also not exactly the same as the one in gesture only condition.

E. Auto-annotation

The auto-annotation [19] refers to the subjects’ autobiographical memory [20] by showing them their experiment videos, referring to their naive point of view concerning their own reactions. As experimenters, we are motivated to discover what gestural strategy the participants used to guide the robot but without ever evoking this idea, we helped the subjects to remember their feelings, intentions, emotions, cognitive processes, etc., all information known as the Feeling of Thinking [21]. To describe these annotation cues, the subjects used their own words and vocabulary. They could even draw or express their description in every modality they wanted, but the experimenter should never suggest any label neither consider himself what can be or not a gesture.

[3] The original command was: “je sers à sauvegarder les données”

For this experiment, we show step by step the video of the implicitly produced gestures recorded with the GoPro Camera. The idea is also to reveal what the subject considers as gestural units, so to understand when a gesture begins and when it finishes. The gesture segmentation is done by the subject himself.

One auto-annotation session was managed by two experimenters. While one of the experimenter transcribed the subjects explanation, another one lead and guided the participant to get the labels for each determined gesture. Each time a subject suggested an accurate description on a precise action at a specific time, the leading experimenter brought him a more detailed description by reusing the employed subjects vocabulary and trying to focus better on the cues (in the present study, the gestures produced) that the experimenter wanted to be labeled but that he never suggested directly. The sessions were recorded with a microphone and the participant was filmed with a GoPro camera to catch possible gestures they can do to explicit their annotation. Finally, to ease the transcription, we used the Snagit software to record the annotation screen using the Elan software (EUDICO Linguistic Annotator).

The Auto-annotation sessions are not finished yet, so we only have some tendencies for the labels. For example, the subjects suggested labels like “come close to me”, “follow me” or “stop”. We also observed static gestures (like pointing an object), or dynamic absolute gestures (for example, the participant draw a virtual line with his finger between the robot’s camera and the aimed object).

F. Label validation through perception test

Knowing that our long term objective is to make Emox aware to recognize and to interpret human gestures, the gestural recognition algorithm must be able to match motion and its meaning. The fact that the robot cannot make sense of any human gestures, while it can capture the movements itself is a semantic gap. To fix that, we firstly need to determine a label for each gesture we have collected during the experiment and to understand what does this gesture suggest. The labels are inferred from the subjects’ auto-annotation that needs to be validated perceptively to objectivate the subjectively auto-annotated label. At this stage of our study, we have not yet performed these perception tests with new naive judges, but ideally it has to be done to: (1) verify if the gestures proposed are relevant gestural units or improperly cut composed parts, (2) have good gestural exemplaries to make a suitable ontology.

III. THE GEE CORPUS

The GEE corpus stands for Gestural Emoz Emox and is composed by 22 subjects (13 men and 9 women) between 18 and 45 years old. They are mainly french, but it’s important to remark that some participants have different cultural origins, such as: italian, spanish, british, and japanese. Each subject spent 3 hours on average to complete the experimentation (including auto-annotation sessions). The IQ game sessions lasted between 11 minutes for the fastest subject and 1 hour and 8 minutes for the slowest one. Auto-annotation sessions lasted between 36 minutes (min) and 3 hours (max). The IQ game provides 30 objects to find, so we can suppose that each subject can do up to thirty predictable interactions with the robot, and few more, knowing that a single interaction can include several gestures. Actually, we noticed, at this stage, around fifty gestures on average for the less communicative subjects at the end of the IQ game, and around a hundred gestures on average for the most communicative ones; mostly due to the repeated similar objects submissions or wrong objects submissions.

A. Implicit vs. explicit gestures

Most of the time, the implicit gestures that are produced spontaneously and the explicit gestures that are produced on demand are quite different.

As we can see in Figure 3, in implicit condition, the subject #1 puts her right hand at the back of Emox to make it go ahead and follow her.

