A critical evaluation of models of gesture and speech production for understanding gesture in aphasia

Jan P. de Ruiter & Carola de Beer

Linguistics, Bielefeld University, Universitätsstrasse 25, Bielefeld, 33615, Germany

Published online: 30 May 2013.

To cite this article: Jan P. de Ruiter & Carola de Beer (2013): A critical evaluation of models of gesture and speech production for understanding gesture in aphasia, Aphasiology, DOI:10.1080/02687038.2013.797067

To link to this article: http://dx.doi.org/10.1080/02687038.2013.797067

Please scroll down for article

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
A critical evaluation of models of gesture and speech production for understanding gesture in aphasia

Jan P. de Ruiter and Carola de Beer

Linguistics, Bielefeld University, Universitätsstrasse 25, Bielefeld 33615, Germany

Background: The background to this study is that aphasiologists have increasingly started to pay attention not only to the speech that people with aphasia produce, but also to their gestures. As there are a number of competing models about the production of gesture and speech in healthy subjects, it is important to evaluate whether, and if so how, these models could be used to guide the research into gesture and speech, and the relationship between these, in speakers with aphasia.

Aims: The aims and objectives of this study are to see how existing models of gesture and speech are able to accommodate the findings regarding the gesture and speech behaviour of speakers with aphasia, in the hope that (1) these models could shed light on the use of gesture in aphasic speakers, and potentially suggest new approaches to therapy for people with aphasia and (2) the aphasia gesture data might help fundamental psycholinguistics to evaluate the adequacy of existing gesture and speech models.

Methods & Procedures: The methodology here was theoretical. Four models of gesture and speech interaction were critiqued and we reviewed their ability to explain some of the central empirical findings in the area of gesture and speech in aphasia.

Outcomes & Results: The outcomes and results of this theoretical analysis were that, with respect to the relationship between gesture and speech in aphasia, (1) the four models under investigation could be reduced to two models, because three of the investigated models were based on the same core assumptions and (2) both of these models adequately explain these findings, but the Growth Point/Sketch/Interface Model is more satisfactory than the Lexical Access Model, because of the better fit with the experimental results on the use of gesture for facilitating word finding, and because it is more compatible with the finding that gestures are also used to enhance communicative efficiency by replacing speech.

Conclusions: The two main conclusions from this study were that both the Growth Point/Sketch/Interface Model and the Lexical Access Model are compatible with data on gesture and speech in aphasia, and that the former appears to be preferable on the basis of the evidence available so far.

Keywords: Gesture; Speech; Adaptation; Broca; Models.
say that at the moment, the data from language users with aphasia are more helpful to our understanding of gesture than gesture theory is to our understanding of aphasia.

As theories about psycholinguistic processes increase in complexity, there comes a point at which it is helpful to formulate a “model” rather than a theory. By “model”, we do not mean a number of boxes with central theoretical concepts in them and some arrows that specify how and in which direction these concepts are interrelated. These kinds of models are helpful in a didactical sense, but rarely more informative than the verbal theory that they are supposed to illustrate. Rather, we mean a computational model, one that specifies at least the central representations and processes, and the communication and temporal interrelations between these processes, as for example in Levelt’s (1989) “Blueprint for the speaker”. A computational model can in principle, possibly in a simplified form, be implemented in an artificial system, or at the very least suggest and/or constrain such an implementation.

Ideally, a computational model allows the generation of predictions that can be tested empirically. These predictions should of course not be about the very phenomena that motivated the theory or assumptions underlying the model itself, because that would be circular reasoning. A good computational model predicts empirical phenomena that are new in the sense that they had not been observed before the model was formulated. These predictions can then be tested against empirical data, and the results of those tests can then be used to reject or adapt the model.

