PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

rstb.royalsocietypublishing.org

Review

Cite this article: Özyürek A. 2014 Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Phil. Trans. R. Soc. B* **369**: 20130296. http://dx.doi.org/10.1098/rstb.2013.0296

One contribution of 12 to a Theme Issue 'Language as a multimodal phenomenon: implications for language learning, processing and evolution'.

Subject Areas:
cognition, language

Keywords:
co-speech gestures, semantics, iconicity, brain, multimodal language

Author for correspondence:
Aslı Özyürek
e-mail: asliozu@mpi.nl

Hearing and seeing meaning in speech and gesture: insights from brain and behaviour

Aslı Özyürek[1,2]

[1]Department of Linguistics, Radboud University Nijmegen, Erasmus Plain 1, 6500 HD, Nijmegen, The Netherlands
[2]Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525 JT, The Netherlands

As we speak, we use not only the arbitrary form–meaning mappings of the speech channel but also motivated form–meaning correspondences, i.e. iconic gestures that accompany speech (e.g. inverted V-shaped hand wiggling across gesture space to demonstrate walking). This article reviews what we know about processing of semantic information from speech and iconic gestures in spoken languages during comprehension of such composite utterances. Several studies have shown that comprehension of iconic gestures involves brain activations known to be involved in semantic processing of speech: i.e. modulation of the electrophysiological recording component N400, which is sensitive to the ease of semantic integration of a word to previous context, and recruitment of the left-lateralized frontal–posterior temporal network (left inferior frontal gyrus (IFG), medial temporal gyrus (MTG) and superior temporal gyrus/sulcus (STG/S)). Furthermore, we integrate the information coming from both channels recruiting brain areas such as left IFG, posterior superior temporal sulcus (STS)/MTG and even motor cortex. Finally, this integration is flexible: the temporal synchrony between the iconic gesture and the speech segment, as well as the perceived communicative intent of the speaker, modulate the integration process. Whether these findings are *special* to gestures or are shared with actions or other visual accompaniments to speech (e.g. lips) or other visual symbols such as pictures are discussed, as well as the implications for a multimodal view of language.

1. Introduction

Since the 1960s and 1970s, research on signed languages has begun to demonstrate clearly that natural languages of deaf communities, even though executed on a very different modality, share many aspects of linguistic structure with spoken languages (e.g. [1,2]) and even recruit brain areas similar to those involved in processing of spoken languages [3]. Since then, our notion of the world's languages has been extended and now comprises two classes, signed and spoken languages, based on the modality through which communicative messages are transmitted: visual–manual versus auditory–vocal.

However, in the past decade, it has become clear that this simple modality distinction does not capture the fundamental multimodal complexity of the human language faculty, especially those of 'spoken' languages [4]. All spoken languages of the world *also* exploit the visual–manual modality for communicative expression and speakers accompany speech with gestures of the hands, face, and body as articulators [5–8]. Kendon [9] defines gestures as visible actions of the hand, body and face that are intentionally used to communicate and are expressed together with the verbal utterance. Co-speech gestures can display semiotic complexity of different types (e.g. points, demonstrations of objects and events (as in so-called iconic gestures)), have different communicative functions (e.g. emphasis, disambiguation and speech acts) and vary in their semantic relation to speech (e.g. conveying redundant or complementary information). Speakers point to the entities they refer to with speech, use iconic gestures as

Royal Society Publishing

they move the fingers of an inverted V-hand in a wiggling manner while saying 'he walked across', use bodily demonstrations of reported actions as they tell narratives, convey different viewpoints of events or use gesture spaces indexing different levels of discourse cohesion parallel to marking similar discourse devices found in speech (e.g. [10,11]). Thus, there has been mounting evidence at the production level that co-speech gestures contribute semantic, syntactic, discursive and pragmatic information to the verbal part of an utterance, forming composite utterances with semiotic diversity [6–8,12]. What is semantically conveyed in gesture can even be specific to the typology of the spoken language (e.g. [13]). Furthermore, speakers in producing composite utterances are sensitive to the temporal overlap of the information conveyed in co-speech gesture and the relevant speech segment they utter [7,14,15].

Research on gestures and their relation to speech has focused mostly on a subset of gestures called iconic or depictive gestures that represent objects and events by bearing partial resemblance to them [7,16]. Much of the capacity of iconic gestures for signification derives from 'perceptual, motoric and analogic mappings that can be drawn between gestures and the conceptual content they evoke' [17, p. 184]. As such, iconic gestures have different representational properties from speech in terms of the meaning they convey. They represent meaning as a whole, not as a construction made out of separate, analytical meaningful components as in speech (or as in sign). Consider, for example, an upward hand movement in a climbing manner when a speaker says: 'the cat climbed up the tree'. Here, the gesture depicts the event as a whole, describing manner ('climb') and direction ('up') simultaneously, whereas in speech the message unfolds over time, broken up into smaller meaningful segments (i.e. different words for manner and direction). Nevertheless, the two modalities convey a unified meaning representation achieved by semantic relatedness and temporal congruity between the two [7]. Note that the relations of iconic gestures and speech are at the level of semantics, due to their formal resemblance to the objects and events they represent. As such, they differ from other visual accompaniments to speech such as lips, where there is a form (but not meaning) matching between lip movements and syllables, and head and eyebrow movements or other hand gestures such as 'beats', meaningless forms of hand movements that are used to increase the prominence of certain aspects of speech or regulate interactions. These will be left out of this review (see [18,19]).

If speakers employ such multimodal utterances where information conveyed in both speech and gesture are semantically and temporally aligned with each other, how do speakers/listeners comprehend them? After all, gestures themselves are not very informative and fuzzy in the absence of speech (i.e. unlike pictures or other informative actions). In this selective review, I will present research regarding whether and how listeners/viewers process the information from co-speech gestures (specifically from iconic gestures) and speech, including behavioral and neurobiological data. This review shows first of all that iconic gestures are processed semantically and that they evoke similar markers of online neural processing and recruit overlapping brain areas to those found in the processing of semantic information from speech. Second, when gestures are viewed in speech context (i.e. accompanying speech), they do not seem to be processed independently but their processing interacts with that of speech. This is evidenced through priming measures, online neural recordings and activations in brain areas known to be sensitive to unification of meaning and crossmodal interactions in the brain. Finally, the interactions between the two modalities further seem to be sensitive to the temporal synchrony of the two channels as well as to the perceived communicative intent of the speakers, and thus seem to be flexible rather than obligatory depending on the communicative context.

2. Co-speech gesture comprehension: behavioral and neural markers of semantic processing

It has been a long-standing finding that addressees pick up information from gestures that accompany speech [20]. That is, gestures are not perceived by comprehenders simply as handwaving or as attracting attention to what is conveyed in speech. Listeners/viewers pay attention to iconic gestures and pick up the information that they encode. For example, Kelly *et al.* [21] showed participants video stimuli where gestures conveyed additional information to that conveyed in speech (gesture pantomiming drinking while speech is 'I stayed up all night') and asked them to write what they heard. In addition to the speech they heard, participants' written text contained information that was conveyed only in gesture but not in speech (i.e. 'I stayed up drinking all night'). In another study, Beattie & Shovelton [22] showed that listeners answer questions about the size and relative position of objects in a speaker's message more accurately when gestures were part of the description and conveyed additional information than speech. McNeill *et al.* [23] presented listeners with a videotaped narrative in which the semantic relationship between speech and gesture was manipulated. It was found that listeners/viewers incorporated information from the gestures in their retellings of the narratives and attended to the information conveyed in gesture when that information complemented or even contradicted the information conveyed in speech (see also [24,25]). Thus, listeners pick up the semantic information conveyed in gesture.

