Mechanisms underlying the beneficial effect of a speaker’s gestures on the listener

Published version:

DOI:10.1016/j.jml.2017.05.004

(Article begins on next page)

Mechanisms underlying the beneficial effect of a speaker's gestures on the listener

Francesco Ianì & Monica Bucciarelli

Dipartimento di Psicologia, Università di Torino

Via Po, 14 –10123 Turin, Italy

e-mail: francesco.iani@unito.it

e-mail: monica.bucciarelli@unito.it

Corresponding author:

Francesco Ianì

Università di Torino

Dipartimento di Psicologia

Via Po, 14

10123 Turin, Italy

e-mail: francesco.iani@unito.it

tel: +39.011.6703038; fax:+39.011.8146231

Abstract

A well-established literature reveals that a speaker's gestures have beneficial effects on the listener's memory for speech. A main assumption of our investigation is that gestures improve memory through the exploitation of the listener's motor system. We tested this prediction in four experiments in which the participants listened to action sentences uttered by a speaker who either stayed still or accompanied the speech with congruent gestures. The results revealed that when the listeners observed gestures their memory for speech improved (Experiment 1), but loading up the listener's motor system during gestures observation cancelled the beneficial effect when the motor task involved the same effectors used by the speaker (arms and hands, Experiments 2-3). The beneficial effect of gestures persisted when the motor task involved different effectors (legs and feet, Experiment 4). These results support the assumption of a main involvement of the motor system in the beneficial effect of observed gestures.

Keywords: gestures; memory for action; mental models; motor system; experimenter-performed task

Introduction

Hand gestures are motor actions that often accompany speech and are intertwined with the spoken content (e.g., Kelly, Manning, & Rodak, 2008; McNeill, 1992). A huge literature revealed that gestures are crucial in communication: both for the speaker and for the hearer (see, for a review Goldin-Meadow, 1999; Goldin-Meadow & Alibali, 2013). Also, most relevant to the present investigation, gestures can improve learning and memory in several ways, both when produced and when observed (see, e.g., Cutica & Bucciarelli, 2008; 2013). Cook, Yip and Goldin-Meadow (2010) found that producing gestures at learning, spontaneously or on demand, makes the information more memorable. Producing gestures is beneficial also in learning from scientific texts (Cutica, Ianì, & Bucciarelli, 2014; Ianì, Cutica, & Bucciarelli, 2016) and in learning math: requiring children to gesture while learning a new mathematical concept helps them to retain the knowledge they had gained during instructions (Cook, Mitchell, & Goldin-Meadow, 2008).

Although the results of some studies seem suggest that performing gestures affects learning more than observing gestures (e.g., Goldin-Meadow, Levine, Zinchenko, Yip, Hemani, & Factor, 2012), gestures may have a pivotal function also for the observer. For instance, children are more likely to learn a task when their teacher accompanies the instructions with congruent gestures than when the instructions do not include gestures (e.g., Church, Ayman-Nolley, & Mahootian, 2004; Cook & Goldin-Meadow, 2006; Ping & Goldin Meadow, 2008), and observing gestures while learning words of a foreign language can improve the level of learning (e.g., Macedonia & Knosche, 2011).

Consistent with these findings, the literature on the so-called *enactment effect* reveals that human memory for action sentences is improved by producing gestures or observing gestures congruent with the action described by the sentences. This effect was formerly detected in 1981 by Cohen: free recall of action phrases like *break the toothpick* was improved when participants, during the learning phase, were asked to perform with gestures the action portrayed in the sentences (*subject-performed task*, SPTs) or when they were asked to observe the speaker performing the action (*experimenter-performed task*, EPTs), as compared to the situation in which the participants just heard or read the sentences (*verbal task*, VTs). The enactment effect has been observed in free recall tasks as well as in recogntion tasks (see Engelkamp, 1998), using entire actions sentences as well as single nouns (e.g., Kormi-Nouri, Nyberg, & Nilson, 1994). Further, the effect has been observed in children (e.g., Thompson, Driscoll, & Markson, 1998) as well as in elderly adults (Feyereisen, 2009).

Although few studies reported an advantage of SPTs on EPTs (e.g., Hornstein & Mulligan, 2004), the beneficial effect of enactment occurs both when the participants themselves perform the gestures and when they simply observe the gestures produced by a speaker (Madan & Singhal, 2012). Cohen (1981), during a free recall task, detected no difference between the SPTs condition and the EPTs one, whereas Engelkamp and Zimmer (1997) detected a recall advantage of SPTs over EPTs. However, as Engelkamp and Dehn (2000) argue, this inconsistency in findings could depend on the length of the list of sentences to recall. Taken together, the studies on the enactment effect suggest comparable recall rates in SPTs condition and EPTs condition (Feyereisen, 2009). An important feature of the enactment effect is that, although several action sentences used in the literature involve external objects (e.g., *open a book*, *play the piano*), it is not necessary to show up the real objects in order to detect a beneficial effect on memory (e.g., Mohr, Engelkamp, & Zimmer, 1989): just a *pantomime*, a gesture performed without using real objects, produces the enactment effect. A study of Engelkamp and Zimmer (1997), where real objects were presented both in VTs, EPTs and SPTs conditions, revealed that introducing the real object didn't improve the enactment effect, thereby suggesting that the object component is not a critical factor in the advantage of SPTs and EPTs over VTs.

Although the positive effects of gestures are robust, their interpretation is still controversial. As Feyereisen points out, it is well established that "enactment adds something to the processing of the verbal material to be memorized […] the problem is to identify what is added" (Feyereisen, 2009, p.374). In particular, a question still waiting for an answer is which mechanisms underlie the beneficial effect of gestures. It has to be excluded an attentional explanation. Indeed, in the enactment effect, as for co-speech gestures (see, e.g., Cutica & Bucciarelli, 2015; Kelly, McDevitt, & Esch, 2009), the beneficial effect of observing gestures depends on their semantic meaning and not by their ability to focus the attention on the word they accompany. In particular, Feyereisen (2006) found that only matching representational gestures facilitated verbal recall, whereas incongruent or beat gestures did not.

In this paper, we test the prediction that observing gestures improves memory through the explotation of the motor system. This prediction is implied by the assumptions of the mental model theory, our theoretical framework, but it is consistent also with alternative theoretical frameworks.

Theories of the beneficial role of gestures in the enactment effect

There exist several theoretical accounts of the enactment effect, all of them not mutually exclusive. Their focus is on the beneficial effect of subject-performed tasks compared to experimenter-performed tasks and pure verbal tasks (hereafter, SPTs, EPTs and VTs, respectively). Among them, the *episodic integration hypothesis* suggests that enacting action sentences reinforces the episodic relationship between the verb portraying the action and the object noun (Kormi-Nouri, 1995). This process results in a stronger association between action and object: these components are encoded in a single memory unit (see, e.g., Kormi-Nouri & Nilsson, 2001; Mangels & Heinberg, 2006). The *distinctiveness hypothesis* (see, e.g., Engelkamp, 1998) suggests that SPTs increase item distinctiveness because planning and executing actions focuses the encoding on item-specific information. According instead to the *multimodality hypothesis* (see, e.g., Engelkamp, 2001), performing an action requires planning and movement control that provide a motor representation which may be reactivated at retrieval (see also Zimmer, 2001). In this view, the classical effect detectable in SPTs condition should arise from the activation and later on the reactivation of information stored in the motor system, thereby enabling a greater elaboration of the action concept in memory. Consistent with this hypothesis, the amount of visual feedbacks does not affect the beneficial effect of gestures: the SPTs effect is detectable also when the persons are blindfolded during the learning phase (Engelkamp, Zimmer, & Biegelmann, 1993) and conversely, memory is not enhanced when a mirror is situated in front of the participants (Hornstein & Mulligan, 2004). These results suggest that the motor information, rather than the visual one, is crucial for the enactment effect.

The multimodal hypothesis gave rise to a series of investigations, and the assumption that stored information is enriched by sensory and motor information during encoding and retrieval resulted in the *reactivation hypothesis*: the motor processes which took place during the study phase should affect the memory and be regenerated during retrieval. These mechanisms would underly the beneficial effect of gestures in SPTs. Consistent with this assumption several neuroimaging studies suggest that the enactment effect results from the possibility to base retrieval on motor information. For example, a PET study revealed a major involvement of the brain motor areas in the verbal retrieval of phrases that the participants formerly accompanied with gestures (Nilsson, Nyberg, Klingberg, Aberg, Persson, & Roland, 2000): remembering action sentences previously accompanied by gestures engages the motor brain areas. An fMRI study of Russ, Mack, Grama, Lanfermann and Knopf (2003) detected a crucial role of postcentral right area (BA2) after the SPTs condition compared to the VTs condition. The area B2 is roughly the equivalent of the primary motor cortex detected in Nilsson et al.'s study (2000).

