# From embodied metaphors to metaphoric gestures

**Margot Lhommet (m.lhommet@neu.edu)**
**Stacy Marsella (s.marsella@neu.edu)**
College of Computer and Information Sciences
Northeastern University
Boston, MA 02115 USA

## Abstract

Humans turn abstract referents and discourse structures into physical gesture using metaphors. The semantic relation between abstract communicative intentions and their physical realization in gesture is a question that has not been fully addressed. Our hypothesis is that a limited set of primary metaphors and image schemas underlies a wide range of gestures. Our analysis of a video corpus supports this view: over 90% of the gestures in the corpus are grounded by a limited set of primary metaphors and image schemas. This further informs the extension of a computational model that grounds various gesture communicative intentions in a physical, embodied context, using those primary metaphors and image schemas. We conclude by discussing the application of this model to automatically generate gesture performances for embodied characters.

**Keywords:** embodied cognition; gesture; metaphor; nonverbal behavior; human-computer interaction

## Introduction

Metaphoric gestures turn abstract ideas and discourse structures into the visual and the embodied. For example, holding or weighting a large object suggests the importance of an idea. Metaphoric gestures also structure the discourse, for example by putting ideas in distinct locations in the physical space to emphasize their difference and also allow referring to them later. The particular locations can have a metaphorical meaning as well: for example, events located on the left are understood as being in the past while events on the right are in the future (Calbris, 2011).

When modeling how speakers select gestures to realize a communicative intention, a key challenge arises: how do gestures, that are physical actions, inherently described in physical terms such as size, location or path, communicate meaningful information about abstract elements that do not have physical features? In other words, where does the semantic relation between abstract referents (such as an important idea) and their physical realization in gesture (a big object) comes from?

There is evidence that the human conceptual system is embodied and structured by metaphors (Tversky & Hard, 2009). We understand abstract concepts by mapping them to image schemas (embodied experiential concepts) (Lakoff & Johnson, 1980). Reasoning processes are actions taken on these image schemas (Barsalou, 2009). For example, we make sense of the sentence "the price rises" by our understanding that an increase in quantity often correlates with an increase in height, then we extend this mapping to abstract quantities. Lakoff and Johnson (1980) and other researchers have studied how these conceptual metaphors are reflected on the verbal channel by textual metaphors, and extracted conventional mappings that associate abstract elements to image schemas (see for example Grady (1997)'s list of Primary Metaphors).

These conceptual metaphors also shape gestures (see (Cienki, 2008) for a review). Our previous work proposed a computational model that uses a restricted set of primary metaphors to map a wide range of communicative intentions to a few highly expressive image schemas (e.g. CONTAINER and OBJECT) that are common to a wide range of metaphoric gestures (Lhommet & Marsella, 2014). The gestures generated by this model communicate information about the referent (e.g. depicting a big object to suggest an important idea) and structure the discourse itself (e.g. contrasting facts by assigning them opposite locations in space). This suggests that it is possible to model a large range (if not the whole range) of gestures using a restricted set of image schemas and primary metaphors.

More specifically, such a generative model of gesturing:

– allows for a large space of gesture communicative intentions to be mapped to a comparatively small space of concrete elements (image schemas)
– can convey complex communicative intentions via composition over this small set of image schemas
– guides how properties in abstract propositions (such as "important idea") can be conveyed by manipulations of the gesture property (size of the gesture)

In this paper, we systematically extend the coverage of our previous model by using a corpus to study how communicative intentions are mapped to gesture elements via primary metaphors. The first section describes the computational model. The second section presents the analysis of the corpus. We then describes the implementation by focusing on two examples. Finally, we conclude by mentioning the advances and limits of this approach as well as discussing future work.

## Model

This work builds on a previous generative model of gesture (Lhommet & Marsella, 2014) that maps gesture communicative intentions (GCIs) to a mental space structured by image schemas using primary metaphors.

As illustrated Figure 1, to be expressed with gestures, a GCI is conceptually grounded, i.e. mapped to image schemas that have physical properties. These properties then inform the generation of a gesture plan that conveys the desired meaning.