Further research has shown that gestures also show semantic priming effects. For example, Yap *et al.* [26] has shown that iconic gestures—shown without speech—(highly conventionalized ones such as flapping both hands on the side meaning bird) prime sequentially presented words.

More evidence for the view that gestures are analysed for meaning comes from studies investigating online processing of co-speech gestures using electrophysiological recordings (event-related potentials, ERPs). These studies focused on the N400 effect known to be responsive to meaningful stimuli. Kutas & Hillyard [27] were the first to observe for words that, relative to a semantically acceptable control word, a sentence-final word that is semantically anomalous in the sentence context, as in 'He spread the warm bread with sock', elicits an N400 effect, a negative-going deflection of the ERP waveform between 300 and 550 ms poststimulus with an enhanced amplitude for incongruous words compared with congruent ones. Additional studies have shown that a semantic violation is not required to elicit an N400 effect. In general, N400 effects are triggered by more or less subtle differences in the semantic fit between the meaning of a word and its context, where the context can be a single word, a sentence or a discourse (e.g. see [28] for a review). A series of studies have shown that gestures used without speech can also evoke similar N400 effects.

Wu & Coulson [29] found that semantically incongruous gestures (shown without speech) were presented after cartoon images elicited a negative-going ERP effect around 450 ms, in comparison to gestures that were congruent with the cartoon image. Furthermore, unrelated words followed by gestures (shown without their accompanying speech) also elicited a more negative N400 than related words [30]. Wu and Coulson interpreted these findings as showing that iconic gestures are subject to semantic processes 'analogous' to those evoked by other meaningful representations such as pictures and words. Wu & Coulson [17] have shown that ERPs for static pictures of gestures (instead of dynamic ones), as well as objects proceeded by matching and mismatching contexts, elicit an N300 effect. (Willems *et al*. [31], however, did not find such a specific effect for pictures' integration to previous sentence context.)

Holle & Gunter [32] extended the use of the ERP paradigm to investigate the semantic processing of gestures in a speech context. They asked whether manual gestures presented earlier in the sentence could disambiguate the meaning of an otherwise ambiguous word presented later in the sentence and investigated the brain's neural responses to this disambiguation. An EEG was recorded as participants watched videos of a person gesturing and speaking simultaneously. The experimental sentences contained an unbalanced homonym in the initial part of the sentence (e.g. *She controlled the* <u>*ball*</u>...) and were disambiguated at a target word in the subsequent clause (*which during the* <u>*game*</u>... versus *which during the* <u>*dance*</u>...). Coincident with the homonym, the speaker produced an iconic gesture that supported either the dominant or the subordinate meaning. ERPs were time-locked to the onset of the target word. The N400 to target words was found to be smaller after a congruent gesture and larger after an incongruent gesture, suggesting that listeners can use the semantic information from gesture to disambiguate upcoming speech.

In another ERP study, Özyürek *et al*. [33] examined directly whether ERPs measured as a response to semantic processing evoked by iconic gestures are comparable to those evoked by words. This ERP study investigated the integration of co-speech gestures and spoken words to a previous sentence context. Participants heard sentences in which a critical word was accompanied by a gesture. Either the word or the gesture was semantically anomalous with respect to the previous sentence context. Both the semantically anomalous gestures and anomalous words to previous sentence context elicited identical N400 effects, in terms of the latency and the amplitude.

Using an functional magnetic resonance imaging (fMRI) method, Straube *et al*. [34] have attempted to isolate the brain's activation in response to iconic gestures to see whether it overlaps with areas involved in processing verbal semantics. fMRI measures brain activity by detecting associated changes in blood flow (i.e. blood-oxygen-level-dependent (BOLD) response), relying on the fact that blood flow and neural activation are coupled. In this study, they compared the brain's activation triggered by meaningful spoken sentences (S+) with sentences from an unknown language (S−), and they also compared activation for co-speech gestures presented without their accompanying speech (G+), and meaningless gestures also without speech (G−). Meaningful iconic gestures activated left inferior frontal gyrus (IFG), bilateral parietal cortex and bilateral temporal areas. The overlap of activations for meaningful speech and meaningful gestures occurred in the left IFG and bilateral medial temporal gyrus (MTG). These findings are consistent with another study by Xu *et al*. [35] showing that left IFG and posterior medial temporal gyrus (MTG) are involved in the comprehension of communicative gestures (i.e. pantomimes such as opening a jar without speech) as well as speech glosses of the same gestures (i.e. open jar) presented separately.

These studies show that iconic gestures, seen without or within a speech context, are analysed for meaning and the brain's neural responses to iconic gestures display similarities to that of speech comprehension. Further research is needed to show whether gestures are special in the way they are processed in the brain and different from the activations observed for actions, pictures or other meaningful representations.

3. Interactions between speech and gesture comprehension

While the above studies have focused on the nature of the processing of iconic gestures, other studies further investigated how comprehenders bring together the semantic information gleaned from the two modalities into a coherent and integrated semantic representation. Are gestures initially processed independently of what is conveyed in speech or are there bidirectional interactions between semantic processing of speech and gestures, in that independent processing of each does not occur? In a priming study [36], participants were presented with action primes (e.g. someone chopping vegetables) followed by bimodal speech and gesture targets. They were asked to press a button if what they heard in speech or gesture depicted the action prime (figure 1*a*). Participants related primes to targets more quickly and accurately when they contained congruent information (speech: 'chop'; gesture: chop) than when they contained incongruent information (speech: 'chop': gesture: twist). Moreover, the strength of the incongruence between overlapping speech and gesture affected processing, with fewer errors for weak incongruities (speech: 'chop'; gesture: cut) than for strong incongruities (speech: 'chop': gesture 'open'). This indicates that in comprehension, the relative semantic relations between the two channels are taken into account, providing evidence against independent processing of the two channels (figure 1*b*). Furthermore and crucially, this effect was bidirectional and was found to be similar when either speech or gesture targets matched or mismatched the action primes. That is, gesture influenced processing of speech processing and speech influenced processing of gesture.

Gestures' influence on accompanying speech was also detected in online measures of comprehending speech. Kelly *et al*. [37] found that ERPs to spoken words (targets) were modulated when these words were accompanied by gestures (primes) that contained information about the size and shape of objects that the target words referred to (e.g. tall, wide, etc.). Compared to matching target words, mismatching words evoked an early P1/N2 effect, followed by an N400 effect, suggesting an influence of gesture on spoken words, first at the level of 'sensory/phonological' processing and later at the level of semantic processing.