Nyberg, Petersson, Nilsson, Sandblom, Åberg and Ingvar (2001) measured and compared the brain activities both at learning and recall in order to investigate more in depth the reactivation hypothesis. They observed a great overlap in brain regions activated in both phases, specifically in the left ventral motor cortex and in the left inferior parietal cortex. Since overlapping regions in motor cortex were activated at both learning and retrieval phases, Nyberg and colleagues concluded that retrieval after enactment in SPTs can depend on motor information and that the function of the motor cortex is not limited to the execution of movements, but it is involved also in non-motor skills (see also Masumoto, Yamaguchi, Sutani, Tsuneto, Fujita, & Tonoike, 2006). In sum, findings in the neurocognitive literature revealing a critical activation of the motor areas during recall or recognition after SPTs condition support the *motor information reactivation hypothesis*.

The role of the motor system within a mental model framework

A central assumption of the mental model theory (Johnson-Laird, 1983; 2006) is that a deep comprehension of a discourse, and the subsequent good recall, is tantamount to the construction of an articulated mental model of the discourse. A *mental model* is an iconic, non-discrete, mental representation that reproduces the state of affairs described in a discourse (see, e.g., Graesser, Millis, & Zwaan, 1997); a model consists of elements, which stand for the entities in the discourse, and the relationship between these elements, which stand for the relationship between the entities. Models encode little or nothing of the linguistic form of the sentences on which they are based, hence the prediction, confirmed by the results of studies in the literature, that individuals recover more information at a semantic level and less information at a verbatim level (e.g., Mani & Johnson-Laird, 1982). In particular, an articulated mental model, compared to a poor mental model, results in a greater number of correct recollections and discourse-based inferences drawn from the information explicitly contained in a given material, along with a poorer retention of the surface information (see, e.g., Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Johnson-Laird & Stevenson, 1970).

Bucciarelli (2007) argued that the information conveyed by the speaker's co-speech gestures, represented in a non-discrete format, are easily included into the discourse mental model, since mental models too are non-discrete representations (see also Cutica & Bucciarelli, 2008; Hildebrandt, Moratz, Rickheit, & Sagerer, 1999). Also, Cutica and Bucciarelli (2013) argue that co-speech gestures, which are spatial in nature, convey information that can be easily incorporated into the text/discourse mental model because mental models themselves are spatially organized (Knauff & Johnson-Laird, 2002).

Hence, co-speech gestures might lead to the construction of representations that are easily incorporated into the discourse model, alongside the representations constructed on the basis of the verbal information, enriching these and completing the mental model. A mental model contains both declarative knowledge, the so called “knowing that”, a set of concepts expressed in the form of propositions, and procedural knowledge, the so called “knowing how”, a set of information that are not represented in an explicit manner, but that allow us to interact effectively with the world. Thus, for example, a mental model contains knowledge about “what is a cup” and about “how grasp a cup”. In a mental model “the difference in representation between declarative knowledge and procedural knowledge disappears. The two types of knowledge continue to be different and are perfectly distinguishable, but they can now be represented in one single general modality, the mental model” (Bara, 1995, p.116). Hence, in production, mental models can both activate abstract propositional representations that may be reflected in speech, but they can also activate motoric representations that may be reflected in gestures; producing or observing gestures could reinforce the encoding of these motor information (Ianì et al., 2016). This assumption is consistent with the proposal that gestures derive from a system of motoric representations of concepts, many of which also come to be reflected in speech (Morrel-Samuels & Krauss, 1992). These motor representations are part of the listener’s mental models, the procedural aspects encoded in them.

With the aim to test the assumption that gestures favor the construction of a discourse mental model, several studies analyzed the effect of producing and observing gestures in order to detect the classical indexes of mental models’ construction. In a series of experiments, Cutica and Bucciarelli (2008) investigated whether listeners exposed to a narrative discourse accompanied by gestures, as compared to a discourse not accompanied by gestures, would build a more articulated mental model. The results of their Experiments 1 and 2 confirmed the predictions: when exposed to a discourse accompanied by gestures the participants produced at recall a higher amount of correct recollections and of discourse-based inferences. The participants in Experiment 3 and 4 of the same study encountered the same videos as those in Experiment 1 and 2, but at the end of each video they encountered a set of sentences with the request to recognize those actually proffered by the speaker: the participants in the no-gesture condition performed better than the participants in the gesture condition in recognizing literal sentences actually spoken by the actor, as well as with paraphrases, thus not endorsing them.

Also the gestures produced while studying a text have a similar effect. Cutica and Bucciarelli (2013) carried out two experiments to test the prediction that the participants, when instructed to represent gesturally the concepts they encountered while studying a scientific text, as compared to a condition where they stay still, recall a greater amount of correct information and draw more discourse-based inferences (recall task, Experiment 1), and have a poorer memory for text verbatim (recognition task, Experiment 2). A similar pattern of results was detected in fifth-grade children invited to enact the concepts encountered while studying a scientific text (Cutica et al., 2014).

The assumption that gestures favor the construction of a discourse mental model holds also for the enactment effect; the facilitating effect of observing the gestures of the speaker on comprehension and memory parallels the EPTs enactment effect, and the facilitating effect of the gestures produced by the learner parallels the SPTs enactment effect. Consistent with our proposal, the actual pattern of movements constituting a SPT is not critical in determining the recall level, as long as the patterns are appropriate to the accompanying speech (e.g., Cohen & Bryant, 1991; Noice & Noice, 2007). For instance, Noice and Noice (2007) detected the so-called *non-literal enactment effect*: performed actions not literally congruent with the verbal material, but related at a higher order level (e.g., action goal level), result in action-enhanced memory for the verbal material. In other words, the meaning of the gesture rather than its literal correspondence to speech is effective for memory.

Further, the proposed mental model framework for the enactment effect can accommodate evidence in the literature embarrassing the motor hypothesis theorists. Several authors argue that the assumption that the enactment leaves a memory-relevant motor code leads to predict that re-enacting the verb increases the probability of recalling the object (see, e.g., Knopf, Mack, Lenel, &

Ferrante, 2005; Mulligan & Hornstein, 2003). As the results of their own studies and those in the literature (Kormi-Nouri, Nyberg, & Nilsson, 1994; Norris & West, 1993) reveal that re-enactment at retrieval has no impact on the recall, they conclude against the motor hypothesis. However, a study by Cutica and Bucciarelli (2011) not only explain, but also predicts this effect. We know from the literature that gestures favor the speaker's mental organization of the discourse by helping to organize the stream of thought (see, e.g., Alibali, Kita, & Young, 2000; Goldin-Meadow & Alibali, 1999; Kendon, 1983; McNeill, 1992). Cutica and Bucciarelli (2011) found that participants who saw the discourse accompanied by gestures were less likely to gesture at recall and they reasoned that, once an individual constructed an articulated mental model of a given material, s/he is less likely to gesture in recalling that material. Hence, the results according to which re-enactment at retrieval is not effective could be explained in terms of an interference of the re-enactment of the gestures on recall.

We argue that the enactment effect relies on the beneficial role of gestures for the construction of a mental model of the material to be learnt through the exploitation of the motor system. In the classical enactment literature, the EPTs condition was formerly introduced as a control condition for SPTs. The underlying assumption was that the activation of motor components was present in SPTs, but not in EPTs, for which the visual component was considered fundamental (see Feyereisen, 2006). Since recall after SPTs is usually slightly better than after EPTs (see the Introduction), some scholars argued that "[…] the recall advantage of SPTs over EPTs is caused by motor processes in SPTs" (Engelkamp & Jahn, 2003, p.150). The implicit assumption underlying their claim is that the motor component is present only in the SPTs condition, whereas it is absent in the EPTs condition. Nevertheless, the relationship and the link between action and perception is more complex. Several investigations detected high degree of overlap between the neural circuits underlying the execution of an action and those underlying the observation of the same action involving the same effectors (see, e.g., Rizzolati & Craighero, 2005). Studies on monkey revealed that during the observation of hand actions there is an increased activation in the area F5, the so called "mirror neuron system" (e.g., Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). A series of studies on the mirror system revealed that the simple observation of an action involves the activation of the same areas devoted to the action production, both in non-human primates (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996) and human primates (e.g., Rizzolatti, 2005). The human homologue to F5 is probably BA44 and BA6, the ventral premotor cortex. The premotor cortex is active during the observation of hand (Nishitani & Hari, 2000) and body movements (Buccino*,* Binkofski, Fadiga et al., 2001). Therefore, it is possible that the observation of the experimenter's pantomime in EPTs activates motor representations in the observer in a "covert" way (Barsalou, Nidenthal, Barbey, & Ruppert, 2003). Thereby the motor component might play a critical role also in the beneficial effect of gestures in EPTs compared to the VTs.