Figure 4. Implicit gesture for “Follow me” (subject #1)

This same subject #1, when she is told, explicitly to execute the command “follow me”, did as in Figure 4. She decides to turn sideways a little on the left, then steps aside, and she slaps her thigh with her right hand.

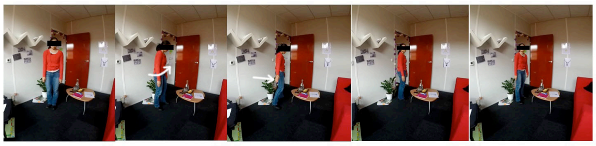

Figure 5. Explicit gesture for “follow me” (subject #1)

In this second example Figure 5, the subject #2 has found an object to show to Emox. To do this, she is almost on her knees, sideways, and she swings her right arm when barely pointing Emox to make it come close to her.

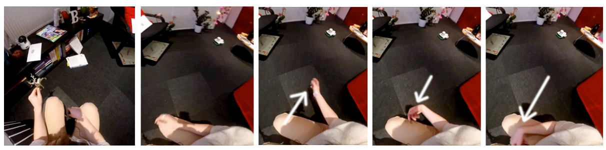

Figure 6. Implicit gesture for “come close to me” (subject #2)

Figure 3. Explicit gesture for “come close to me” (subject #2)

This subject #2, for the explicit command "come close to me", faces the robot and she raises her right arm a little in front of her as in Figure 6. Next she snaps her fingers and then puts her hand behind her back, and waits.

The third example shows another implicit gesture for "come close to me" produced by the subject #3, who is standing sideways from Emox, but looking at it, his hands in his pockets (see Figure 7). When he gets its attention (which means its camera is directed to him), he slowly side-steps to the direction he wants to go, waiting for the robot.

Figure 7. Implicit gesture for "come close to me" (subject #3)

The subject #3, in explicit condition, takes his hands out of his pockets, he raises his arms in front of him and points Emox with his two outstretched index fingers. From that position he bends his elbows until his hands are against his stomach. Finally he puts his hands back in his pockets as in Figure 8.

Furthermore, big differences appear when a subject executes the same gesture within an implicit situation, and then during an explicit situation. For example, subject #3 never used his arms or even his hands, but then, to express the exact same order within an explicit situation, he only used his arms, hands, and fingers, and never his legs.

Figure 8. Explicit gesture for "come close to me" (subject #3)

Moreover for some explicit commands from our list, the subjects considered them as similar and redundant (e.g. do a u-turn vs. turn around), whereas this redundancy was not real in the ecological context. Sometimes commands proposed explicitly were never addressed to Emox while the subjects had some behaviors revealing this command attitude in the spontaneous context (e.g. "I don't agree" which was never produced for the robot but always for the subject himself). On the contrary there are some attitudes never implicitly produced by gestures while they were frequently expressed by the subjects (e.g. "go away" never appeared in gestures while the subjects said many time to be troubled by the robot which was "staying in their legs"). Finally the subjects implicitly produced some gestures that we have not planned in our commands list (e.g. "catching the robot attention" particularly produced while the robot was not looking at the subject or before he gives a command).

B. Gesture changes between subjects, through time and along the "glue" dynamics

In the ecological context implying implicit gestures, we noticed some subtle changes for some of the same command gestures in the conditions where we included a positive *gluing* process through the Emox robot feedbacks and reactions. We tested this phenomena only on a few people and we intend to extend this kind of experimentation in future studies. We have not yet analyzed precisely the gestural changes, but we still observed differences for the command "follow me". In fact, at the beginning of the experiment, the subjects produced accurate and detailed gestures, whereas at the end of the spontaneous gestures phase, the same gesture is quicker and shorter which can be due to the fact that the robot and the subjects get used to each other and changed their reactions through time. Some subjects also tend to do more frequently and different "follow me" gestures at the end to have the robot close to them. This can be explained by the pretext task constraints, as the subjects do not have to look for the robot when they want to point it an object, but it also can be possible that they want to have Emox close to them because their relationship changed. In both cases, we can clarify this phenomenon with the auto-annotations associated to these gestures. It can be supposed that the gestures would not become subtel if the user would be aware about gesture recognition constraints, but it must be noted that many subjects have supposed poor and constrained visual recognition by the recognition, and have however produced subtel gestures as soon as the glue is installed (as it has been already observed in different ECA studies [22]).