The four gesture models that we will discuss in this article have all been formulated mainly on the basis of observations and experimental data obtained from speakers without language impairments. These models aim to explain gesture and speech behaviour in normal, everyday conversation by healthy individuals. This is not to say that those who have formulated those models have never looked at or thought about language impairment, but just that they have not primarily developed their models to explain the linguistic behaviour of individuals with language impairments. However, it is important that these models are tested on the data different from those that were used to formulate them. This is precisely why aphasia research provides us with an excellent source of data to evaluate computational models. Those models that can, at the very least, plausibly explain aphasia phenomena might then be good candidates for use in clinical contexts, for instance by suggesting or motivating certain therapies for aphasia. But it is important to validate the models in the light of the existing (aphasia) data, before trying to derive therapeutic ideas or new predictions from them.

In the following discussion, we will primarily focus on the class of so-called *iconic* gestures, because three of the four models we will discuss explain only the origin of gestures of this type (although they are sometimes labelled differently). Another self-imposed restriction here is that we will focus on people with nonfluent aphasia (including what used to be called Broca’s type aphasia and global aphasia), characterised by severe disturbances in speech production. More specifically, the type of aphasia we focus on in this article (and that we will refer to as “nonfluent aphasia” in the remainder of the article) has the following characteristics: (1) the produced speech shows a reduced phrase length and a high amount of pauses, (2) the utterances, even if they are reduced to single words, are used in a coherent way, (3) speech

---

1For instance, Bühler’s (1934) famous “Organon” model.
perception is less affected than production and (4) despite the severe loss of fluency and word-finding problems, the communicative intention is achieved.

**BASIC FINDINGS**

There are a number of basic findings from research into gesture and nonfluent aphasia that we would like to focus on in this discussion. The first (Finding I) is that people with nonfluent aphasia try to communicate, and the message they try to convey is not impaired as such. It means that they try to communicate something appropriate and felicitous, but they have trouble doing so effectively. The extensive and outstanding case study by Goodwin (1995, 2000) serves as a powerful, but by no means unique, illustration here. The case study of Goodwin (1995, 2000) provides a detailed description of multimodal communication in severe nonfluent aphasia. A multiple case study by Wilkinson, Beeke, and Maxim (2010) supports the notion of the use of multimodal communication to convey complex content. Goodwin’s observation that gestures are used to compensate for the affected verbal skills has also been confirmed by the results of group studies (e.g., Herrmann, Reichle, Lucius-Hoene, Wallesch, & Johannsen-Horbach, 1988; Rousseaux, Daveluy, & Kozlowski, 2010).

The second finding (Finding II) is that people with agrammatism (which is a subgroup of nonfluent aphasic speakers) produce syntactically reduced but (in some contexts) adequate or “elliptic” speech, arguably as a strategy to adapt to the underlying deficit (see Kolk & Heeschen, 1990). This adaptation theory has been “expanded” by considering interaction with others as a communicative resource for agrammatic speakers (Beeke, Wilkinson, & Maxim, 2003, 2007; Heeschen & Schegloff, 1999). Adaptation to the verbal deficit seems to be a process of collaboration between agrammatic speakers and their interlocutors. Interaction enables agrammatic speakers to communicate complex content by integrating the interlocutor’s linguistic resources with their own multimodal utterances.

The third finding of relevance (Finding III) is that these aphasic speakers still produce gestures, including iconic ones, but at a lower rate per time unit, just as with speech (Cicone, Wapner, Foldi, Zurif, & Gardner, 1979; McNeill, 1992). Interestingly (and this is Finding IV), the gesture rate per word is higher than in normal controls (Carlomagno & Cristilli, 2006; Feyereisen, 1983), and this is also the case specifically for iconic gestures (De Beer, 2012; Sekine, Rose, Foster, Attard, & Lanyon, this issue). So, while both the speech rate and the gesture rate (per minute) are reduced, the gesture rate is less reduced than the speech rate in words per minute, resulting in a higher ratio of gestures per word. This is important, because if the gesture rate and the speech rate were both reduced by the same proportion, the relationship between gesture and speech might well be the same as in healthy speakers. But if they are different, as seems to be the case, the content relationship between gesture and the related speech must somehow have changed.