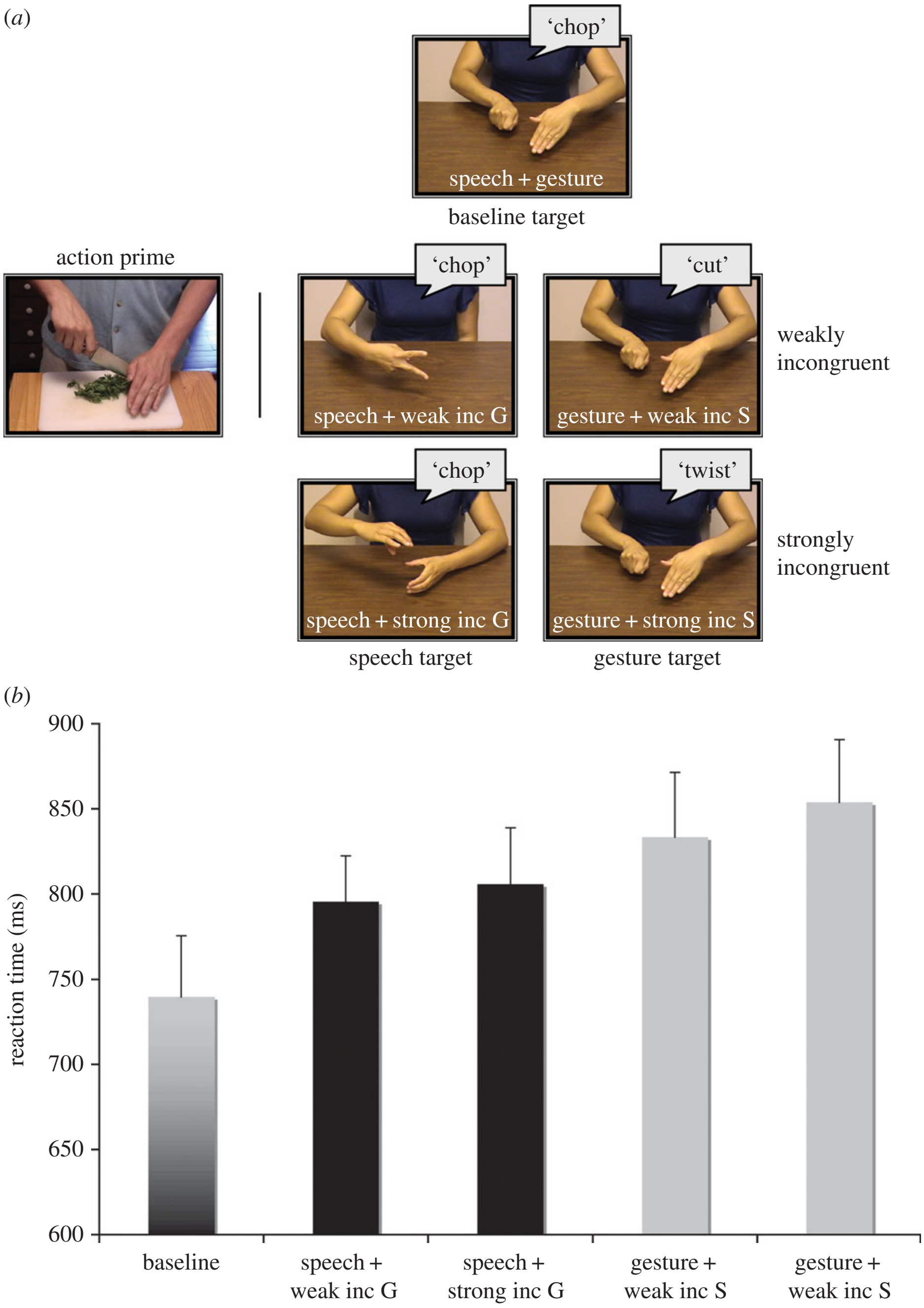

Figure 1. (*a*) Action primes, speech–gesture targets and congruency manipulation between the channels (*b*) Proportion of errors with decrease in semantic overlap between speech and gesture, shown separately for speech and gesture targets [36]. Error bars show the standard errors. inc G, incongruent gesture; inc S, incongruent speech.

Further fMRI studies have attempted to locate the brain areas involved in integrating information from speech and gesture. In order to locate areas of integration between the two modalities, they compared multimodal stimuli to unimodal ones, gestures coupled with degraded to those with clear speech or manipulated the semantic relations between the two channels (i.e. speech and gesture match, mismatch or complement each other). Even though these studies find left frontal and left posterior temporal cortices to be implicated in integrating gestures with speech, they vary with respect to whether they consistently find co-speech gesture-related activation in the following regions: left IFG, bilateral posterior superior temporal sulcus (STSp) and middle temporal gyrus (MTGp) [34,38–49]. Interestingly, these are the areas that are also involved when increased semantic processing is required during speech comprehension. Studies examining increased semantic processing (i.e. ambiguity, mismatch, etc.) in spoken language alone as well as studies examining co-speech gesture in the context of speech have found activities in similar brain regions, as illustrated in figure 2 [39]. Furthermore, the temporal areas, especially STS that are sensitive to speech–gesture integration, are also known to be implicated in integration of other types of multimodal stimuli such as lips and syllables (e.g. [18]).

However, complete consensus has not been achieved concerning the nature of the participation of these brain regions in gesture–speech integration. The contribution of left IFG to semantic integration of speech and gesture was first reported by Willems *et al*. [48]. In that study, participants heard sentences in which a critical word was accompanied by a gesture (the same stimuli as in [33] were used). Either the word or the gesture could be semantically anomalous with respect to the context

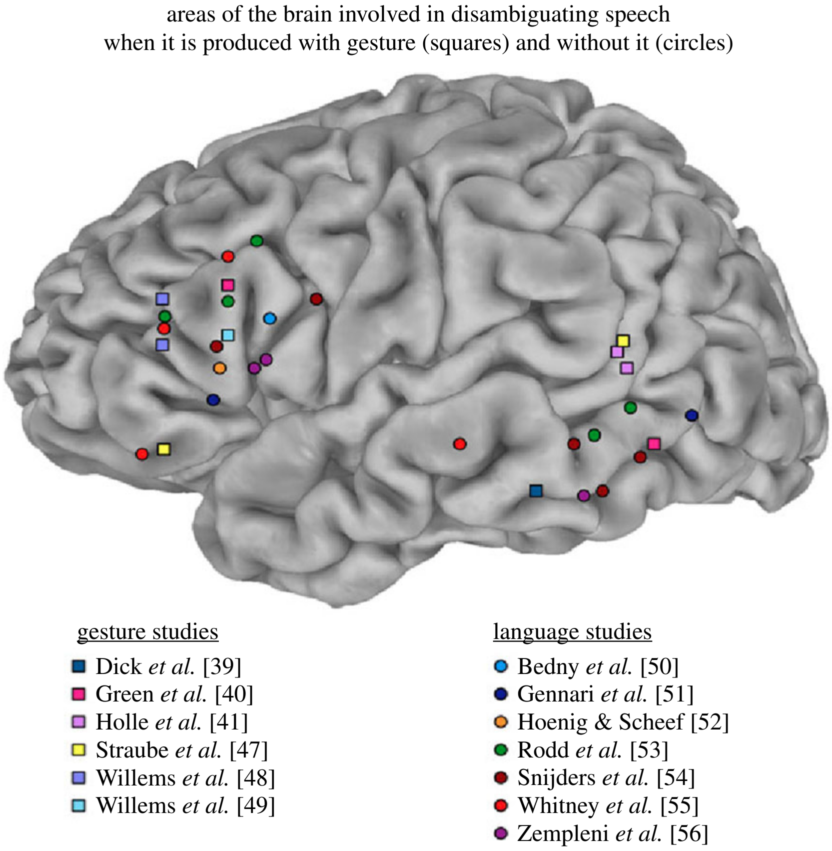

Figure 2. Overlap of areas in left hemisphere of the brain activated sensitive to processing of co-speech gestures in the context of speech (gesture studies) as well as to increased semantic processing of speech alone (language only studies) [57].

set-up by the sentence, with anomalous (incongruent) gestures demanding more semantic processing. Incongruent conditions involving either the word or the gesture elicited greater activity than congruent conditions in left IFG, pars triangularis. Similarly, Willems *et al.* [49] reported increased activation in left IFG for incongruent co-speech gestures compared with congruent co-speech gestures, using naturalistic (rather than acted out) co-speech gestures as stimuli.