Buccino and colleagues (2001) compared the activation deriving from the observation of different actions involving different effectors. The observation of mouth actions activated the ventral area 6 and area 44, the observation of hand actions, in comparison, resulted in a greater activation of the dorsal part of the ventral area 6 and of a dorsal sector of area 44, and the observation of foot actions activated a dorsal sector of area 6. Because of the shift in the premotor cortex activation, from ventral to dorsal, when the effector observed were mouth, arm and hand, and foot, respectively, the authors concluded in favor of a somatotopic organization of the brain activity in this area during action observation. Buccino and colleagues (2001) concluded that "during action observation there is a recruitment of the same neural structures which would be normally involved in the actual execution of the observed action" (Buccino et al., 2001, p. 404). This somatotopic organization of PMC is much in line with the classic homunculus of the motor system: foot movements are located at superior areas, mouth movements at inferior areas and other body parts in between of them (see, e.g., Van Overwalle & Baetens, 2009). Over the years, several studies highlighted the somatotopic organization of the pre-motor cortex. For instance, Wheaton, Thompson, Syngeniotis, Abbott and Puce (2004) compared the passive observation of videos portraying movements performed with different types of effectors and the observation of static images of the same stimuli and found a somatotopic organization in the premotor cortex. Consistent with former findings in the literature the activation deriving from the observation of legs' movements, compared to the observation of hands' and mouth's movements, was more associated to the dorsal activation, whereas the observation of face movements was associated more to a ventral activation. Similar findings were reported in Sakreida, Schubotz, Wolfensteller and von Cramon (2005).

Two meta-analyses of Van Overwalle and Batens (2009) and Caspers, Zilles, Laird and Eickhoff (2010) supported these findings. In particular, more than 100 studies on action observation detected a clear network: across both hemispheres action observation involves the BA 44/45 (overlapping with the ventral BA6), the dorsal premotor cortex (dPMC, BA6), the supplementary motor area (SMA, overlapping the more dorsal part of BA6), the inferior parietal lobule (IPL) and the primary somatosensory cortex (SI, BA1, 2). As regards the premotor cortex, the observation of face and mouth movements were more associated to the activation of the ventral part of the PMC, whereas the observation of hand actions was more associated with the activations of a more dorsal part of the PMC. Caspers and colleagues conclude that there is "evidence for a somatotopical organization of activations within the lateral premotor cortex" (Caspers et al., 2010, p.1164). Neuroscientific evidences for a somatotopic organization of the premotor areas and their causal role in action understanding come from a study of Michael, Sandberg, Skewes, Wolf, Blicher, Overgaard and Frith (2014).

Do the processes described above play a causal role in the beneficial effect of gestures? More specifically, do such processes play a causal role in the EPTs enactment effect? One could argue that the nature of the pantomime (commonly used in the enactment paradigm) is slightly different from the real actions usually investigated in the study on action observation. Further, pantomime differs from co-speech gestures in that they comprehend a wider variety of arms' and hands' movements; it might be improper to extend the findings concerning gestures to pantomimes. However, the assumption that the motor system is involved in co-speech gestures understanding is plausible in the light of the results of a study by Lotze, Heymans, Birbaumer, Veit, Erb, Flor and Halsband (2006), who compared the activation of brain areas during the observation of different types of gestures: real hand movements (e.g., pouring in a cup), body-related movements (e.g., combing hair) and expressive gestures (e.g., make hallo). Although some areas were selective for some specific type of gestures, the mirror system in the premotor cortex was active in all conditions. Further, our assumption is consistent also with recent findings by Ping, Goldin-Meadow and Beilock (2014), which suggest that the observer's motor system plays a role in understanding the gestures of a speaker. In their Experiment 1, the task of the participants was to carefully observe a series of videos in which an actor utters a series of sentences, each one accompanied by a gesture. The participants were not instructed to attend to gesture. Immediately after each photo, the participants saw a figure of an object that could be in a position *congruent* or *incongruent* with the gesture observed (for instance, a nail in vertical or horizontal position), and their task was to respond "yes" if the name of the object in the figure was mentioned in the sentence and "no" if it was not (filler trials). When the object in the figure was in a position congruent with the gesture observed in the video (in other words, when the information conveyed in gesture matched the information conveyed in the figure), the participants were faster in responding correctly compared to when the object in the figure was in an incongruent position. Ping and colleagues conclude that the participants *automatically incorporated* the information coming from actor's hand gestures into their mental representation of the speaker's message. In Experiment 2 Ping and colleagues (2014) investigated whether the automatic incorporation of the information conveyed by the gesture in the model of the speaker's message involves the listener's motor system. The task of the participants was the same as in Experiment 1, with the exception that they were invited, during the observation phase, to perform a motor secondary task, either with their arms and hands (the same effectors used by the actor in the video) or with their legs and feet (different effectors from those used by the actor). The rationale of the dual task was to discover whether loading up the motor system of the participants by asking to move their arms and hands would result in a difficulty in gestures understanding with the consequence that the congruence effect should disappear. Consistently with the somatotopic organization of the premotor activation during action observation, this interference should be specific to motor resources controlling the effectors used in producing gesture - in this case, arms and hands. To test this prediction, half of the participants in Experiment 2 were asked to plan and execute movements with their arms and hands (arm movements condition), while watching the videos and giving their responses. The other half of the participants was asked to plan and execute movements with their legs and feet (leg movements condition). The results revealed that the participants in the arms movements condition didn't show the congruent effect in the picture judgment task: there was no difference in response times to congruent versus incongruent gesture-picture combinations. In contrast, the congruence effect persisted when the participants were asked to move their legs and feet: they responded more quickly to pictures that were congruent with the speaker's gestures than to pictures that were incongruent. Ping and colleagues concluded that the listener's motor system is involved in gesture understanding.

Experimental hypothesis

Since there is a substantial overlap between the neural areas activated in perceiving someone who performs an action and the neural areas activated when we ourselves perform that same action, we assumed that the pantomimes used in EPTs condition are motor acts that automatically and somatotopically activate the listener's motor system, and that this process is crucial for the beneficial effect of the enactment in EPTs. We tested the predictions deriving from these assumptions in a series of experiments. In all the experiments the participants observed the phrases uttered by an actress in two conditions: in the gesture condition the actress performed iconic gestures which represent the actions described, in the no-gesture condition the actress keept her hands still. The speaker's enactment of phrases should:

- improve memory in the listener who stays still during the observation (Experiment 1);

- not improve memory in the listener who moves her arms during the observation, i.e. the same motor effectors moved by the speaker (consistently with the somatotopic organization of the premotor activation during action observation) (Experiments 2 and 3);
- improve memory in the listener who moves her legs during the observation, i.e., different motor effectors from those moved by the speaker (Experiments 4).

We devised the experimental material of the experiments through a normative study, whose aim was to identify a series of sentences eliciting a great motor activity in the listener. As an example, consider the sentences *rowing a galley-ship* and *looking the wristwatch* from our pilot study; the former elicits in the listener more motor activity than the latter.

Normative study: Sentences eliciting great motor activity in the listener

The participants in the study read 60 sentences representing actions and they rated each sentence according to how strongly each one elicited movement on a 7-point Likert scale.

Method

Participants

The participants in the study were 40 Italian adults (17 males and 23 females, mean age: 29.1 years, *SD* = 7.3). They were recruited among people around the university campus. They took part in the study voluntarily.

Materials

The experimental material consists of a series of 60 sentences portraying actions. Forty-eight of the sentences were extracted from the normative study by Molander and Arar (1998), in which the participants rated 439 actions on a 7-points scale according to the motor activity dimension. Specifically, we selected phrases that received a mean rating above 3 on the motor activity dimension, and involving objects (e.g., *drinking a glass of water*): the rationale was that this kind of sentences elicits movement more than sentences not involving objects (e.g., *blowing in the air*). At the same time, we selected only the sentences that can be easily represented by arms' and hands' movements (for instance we didn't select sentences involving the entire body like *standing up and sitting down*, or sentences that cannot be easily represented by a pantomime like *flying a kat*). In order to devise a set of 60 sentences we added 12 sentences, still referring to an action, on an object, that can be easily represented with a pantomime of hands and arms. Each sentence was printed on a sheet, and just below the sentence there was a Likert scale whose values ranged from 1 to 7. In correspondence of 1 there was the label "It elicits little movement", in correspondence of 4 there was the label "50:50", and in correspondence of 7 there was the label "It elicits great movement". The sheets containing the 60 sentences were assembled in a booklet, in a random order for each participant in the study.