**GESTURE MODELS**

In this section, we discuss four influential papers about gesture models (or theoretical approaches) in historical order, focusing on how these models can accommodate the four findings that are discussed in the previous section, and what we can learn from these models about gesture behaviour in speakers with nonfluent aphasia.
Growth point theory

Growth Point (GP) theory (McNeill, 1992; McNeill & Duncan, 2000) has been so influential in gesture research that we will discuss it here first. Even though it is not a computational model, some of its central ideas have strongly influenced computational models, as we will make clear later in this article. The core of GP theory is that a GP is a holistic representation that forms the initial “seed” from which an utterance grows. Importantly, this holistic representation contains both imagistic (analogical) and symbolic (discrete) information, constituting a proto-utterance that contains the information that is new (in the linguistic information state sense of the word, so new here means the same as not given), relative to a background of previously established discourse context. Producing an utterance is then described as a GP “unfolding” into separate symbolic and imagistic components. Importantly, during speech production, the symbolic part of a GP is turned into speech, and the imagistic part into a gesture. GP theory therefore explains why there is temporal and semantic synchrony between (iconic) gestures and speech: on this account, they come from the same representation.

Under these assumptions, one possible cause for the aphasia symptoms that we focus on in this article would be that the creation of a GP is disturbed; the GP could be “wrong” in the sense of not expressing the proper information. Note that we cannot specify whether a disturbed GP would affect gesture or speech, as an unexpanded GP is itself not yet specified in terms of either gesture or speech. However, the behaviour of aphasic speakers whose communicative intentions are still felicitous (the group we focus on in this article) rules out this account. We also know that the speech produced by people with nonfluent aphasia is (1) often accompanied by gestures (including iconic gestures, see Lanyon & Rose, 2009; McNeill, 1992) and (2) clearly tries to communicate new information to interlocutors (Goodwin, 1995). So there is a GP, but it is not unfolded in the way it would be in a healthy control. At this point we would like to examine GP theory’s predictions as to what happens to the GP in language users with aphasia. However, this is a processing question and is not directly addressed by GP theory. Nevertheless, the theory has explicitly or implicitly inspired two of the gesture models that we will discuss later.

McNeill (1992, p. 341) writes about accommodating Findings II, III and IV that people with Broca’s type aphasia have GPs that are “tiny and unpackable” (probably “un-unpackable” was intended here), and that these GPs have “excessive contacts” with imagery, which is consistent with our interpretation that these aphasic speakers do have GPs, but experience trouble “unpacking” them into speech. McNeill also notes that people with Broca’s aphasia have trouble expanding the symbolic part of GPs into language because they “cannot activate the necessary closed-class items” and “they cannot find the parts with which to build up a linguistic representation of holistic images.” Consistent with arguments proposed by Grodzinsky (1988) and Ben-Shachar, Hendler, Kahn, Ben-Bashat, and Grodzinsky (2003), we suggest that problems in producing grammatical speech could, instead of being a problem of access to closed words, be related to the process of properly combining different constituent in syntactic combinations. This may be related to the so-called binding problem. This problem is the general computational problem of keeping track of the relationships (binding) between representations within and between levels while reordering some of these elements. Hagoort (2005) describes this problem concisely as “The need for combining independent elements into a coherent overall representation” (but see Treisman, 1996 for a more in-depth discussion of the binding problem in the context of neurobiological models). An elegant computational theory about how the brain
might solve this problem can be found in Vosse and Kempen (2000), which in turn has inspired neurocognitive researchers like Hagoort (2005) to suggest that Broca’s area is generally involved in solving the binding problem.