Left IFG has also been found to respond more strongly to metaphoric gestures, that is, gestures with abstract meaning (e.g. a 'high' gesture accompanying speech like 'the level of presentation was high'), compared with iconic gestures accompanying the same speech ([47]; also see [43]). These results indicate that gestures that carry more semantic load due to their metaphoric content activate left IFG (and not just gestures that are incongruent with speech). In the Straube *et al.* [47] study, iconic gestures as well as grooming movements, even though used as control, activated left IFG, when compared with no movement. Dick *et al.* [39] also found left IFG to be sensitive to meaning modulation by iconic gestures; that is, more activation in this area for complementary (speech: 'I worked all night'; gesture: type) than redundant gestures accompanying speech (speech: 'I typed all night'; gesture: type). Complementary gestures add information and require more semantic processing than redundant gestures. Finally, Skipper *et al.* [44] found that when hand movements (iconic gestures) were related to the accompanying speech, left IFG (pars triangularis and pars opercularis) exhibited a weaker influence on other motor- and language-relevant cortical areas compared with when the hand movements were meaningless (i.e. grooming gestures or 'self-adaptors') or when there were no accompanying hand movements.

Thus, left IFG is responsive to increased semantic processing load of integration of iconic gestures to speech: that is, when gestures are difficult to integrate into the previous or overlapping co-speech context (in the case of incongruent gestures) and for metaphoric or complementary iconic gestures that require more semantic processing compared with gestures that simply convey redundant or similar information to that in speech.

Researchers have also examined the role of posterior temporal regions—STSp and MTGp in particular—in the semantic integration of gesture and speech. While MTG has been more frequently found to be involved in speech and gesture integration, the role of STS has been more controversial. Holle *et al.* [41] was the first to suggest that activity in STSp reflects sensitivity to the semantic integration of gesture and speech. In this study, STSp (but not left IFG) was more active for ambiguous words (dominant or subordinate homonyms such as *mouse*) accompanied by meaningful iconic gestures than to speech accompanied by non-meaningful grooming movements. This result was replicated in a second study in which brain activations to iconic action gestures coupled with action speech that conveyed similar information were compared in two situations: where speech was degraded versus not degraded [42]. STS was more active when gestures accompanied degraded speech compared with clear speech. However, not all studies have found greater activation in STSp that reflected a specific role in semantic integration of speech and gesture. For example, Dick *et al.* [39] did not find activation in this area when the semantic relation of the gesture to speech was manipulated (i.e. complementary versus redundant to speech; see also [48,49] for lack of activation of STSp). Dick *et al*. [39] argued that STSp is not involved in semantic integration *per se* but may be involved in connecting information from the visual and auditory modalities in general.

A stronger consensus has been achieved with regard to activation of left and/or right MTGp, which is anatomically close to STSp, in relation to semantic integration of speech and gesture. For example, Green *et al*. [40] found that, in German speakers, left MTGp responded more strongly to sentences accompanied by unrelated gestures (hard to make sense of in relation to speech) than to the same sentences accompanied by related gestures. Dick *et al*. [39] also found this area to be sensitive to complementary gestures, in comparison to redundant gestures.

Willems *et al*. [49] found that the left and right MTGp (as well as left STSp) responded more to speech accompanied by incongruent pantomimes (conventionalized actions with objects such as ironing, twisting, etc., the meaning of which would be clear without speech) than to the same speech accompanied by congruent pantomimes. However, MTGp was not activated for incongruent speech–co-speech gesture pairs (gestures that are ambiguous without speech; i.e., hands moving back and forth in an undefined manner in co-speech gesture while speech is 'I packed up my clothes') compared with congruent pairs. Incongruent speech–gesture pairs activated only left IFG and not MTGp. On the basis of these findings, the authors suggest that bilateral MTG is more likely to be involved in matching two input streams for which there is a relatively stable common object representation (i.e. 'twist' in speech with a twisting gesture). This idea is parallel to the notion that both the sight of a dog and the sound of its barking form part of a representation of our knowledge about dogs [58]. By contrast, when integration of gesture and speech requires a new and unified representation of the input streams, the increased semantic processing of iconic gestures results in the increased activation of left IFG (e.g. [59]). Note that at this point, these characterizations should be seen more as tendencies rather than exclusive functions of left IFG and MTG's contributions to speech and gesture integration at the semantic level (see [39] for further discussion).

Finally, while the above imaging studies have provided information about the use of speech and gesture integration at the macro-anatomical level (in terms of brain regions using event-related measurements), a recent study by Josse *et al*. [60] has used a repetition suppression paradigm to address this question at the neuronal level. This paradigm is based on the principle that repetition of the same stimuli is associated with a decrease in both neuronal activity and BOLD signal. In this study, subjects were first shown words alone and then words with congruent gestures as well as with the same words with incongruent gestures. While words with congruent gestures (speech: 'grasp'; gesture: grasp) have shown repetition suppression (i.e. decrease in activation) in relation to words alone, this suppression has not been observed when words were repeated with incongruent gestures (speech: 'grasp': gesture: sprinkle). Thus, the suppression effect shows that words and gestures activate the same neural population. The suppression effects were found in the 'dorsal' route of the brain in premotor cortex and the temporal–parietal areas (left and right STS), flagging these areas as major sites for speech and gesture integration and semantic processing, when both word and gesture tap into the same conceptual representation. These findings support the view that STS (as found in [41]) (as well as motor cortex) is also involved in matching two input streams for which there is a relatively stable common object representation, as found for MTG [49].

4. Interactions between speech and gesture: obligatory or flexible?

While the above-mentioned studies have shown interactions between speech and gesture in behavioural as well as in neural responses, some recent studies have tapped further into questions about to what extent this integration is obligatory and automatic or flexible. After all, spontaneous speech is not always accompanied by gestures; gestures might sometimes be asynchronous with the relevant speech segment [61], and the frequency or the informativeness of the representations in gestures can vary depending on the communicative nature of the situation (i.e. whether there is shared common ground between the listener and the addressee or not, etc. (e.g. [62]). Even though Kelly *et al*. [36] have argued that the interactions between speech and gesture are obligatory, some of his own work and that of others has shown that semantic processing from gestures as well as their interactions might be modulated depending on the level of synchrony between the channels and the perceived communicative intent of the speaker.