Procedure

The study was run individually, in a quiet room, at the sole presence of the experimenter. The experimenter invited the participant to read carefully the instructions written on the first page of the booklet: "Thank you for your participation in this experiment. This booklet contains a series of sentences representing actions. Your task is to carefully read each sentence one at time and rate, using a scale ranging from 1 to 7, how strongly the written action elicits movement. For instance, the sentence *thinking a number* elicits a low level of motor activity whereas the sentence *climbing the stairs* elicits a great level of motor activity. Feel free to use the entire scale (1-7) but please mark a number and not the space between two numbers. Don't worry if you use the same number for some subsequent sentences' evaluations."

Results

The Appendix shows the mean rating of the 24 sentences that most elicited movement, from the sentence eliciting the greatest motor activity to the sentence eliciting the least motor activity. The rationale for choosing 24 sentences was that the beneficial effect of EPTs on memory is optimal when the lists of sentences are short (12-18 items, see Feyereisen, 2006), and for the experiments proper we needed two lists of sentences.

Stimuli preparation

For each of the 24 sentences we created a couple of videos: in one video an actress accompanies the sentence with congruent gestures (gesture condition) and in the other she keeps her hands still (no-gesture condition). In the gesture condition the actress was instructed to accompany each sentence with a congruent arms' and hands' movement. In other words, she produced a pantomime of the action performed on the object, as mentioned in the sentence. Further, the actress was asked to gesture at the same time in which she started to pronounce the sentence. In the no-gesture condition the actress was instructed to pronounce the sentence keeping her arms still on the knees. In both conditions she was invited to use the same intonation while uttering the sentence. In Figures 1, as examples, two frames from the videos created for the sentence *throwing a stone*. Then, we created two experimental protocols. In Protocol 1 half of the sentences occurred in the gesture condition (number of words in the list of sentences: 42) and the other half in the no-gesture condition (number of words in the list of sentences: 42). In Protocol 2, the sentences occurring in the gesture condition in Protocol 1 occurred in the no-gesture condition, and the sentences occurring in the no-gesture condition in Protocol 1 occurred in the gesture condition. Further, within each protocol we balanced the order of presentation of the gesture and the no-gesture condition.

(a) (b)

Figure 1. The actress utters the sentence *throwing a stone* in the gesture (a) and in the no-gesture (b) conditions.

The videos were used in the experiments proper, described below. All the experiments were approved by the Ethical Committee of the University of Turin.

Experiment 1. Gestures favor memory of action phrases

Several studies reveal that action phrases are recalled better by the listener when the speaker accompanies them with coherent gestures (see, e.g., Feyereisen, 2006). The aim of the experiment was to replicate this usual enactment effect detectable in EPTs condition with our new experimental material. In particular, we investigated whether sentences depicting actions are learnt and memorized better if accompanied by congruent gestures at learning phase (EPTs – experiment performed tasks or gesture condition) compared to when they are presented alone (VTs – verbal tasks or no-gesture condition). The participants in the experiment were invited to watch and listen to a series of videos. From our assumptions derives the prediction that the participants recall better the sentences accompanied by gestures than those uttered by the speaker without gesturing. To sum up, gestures should improve memory performance in the listener.

Method

Participants

The participants in the experiment were 28 students at the University of Turin (8 males and 19 females, mean age 22.9 years, SD = 2.4). They took part in the experiment voluntarily, in exchange of course credits. All the participants gave their written consent prior to their participation in the study.

Material and procedure

The experimental material consisted of the videos constructed from the assertions from the normative study. The participants were randomly assigned to Protocol 1 and Protocol 2. The experiment has a block design: in one block each participant encountered 12 sentences in the gesture condition (actress gestures while uttering the sentences) and in the second block they encountered the other 12 sentences in the no-gesture condition (the actress keeps her hands still). The order of presentation of the two blocks was counterbalanced in each protocol.

The experiment was in a single session and took place at the sole presence of the experimenter, who invited the participant to sit down on a seat in front of a desk where a computer was placed (approximately at 7 inches from the desk's border). The instructions in both conditions were as follows: "Thanks for your participation and for your time. Your task in the experiment is to carefully watch and listen to a series of videos in which an actress utters a series of sentences representing actions. At the end of the last video, when the word 'Now' will appear on the screen, repeat as accurately as you can the sentences you heard. The order of the recollections doesn't matter, and you can repeat two times the same sentence. The word 'Now' will remain on the screen for 90 seconds, that is the time you have at disposal for your free recall. After 90 seconds the word 'End' will appear on the screen".

Coding of the recollections

Classically, the codings adopted in the literature include recollections consistent in meaning with the original sentences, in which the participants recalled the gist of the sentences with some missing details (e.g., Feyereisen, 2006). We distinguished correct recollections in two categories:

- *Literal recollection*: a sentence recalled exactly in its literality.
- *Paraphrase*: a sentence recalled using different words or different prepositions, but with the same meaning of the original sentence. Examples are sentences recalled paraphrasing some elements: plural/singular (*wringing out the cloth* instead of *wringing out the clothes*), article (e.g., *hammering the nail into the wall* instead of *hammering a nail into the wall*), verb (e.g., *hurling a stone* instead of *throwing a stone*).

All the other types of recollection were considered errors:

- *Erroneous recollection*: a recollection inconsistent in meaning with any of the original sentences, e.g., the sentence *opening the fridge*, absent from the list of sentences, or the sentence *cleaning the car* which is a blending of two original sentences, *cleaning the window* and *driving the car*.

Results and discussion

Three participants whose number of correct recollections was 2 standard deviations above or below the mean, in at least one experimental condition, were removed from the analysis. Hence, we present the results for the remaining twenty-five participants. Table 1 illustrates the mean number of recollections per typology in the two experimental conditions.

	Correct	Errors
Gesture condition	8.6 (*1.6*) Literal: 5.4 (*1.9*) Paraphrase: 3.2 (*1.9*)	0.04 (*0.2*)
No-gesture condition	7.5 (*2.0*) Literal: 5.5 (*2.3*) Paraphrase: 2.0 (*1.7*)	0.3 (*0.5*)

Table 1. Means (and standard deviations) for each type of recollection in the gesture and no-gesture conditions of Experiment 1 (N = 25).

We verified whether observing gestures at encoding phase had an impact on participants' recollections using mixed-effect logistic regression model implemented with the glmer() function from the lme4 package (version 1.1-12; Bates, Maechler, Bolker, & Walker, 2015) in the R statistical programming environment (version 3.3.1; R Development Core Team, 2016). We carried out three different models: the first one with correct recollections (literal + paraphrases) as the outcome of the model, and two further models, one for literal recollections and one for paraphrases, considered separately.

These models included Gesture (Gesture vs. No Gesture) as fixed factor of interest and Subjects and Item (i.e., sentences) as crossed random effects, thereby simultaneously taking into account differences among participants and among items. Following the current guidelines in the psycholinguistics literature, in all models we started including the maximal structure of random effects supported by the design (Barr, Levy, Scheepers & Tily, 2013). In particular, given the within-subjects and the within-item design (we had multiple observations per treatment level both per subjects and item), we included random intercepts for both participants and items, as well as random by-subjects and random by-item random slopes for the fixed effect (Barr et al., 2013). The resulting model specification was as follows: `Recalled ~ Gesture + (1+Gesture|Subjects) + (1+Gesture|Item).`

Although we started including the maximal structure of random effects supported by the design, the model with the correct recollections as the outcome failed to converge and in this case we used a model that did not estimate the correlations of the random effects, as these terms have negligible influence on results and a maximal model with no random correlations or even missing within unit random intercepts is preferable to one missing the critical random slopes (Barr et al., 2013). In this case the resulting model specification was as follows: `Recalled ~ Gesture + (1|Subjects) + (0+Gesture|Subjects) + (1|Item) + (0+Gesture|Item).` The models with paraphrases and literal recollections as outcomes models converged with the full random-effects specification.

We detected a significant effect of Gesture on correct recollections ($\beta = .47$, $SE = .20$, $z = 2.35$, $p = .02$); the same results held for paraphrases ($\beta = .91$, $SE = .34$, $z = 2.68$, $p = .007$), but not for literal recollections ($\beta = .06$, $SE = .19$, $z = .34$, $p = .73$).