Finally, McNeill (1992) observes that people with Broca’s aphasia often use gesture to compensate for difficulties in the speech modality (pp. 337–338). The observation that important content is expressed exclusively in iconic gestures by aphasics, when word retrieval fails, is also confirmed by a differentiated case study of an anomic aphasic speaker retelling motion events (Kemmerer, Chandrasekaran, & Tranel, 2007). This explanation is in fact very similar to the Mutually Adaptive Modalities hypothesis in De Ruiter (2006). However, it is unclear whether these compensatory gestures are interpreted as being part of the GP. On the one hand, they express a similar content to that which the (failed) speech component would have expressed, but on the other, they are not temporally coordinated with speech (as there is trouble with speech) which is one of the primary properties of GPs, and the way McNeill proposes that we can identify them.

In summary, GP theory can accommodate the relevant findings by assuming that healthy GPs are in place, but that the speech part of the GP is not expanded properly. Perhaps to compensate for the trouble, aphasics will adapt either by making smaller GPs with the gesture being produced accompanying an elliptic or otherwise incomplete utterance (Kolk & Heeschen, 1990), and/or by producing compensatory gestures. This explains Findings I–III and importantly, also Finding IV, as both these adaptations will cause the ratio of gestures to words to increase.

The lexical facilitation model

According to a model by Krauss, Chen, and Gottesmann (2000), the primary function of producing iconic gestures, which they call lexical gestures for reasons that will become clear, is to facilitate lexical access in speech production. In other words, the primary function of iconic gestures is not to convey imagistic information to an interlocutor, but rather to facilitate the speaker-internal process of word form retrieval (an idea originally put forth by Hadar & Butterworth, 1997). Krauss et al.’s model is designed to explain how the claim that lexical gestures facilitate word form retrieval could be unified with and/or explained in the framework of Levelt’s (1989) production model. The basic idea is that the motor program for the gesture of a certain visuo-spatial representation, created from a number of its spatial features, functions as a cross-modal prime for word form retrieval.

While this model does not have much to say about Findings I and II, it can accommodate Findings III and IV, if we assume that lexical gestures are at least roughly similar to iconic gestures. The higher rate of iconic gestures per word may be caused by attempts to facilitate the word retrieval that is difficult for speakers with nonfluent aphasia. Assuming that these speakers have more difficulty retrieving words, they will then generate more gestures to counteract this difficulty.

There are a number of caveats here. First, the model by Krauss et al. assumes that lexical gestures, as the name suggests, are tightly connected to words. However, there is ample evidence that this is not generally the case for iconic gestures. The classic “roll down the hill” example (see e.g., Kita & Özyürek, 2003; McNeill, 1992) is a good illustration of this. If performing a gesture indicative of “rolling down the hill” facilitates the retrieval of phonological form in an English speaker, which of the words making up the clause “rolls down the hill” is facilitated, given that the spatial features
in the gesture can only be linked to meaning of the entire clause? Another example is Kendon’s example of someone saying “New York Times”, while indicating the thickness of this publication using thumb and index finger. Here, there is no word in the verbal utterance that is linked to the gesture. This is why De Ruiter (2000) proposed that gestural affiliates are conceptual, rather than lexical.

However, even if we limit ourselves to iconic gestures that are linked to a specific word, there is a methodological problem. For every instance where we suspect that a speaker is internally facilitating word form retrieval, there are two other explanations that are at least as plausible. The first is that the speaker, experiencing trouble finding the word, is gesturing to keep the floor (in the turn-taking sense) while trying to find the appropriate verbal form (Beattie & Coughlan, 1999). The second is that the speaker is compensating for the failing verbal communication by using gesture. The very gesture that would, due to its spatial features, facilitate retrieval of the correct word for the speaker him/herself would do exactly the same for the interlocutor. A classic counterargument to this point is that people do gesture while being on the phone. This is true, but people also produce appropriate facial expressions on the phone, or even bow to their telephonic interlocutor, so it might very well be that this behaviour is so automated that it is not worth the effort to suppress these natural aspects of communication. It is important to stress that just establishing an association between speech difficulty and the production of gesture is not sufficient evidence for the lexical facilitation hypothesis as long as there are at least two plausible alternative theories (communicative compensation and floor holding) that would explain this association equally well. Discussing all these studies would take up too much space, but we refer the interested reader to De Ruiter (2006) for an extensive discussion of these correlational studies. The only valid and conclusive way to prove or disprove the existence of a causal relationship is to use controlled experiments.