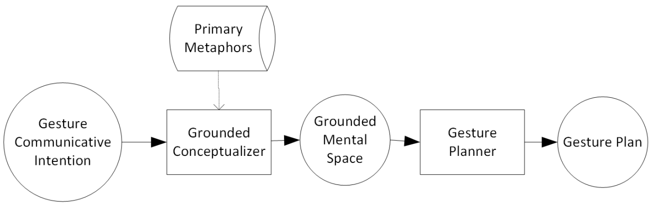


Figure 1: Our model grounds a GCI using using primary metaphors then derives a gesture plan

**Gesture Communicative Intention** Gesture can express a wide range of information that can complement, reinforce or contradict the information communicated via other modalities (Kendon, 2000). Our model takes as input a GCI that describes the meaning that a speaker wants to convey via gesture. This GCI contains the minimal set of information required to generate a gesture performance that communicates the intended meaning. For example, a speaker can have the intent to give information about the social status of an individual or about an action.

**Grounded Conceptualizer** The *Grounded Conceptualizer* maps the elements of the GCI to image schemas using a set of primary metaphors. This mapping is a systematic projection of the objects, properties and relations from one domain to another[1]. For example, the primary metaphor SOCIAL STATUS IS VERTICAL ELEVATION links the social status of an individual (or entity) to a location on a vertical scale. Individuals with a lower social status will have a lower location in space.

**Grounded Mental Space** A grounded mental space is structured by image schemas and actions taken on them. This is in line with the work of Barsalou (2009) and others, that show evidence that the brain regions responsible for perception and action coordinate during meaning creation and comprehension to create "embodied simulations" of linguistic content. This suggests that thought and reasoning processes are actually actions taken on the objects of the grounded mental space. Using the previous example, we can move individuals up and down the social status scale and infer how it impacts their social status.

Since the grounded mental space informs the generation of gesture, it should contain elements that have gesture correlates. For example, the representation of an individual on a social status scale suggests the existence of a concrete OBJECT with physical location in space, and actions to simulate moving up or down a scale.

**Gesture Planner** Finally, the *Gesture mapper* combines elements of a grounded mental space into a gesture plan that conveys the speaker's GCI. This FML-like output (Heylen, Kopp, Marsella, Pelachaud, & Vilhjálmsson, 2008) can be forwarded to a nonverbal behavior generator (Marsella et al., 2013) to generate a multimodal performance.

[1]This process can be seen as a simplified blending (Fauconnier & Turner, 2008)

# Corpus

To help quantify the primary metaphors and image schemas that play a significant role in the generation of metaphoric gestures, we created an annotated corpus of human gesturing. Several criteria were taken into account: 1. the gesturers should be "good gesturers", 2. have both of their hands visible and free, 3. the discussion topic should be abstract to elicit metaphoric gestures and 4. the discussion should be improvised instead of rehearsed.

**Description** We used a video[2] from the footage of the Working Families Summit (Washington D.C., June 23rd 2014). It portrays 6 female speakers (a journalist, a politician, two professors, a CEO and an activist) who discuss abstract concepts such as time, flexibility, income and social status. The 50 minutes video was chunked into segments that portray only one speaker at a time. We discarded pauses and segments where the journalist holds a pen and a notebook, leaving a total of 22 videos with a mean duration of 1min 42s (SD=50s), for a total of 37min 32s. 740 gestures were annotated, which gives an average of one gesture every 3 seconds.

**Annotations** 2 coders annotated the corpus with VideoAnt[3]. They selected the GCI reflected by each gesture from the following list:

- Generic reference: simple reference to an object or fact
- Specialized reference: reference to an object or fact and depiction of one or several of its properties
- Action: reference to an action
- Discourse structure: enumeration, contrast or causal relationship
- Other: none of the previous categories seems appropriate

They also annotated which element(s) of the gesture convey the desired meaning (e.g. the size of the object depicted, the shape of the motion) and selected the primary metaphor(s) that underlies this association using Grady's (1997) list 100 primary metaphors.