On the other hand, gestures change between subjects in both implicit and explicit condition. For instance, among implicit gestures, the subjects #2 and #3 used their own strategy to command Emox in two very different ways, to gesturally say "come close to me", and other means have also emerged. However, some gestures seemed to remain on one participant to another, despite their individual specificities. Nevertheless, we haven't yet analyzed deeper enough at this point, to place a label on these types of gestures. The variation of these gestures, auto-annotated and validated, leads to determine gesture prototypes. The gesture ontology will be reproduced by imitating many time the prototypes to build the data corpus constrained by the gestural machine learning system (LIRIS partner).

IV. Training Dataset Definition and Collection by Motion Imitation

Once the dictionary of targeted gestures is defined, including the exact semantic meanings of each gesture class, a training set will be constructed for off-line training of the gesture recognizer. Supervised training of deep visual models, as the ones we employ in our system, requires very large amounts of training data. In principle, the training data gesture samples should contain all possible gesture classes and widely cover variations in input variables and conditions, such as the following: (1) camera viewpoint with respect to the person performing gestures; (2) eventual direction pointed by the person, if the gesture class concerns an object in the scene; (3) differences in the morphology of the person's body (size, body mass and proportions etc.); (4) speed and rhythm of the gesture performance; (5) sex and age of the person.

Differences in viewpoints are due to different positions in which people can address the robot, but also on the type of the robot itself. Different robots can result in different camera height, which result in viewpoints changes and significantly modify gesture appearances. Different cameras are also characterized by different intrinsic parameters such as focal length etc. The deep multi-modal system developed for this project has compared very favorably with respect to the state of the art when trained on the public dataset ChaLearn 2014 looking at people: gesture recognition comprising 14.000 upper body gestures of 20 classes performed by 20 different people [18]. Our system uses different modalities:

- the full body pose estimated by a middle ware, such as MS Kinect SDK;
- deep learning on short sequences of the right and left hand;
- deep learning on short audio-sequences.

Features from the full body (the full body skeleton) have been made invariant explicitly in order to compensate for differences in viewpoints, morphology, size etc. However, features from handpose, which significantly contribute to the systems performance, are not invariant, which is one of the drawbacks of methods based on deep learning of representations. As a consequence, the training set should cover as widely as possible all variations in input data expected during test time.

V. CONCLUSION

In this preliminary paper we used an IQ game-like Wizard of Oz scenario to collect spontaneous gestures produced in an ecological situation to propose an ontology of gesture that can be used to develop a gesture recognition tool for HRI context. We started to compare these implicitly produced gestures to explicit gestures to get the relevance of the gestural prototypes. Thus these gestures can be the base for an intelligent, incremental and long-life machine learning system which can evolve with the *gluing* process established between the robot and its user. In order to take into account the LIRIS system's big data constraints, the ontology determined by our analyses will be imitated to build a huge data corpus of relevant gestures.

ACKNOWLEDGEMENT

This work was partially funded by French grants Interabot parts of BGLE no2 Investissements d'Avenir and it has been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01). We would like to thank Brigitte Meillon, Jander Botelho-Nascimento, Jean-Philippe Guilbaud and Sylvie Humblot for their support during the experiments.

REFERENCES

[1] V.Aubergé, Y.Sasa, T.Robert, N.Bonnefond, and B.Meillon,"Emoz: a wizard of oz for emerging the socio-affective glue with a non humanoid companion robot," in WASSS - Workshop on Affective Social Speech Signals, Grenoble, France, 2013.