Frick-Horbury and Guttentag (1998) performed such a controlled experiment, by asking participants to produce the words described by presented definitions. They used a $2 \times 2$ between-participants design, controlling (1) whether or not the hands of the participants were restricted and (2) whether the Scholastic Aptitude Test (SAT) score of the participant was in the high (above median) or the low (below median) group. They found that the participants in the unrestricted group on average successfully retrieved significantly more words than those in the restricted group. This suggests that those who were allowed to gesture had more successful lexical access, and hence appears to support the idea that people gesture to facilitate the retrieval of words. However, while this is certainly the type of evidence that would perhaps be necessary for the claim that gesture performs the function of internal facilitation, it is not sufficient. What this shows is that not being able to use the hands makes lexical access worse, but not that using the hands makes it better. Note that the first part of the previous sentence mentions ability, and the second part behaviour. To put it another way, it is not implausible that not being able to behave as one would normally do, and remembering not to gesture (or, in the case of the experiment by Frick-Horbury and Guttentag, to hold onto the bar participants were given during the experiment) will take cognitive and/or attentional resources away from the main task. This effect is interesting and probably worth investigating further, but the crucial prediction generated from the lexical access hypothesis is that performing gestures facilitates retrieval,

This is also very clear from looking at the model by Krauss et al. (2000, Figure 1)
due to the priming effect of the gestural motor program. Evidence for this would be, for instance, if the participants in this study were able to resolve tip-of-the-tongue (TOT) states more often by gesturing than they could when they did not gesture. However, Frick-Horbury and Guttentag themselves reported that “participants were not more likely to retrieve an item correctly on trials when a gesture was produced than when a gesture was not produced” (1998, p. 53).

Hence, a crucial question in experimental work on the lexical access hypothesis is whether gesturing affects the resolution rate of TOT states. Beattie and Coughlan (1999) measured and reported this in an experiment in which they also deliberately invoked TOT states in subjects that were either allowed to gesture or had their arms folded. Their principal findings were that, (1) contrary to Frick-Horbury and Guttentag (1998), the participants who had their arms folded were more successful in finding the correct words than those who were allowed to gesture, but also, and more importantly, (2) for the participants who were allowed to gesture and were in a TOT state, those who performed an iconic gesture were just as likely to resolve the TOT state as those who did not perform such a gesture. This is very strong evidence that performing iconic gestures does not help alleviate word-finding problems.

Focusing on aphasic speakers again, we need to distinguish between training studies (including the production of gestures) that concern the production of object words and those which concern verbs. Investigations that focused on the facilitation of verb production by the use of gestures (e.g., Boo & Rose, 2011; Rodriguez, Raymer, & Gonzalez Rothi, 2006; Rose & Sussmilch, 2008) did not produce clear indications of a facilitating effect of gesture production on verb retrieval. The work by Rose and colleagues (e.g., Boo & Rose, 2011; Rose & Douglas, 2001, 2008; Rose & Sussmilch, 2008) shows that for some people with aphasia, training of picture naming involving gestures does have a beneficial effect on word production. Rose and Douglas (2001) reported that aphasic speakers with a primary deficit in phonological encoding benefit from the production of iconic gestures in word (form) retrieval. But this subgroup only represents a small proportion of the people with nonfluent aphasia considered here. A follow-up study (Rose & Douglas, 2008) produced inconclusive results concerning the facilitation effects of gestures on word retrieval. The authors pointed out that it might have been the training of semantic features that resulted in improved naming performance, rather than the gesture training. This raises the interesting issue of to what degree TOT states are comparable to the state of a speaker with aphasia experiencing word-finding problems. In the TOT state subjects typically know that there is a word, and often even its gender, the stress pattern or the number of syllables (Brown, 1991). This might be different in the case of nonfluent aphasia. Perhaps it is the case that some aphasic speakers have stronger associative links between certain motor patterns and concepts than others, which would make them more responsive to combined gesture/semantic training, as in Rose and Douglas (2001). However, this therapy would then, in the terminology of Levelt’s (1989) production model, affect the level of lemma retrieval, not word form retrieval, as is assumed to be the case for TOT states.