## Analysis

Figure 2 describes the list of the GCIs that results from the analysis of the corpus.

**Generic references** 40% of the gestures are known of gesture researchers under the name of Conduits (Reddy, 1979). The primary metaphor ABSTRACT IS CONCRETE instantiates an object in space to represent a concrete or an abstract object or an element of the discourse. The hand, facing up with an open palm, presents an immaterial object for the viewer to see. The type of the referent has little impact on the gesture shape (McNeill, 2005).

[2]https://www.youtube.com/watch?v=cQBlciBr_3w

[3]VideoAnt is a web-based annotation tool developed by the College of Education & Human Development at the University of Minnesota, available at http://ant.umn.edu

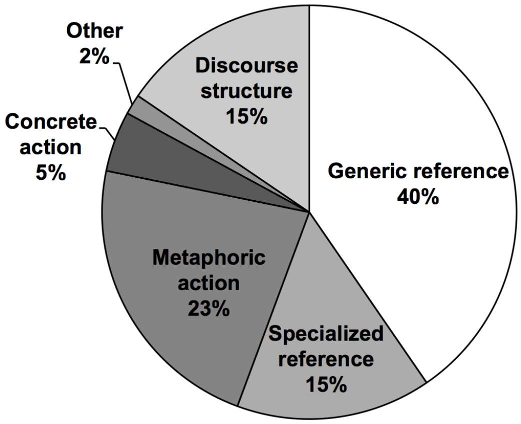

Figure 2: Distribution of the GCIs in the corpus.

**Specialized references** 15% of the GCIs consist in illustrating abstract properties of referents. The following primary metaphors are used in the corpus to map abstract properties to physical properties expressed with gestures:

- *Object location* (46%): location of the object in the physical space
  - SOCIAL STATUS IS VERTICAL ELEVATION (25%)
  - MOMENT IN TIME IS LOCATION (15%)
  - KNOWLEDGE IS LOCATED IN THE HEAD (6%)
- *Object size* (25%): e.g. the distance between two hands or the size of the gap between two fingers
  - QUANTITY IS SIZE (15%)
  - IMPORTANCE IS SIZE (10%)
- *Object shape* (29%): the shape of the hands reflects the shape of the referent
  - ESSENTIAL IS INTERNAL (23%): e.g. palms oriented towards the speaker
  - CERTAIN IS FIRM (6%): e.g. hand shape is a fist

**Depicting actions** 25% of the gestures represent actions. 20% are metaphoric actions, i.e. prototypical actions that have meaning based on a underlying metaphor (such as depicting improving one's social status by moving an object up in the physical space). Figure 3 represents the distribution of primary metaphors that underly the generation of metaphoric gesture in the corpus. 5% of the GCIs are concrete actions that mimic an actor acting in the physical space (such as a woman lifting a brick over her head).

The reader familiar with gesture studies may notice that the distribution between concrete and abstract actions differs from what is typically reported, with comparatively few concrete actions and a lot of metaphoric actions. Our view is that this difference is largely due to the nature of the corpora used. Most research on gesture have used corpora about physical phenomena (e.g. retelling a scene from a cartoon (McNeill, 1992) or explaining how to navigate a city (Bergmann & Kopp, 2009)). Therefore, gestures in these corpora reflect concrete actions. Our corpus focuses on abstract topics that do not have concrete features, so most gestures depict metaphoric actions.