[2] C. Breazeal, "Toward sociable robots," Robotics and Autonomous Systems, vol. 42, no. 3-4, pp. 167–175, 2003.

[3] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," Robotics and autonomous systems, vol. 42, no. 3, pp. 143–166, 2003.

[4] S. Turkle, Alone together: why we expect more from technology and less from each other. s.l.: Basic Books, 2013.

[5] K. Dautenhahn, "Socially intelligent robots: dimensions of human-robot interaction," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 362, no. 1480, pp. 679–704, 2007.

[6] K. Darling, "Extending legal rights to social robots," Social Science Research Network, Rochester, NY, Tech. Rep., 2012.

[7] N. Kitano, "Animism, rinri, modernization: the base of japanese robotics," in ICRA'07 - IEEE International Conference on Robotics and Automation, Rome, Italy, 2007.

[8] M. Chetouani, "Role of inter-personal synchrony in extracting social signatures: Some case studies." ACM Press, 2014, pp. 9–12.

[9] Y. Sasa and V. Aubergé, "Socio-affective interactions between a companion robot and elderly in a smart home context: prosody as the main vector of the socio-affective glue," in Speech Prosody 7, Dublin, Ireland, 2014, pp. 86–90.

[10] V. Aubergé, Y. Sasa, N. Bonnefond, B. Meillon, T. Robert, J. Rey-Gorrez, A. Schwartz, L. Batista Antunes, G. De Biasi, S. Caffiau, and F. Nebout, "The eee corpus: socio-affective "glue" cues in elderly-robot interactions in a smart home with the emoz platform," in 5th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data, Reykjavik, Iceland, 2014, pp. 27–34.

[11] D. McNeill, "So you think gestures are nonverbal?" Psychological Review, vol. 92, no. 3, pp. 350–371, 1985.

[12] A. Kendon, "Do gestures communicate? a review," Research on Language and Social Interaction, vol. 27, no. 3, pp. 175–200, 1994.

[13] S. Goldin-Meadow, "The role of gesture in communication and thinking," Trends in cognitive sciences, vol. 3, no. 11, pp. 419–429, 1999.

[14] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 19, no. 7, pp. 677–695, 1997.

[15] J. Triesch and C. Von Der Malsburg, "A gesture interface for human-robot-interaction," in 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. IEEE Computer Society, 1998, pp. 546–546.

[16] S. Waldherr, R. Romero, and S. Thrun, "A gesture based interface for human-robot interaction," Autonomous Robots, vol. 9, no. 2, pp. 151–173, 2000.

[17] N. Neverova, C. Wolf, W. T. Graham, and F. Nebout, "Mod- drop: adaptive multi-modal gesture recognition," ArXive pre-print, http://arXiv:1501.00102, 2015.

[18] N. Neverova, C. Wolf, W. T. Graham, and F. Nebout, "Multi-scale deep learning for gesture detection and localization," in ECCV ChaLearn Workshop on Looking at People, 2014.

[19] V. Aubergé, N. Audibert, and A. Rilliard, "Auto-annotation: an alternative method to label expressive corpora," in LREC, 2006.

[20] M. A. Conway and C. W. Pleydell-Pearce, "The construction of autobiographical memories in the self-memory system," Psychological Review, vol. 107, no. 2, pp. 261–288, 2000.

[21] A.Vanpé and V.Aubergé, "Pertinence perceptive d'icônes gestuelles du feeling of thinking," in Waca, 2006, pp. 55–59.

[22] R. Niewiadomski, S. Hyniewska, C. Pelachaud, Computational Models of Expressive Behaviors for a Virtual Agent. In Social emotions in nature and artifact: Emotions in human and human-computer interaction. Eds Jonathan Gratch, Stacy Marsella, Oxford University Press, to appear