As regards the erroneous recollections, since an error is not related to a specific item (it does not refer to a sentence in the list) we could not simultaneously modeling crossed participants and item effects by using linear mixed models. For this reason we report a plain paired-samples comparison between the conditions: errors were fewer in the gesture than in the no-gesture condition (Wilcoxon test: $z = 2.12$, $p = .03$; *Cliff's* $\delta = .24$).

As predicted, the results of Experiment 1 revealed that the accuracy in recollection was greater in the gesture than in the no-gesture condition. Interestingly, the paraphrases but not the literal recollections were greater in the gesture condition compared to the no-gesture condition. This result is consistent with our assumption that gestures favor the construction of a discourse mental model, with the consequence to loose memory for the surface form (i.e., memory for literality). Further, the results revealed that errors were fewer in the gesture compared to the no-gesture condition. Although this result was not predicted, it is in line with the assumption that gestures accompanying speech increase the quality of the speech's mental model.

Experiment 2. Eliciting continuous harms' movements during gestures' observation cancels the enactment effect

The aim of the experiment was to test a prediction deriving from our assumptions: if the beneficial effect of gestures relies on the activation of motor areas in the observer, specifically those areas that would be active if the observers themselves would gesture with the same effectors used by the speaker to gesture, then the effect should disappear when the observers, while watching the videos with the speaker gesturing, accomplish a motor task with the same effectors used by the speaker. The task of the participants in the experiment was identical to the task in Experiment 1, with the exception that in the observation phase of both the gesture and the no-gesture condition the participants were invited to move their arms and hands, namely the same effectors moved by the actress in the video. In particular, the participants were invited to perform not repetitive movements; the rationale was to engage them in a continuous planning requiring an engagement of the premotor resources. We predicted the disappearance of the enactment effect.

Method

Participants

The participants in the experiment were 28 adults, students at the University of Turin (13 males and 15 females, mean age 23.9 years, *SD* = 3.9). They took part in the experiment voluntarily after written informed consent.

Material and procedure

The material was the same of the first experiment. We used also the same procedure with just an addition in the instruction: "During the observation of these videos, starting with your hands placed on your knees, please alternately touch with your index fingers the two marks placed on the table in front of the computer while watching the video clips. It is important that your movements be continuous and alternate. Start with an arm's movement (left or right), only after the other hand has come back on the knee". This task was meant both to avoid that the participants imitated the observed action and to engage their premotor resources. The experimenter made same examples of possible movements and specified that the movements should stop at the end of the presentation of the last video. The experimenter was sitting down behind the participants, but in a position (approximately at 45 degrees) where he was able to check and keep track of the correct performance of the motor secondary task. We coded the recollections as in Experiment 1.

Results and discussion

All the participants performed correctly the motor task. One participant whose number of correct recollections was 2 standard deviations above the mean, in both experimental condition, was removed from the analysis. Hence, we present the results for the remaining twenty-seven participants. Table 2 illustrates the mean scores for types of recollection in the two experimental conditions.

	Correct	Errors
Gesture condition	7.4 (*1.9*) Literal: 4.4 (*1.7*) Paraphrase: 3.0 (*1.6*)	0.4 (*0.7*)
No-gesture condition	7.1 (*1.4*) Literal: 5.0 (*1.6*) Paraphrase: 2.1 (*1.5*)	0.3 (*0.5*)

Table 2. Means (and standard deviations) for each type of recollection in the gesture and no-gesture conditions of Experiment 2 (N = 27).

As for Experiment 1, we carried out three different mixed-effect logistic regressions, each one with a different outcome (correct recollections, literal recollections and paraphrases). All these models included Gesture (Gesture vs. No Gesture) as fixed factor of interest and Subjects and Item (i.e., sentences) as crossed random effects. We started including the maximal structure of random effects supported by the design. However, the models with literal recollections and paraphrases recollections as the outcomes failed to converge. In these cases we used a model that did not estimate the correlations of the random effects. In line with our prediction, all types of recollection occurred to the same extent in the gesture and the no-gesture conditions: we didn't detect a significant effect of Gesture on correct recollections ($\beta = .11$, $SE = .24$, $z = .46$, $p = .65$), literal recollections ($\beta = .22$, $SE = .20$, $z = 1.11$, $p = .27$) and paraphrases ($\beta = .41$, $SE = .23$, $z = 1.78$, $p = .07$).

Further, the number of erroneous recollections was comparable in the gesture and the no-gesture conditions (Wilcoxon test: $z = .66$, $p = .51$; *Cliff's* $\delta = .05$).

The results of Experiment 2 reveal that the enactment effect disappears when the participants, during the observation phase, perform a motor secondary task involving the same effectors moved by the actress in the videos. However, the presence of the two markers on the table may have diverted the visual attention of the participants to the actress' gestures. Hence, the absence of the enactment effect could be ascribed to a decreased visual attention for the actress' gestures rather than to an overload of the observer's motor system. In order to exclude this alternative explanation of the results we carried out a new experiment, without the two markers on the table.

Experiment 3. A replication of Experiment 2 with a different motor secondary task

The task of the participants in the experiment was identical to the task in Experiment 2, with the exception that in the observation phase they were asked to touch alternately with their hands a casual point on the table. We predicted the disappearance of the enactment effect.

Method

Participants

The participants in the experiment were 32 adults, students of the University of Turin (12 males and 20 females, mean age 23.7 years; *SD* = 3.0). They took part in the experiment voluntarily, in exchange of course credits, and after informed consent.

Material and procedure

The material was the same of Experiment 2. We used also the same procedure but with different instructions: "During the observation of these videos, starting with your hands placed on your knees, please alternately touch with your index fingers a casual point on the table in front of the computer while watching the video clips. It is important that your movements be continuous and alternate. Start with an arm's movement (left or right), only after the other hand has come back on the knee". We coded the recollections as in Experiment 2.

Results and discussion

We excluded from the analyses three of the participants, who didn't perform correctly the secondary task. In particular, one of them remained still and didn't perform any harms' movements, another one didn't perform continuous movements and one didn't come back with his hands on the knees. Also, two participants whose number of correct recollections was 2 standard deviations above or below the mean, in at least one experimental condition, were removed from the analysis. Hence, we present the results for the remaining twenty-seven participants. Table 3 illustrates the mean scores for types of recollection in the two experimental conditions.

	Correct	Errors
Gesture condition	7.6 (*1.3*) Literal: 4.7 (*1.2*) Paraphrase: 2.9 (*1.4*)	0.3 (*0.5*)
No-gesture condition	7.1 (*1.5*) Literal: 4.8 (*1.7*) Paraphrase: 2.3 (*1.2)*	0.3 (*0.5*)

Table 3. Means (and standard deviations) for each type of recollection in the gesture and no-gesture conditions of Experiment 3 (N = 27).

As for Experiments 1 and 2 we carried out three different mixed-effect logistic regressions each one with a different outcome (correct recollections, literal recollections and paraphrases). All these models included Gesture (Gesture vs. No Gesture) as fixed factor of interest and Subjects and Item (i.e., sentences) as crossed random effects. All the three models converged with the full random-effects specifications.

In line with our prediction we didn't detect a significant effect of Gesture on correct recollections (β = .16 , *SE* = .18, *z* = .91, *p* = .36), literal recollections (β = .05 , *SE* = .18, *z* = .28, *p* = .78) and paraphrases (β = .27 , *SE* = .21, *z* = 1.31, *p* = .19).

Further, we didn't find a significant statistical difference between the amount of errors in the gesture and no-gesture conditions (Wilcoxon test: *z* = .26, *p* = .80; *Cliff's* δ = .00).

The results of Experiment 3 suggest that a motor dual task involving the same effectors involved in the observed actions (hands and arms) cancels the enactment effect. The beneficial effect of observing gestures disappears when the observer is involved in a motor secondary task, suggesting a pivotal role of the observer's motor system. However, at this point we are not able to discern whether the enactment effect disappears as a result of whatever motor secondary task or whether the absence of the beneficial effect depends by the specific motor task we used. Experiment 4 was designed to exclude this alternative explanation of our results.

Experiment 4. Eliciting repetitive legs' movements during gestures' observation does not cancel the enactment effect

The primary task of the participants in the experiment was the same as in Experiments 2 and 3, but the motor secondary task was different. The task involved different effectors respect to those used by the actress in the videos, specifically the participants' legs and feet. Since we know that the activation deriving from action observation follows a somatotopic organization, the interference of the motor secondary task should be specific to the part of body involved in the motor task: we predict that when the observer moves effectors different from those observed in the videos, the enactment effect persists.