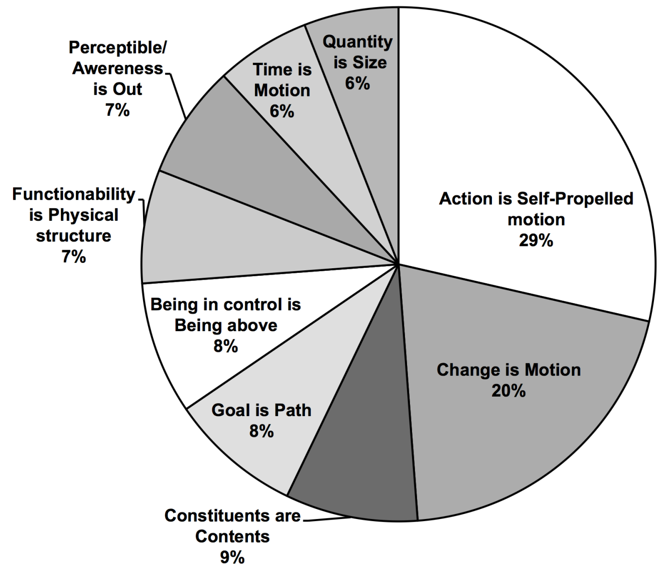

Figure 3: Distribution of the primary metaphors underlying metaphoric actions in the corpus.

**Discourse structures** 15% of the gestures structure and organize the discourse. Among them, enumerations, contrasts and expression of causality are equally distributed. Half of the enumerations in the corpus are represented as objects sequentially taken out of a container. The other half by counting on fingers. Expression of causality relies on the primary metaphor EFFECTS ARE OBJECTS WHICH EMERGE FROM CAUSES. Contrasting objects over a property relies on the metaphor SIMILARITY IS PROXIMITY where the distance between objects represents how much they differ regarding this property. The property itself can influence elements of the gesture; for example, comparing the social status of two individuals uses the vertical scale while comparing events in time uses the horizontal scale. Our previous work offers additional detail on discourse structures and their relation to primary metaphors (Lhommet & Marsella, 2014).

**Others** 2% of the gestures communicate intentions that do not fit the annotation scheme. These are gestures that express uncertainty (shrugs combined to stereotyped facial expressions) as well as emblem gestures (in particular, the corpus counts two occurrences of "quotes" traced in the physical space).

## Implementation

The computational model is implemented as a system that leverages the Cyc architecture[4]. Cyc uses a reasoning engine based on first-order logic that runs forward and backward inferences. Its knowledge base contains over 500,000 terms and 7 millions of assertions (facts and rules) relate those terms. The knowledge is hierarchically organized so properties and rules can be propagated along the inheritance links.

In (Lhommet & Marsella, 2014), we described the implementation of this computational model into a framework that

[4] `http://www.cyc.com`

(a) A continuing action (b) The shape of an action

Figure 4: Gestures can depict actions at two levels

derives gesture performances for several communicative intentions. They include (a) depicting generic referents, (b) depicting properties of object using elaborate metaphors and (c) realizing enumerations and (d) contrasts .

The following examples illustrate how our implementation generates gestures that communicate information about actions. Our corpus analysis showed that gestures can communicate two kinds of information about actions: (a) Gestures can depict the status of actions. For example, the speaker on Figure 4a says "In this country we have to continue to do that" while making a loop in the physical space. (b) Gestures can also depict the shape of physical actions in space. Another speakers says "a lot of countries have horrible cultural mores that are suppressing women" while making the gesture depicted by Figure 4b. This gesture suggests a force applied downwards that represents the control applied on women. This gesture seems to be driven by the primary metaphor BEING IN CONTROL IS BEING ABOVE.

**Gesture communicative intentions** are specified using CycL, Cyc's declarative language based on first-order logic. Script 1 represents[5] the GCIs associated to the gestures depicted in Figures 4a-4b.

**Script 1** Communicative intentions to depict actions

(a) Depict a continuing action: "We have to continue to do that"

```
(intention depictAction a)
(isa Continuation a)
```

(b) Depict the shape of an action: "Mores are suppressing women"

```
(intention depictAction b)
(isa ExercisingAuthoritativeControl b)
(performedBy b mores) (objectControlled b women)
```

Cyc's high-level term *Action*, and specializations of this term with more refined meanings, are used to model the GCIs. In Script 1(a), *Continuation* specifies that an action previously initiated continues. In Script 1(b), the action is typed as *ExercisingAuthoritativeControl*, a specialization of *ControllingSomething*, which itself inherits from *PurposefulPhysicalAction*. The actor (the mores) and object (the women) of the action are associated to the action using predicates.