To summarise, the model by Krauss et al. can accommodate the Findings III and IV, but the relevant phenomena admit other plausible explanations (communicative compensation and floor-keeping), and the experimental evidence casts serious doubt on this account’s central assumptions. The often reported association between speech difficulty and iconic gesture frequency is insufficient evidence for the lexical facilitation hypothesis, as there are at least two alternative and plausible explanations for why people would gesture more when they have trouble speaking (see also De Ruiter, 2006).
Nevertheless, the (inconclusive) findings from gesture-based therapies with aphasic speakers suggest that there might be a beneficial effect of gesture on fluency, albeit at another representational level than that assumed by the Krauss et al. model.

The sketch model

The Sketch Model (De Ruiter, 2000) is a general model for gesture and speech production, based on Levelt’s (1989) architecture for the speaker. In contrast to the other models discussed here, it does not focus exclusively on iconic/metaphoric gestures, but explicitly includes accounts of deictic, emblematic and pantomimic gestures as well. It means that it gives a computational account for the production of speech in combination with all types of gestures, with the exception of beats. It is beyond the scope of this article to give a detailed description of the workings of the model, but we will mention a number of crucial assumptions relevant for the processing of iconic gesture in combination with speech. See Figure 1 for a graphical depiction of the model.

First, the model assumes that both a gesture and the speech it accompanies have a communicative function. Second, it is assumed that both the speech and the accompanying gesture originate from one and the same communicative intention. Barring some hair-splitting about terminology, these two assumptions are actually very similar, if not the same, to the central assumptions in GP theory. This should not be too surprising, as the Sketch Model has been influenced strongly by GP theory. A third relevant assumption is that the distribution of “communicative load” over a gesture and speech channel is performed by the conceptualiser. This is a process (or module) taken over from Levelt’s (1989) Blueprint for the Speaker, but modified to not only generate a preverbal message, but also a so-called Sketch, which is an abstract spatio-temporal trajectory that is as yet underspecified with respect to concrete motoric parameters (such as size, speed and location). This Sketch is the gestural equivalent of the Preverbal Message in speech. A fourth assumption is that, in the event that the conceptualiser is hampered in the generation of either a Preverbal Message or a Sketch, it will compensate by shifting a higher load to the respective other channel. This is the so-called Mutually Adaptive Modalities assumption (De Ruiter, 2006), later also called the Trade-off Hypothesis (De Ruiter, Bangerter, & Dings, 2012; Van der Sluis & Krahmer, 2007). A fifth assumption is that after the Sketch and the Preverbal Message have been generated, they are sent to, respectively, a Gesture Planner and a Formulator (the latter being as in Levelt’s model). The Gesture Planner and the Formulator then operate independently from one another, apart from a few discrete stop/go signals to assure temporal synchronisation of the initiation of the speech and gesture fragments. Importantly for present purposes, the Sketch Model “inherits” a sixth assumption from Levelt’s model, namely that the conceptualiser breaks down a communicative intention into smaller units that are sent, in sequence, to lower modules. This breaking down into units (corresponding, roughly speaking, to a clause, Bock, 1982; Levelt, 1989), also called “linearisation”, is not a trivial process, as the

---

In the Sketch Model, there is no distinction made between metaphoric and iconic gestures, because for processing in a production model it is not relevant whether the imagistic representation underlying a gesture is of a metaphoric or a concrete origin.