Habets *et al*. [63] investigated the degree of synchrony in speech and gesture onsets that is optimal for semantic integration of the concurrent gesture and speech. Videos of a person gesturing were combined with speech segments that were either semantically congruent or incongruent with the gesture. The onset of the gesture strokes (i.e. the meaningful part of the gesture, but not the preparation) and speech were presented with three different degrees of synchrony: a stimulus onset asynchrony (SOA) 0 condition (the gesture stroke onset and the speech onset were simultaneous) and two delayed SOAs, where speech was delayed by 160 ms (partial overlap with speech) or 360 ms (speech onset presented after gesture stroke was executed; no overlap between the two) in relation to the gesture stroke onset. ERPs time-locked to the speech onset showed a significant difference between semantically congruent versus incongruent gesture–speech combinations for the N400 component with SOAs of 0 and 160 ms, respectively, but not for the 360 ms SOA. Therefore, the closer speech and gesture are temporally to each other (or at least when some temporal overlap is possible), the more likely they are to be integrated with each other (figure 3). It is important to note that in this study, gestures used as stimuli, when viewed without speech, were ambigious, as they are in most co-speech gestures. Thus, mutual influence between speech and gesture is crucial (i.e. possible only with total or partial temporal overlap between the two) for speech and gesture integration to take place (Kelly *et al*. [36]).

Similar results were also found by Obermeier *et al*. [64] who used the same design as in the Holle & Gunter [32] study mentioned above and changed the temporal synchrony between the homonyms and gestures. He found that when gestures (actually gesture fragments used in this study) did not temporally overlap with the homonyms and when subjects were not explicitly asked to pay attention to gestures,

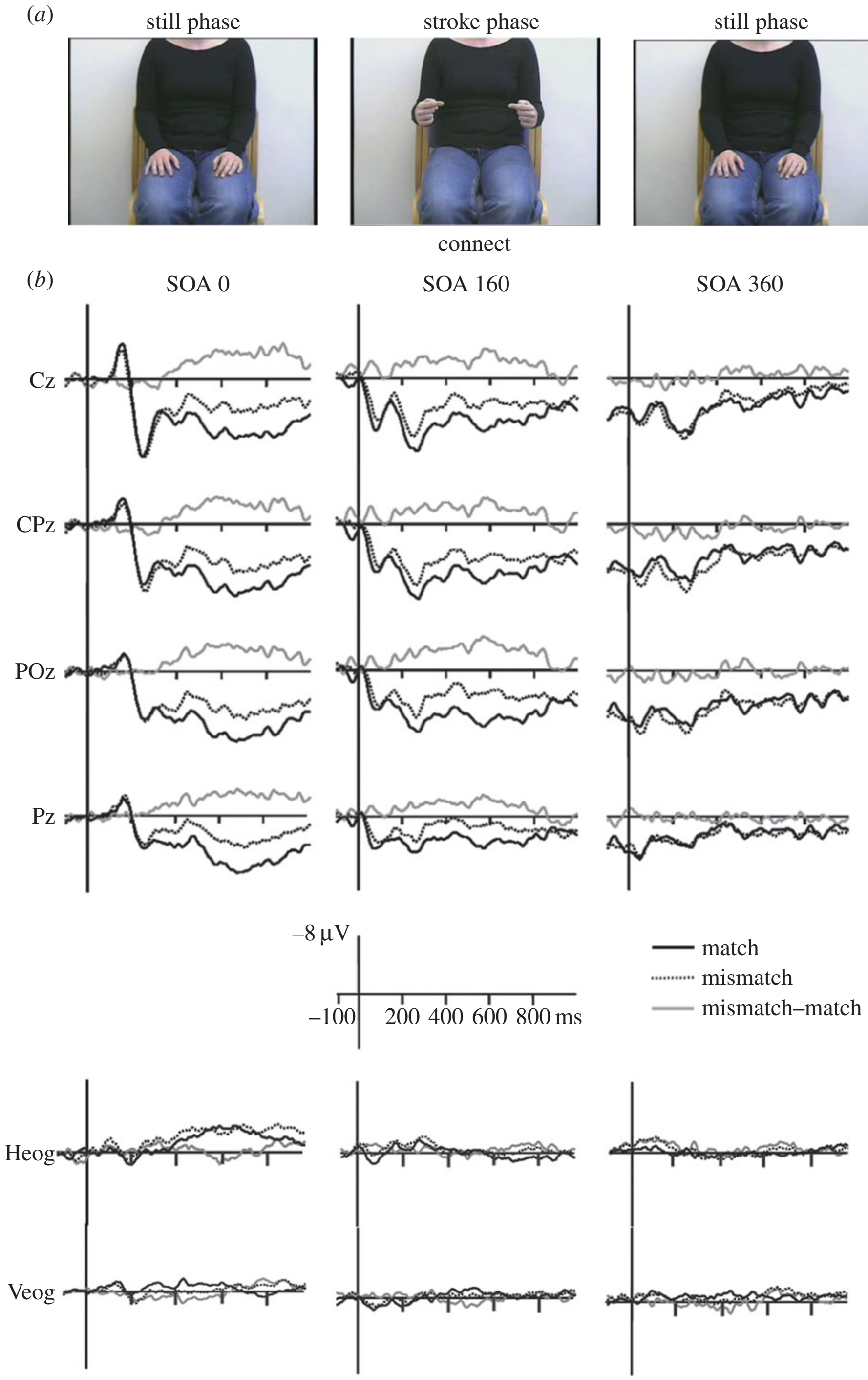

Figure 3. (*a*) Example of a gesture stroke and speech segment. (*b*) ERP waveforms time-locked to the onset of speech and gesture with different SOAs in relation to speech and different semantic relations between speech and gesture; match versus mismatch [63].

speech–gesture integration did not occur. Further research showed that when, for the same stimuli, participants are presented with degraded speech or had hearing impairments, gestures in the same asynchronous contexts *were* integrated with speech. This shows again that integration can be modulated by the aspects of the communicative situation [65].

Not only the asynchrony but also the perceived communicative intent of the speakers seems to modulate the speech–gesture integration or the semantic processing of gestures. ERP studies by Kelly *et al.* [66] have demonstrated that our brain integrates speech and gesture less strongly when the two modalities are perceived as not intentionally coupled (i.e. gesture and speech being produced by two different persons) than when they are perceived as being produced by the same person. In this study, adults watched short videos of gesture and speech that conveyed semantically congruous and incongruous information. In half of the videos, participants were told that the two modalities were intentionally coupled (i.e. produced by the same communicator), and in the other half, they were told that the two modalities were not intentionally coupled (i.e. produced by different communicators). When participants knew that the same communicator produced the speech and gesture, there was a larger bilateral frontal and central N400 effect to words that were semantically incongruous versus congruous with gesture. However, when participants knew that different communicators produced the speech and gesture—that is, when gesture and speech were not intentionally meant to go together—the N400 effect was present only in right-hemisphere frontal regions. The results demonstrate that pragmatic knowledge about the intentional relationship between gesture and speech modulates neural processes during the integration of the two modalities.