[5]For the sake of clarity, we present simplified pseudocode instead of raw CycL code.

**Primary metaphors** are modeled as inference rules that map terms from the GCIs to concrete terms that represent image schemas, using Cyc's forward chaining engine. During the grounding phase, all primary metaphors rules are tested against the contents of a given GCI. If the condition side of the rule (i.e. the tuples before the '`->`' symbol) matches the input, then the predicates in the action side of the rule (i.e. the tuples after the '`->`' symbol) are set as true. The grounded mental space is created with all the predicates that are true when quiescence occurs (i.e. no rule matches anymore).

Script 2 details the implementation of the primary metaphor BEING IN CONTROL IS BEING ABOVE. Applying this rule to the GCI defined by Script 1b) results in adding to the grounded mental space two *Concrete Objects* that represent the mores and the women, and assigning them locations on a vertical scale such as the object representing the mores is located above the object representing the women. Another rule, not depicted here, matches with the fact that the action is a *PurposefulPhysicalAction* and adds a `(shape act forceful)` predicate to the grounded mental space.

**Script 2** Primary metaphor: BEING IN CONTROL IS BEING ABOVE

```
(isa ControllingSomething act)
(performedBy act actor) (isa actor Agent)
(objectControlled a object) (isa object Thing)
->
(isa ConcreteObject actor2)
(isa ConcreteObject object2)
(location actor2 locA) (location object2 locO)
(> locA locO s) (isa s VerticalScale)
```

**Gesture plans** are derived by another set of inference rules. They convert the grounded mental space into a gesture plan that reflects the physical properties using a FML-like formalism (Heylen et al., 2008). The gesture plans for the mentioned examples are described by Script 3. The system proposed by Xu, Pelachaud, and Marsella (2014) converts this formalism into the standard BML format (Kopp et al., 2006) to be rendered by the SmartBody animation system (Thiebaux, Marsella, Marshall, & Kallmann, 2008).

## Related Work

Researchers have explored several techniques to automate the generation of virtual humans' nonverbal behaviors that realize communicative intentions. Earlier systems used manual annotations of the information to convey nonverbally (e.g. (Kopp & Wachsmuth, 2002)). Some systems learn the mapping from speech input to specific classes of nonverbal behaviors (e.g. prosody to beat gestures (Levine, Krähenbühl, Thrun, & Koltun, 2010), text to head movements (Lee &

**Script 3** Gesture plans

(a) Depict a continuing action

```
<goal=depictShape shape=cycle/>
```

(b) Depict the shape of an action: “Mores are suppressing women”

```
<goal=depictShape shape=force source=locA target=locB
scale=vertical constraints=[locA>locB]/>
```

Marsella, 2010) or text to gesturing style (Neff, Kipp, Albrecht, & Seidel, 2008). Other approaches rely on expert rules that infer information from the speech. BEAT infers rheme and theme from the text to generate intonation and emphasis (Cassell, Nakano, Bickmore, Sidner, & Rich, 2001). NVBG detects communicative intentions in the text (e.g. affirmation, emphasis, disfluencies) using a keywords mapping (Lee & Marsella, 2006). Cerebella integrates acoustic, syntactic and semantic analyses to infer communicative intentions and elements of the mental state (emotional state, energy, emphasis,. . . ) (Marsella et al., 2013; Lhommet & Marsella, 2013). Approaches that take speech as input generate nonverbal behavior that is limited in the range of what can be inferred from the speech utterance only.

Some work address the production of speech and gesture from a joint representation. Bergmann, Kahl, and Kopp (2013) studies how linguistic and cognitive constraints impact the coordination of speech and gesture. Lascarides and Stone (2009) formalize the relation of gesture and speech with a logical form of multimodal discourse, in particular between discourse elements and deictic gestures. In the Gestures as Simulated Action framework, perceptual and motor representations automatically become active during language production and, under certain conditions are sources of gestures (Hostetter & Alibali, 2008).