Assumptions three to five are not addressed in GP theory, as they are concerned with computational implementation, an aspect that lies outside the scope of GP theory.
total amount of information that a person wants to share with an interlocutor can, and often is, much larger than can fit in one speech unit.

An important aspect of the Sketch Model, which it inherits from Levelt’s (1989) model, is that there are three feedback mechanisms that allow speakers to monitor their own utterances. The first is at the level of the Conceptualiser, where planned preverbal messages are fed back into the Conceptualiser to check if they correspond with the communicative intention(s). The second feedback loop is that the phonetic plan produced by the phonological encoding processes is fed back into
the comprehension system (not visible in Figure 1, see Levelt (1989) for details). The third and last feedback loop is after articulation, when a speaker hears his or her own speech, also of course interpreted through his or her own comprehension system. These feedback loops are essential for explaining phenomena like self-repair (Levelt, 1983), and model the capabilities of language users to keep track of their performance during speaking.

Looking again at the findings from aphasia research, we can first note that Finding I, the finding that the people with nonfluent aphasia (as we have defined them in our introduction) have felicitous communicative intentions, corresponds with the first assumption of the Sketch Model. To accommodate Finding II (elliptic speech), one would have to assume that the Formulator is damaged or hindered in its proper operation. This is a natural assumption, because the impoverished syntax and word finding problems in aphasic speakers suggest that the modules responsible for syntactic encoding and word-form encoding (i.e., the two sub-processes of Levelt’s Formulator) are impaired. Also, in the Sketch Model framework, the impairment cannot be in the conceptualiser or in the articulator. It can’t be located in the conceptualiser because that would then cause the gestures to be impoverished to the same degree as the speech (and there is no evidence for that for the group of people with aphasia that we focus on here) and it can’t be in the articulator because if it were there, it would mean that syntactically correct and nonelliptic speech would be produced, albeit with nonstandard phonetic realisation. Both Finding III (the lower gesture rate per time unit) and Finding IV (the higher gesture per word rate) can be accommodated by assuming that the conceptualiser adapts to the problems in the Formulator by producing smaller units.

This is the central assumption that connects the Sketch Model with the aphasia data, so we will elaborate on it in more detail. Because of the more cumbersome and slower formulation and production of these (smaller) units, the gesture rate per time unit will go down, but the word rate per time unit will drop even more. The resulting generated speech, being of a linear-symbolic nature, will be shorter, but the more “holistic” gesture that corresponds to the utterance will still be intact. So for those communicative intentions that also involve gestures, the same number of gestures will be produced, but fewer words will be, due to the smaller speech unit and the correspondingly smaller and more elliptic utterance from the impeded Formulator. This causes the number of words to be reduced much more than the number of gestures, while still lowering the rate (per time unit) of both gesture and speech.

A potential criticism would be to ask how the conceptualiser could know that the Formulator is in trouble. After all, it is supposed to be a module, meaning it is not exchanging state information with the Formulator. The reason that the conceptualiser “knows” about the problems in the Formulator is because of the self-monitoring processes mentioned above.5

This account linking the Sketch Model to the aphasia gesture data is very similar to the adaptation hypothesis by Kolk and Heeschen (1990). These authors assume that the elliptic speech frequently observed in Broca’s type aphasia is not directly caused by the aphasia itself, but by an adaptation strategy to aphasia-related problems (such as problems with the computation of morphosyntactic agreement relations for instance). The Sketch Model account is similar. By producing smaller units, the morphosyntactic

5An intriguing proposal for a faster and therefore arguably more plausible self-monitoring architecture based on forward modelling can be found in Pickering and Garrod (in Press).
and word-finding load for the Formulator are reduced, which leads to smaller units with fewer words, and therefore to a higher number of gestures per word.