Method

Participants

The participants in the experiment were 28 adults, students at the University of Turin (11 males and 17 females, mean age 24.3 years, *SD* = 3.8). They took part voluntarily in the experiment. All the participants had given their written consent prior to their participation in the study.

Material and procedure

The material and the procedure were identical to those of Experiments 2 and 3. The secondary task, as well as the relative instructions, were different: "During the observation of these videos, starting from the sitting position, please alternately stretch your legs in front of you and touch with only your heels the floor while watching the video clips. It is very important that your movements are continuous and alternate. Start with a leg's movement (left or right) only after the other leg has come back to the starting position". The experimenter showed the motor movements requested and specified that this motor secondary task should stop at the end of the observation of the last video.

Results

One participant whose number of correct recollections was 2 standard deviations belove the mean in one experimental condition was removed from the analyses. Hence, we present the results for the remaining twenty-seven participants.

Table 4 illustrates the mean scores for types of recollection in the experimental conditions.

	Correct	Errors
Gesture condition	7.8	0.1
	(*1.8*)	(*0.4*)
	Literal: 5.0 (*1.7*)	
	Paraphrase: 2.8 (*1.6*)	
No-gesture condition	6.9	0.6
	(*1.7*)	(*0.6*)
	Literal: 5.1 (*1.9*)	
	Paraphrase: 1.7 (*1.2*)	

Table 4. Means (and standard deviations) for each type of recollection in the gesture and no-gesture condition of Experiment 4 (N = 27).

As for Experiments 1, 2 and 3 we carried out three different mixed-effect logistic regressions each one with a different outcome (correct recollections, literal recollections and paraphrases). All these models included Gesture (Gesture vs. No Gesture) as fixed factor of interest and Subjects and Item (i.e., sentences) as crossed random effects. We started including the maximal structure of random effects supported by the design. However, the model with the correct recollections as the outcome failed to converge. In this case we used a model that did not estimate the correlations of the random effects.

We detected a significant effect of Gesture on correct recollections ($\beta = .33$, $SE = .17$, $z = 1.98$, $p < .05$); the same results held for paraphrases ($\beta = .64$, $SE = .30$, $z = 2.13$, $p = .03$), but not for literal recollections ($\beta = .08$, $SE = .17$, $z = .46$, $p = .65$).

Further, errors in the gesture condition were fewer than in the no-gesture condition (Wilcoxon test: $z = 2.40$, $p < .02$; *Cliff's* δ = .xx).

The global results of the present investigation enforce the assumption that the motor system plays a crucial role in the enactment effect; the participants' memory for speech improved when they observed gestures, but not when during observation their motor system was loaded up through a motor task involving the same effectors used by the speaker.

General discussion and conclusions

We assumed that a speaker's gestures improve memory for speech in the listeners through the exploitation of their motor system. We also assumed that gestures favor the construction of a mental model of the material to be learnt in that they convey motoric information that are coded in the model at a procedural level. Indeed, a mental model contains both declarative knowledge (the so called "knowing that", a set of concepts expressed in the form of propositions) and procedural knowledge (the so called “knowing how”). In other words, in a mental models framework our memory employs more than one format of knowledge representation (e.g., visuo-spatial, motoric). Since gestures activate and reinforce the motoric information contained in a mental model, they play a significant role in comprehension and learning. These assumptions, along with evidence in the literature that when people observe an action their motor system is automatically and somatotopically activated, motivated our investigation on the role of the motor system in the beneficial effect of the gestures observed on memory for action sentences (i.e., the enactment effect in EPTs, namely in experimenter performed tasks, compared to VTs, namely plain verbal tasks in which the speaker does not accompany the sentences with gestures).

We carried out a series of behavioral experiments to falsify the predictions implied by our assumptions. In particular, we focused on the beneficial effect of gestures in EPTs and from our assumptions we derived the prediction that EPTs, as compared to VTs, favor memory for actions sentences through the exploitation of the listener’s motor system. In Experiment 1 we replicated the classical enactment effect for our list of Italian action sentences: the participants’ recollection of the sentences was more accurate in the EPTs condition compared to the VTs condition. The sentences were used in all the subsequent experiments. An intriguing result was that the greater accuracy of the participants in the gesture condition was due to a greater number of paraphrases.

From our assumptions derives the prediction that a motor dual task involving the same effectors involved in the observed gestures (in our case, hands and arms) should erase the usual enactment effect. The results of Experiments 2 and 3 confirmed the prediction: recollections in form of paraphrases occurred to the same extent in the gesture and the no-gesture conditions. On the other hand, from our assumptions also derives the prediction that a motor dual task involving different effectors from those involved in the observed gestures (in our case, legs and feet) should not erase the usual enactment effect. The results of Experiment 4 confirmed the prediction: recollections in the form of paraphrases were greater in the gesture compared to the no-gesture condition.

The results of Experiments 2 and 3 are consistent with an fMRI study by Straube, Green, Weis and Chatterjee (2009); the task of the participants was to observe a large set of sentences accompanied by either metaphoric gestures congruent with the abstract meaning of the verbal message, or incongruent gestures or no gestures, and later on they were tested for sentence memory. The activations in the left inferior frontal gyrus (IFG), the premotor cortex (PMC) and the middle temporal gyrus (MTG) positively correlated with memory performance for sentences presented with congruent gestures, but not for sentences presented with incongruent gestures or no gestures. These findings suggest that the activation of the gesture-speech neural network is a predictor of effective memory consolidation.

Further, the results of Experiments 2, 3 and 4 are consistent with those by Ping and colleagues (2014). The authors employed a motor secondary task similar to the one used in our experiments, with the aim to ascertain whether the listener's motor system is involved in gesture's understanding. The results of their experiments suggest that the motor system plays a crucial role in the incorporation of the information coming from the speaker's hands' gestures in the mental representation of the speaker's message, thereby suggesting that the listener's motor system is involved in gestures understanding (see also Michael et al., 2014). Our proposed theoretical framework can subsume Ping and colleagues' assumptions, and leads to the prediction that the listener's motor system is involved also in memory for action sentences; gestures convey procedural knowledge that favors the construction of a mental model of the information conveyed by the speech, thus favoring memory for speech. In particular, our assumption that gestures favor the construction of a mental model of the information in the speech implies that starting from such model it is possible to use at recall a variety of linguistic descriptions, which paraphrase the sentence originally accompanied by the gestures.

Our proposal according to which the observation of gestures favors the construction of a mental model through the exploitation of the motor system is consistent with the assumption that viewing gestures induces in the listeners the simulation of the corresponding actions and it can accommodate the results of a study by Cook and Tanenhaus (2008). In line with our assumption that gestures reflect the procedural knowledge of a mental model, they argue that gestures can be of help to speakers in communicating information about how to perform actions in the world – for example, explaining how to tie one's shoes or ride a bicycle. The authors invited some of the participants in their study to solve the Tower of Hanoi task either on real object or on a computer, then to explain to other participants, the listeners, how they solved the task. The results revealed that the speakers expressed information about the particular actions that they had performed in their gesture, and that listeners were sensitive to this action information in gesture. In other words, speakers' hand gestures reflected their procedural knowledge of the task, demonstrating that gesture can be a vehicle for the expression of information that is unlikely to be expressed in the accompanying speech. The results are also consistent with our assumption that the gestures of a speaker favor in the listener the construction of a mental model of the speech information.

More in general, the mental model account is consistent with the assumption that gestures derive from simulated actions or perceptual states ("Gestures as simulated action" framework; e.g., Hostetter & Alibali, 2008). However, we envisage two main differences.

First, Hostetter and Alibali (2008) suggest that gestures derive from "mental images" instead of "mental models" (see also Cook & Tenenhaus, 2008). They argue that a series of evidences suggest that gestures reflect and stem from speaker's mental images. For each evidence we have a different interpretation, consistent with well-established findings in the literature on the inferential processes involved in comprehension and reasoning. For example, Hostetter and Alibali argue that images help in certain cognitive tasks, such as mental rotation task, and takes this evidence as enforcing their assumption that the beneficial role of gestures is due to their tight relation with mental images. However, contrary to what the classical studies on mental images would suggest, studies in the literature revealed that mental images can be an obstacle to the inferential processes (see, e.g., Knauff & Johnson-Laird, 2002). Further, Hostetter and Alibali argue that gestures are global and synthetic in the sense that they do not rely on analytic rules, like mental images, and that people often use their bodies to gesture when they describe mental images. However, it has to be noticed that mental models, as well as images, are iconic mental representations and that people often gesture also when they run kinematic mental models (Bucciarelli, Mackiewicz, Khemlani, & Johnson-Laird, 2016). Further, evidence interpreted in favor of the connection gesture-images is their spatial common nature; however, this does not demonstrate that gestures stem from imagistic representations, because also models are spatial in nature, and gestures could stem from mental models.