## Discussion

In this paper, we presented a computational model of gesture generation informed by embodied cognition that turns various communicative intentions into gesture by grounding them in a physical, embodied context. Using the analysis of a video corpus, we showed that most GCIs present in the corpus can be conveyed using a very limited set of primary metaphors (at the exception of a few stereotyped gestures -6 occurrences over 740 gestures- that could easily be integrated by providing a direct mapping from specific GCIs to these emblem gestures.)

A possible application of this model is the automatic generation of multimodal performances for virtual humans. Virtual humans are autonomous agents that engage users in face-to-face interactions, ideally using the same verbal and nonverbal behaviors as humans (Cassell, 2000). They have proven to be effective in a wide range of applications, for example to persuade patients to adhere to health regimen (Bickmore & Cassell, 2005) or to train cross-cultural negotiation skills (Kim et al., 2009). Metaphoric gestures improve message understanding and impact how a speaker is perceived in particular in terms of persuasiveness and competence (Cohen, Beattie, & Shovelton, 2011; Beaudoin-Ryan & Goldin-Meadow, 2014). This may be another reason why metaphoric gestures dominate in this corpus since all the speakers are professional public speakers. Given that good communication skills, persuasiveness and competence are critical in health interventions and training, metaphoric gestures should therefore be an important capability of virtual humans designed for these applications.

Furthermore, this computational model provides a more controlled yet flexible methodology to experiment with social and psychological constructs; for example, virtual humans can serve as confederates in psychology and social psychology experiments to study the impact of nonverbal behaviors.

A limit to the broad application of this work is the need to manually specify the gesture communicative intent of the speaker. A promising avenue here is the Embodied Construction Grammar (ECG) framework (Bergen & Chang, 2005) that represents a speaker’s intended meaning based on image schemas, along with the mental simulation of these representations using executing schemas (S. S. Narayanan, 1997). Our future work will investigate the integration of our computational model into the ECG framework, in particular applying the work of S. Narayanan (1999) on inferring and reasoning on conceptual metaphors from speech onto gesture.

## References

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281-1289.

Beaudoin-Ryan, L., & Goldin-Meadow, S. (2014). Teaching moral reasoning through gesture. *Developmental Science*, *17*(6), 984-990.

Bergen, B. K., & Chang, N. (2005). Embodied construction grammar in simulation-based language understanding. In J.-O. Östman & M. Fried (Eds.), *Construction grammars: Cognitive grounding and theoretical extensions* (p. 147-). John Benjamins Publishing.

Bergmann, K., Kahl, S., & Kopp, S. (2013). Modeling the semantic coordination of speech and gesture under cognitive and linguistic constraints. In *Intelligent virtual agents* (p. 203-216). Edinburgh, UK.

Bergmann, K., & Kopp, S. (2009). Increasing the expressiveness of virtual agents: autonomous generation of speech and gesture for spatial description tasks. In *Proceedings of the 8th international conference on autonomous agents and multiagent systems - volume 1* (p. 361-368). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Bickmore, T., & Cassell, J. (2005). Social dialongue with embodied conversational agents. In *Advances in natural multimodal dialogue systems* (p. 23-54). Springer.

Calbris, G. (2011). *Elements of meaning in gesture* (Vol. 5). John Benjamins Publishing.

Cassell, J. (2000). *Embodied conversational agents*. MIT Press.

Cassell, J., Nakano, Y., Bickmore, T., Sidner, C., & Rich, C. (2001). Annotating and generating posture from discourse structure in embodied conversational agents. In *Workshop on representing, annotating, and evaluating non-verbal and verbal communicative acts to achieve contextual embodied agents.*

Cienki, A. (2008). Why study metaphor and gesture. In A. Cienki & C. Müller (Eds.), *Metaphor and gesture* (p. 5-25). Amsterdam; Philadelphia: John Benjamins Pub. Co.

Cohen, D., Beattie, G., & Shovelton, H. (2011). Tracking the distribution of individual semantic features in gesture across spoken discourse: New perspectives in multi-modal interaction. *Semiotica*, *2011*(185), 147-188.

