The potential of Image Schemas for computing automatically metaphoric gestures for embodied conversational agents

Brian Ravenet[1], Chloé Clavel[1], and Catherine Pelachaud[2]

[1] LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France
[2] CNRS - ISIR, Université Pierre et Marie Curie, Paris, France

Abstract. Embodied conversational agents are virtual characters capable of displaying and understanding human communication in order to create natural interactions between the user and the machine. In our work, we aim at giving these agents the ability to automatically produce the metaphorical gestures accompanying their discourse in order to produce a coherent multimodal communicative behavior. Our approach relies on two steps: identifying a common representation used in both verbal and nonverbal channels while talking about abstract concepts and manipulating this representation to make the connection between speech and gestures. In this paper, we explain how we use *Image Schemas* to represent meanings from both verbal and nonverbal channels.

1 Introduction

Our objective is to give to an embodied conversational agent the ability to compute and to produce metaphoric gestures automatically, given only the text transcription of the speech it is about to pronounce. In this paper, we start by presenting the main theoretical concepts of our approach. We first describe the roles of metaphoric gestures, along with the links between thoughts, gestures and speech by articulating them around *Image Schemas*. We also divide our task into two stages, extracting *Image Schemas* from the text and using these *Image Schemas* to produce and combine the corresponding gestures. A first version of the gesture generation process is proposed in the last section.

2 What are metaphoric gestures?

When humans talk, they usually accompany their discourse with co-speech gestures that contribute to convey the desired communicative intentions. According to Mc Neill, it exists 4 categories of gestures: iconic, deictic, beat and metaphoric [8]. In our work, we focus on metaphoric gestures. While iconic gestures aim at reproducing physical properties of concrete concepts (for instance mimicking the spherical shape of a ball with the hands), metaphoric gestures tend to give similar physical properties to abstract ideas. They can be viewed as deriving from concrete actions [1]. For instance, while talking about an important discovery, one can represent this importance by describing an object ('discovery') that is heavy ('important') with his hands. In his *Growth Point* theory, Mc Neill explained that there are moments in a person's discourse, such as pauses, where gestures and speech are planned and combined together around a common mental imagery [9]. Therefore, in order to replicate the mechanisms humans use to produce metaphorical gestures, it is interesting to explore this imagery and to understand how they can connect the speech and the gestures.

3 Metaphors and Image Schemas

In [6], the authors describe how interactions in the physical environment shape the metaphors people use in their language production. Following that idea of metaphors being embodied concepts, built from our personal physical experience, Johnson suggested that humans use recurring patterns of reasoning, called *Image Schemas*, to map these metaphors from an entity to another [5]. This would give an explanation on how humans transfer their reasoning about their physical reality onto abstract concepts, thus giving physical attributes to abstract entities. In [2], Cienki and Müller gather different works aiming at understanding the links between metaphorical reasoning and gestures and *Image Schemas* are mentioned several times as a potential underlying structure of this reasoning as well. For instance, in [10], the author describes how a gesture (mimicking the shape of a box in the example) can represented the *Image Schema* OBJECT or CONTAINER underlying the conceptual metaphor IDEAS ARE OBJECTS.

4 Image Schemas extraction

In our approach, we envision to develop an extraction mechanism that uses the text of the agent's speech to retrieve a wide scope of *Image Schemas* to be used in the gesture production. The only automated method for *Image Schemas*

detection we are aware of is the work of Gromann and Hedblom [3]. In this work, the authors use a clustering method on the Europarl corpus to obtain clusters of *verb-preposition* couples that they assign to *Image Schemas*. They apply a semi-automatic approach using a semantic role labelling tool [11] as well as asking two annotators to label the clusters. The authors focus on on type of *Image Schemas*, the spatial one. As a first step, their results are convincing and highlight the potential of finding *Image Schemas* in the text.
We aim to develop fully automatic approach to detect *Image Schemas*. We are currently annotating the NOXI corpus[3] in terms of *Image Schemas*. Our objective is then to apply Machine Learning techniques on it. Sequential learning has proven to be effective to extract semantic information from text [4]. We will explore its potential for extracting directly *Image Schemas*.

5 Metaphoric gesture generation

In [7], the authors proposed a model that maps the communicative intentions of an agent to primary metaphors in order to build a mental state in terms of *Image Schemas*. This mental state is then used to produce the corresponding gestures using a second layer of reasoning. Additionally, they used the system developed in [12] to render their final animation. However, this work limited themselves to a subset of *Image Schemas* and of gestures. In our approach, inspired by Calbris description of gestures components [1], we are looking to establish an association between *Image Schemas* and gesture primitives. Gestures are described in term of elements such as shape, direction, path. Following Calbris' terminology, some core elements of a gesture carries meaning; they are refered to as *invariant*. For example the hand shape of a 'cup' is an invariant of a gesture to refer to an object (eg 'idea'). While other elements of a gesture, refered to as variants, may vary (e.g. position of wrist in space).
Ideational Units are a decomposition of a person speech by rythm and meaning [1]. During an Ideational Unit, the gestures transfer properties from one to another (like the orientation of the hand or the shape of the hand for instance) in order to maintain the dynamic and the meaning being expressed. By manipulating gesture variants and invariants, we want to reproduce this transfer mechanism between gestures according to the underlying *Image Schemas*. A review of the literature on the topic is used to identify the different gesture primitives as a basis for reasoning and later on, annotated data of gestures will be used to learn the decomposition of primitives associated with each *Image Schema*.

References

1. Calbris, G.: Elements of meaning in gesture, vol. 5. John Benjamins Publishing (2011)
2. Cienki, A., Müller, C.: Metaphor and gesture, vol. 3. John Benjamins Publishing (2008)
3. Gromann, D., Hedblom, M.M.: Kinesthetic mind reader: A method to identify image schemas in natural language. In: Advances in Cognitive Systems. vol. 5, pp. Paper–9. Cognitive Systems Foundation (2017)
4. He, L., Lee, K., Lewis, M., Zettlemoyer, L.: Deep semantic role labeling: What works and what's next. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics (2017)
5. Johnson, M.: The body in the mind: The bodily basis of meaning, imagination, and reason. University of Chicago Press (2013)
6. Lakoff, G., Johnson, M.: Conceptual metaphor in everyday language. The journal of Philosophy 77(8), 453–486 (1980)
7. Lhommet, M., Marsella, S.: From embodied metaphors to metaphoric gestures. In: Proceedings of the 38th Annual Meeting of the Cognitive Science Society. COGSCI2016 (2016)
8. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press (1992), http://books.google.fr/books?id=3ZZAfNumLvwC
9. McNeill, D.: So you think gestures are nonverbal? Psychological review 92(3), 350 (1985)
10. Mittelberg, I.: Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. Metaphor and gesture 3, 115–154 (2008)
11. Punyakanok, V., Roth, D., Yih, W.t.: The importance of syntactic parsing and inference in semantic role labeling. Computational Linguistics 34(2), 257–287 (2008)
12. Xu, Y., Pelachaud, C., Marsella, S.: Compound gesture generation: A model based on ideational units. In: International Conference on Intelligent Virtual Agents. pp. 477–491. Springer (2014)

[3] https://noxi.aria-agent.eu/