Second, our assumptions on the role of mental models in speech comprehension and their detrimental effect on memory for discourse verbatim is consistent with a vast literature on discourse comprehension, whereas we ignore the existence of theoretical frameworks that can accommodate this evidence. In particular, the detection of the enactment effect in free recall tasks in terms of paraphrases, but not in terms of literal recollections, is a result in line with findings on co-speech gestures revealing that they favor the deep comprehension of the speech they accompany at the expense of poor recognition for verbatim (e.g., Cutica & Bucciarelli, 2008; 2013; Cutica et al., 2014). In particular, it is well-established that deep comprehension relies on the construction of a "mental model" (Johnson-Laird, 1983, 2006) or "situation model" (van Dijk & Kintsch, 1983; an extension is the Construction Integration Model of Comprehension - Kintsch, 1998); the two terms can be considered equivalent, disregarding their different theoretical roots (see also Kaup, Kelter, & Habel, 1999). Enriched models of the discourse lead to a poor retention of the surface form of the text (see Garnham, Oakhill, & Cain, 1998; Johnson-Laird & Stevenson, 1970). Obviously, a distinction should be made between the effect of co-speech gestures on memory for sentences and on memory for discourse. In the case of co-speech gestures accompanying connected sentences (i.e., discourse) co-speech gestures favor the construction of a mental model of the discourse, as revealed by a great production of discourse-based inferences at recall. In the case of gestures accompanying a sentence, instead, only the paraphrase of the sentence can be considered an index of model's construction. Future studies might further explore the possibility that the enactment effect exploits the construction of mental models through a recognition task, the complementary task necessary to conclude that gesture favor recall in form of paraphrases at the expense of poor recognition for verbatim.

Although we assumed that mental models are the mental representations incorporating the information coming from the speaker's hands' gestures, the main focus of the present investigation was on the role played by the listener's motor system in the beneficial effect of the gestures observed rather than on the nature of the mental representations resulting from this process. Future studies might be specifically devised in order to investigate more in depth the nature of such mental representations.

Globally considered, the results of our four experiments enforce our assumption that the listener's motor system plays a pivotal role in the EPTs enactment effect; they are consistent with the claim that the visual inputs presented in the EPTs condition activate "covert motor representations in the absence of any task demands" (Wilson, 2002, p. 631). Also, the results suggest that the motor system has a causal role in the subsequent beneficial effect on memory. In other words, the motor coding involved in observing gestures is particularly efficient for encoding information into memory and loading up the listener's motor system interferes with this process.

Since we have ascertained the role of the motor system in the enactment effect, future studies could shed light on the specific moment in which the motor component plays a main role. According to our results, the listener's motor system plays a crucial role for the *formation* of the memory trace. Indeed, we disrupted the beneficial effect of gestures by interfering with the listener's motor system during the observation phase. Moving the dual task after the encoding phase but before recall, or during recall, could help us to understand whether the motor component is crucial also for the *consolidation* or the *retrieval* of the memory trace, respectively.

An interesting open question is whether the motor system is involved also in the beneficial effect of gestures that accompany an entire discourse (e.g., Cutica & Bucciarelli, 2008). This possibility is consistent with the results of a study by Skipper, Goldin-Meadow, Nusbaum and

Small (2007). The participants in their study listened to entire stories (adapted Aesepo's Fables and lasting approximately 45-50s) while watching meaningful speech-associated gestures, speech-irrelevant gestures (self-grooming) or no hand/arm movements. Skipper and colleagues detected the activations of the supramarginal gyrus (SMG), the ventral premotor cortex (vPMC) and the dorsal premotor cortex (dPMC) in the meaningful speech-associated gestures condition. Since these areas, reflect the activity of the human mirror system they conclude that the system is involved in the extraction of semantic information from the co-speech gestures. It is possible that the speaker's arms and hands gestures, while making a discourse or telling a story, improve memory in the listener who stays still or moves her legs and feet (i.e., different motor effectors from those moved by the speaker), but it does not improve memory in the listener who moves her harms and hands (i.e. the same motor effectors moved by the speaker). Future studies might explore this possibility, along with the possibility that individuals suffering from motor hyperactivity might not benefit from the observation of a speaker's gestures.

Appendix

The sentences selected for the experiments proper (translated from Italian) and their mean rating in the normative study.

Rowing a boat 5.7
Conducting an orchestra 5.2
Playing the violin 5.1
Dribbling with a basketball 5.0
Playing the piano 4.9
Cleaning a window 4.7
Driving the car 4.6
Painting a painting 4.4
Ironing a shirt 4.4
Beating eggs 4.3
Wringing out the clothes 4.3
Throwing a stone 4.2
Getting shampoo 4.2
Polishing silver 4.0
Hammering a nail into the wall 4.0
Brushing the teeth 4.0
Creaming the body 4.0
Laying some blocks one above another 3.9
Sewing by hand 3.7
Typing 3.7
Hugging someone 3.7
Shooting with the gun 3.5
Rolling up the ball of yearn 3.5
Sharpening a knife 3.5

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, *15* , 593-613.

Bara, B. G. (1995). *Cognitive science: A developmental approach to the simulation of the mind*. Psychology Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structures for confitmatory hypothesis testing: Keep it maximal. *Journal of Memory & Language, 68*, 255-278.

Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. A. (2003). Social embodiment. *Psychology of Learning and Motivation, 43*, 43-92.

Bates, B., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

Bucciarelli, M. (2007). How the construction of mental models improves learning. *Mind & Society*, *6*, 67-89.

Bucciarelli, M., Mackiewicz, R., Khemlani, S.S., & Johnson-Laird, P.N. (2016). Children's creation of algorithms: Simulations and gestures. *Journal of Cognitive Psychology*, *28*, 297-318.

Buccino, G., Binkofski, F., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., & Freund, H.J. (2001). Action observation activates premotor and parietal areas in somatotopic manner: An fMRI study. *European Journal of Neuroscience*, *13*, 400-404.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage*, *50*, 1148-1167.

Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *International Journal of Bilingual Education and Bilingualism*, *7*, 303-319.

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology*, *22*, 267-281.

Cohen, R. L., & Bryant, S. (1991). The role of duration in memory and metamemory of enacted instructions (SPTs). *Psychological Research*, *53*, 183-187.

Cook, S. W., & Goldin-Meadow, S. (2006). The role of gesture in learning: Do children use their hands to change their minds? *Journal of Cognition and Development*, *7*, 211-232.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*, 1047-1058.

Cook, S. W., & Tanenhaus, M. K. (2008). Embodied communication: Speakers' gestures affect listeners' actions. *Cognition, 113*, 98-104.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of Memory & Language, 63*, 465-475.

Core Team, R. (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>.

Cutica, I., & Bucciarelli, M. (2015). Non-determinism in the uptake of gestural information. *Journal of Nonverbal Behavior*, *39*, 289-315.

Cutica, I., & Bucciarelli, M. (2008). The deep versus the shallow: Effects of co-speech gestures in learning from discourse. *Cognitive Science*, *32*, 921-935.

Cutica, I., & Bucciarelli, M. (2011). "The more you gesture, the less I gesture": Co-speech gestures as a measure of mental model quality. *Journal of Nonverbal Behavior*, *35*, 173-187.

Cutica, I., & Bucciarelli, M. (2013). Cognitive change in learning from text: Gesturing enhances the construction of the text mental model. *Journal of Cognitive Psychology*, *25*, 201-209.

Cutica, I., Ianì, F., & Bucciarelli, M. (2014). Learning from text benefits from enactment. *Memory & Cognition*, *42*, 1026-1037.

Engelkamp, J. (1998). *Memory for actions*. Psychology Press/Taylor & Francis (UK).

Engelkamp, J. (2001). Action memory: A system-oriented approach. In H. D., Zimmer, R. L., Cohen, M. J., Guynn, J., Engelkamp, R., Kormi-Nouri, & M. A., Foley (Eds.), *Memory for action: A distinct form of episodic memory?* (pp. 49-96). New York: Oxford University Press.

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 671-682.

Engelkamp, J., & Jahn, P. (2003). Lexical, conceptual and motor information in memory for action phrases: A multi-system account. *Acta Psychologica*, *113*, 147-165.