Fauconnier, G., & Turner, M. (2008). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.

Grady, J. E. (1997). *Foundations of meaning: Primary metaphors and primary scenes*. Doctoral dissertation, University of California, Berkeley, United States – California.

Heylen, D., Kopp, S., Marsella, S., Pelachaud, C., & Vilhjálmsson, H. (2008). The next step towards a function markup language. In *Intelligent virtual agents* (p. 270-280). Tokyo, Japan.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic bulletin & review*, *15*(3), 495-514.

Kendon, A. (2000). Language and gesture: Unity or duality. In D. McNeill (Ed.), *Language and gesture* (p. 47-63). Cambridge University Press.

Kim, J. M., Hill Jr., R. W., Durlach, P. J., Lane, H. C., Forbell, E., Core, M., ... Hart, J. (2009). Bilat: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education (IOS Press)*, *19*(3), 289-308.

Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., ... Vilhjálmsson, H. (2006). Towards a common framework for multimodal generation: The behavior markup language. In *Intelligent virtual agents* (p. 205-217). Marina del Rey, CA.

Kopp, S., & Wachsmuth, I. (2002). Model-based animation of co-verbal gesture. In *Computer animation, 2002. proceedings of* (p. 252-257). Geneva, Switzerland.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by* (2003rd ed.). Chicago: University of Chicago Press.

Lascarides, A., & Stone, M. (2009). A formal semantic analysis of gesture. *Journal of Semantics*, *26*(4), 393-449.

Lee, J., & Marsella, S. C. (2006). Nonverbal behavior generator for embodied conversational agents. In *6th international conference on intelligent virtual agents.* Marina del Rey, CA.

Lee, J., & Marsella, S. C. (2010). Predicting speaker head nods and the effects of affective information. *Multimedia, IEEE Transactions on*, *12*(6), 552-562.

Levine, S., Krähenbühl, P., Thrun, S., & Koltun, V. (2010). Gesture controllers. In *Acm siggraph 2010 papers* (p. 124:1-124:11). New York, NY, USA: ACM.

Lhommet, M., & Marsella, S. (2014). Metaphoric gestures: Towards grounded mental spaces. In *Intelligent virtual agents* (Vol. 8637, p. 264-274). Boston, MA: Springer International Publishing.

Lhommet, M., & Marsella, S. C. (2013). Gesture with meaning. In *Intelligent virtual agents* (Vol. 8108, p. 303-312). Edinburgh, UK: Springer Berlin Heidelberg.

Marsella, S., Xu, Y., Lhommet, M., Feng, A., Scherer, S., & Shapiro, A. (2013). Virtual character performance from speech. In *Proceedings of the 12th acm siggraph/eurographics symposium on computer animation, anaheim, ca* (p. 25-35). New York, NY: ACM.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

Narayanan, S. (1999). Moving right along: A computational model of metaphoric reasoning about events. In *In proceedings of the national conference on artificial intelligence (aaai'99.* Orlando, FL.

Narayanan, S. S. (1997). *Knowledge-based action representations for metaphor and aspect (karma)*. Unpublished doctoral dissertation, UNIVERSITY of CALIFORNIA.

Neff, M., Kipp, M., Albrecht, I., & Seidel, H. P. (2008). Gesture modeling and animation based on a probabilistic recreation of speaker style. *ACM Transactions on Graphics (TOG)*, *27*(1).

Reddy, M. J. (1979). The conduit metaphor: A case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and thought* (Vol. 2, p. 164-201). Cambridge: Cambridge University Press.

Thiebaux, M., Marsella, S., Marshall, A. N., & Kallmann, M. (2008). Smartbody: behavior realization for embodied conversational agents. In *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems* (Vol. 1, p. 151-158). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, *110*(1), 124-129.

Xu, Y., Pelachaud, C., & Marsella, S. (2014). Compound gesture generation: A model based on ideational units. In T. Bickmore, S. Marsella, & C. Sidner (Eds.), *Intelligent virtual agents* (p. 477-491). Springer International Publishing.