Finally, Holler *et al*. [67] has investigated how listeners/viewers comprehend speech–gesture pairs in a simulated triadic communication setting where the speakers' eye gaze is directed at them versus to another addressee (i.e. away from them). Participants were scanned (fMRI) while taking part in triadic communication involving two recipients and a speaker. The speaker uttered sentences that were accompanied by complementary iconic gestures (speech: 'she cleaned the house' gesture: mopping) or with speech only. Crucially, the speaker alternated her gaze direction towards or away from the participant in the experiment, thus rendering him/or in two recipient roles: addressed (direct gaze) versus unaddressed (averted gaze) recipient. 'Speech and gesture' utterances, but not 'speech only' utterances, produced more activity in the right MTG, one of the brain areas found consistently involved in speech–gesture integration, when participants were addressed than when not addressed. Thus, when the eye gaze of the speaker is averted away from the listener/viewer, indexing decrease in the perception of communicative intent, integration of the two channels and/or semantic processing gesture might be reduced (also see [68] for similar effects shown by behavioural measures).

5. Summary and conclusion

Even though so-called 'spoken' languages are traditionally characterized as auditory–vocal languages (as opposed to 'signed languages' that are visual–gestural [69]), they are essentially multimodal in nature and *also* exploit the visual–gestural modality for communicative expression, as well as other non-manual visual articulators such as lips, face, eye gaze or head movements. This review shows that at least a subset of these gestures, the iconic gestures that convey semantic information by virtue of their form–meaning resemblance to the objects and events that they represent, are not perceived as mere incidental accompaniments to the speech channel (e.g. to increase attention to speech or contribute to the evaluation of the speaker [70]). They are processed semantically during comprehension and as an integrated part of the speaker's communicative message. Listeners/viewers do not perceive gestures automatically but take the communicative intent of the speaker into account, relying on other visible cues as such as eye gaze direction and also depending on their temporal synchrony with the speech channel. Thus, iconic gestures are processed as 'communicative' meaning representations.

An important conclusion of this review is that the brain areas (left IFG, bilateral MTG/STS) involved in the processing of iconic gestures with or without speech overlap with those brain areas that are also involved in processing semantic information from speech and higher level 'unification' processes of meaning [71]. Gestures activate similar brain areas to those involved in processing semantic information from speech (i.e. similar latency and amplitude of the N400). These areas (left IFG, MTG/STS) seem to be playing different roles in hearing and seeing meaning and are sensitive to different levels and types of semantic relations between the two modalities; for example, while left IFG is sensitive to the increase in the semantic load required to process iconic gestures and unification of new meaning representations, MTG is activated when similar information is conveyed in the two input streams. It is also important to point out that in some cases, right-hemisphere homologues of these areas have also been found, showing modality specificity of gestural representations (yet, currently it is not known whether different lateralization of these areas implies different processes; see [68], for some indications). Given that STG/S is known to be involved in audiovisual speech integration (e.g. lips/syllables [72]), this region may be engaged in the integration of gesture and speech at the audio–visual binding level, in addition to playing a possible role in meaning integration.

6. Are gestures special?

The parallels in brain activation for gesture and speech semantics do not necessarily mean that gestures are *special*, even though some have claimed that speech and gesture share the same communication system ([7,8,73] mostly based on production data). After all, their processing during comprehension shows overlaps with observing action, pictures or other meaningful representations that do not usually or necessarily coupled with speech (see [31,74–76] for reviews). However, studies directly comparing brain activations across different domains of meaningful representations and their integration with speech are lacking. Furthermore, the brain activations that are involved in speech and gesture integration show a lot of overlaps with those of other sound–meaning couplings (MTG) (e.g. sight of a dog–barking of a dog as in [58]), audiovisual integration such as between lips and syllables (STS) [18,72] and body motion light displays and speech [77], and integration of information from multiple non-linguistic sources such as world knowledge, speaker identity, etc. (left IFG) [28]. For instance, it is unclear whether crossmodal interactions at the form-matching level (lips/syllables) recruit similar areas to those in meaning-matching such as in speech and gesture, or how three-way interactions among these modalities occur. Finally, it is also crucial to find out whether processing of all crossmodal interactions between different channels of communication and other types of meaningful representations such as actions and pictures is modulated by temporal asynchronies or the perceived communicative intent or goal of the speaker or the listener. Answers to these will shed further light onto the differential roles that brain areas play and their domain specificity in understanding spoken languages as composite utterances that orchestrate multiple channels of communication. These will have also important implications for understanding information uptake in hearing impairments, cochlear implantation, second language learners and other communication disorders where gestures seem to help as alternative ways of communication, such as in autism, aphasia, etc.

Thus, as we gain a broader, more multimodal view on language and communication, it is becoming increasingly clear that visible meanings, the iconically motivated form–meaning mappings available through the affordances of our body for communicative expression, are an integral aspect of our language faculty; not only for signed but also for spoken languages [4,78,79].

Acknowledgement. I would like to thank Prof. Peter Hagoort for comments on this paper. Ideas in this paper were presented at the Vth Cognitive

Neuroscience Summer School, Lake Tahoe and were commented upon by Prof. Karen Emmorey. I would also like to thank Dr David Vinson for his valuable constructive ideas in an earlier version as an editor.

Funding statement. Some of the research conducted in this review was funded by Dutch Science Foundation, NWO grant and a Marie Curie Fellowship (255569).