Engelkamp, J., & Zimmer, H. D. (1983). Zum einflu–s von wahrnehmen und tun auf das behalten von verb–objekt-phrasen. *Sprache & Kognition*, 2, 117-127.

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica*, *96*, 43-60.

Engelkamp, J., Zimmer, H. D., & Biegelmann, U. E. (1993). Bizarreness effects in verbal tasks and subject-performed tasks. *European Journal of Cognitive Psychology*, *5*, 393-415.

Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology*, *18*, 185-205.

Feyereisen, P. (2009). Enactment effects and integration processes in younger and older adults' memory for actions. *Memory*, *17*, 374-385.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593-6.

Garnham, A., Oakhill J., & Cain K., (1998). Selective retention of information about the superficial form of text: Ellipses with antecedents in main and subordinate clauses. *The Quarterly Journal of Experimental Psychology: Section A, 51*, 19–39.

Goldin-Meadow, S. (1999). The role of gestures and communication in thinking. *Trends in Cognitive Sciences, 3*, 419-429.

Goldin-Meadow, S., & Alibali, M. W. (1999). Does the hand reflect implicit knowledge? Yes and no. *Behavioral and Brain Sciences, 22*, 766-767.

Goldin-Meadow, S., & Alibali, M.W. (2013). Gestures' role in speaking, learning, and creating language. *Annual Review of Psychology, 123*, 448-453.

Goldin-Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science*, *15*, 876-884.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163-189.

Hildebrandt, B., Moratz, R., Rickheit, G., & Sagerer, G. (1999). 10 Cognitive modelling of vision and speech understanding. *Advances in Psychology*, *128*, 213-236.

Hornstein, S. L., & Mulligan, N. W. (2004). Memory for actions: Enactment and source memory. *Psychonomic Bulletin & Review*, *11*, 367-372.

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

Ianì, F., Cutica, I., & Bucciarelli, M. (2016). Timing of gestures: Gestures anticipating or simultaneous with speech as indexes of text comprehension in children and adults. *Cognitive Science*. doi: 10.1111/cogs.12381

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Johnson-Laird, P. N. (2006). *How we reason*. Oxford University Press, USA.

Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction*. Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P. N., & Stevenson, R. (1970). Memory for syntax. *Nature, 227*, 412.

Kaup, B., Kelter, S., & Habel, C. (1999). Taking the functional aspect of mental models as a starting point for studying discourse comprehension. In G. Rickheit & C. Habel (Eds.), *Mental models in discourse processing and reasoning* (pp. 93–112). New York: Elsevier.

Kelly, S. D., Manning, S. M., & Rodak, S. (2008). Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education. *Language and Linguistics Compass*, *2*, 569-588.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*, 313-334.

Kendon, A. (1983). Gesture and speech: How they interact. In J. M. Weimann & R. E Harrison (Eds.), *Sage annual reviews of communication research: Nonverbal interaction* (Vol. 11, pp. 13-45). Beverly Hills, CA: Sage.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Knopf, M., Mack, W., Lenel, A., & Ferrante, S. (2005). Memory for action events: Findings in neurological patients. *Scandinavian Journal of Psychology*, *46*, 11-19.

Kormi-Nouri, R. (1995). The nature of memory for action events: An episodic integration view. *European Journal of Cognitive Psychology*, *7*, 337-363.

Kormi-Nouri, R., & Nilsson, L. G. (2001). The motor component. In Zimmer, H. D. Cohen, R. L. Guynn, M. J. Engelkamp, J. Kormi-Nouri, R. & M. A. Foley (Eds.), *Memory for action: A distinct form of episodic memory?* (pp. 49-96). New York: Oxford University Press.

Kormi-Nouri, R., Nyberg, L., & Nilsson, L. G. (1994). The effect of retrieval enactment on recall of subject-performed tasks and verbal tasks. *Memory & Cognition, 22*, 723-728.

Kush, D., Lidz, J., & Phillips, C. (2015). Relation-sensitive retrieval: Evidence from bound variable pronoums. *Journal of Memory and Language*, *82*, 18-40.

Lotze, M., Heymans, U., Birbaumer, N., Veit, R., Erb, M., Flor, H., & Halsband, U. (2006). Differential cerebral activation during observation of expressive gestures and motor acts. *Neuropsychologia*, *44*, 1787-1795.

Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education*, *5*, 196-211.

Madan, C. R., & Singhal, A. (2012). Using actions to enhance memory: effects of enactment, gestures, and exercise on human memory. *Frontiers in Psychology*, *3*, 507-511.

Mangels, J. A., & Heinberg, A. (2006). Improved episodic integration through enactment: Implications for aging. *The Journal of General Psychology, 133*, 37-65.

Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition*, *10*, 181-187.

Masumoto, K., Yamaguchi, M., Sutani, K., Tsuneto, S., Fujita, A., & Tonoike, M. (2006). Reactivation of physical motor information in the memory of action events. *Brain Research*, *1101*, 102-109.

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

Michael, J., Sandberg, K., Skewes, J., Wolf, T., Blicher, J., Overgaard, M., & Frith, C. D. (2014). Continuous theta-burst stimulation demonstrates a causal role of premotor homunculus in action understanding. *Psychological Science*, 963-972.

Mohr, G., Engelkamp, J., & Zimmer, H. D. (1989). Recall and recognition of self-performed acts. *Psychological Research*, *51*, 181-187.

Molander, B., & Arar, L. (1998). Norms for 439 action events: Familiarity, emotionality, motor activity, and memorability. *Scandinavian Journal of Psychology*, *39*, 275-300.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 615-622.

Mulligan, N. W., & Hornstein, S. L. (2003). Memory for actions: Self-performed tasks and the reenactment effect. *Memory & Cognition*, *31*, 412-421.

Nilsson, L. G., Nyberg, L., Klingberg, T., Åberg, C., Persson, J., & Roland, P. E. (2000). Activity in motor areas while remembering action events. *NeuroReport*, *11*, 2199-2201.

Nishitani, N., & Hari, R. (2000). Temporal dynamics of cortical representation for action. *Proceedings of the National Academy of Sciences*, *97*, 913-918.

Noice, H., & Noice, T. (2007). The non-literal enactment effect: Filling in the blanks. *Discourse Processes, 44*, 73-89.

Norris, M. P., & West, R. L. (1993). Activity memory and aging: The role of motor retrieval and strategic processing. *Psychology and Aging*, *8*, 81-86.

Nyberg, L., Petersson, K. M., Nilsson, L. G., Sandblom, J., Åberg, C., & Ingvar, M. (2001). Reactivation of motor brain areas during explicit memory for actions. *Neuroimage, 14*, 521-528.

Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology*, *44*, 1277-1287.

Ping, R. M., Goldin-Meadow, S., & Beilock, S. L. (2014). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General*, *143*, 195-204.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology*, *210*, 419-421.

Rizzolatti, G., & Craighero, L. (2005). Mirror neuron: a neurological approach to empathy. In *Neurobiology of human values* (pp. 107-123). Springer Berlin Heidelberg.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131-141.

Russ, M. O., Mack, W., Grama, C. R., Lanfermann, H., & Knopf, M. (2003). Enactment effect in memory: evidence concerning the function of the supramarginal gyrus. *Experimental Brain Research*, *149*, 497-504.

Sakreida, K., Schubotz, R. I., Wolfensteller, U., & von Cramon, D. Y. (2005). Motion class dependency in observers' motor areas revealed by functional magnetic resonance imaging. *The Journal of Neuroscience*, *25*, 1335-1342.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca’s area, and the human mirror system. *Brain and Language*, *101*, 260-277.

Straube, B., Green, A., Weis, S., & Chatterjee, A. (2009). Memory effects of speech and gesture binding: cortical and hippocampal activation in relation to subsequent memory performance. *Journal of Cognitive Neuroscience*, *21*, 821-836.

Thompson, L. A., Driscoll, D., & Markson, L. (1998). Memory for visual-spoken language in children and adults. *Journal of Nonverbal Behavior*, *22*, 167-187.

van Dijk, I. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*, 564-584.

Wheaton, K. J., Thompson, J. C., Syngeniotis, A., Abbott, D. F., & Puce, A. (2004). Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. *Neuroimage*, *22*, 277-288.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625-636.

Zimmer, H. D. (2001). Why do actions speak louder than words: Action memory as a variant of encoding manipulations or the result of a specific memory system. In H.D. Zimmer, R.L. Cohen, M.J. Guynn, J. Engelkamp, R. Kormi-Nouri, & M.A. Foley (Eds.), *Memory for action: A distinct form of episodic memory?* (pp. 49-96). New York: Oxford University Press.