References

1. Klima E, Bellugi U. 1979 *Signs of language*. Cambridge, MA: Harvard University Press.
2. Stokoe WC. 1960 Sign language structure: an outline of the visual communication systems of the American deaf. *Stud. Linguist. Occas. Pap.* **8**, 1–78.
3. Poizner H, Klima ES, Bellugi U. 1987 *What the hands reveal about the brain*. Cambridge, MA: MIT Press.
4. Kendon A. 2014 Semiotic diversity in utterance production and the concept of 'language'. *Phil. Trans. R. Soc. B* **369**, 20130293. (doi:10.1098/rstb.2013.0293)
5. Goldin-Meadow S. 2003 *Hearing hands: how our hands help us think*. Cambridge, MA: Harvard University Press.
6. Kendon A. 2004 *Gesture: visible action as utterance*. Cambridge, UK: Cambridge University Press.
7. McNeill D. 1992 *Hand and mind: what gestures reveal about thoughts*. Chicago, IL: University of Chicago Press.
8. McNeill D. 2005 *Gesture and thought*. Chicago, IL: University of Chicago Press.
9. Kendon A. 1980 Gesticulation and speech: two aspects of the process of utterance. In *The relationship of verbal and nonverbal communication* (ed. M Ritchie Key), pp. 207–227. The Hague, The Netherlands: Mouton.
10. Debreslioska S, Özyürek A, Gullberg M, Perniss PM. 2013 Gestural viewpoint signals referent accessibility. *Discourse Process.* **50**, 431–456. (doi:10.1080/0163853X.2013.824286)
11. So WC, Kita S, Goldin-Meadow S. 2009 Using the hands to identify who does what to whom: gesture and speech go hand-in-hand. *Cogn. Sci.* **33**, 115–125. (doi:10.1111/j.1551-6709.2008.01006.x)
12. Enfield NJ. 2009 *The anatomy of meaning: speech, gesture, and composite utterances*. Cambridge, UK: Cambridge University Press.
13. Kita S, Özyürek A. 2003 What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* **48**, 16–32. (doi:10.1016/S0749-596X(02)00505-3)
14. Church B, Kelly S, Holcomb D. 2013 Temporal synchrony between speech, action and gesture during language production. *Lang. Cogn. Process.* **29**, 345–354.
15. Loehr D. 2007 Aspects of rhythm in gesture and speech. *Gesture* **7**, 179–214. (doi:10.1075/gest.7.2.04loe)
16. Clark H. 1996 *Using language*. Cambridge, UK: Cambridge University Press.
17. Wu Y, Coulson S. 2011 Are depictive gestures like pictures? Commonalities and differences in semantic processing. *Brain Lang.* **119**, 184–195. (doi:10.1016/j.bandl.2011.07.002)
18. Campbell R. 2008 The processing of audio-visual speech: empirical and neural bases. *Phil. Trans. R. Soc. Lond. B* **363**, 1001–1010. (doi:10.1098/rstb.2007.2155)
19. Wagner P, Malisz Z, Kopp S. 2014 Speech and gesture in interaction: an overview. *Speech Commun.* **57**, 209–232. (doi:10.1016/j.specom.2013.09.008)
20. Graham JA, Argyle M. 1975 A cross-cultural study of the communication of extra-verbal meaning by gestures. *Int. J. Psychol.* **10**, 57–67. (doi:10.1080/00207597508247319)
21. Kelly SD, Barr D, Church RB, Lynch K. 1999 Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. *J. Mem. Lang.* **40**, 577–592. (doi:10.1006/jmla.1999.2634)
22. Beattie G, Shovelton H. 1999 Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica* **123**, 1–30. (doi:10.1515/semi.1999.123.1-2.1)
23. McNeill D, Cassell J, McCullough K-E. 1999 Communicative effects of speech-mismatched gestures. *Res. Lang. Soc. Interact.* **27**, 223–238. (doi:10.1207/s15327973rlsi2703_4)
24. Goldin Meadow S, Momeni Sandhofer C. 1999 Gestures convey substantive information about a child's thoughts to ordinary listeners. *Dev. Sci.* **2**, 67–74. (doi:10.1111/1467-7687.00056)
25. Singer MA, Goldin Meadow S. 2005 Children learn when their teacher's gestures and speech differ. *Psychol. Sci.* **16**, 85–89. (doi:10.1111/j.0956-7976.2005.00786.x)
26. Yap DF, So WC, Yap M, Tan YQ. 2011 Iconic gestures prime words. *Cogn. Sci.* **35**, 171–183. (doi:10.1111/j.1551-6709.2010.01141.x)
27. Kutas M, Hillyard SA. 1980 Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* **207**, 203–205. (doi:10.1126/science.7350657)
28. Hagoort P, Van Berkum JJA. 2007 Beyond the sentence given. *Phil. Trans. R. Soc. B* **362**, 801–811. (doi:10.1098/rstb.2007.2089)
29. Wu Y, Coulson S. 2005 Meaningful gestures: electrophysiological indices of iconic gesture comprehension. *Psychophysiology* **42**, 654–667. (doi:10.1111/j.1469-8986.2005.00356.x)
30. Wu Y, Coulson S. 2007 Iconic gestures prime related concepts: an ERP study. *Psych. Bull. Rev.* **14**, 57–63. (doi:10.3758/BF03194028)
31. Willems RM, Özyürek A, Hagoort P. 2008 Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *J. Cogn. Neurosci.* **20**, 1235–1249. (doi:10.1162/jocn.2008.20085)
32. Holle H, Gunter TC. 2007 The role of iconic gestures in speech disambiguation: ERP evidence. *J. Cogn. Neurosci.* **19**, 1175–1192. (doi:10.1162/jocn.2007.19.7.1175)
33. Özyürek A, Willems RM, Kita S, Hagoort P. 2007 On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cogn. Neurosci.* **19**, 605–616. (doi:10.1162/jocn.2007.19.4.605)
34. Straube B, Green A, Weis S, Kircher T. 2012 A supramodal neural network for speech and gesture semantics: an fMRI study. *PLoS ONE* **7**, e51207. (doi:10.1371/journal.pone.0051207)
35. Xu J, Gannon P, Emmorey K, Smith JF, Braun AR. 2009 Symbolic gestures and spoken language are processed by a common neural system. *Proc. Natl Acad. Sci.* **106**, 20 664–20 669. (doi:10.1073/pnas.0909197106)
36. Kelly SD, Özyürek A, Maris E. 2010 Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychol. Sci.* **21**, 260–267. (doi:10.1177/0956797609357327)
37. Kelly SD, Kravitz C, Hopkins M. 2004 Neural correlates of bimodal speech and gesture comprehension. *Brain Lang.* **89**, 253–260. (doi:10.1016/S0093-934X(03)00335-3)
38. Dick AS, Goldin-Meadow S, Hasson U, Skipper J, Small SL. 2009 Co-speech gestures influence neural responses in brain regions associated with semantic processing. *Human Brain Mapping* **30**(11), 3509–3526.
39. Dick AS, Goldin-Meadow S, Solodkin A, Small SL. 2012 Gesture in the developing brain. *Dev. Sci.* **15**, 165–180. (doi:10.1111/j.1467-7687.2011.01100.x)
40. Green A, Straube B, Weis S, Jansen A, Willmes K, Konrad K, Kircher T. 2009 Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. *Hum. Brain Map.* **30**, 3309–3324. (doi:10.1002/hbm.20753)
41. Holle H, Gunter TC, Rüschemeyer SA, Hennenlotter A, Iacoboni M. 2008 Neural correlates of the processing of co-speech gestures. *NeuroImage* **39**, 2010–2024. (doi:10.1016/j.neuroimage.2007.10.055)
42. Holle H, Obleser J, Rueschemeyer S-A, Gunter TC. 2010 Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage* **49**, 875–884. (doi:10.1016/j.neuroimage.2009.08.058)
43. Kircher T, Straube B, Leube D, Weis S, Sachs O, Willmes K, Konrad K, Green A. 2009 Neural interaction of speech and gesture: differential activations of metaphoric co-verbal gestures. *Neuropsychologia* **47**, 169–179. (doi:10.1016/j.neuropsychologia.2008.08.009)

44. Skipper JI, Goldin-Meadow S, Nusbaum HC, Small SL. 2007 Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang.* **101**, 260–277. (doi:10.1016/j.bandl.2007.02.008)
45. Skipper JI, Goldin-Meadow S, Nusbaum HC, Small SL. 2009 Gestures orchestrate brain networks for language understanding. *Curr. Biol.* **19**, 661–667. (doi:10.1016/j.cub.2009.02.051)
46. 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. *J. Cogn. Neurosci.* **21**, 821–836. (doi:10.1162/jocn.2009.21053)
47. Straube B, Green A, Bromberger B, Kircher T 2011 The differentiation of iconic and metaphoric gestures: common and unique integration processes. *Hum. Br. Mapp.* **32**, 520–533. (doi:10.1002/hbm.21041)
48. Willems RM, Özyürek A, Hagoort P. 2007 When language meets action: the neural integration of gesture and speech. *Cerebral Cortex* **17**, 2322–2333. (doi:10.1093/cercor/bhl141)
49. Willems RM, Özyürek A, Hagoort P. 2009 Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage* **47**, 1992–2004. (doi:10.1016/j.neuroimage.2009.05.066)
50. Bedny M, McGill M, Thompson-Schill SL. 2008 Semantic adaptation and competition during word comprehension. *Cerebr. Cortex* **18**, 2574–2585. (doi:10.1093/cercor/bhn018)
51. Gennari SP, MacDonald MC, Postle BR, Seidenberg MS. 2007 Context-dependent interpretation of words: evidence for interactive neural processes. *Neuroimage* **35**, 1278–1286. (doi:10.1016/j.neuroimage.2007.01.015)
52. Hoenig K, Scheef L. 2009 Neural correlates of semantic ambiguity processing during context verification. *Neuroimage* **45**, 1009–1019. (doi:10.1016/j.neuroimage.2008.12.044)
53. Rodd JM, Davis MH, Johnsrude IS. 2005. The neural mechanisms of speech comprehension: FMRI studies of semantic ambiguity. *Cerebr. Cortex* **15**, 1261–1269. (doi:10.1093/cercor/bhi009)
54. Snijders TM, Vosse T, Kempen G, Van Berkum JJA, Petersson KM, Hagoort P. 2009 Retrieval and unification of syntactic structure in sentence comprehension: an FMRI study using word-category ambiguity. *Cerebr. Cortex* **19**, 1493–1503. (doi:10.1093/cercor/bhn187)
55. Whitney C, Kirk M, O'Sullivan J, Lambon Ralph MA, Jefferies E. 2011 The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cerebr. Cortex* **21**, 1066–1075. (doi:10.1093/cercor/bhq180)
56. Zempleni MZ, Renken R, Hoeks JC, Hoogduin JM, Stowe LA. 2007 Semantic ambiguity processing in sentence context: evidence from event-related FMRI. *Neuroimage* **34**, 1270–1279. (doi:10.1016/j.neuroimage.2006.09.048)
57. Dick AS, Mok E, Raja Beharelle A, Goldin-Meadow S, Small SL. 2014 Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. *Hum. Br. Mapp.* **35**, 900–917. (doi:10.1002/hbm.22222)
58. Hein G, Doehrmann O, Muller NG, Kaiser J, Muckli L, Naumer MJ. 2007 Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci.* **27**, 7881–7887. (doi:10.1523/JNEUROSCI.1740-07.2007)
59. Hagoort P, Baggio G, Willems RM. 2009 Semantic unification. In *The cognitive neurosciences*, 4th edn (ed. MS Gazzaniga), pp. 819–836. Cambridge, MA: MIT Press.
60. Josse G, Joseph S, Bertasi E, Giraud A. 2012 The brain's dorsal route for speech represents word meaning: evidence from gesture. *PLoS ONE* **7–9**, e46108. (doi:10.1371/journal.pone.0046108)
61. Chui K. 2005 Temporal patterning of speech and iconic gestures in conversational discourse. *J. Pragmat.* **37**, 871–887. (doi:10.1016/j.pragma.2004.10.016)
62. Holler J, Wilkin K. 2009 Communicating common ground: how mutually shared knowledge influences the representation of semantic information in speech and gesture in a narrative task. *Lang. Cogn. Process.* **24**, 267–289. (doi:10.1080/01690960802095545)
63. Habets B, Kita S, Shao Z, Özyürek A, Hagoort P. 2011 The role of synchrony and ambiguity in speech–gesture integration during comprehension. *J. Cogn. Neurosci.* **23**, 1845–1854. (doi:10.1162/jocn.2010.21462)
64. Obermeier C, Holler H, Gunther T. 2011 What iconic gesture fragments reveal about gesture–speech integration when synchrony is lost: memory can help. *J. Cogn. Neurosci.* **23**, 1648–1663. (doi:10.1162/jocn.2010.21498)
65. Obermeier C, Dolk T, Gunther T. 2012 The benefit of gestures during communication: evidence from hearing and hearing-impaired individuals. *Cortex* **48**, 857–870. (doi:10.1016/j.cortex.2011.02.007)
66. Kelly SD, Ward S, Creigh P, Bartolotti J. 2007 An intentional stance modulates the integration of gesture and speech during comprehension. *Brain Lang.* **101**, 222–233. (doi:10.1016/j.bandl.2006.07.008)
67. Holler J, Kokal I, Toni I, Hagoort P, Kelly S, Özyürek A. 2014 Eye'm talking to *you*: speakers' gaze direction modulates co-speech gesture processing in the right MTG. *Soc. Cogn. Affect. Neurosci*. (doi:10.1093/scan/nsu047)
68. Holler J, Kelly S, Hagoort P, Özyürek A. 2012 When gestures catch the eye: the influence of gaze direction on co-speech gesture comprehension in triadic communication. In *Proceedings of the 34th annual meeting of the cognitive science society* (eds N Miyake, D Peebles, RP Cooper), pp. 467–472. Austin, TX: Cognitive Society.
69. Emmorey K, Özyürek A. In press. Language in our hands: neural underpinnings of sign language and cospeech gesture. In *Handbook of cognitive neuroscience*, 5th edn (ed. MS Gazzaniga). Cambridge, UK: MIT Press.
70. Maricchiolo F, Gnisci A, Bonaiuto M, Ficca G. 2009 Effects of different types of hand gestures in persuasive speech on receivers' evaluations. *Lang. Cogn. Process.* **24**, 239–266. (doi:10.1080/01690960802159929)
71. Hagoort P. 2005 On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* **9**, 416–423. (doi:10.1016/j.tics.2005.07.004)
72. Calvert GA. 2001 Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex* **11**, 1110–1123. (doi:10.1093/cercor/11.12.1110)
73. Bernardis P, Gentilucci M. 2006 Speech and gesture share the same communication system. *Neuropsychologia* **44**, 178–190. (doi:10.1016/j.neuropsychologia.2005.05.007)
74. Aziz-Zadeh L, Wilson SM, Rizzolatti G, Iacoboni M. 2006 Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.* **16**, 1818–1823. (doi:10.1016/j.cub.2006.07.060)
75. Molnar-Szakacs I, Iacoboni M, Koski L, Mazziotta JC. 2005 Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. *Cerebral Cortex* **15**, 986–994. (doi:10.1093/cercor/bhh199)
76. Willems RM, Hagoort P. 2007 Neural evidence for the interplay between action, gesture and language: a review. *Brain Lang.* **101**, 278–289. (doi:10.1016/j.bandl.2007.03.004)
77. Meyer GF, Harrison NR, Wuerger S. 2013 The time course of auditory–visual processing of speech and body actions: evidence for the simultaneous activation of an extended neural network for semantic processing. *Neuropsychologia* **51**, 1716–1725. (doi:10.1016/j.neuropsychologia.2013.05.014)
78. Perniss P, Thompson RL, Vigliocco G. 2010 Iconicity as a general property of language: evidence from spoken and sign languages. *Front. Psychol.* **1**, 227. (doi:10.3389/fpsyg.2010.00227)
79. Perniss P, Vigliocco G. 2014 The brdige of iconicity: from a world of experience to the experience of language. *Phil. Trans. R. Soc. B* **369**, 20130300. (doi:10.1098/rstb.2013.0300)