The interface model

Kita and Özyürek (2003) have published intriguing data that have shed much light on the relationship between gesture and speech. Their principal findings were twofold. First, the availability of words in a language expressing a certain concept has an influence on the gesture that accompanies a speech fragment expressing that concept. So for instance, as Kita and Özyürek found, speakers of Japanese describing someone swinging on a rope (who do not have a word corresponding to the verb “to swing” in English) will not only use other words to express this particular motion in their speech, but also produce fewer gestures that exhibit an arc-like motion. It is as if the content of the gesture is adapting itself to the content of the speech—if there is no word for “swing”, the gesture will less often encode the arc-trajectory.

Second, there is a relationship between the type of gesture and the clausal structure of manner–path combinations of a language. For instance, some languages, such as English, can conflate manner (e.g., “rolling”) and path (e.g., “down”) into one clause combining manner and path (“rolling down”) whereas other languages, such as Turkish, require two separate clauses for this (e.g., “moving down” (encoding path) “in a rolling fashion” (encoding manner)). What Kita and Özyürek found was that speakers of English who say “rolling down” augment this phrase with a gesture that also combines the path and the manner; for instance, by producing a spiralling gesture in a downward slope, which expresses both manner (the spiralling) and path (the downward slope). Conversely, speakers of Turkish, whose grammar does not allow for such a combination, tended to say something like “rolling, in a downward direction”, and tended to produce two gestures, a rotating gesture that accompanied the phrase that meant “rolling” and a gesture straight down that accompanied the phrase that meant “in a downward fashion”. Both these phenomena show that both the lexical as well as the grammatical properties of a language influence the nature of the gestures that are produced while speaking that language. Although these are probabilistic rather than deterministic phenomena, they clearly show how the nature of speech at least partially determines the nature of gesture. To accommodate these findings, Kita and Özyürek formulated a speech and gesture production model, also based on Levelt’s (1989) architecture, in which the speech generating and the gesture generating processes negotiate with each other, or engage in a constraint-satisfaction operation, to ascertain that the resulting gesture and the speech express similar information.

Before discussing how the Interface Model can accommodate the aphasia findings described in the introduction, it is important to evaluate to what degree the Interface Model actually differs, in a strict computational sense, from the Sketch Model. We will argue that these differences are much smaller and less central than a superficial inspection of their respective diagrammatic representations may suggest. The apparent differences between the models might well be due to differences in terminology. For instance, Kita and Özyürek defined the so-called “Free Imagery Hypothesis”, which states that “[gestures] are generated ‘prelinguistically’, that is, independently from the representational potential of the language” (p. 17). They then claim that both the Lexical Access Model and the Sketch Model are examples of models that are based on the Free Imagery Hypothesis. For the Lexical Access model of Krauss
et al. (2000), discussed above, this is indeed correct, but not for the Sketch Model. To quote De Ruiter (2000, p. 293), “If imagistic information from working memory has to be expressed in an iconic gesture, the shape of the gesture will be largely determined by the content of the imagery” (italics added). To further illustrate the similarity of the models, on p. 17, Kita and Özyürek write, “... the Free Imagery Hypothesis predicts that the information encoded in a gesture is not influenced by how the information could be verbally expressed.” But both the Sketch Model and the Interface Model follow GP theory in the assumption that “The communicative intention is split in two parts: a propositional part that is transformed into a preverbal message, and an imagistic part that is transformed into a sketch” (De Ruiter, 2000, p. 292). In the Sketch Model, there is ample opportunity for the conceptualiser to make sure that the content of the gesture matches the content of the speech, and this is important given that both the gesture and the speech are part of one and the same communicative intention.

The reason Kita and Özyürek assume that there are substantial differences between the models may be due to a different interpretation of (1) what a Preverbal Message is and (2) what the Formulator does. The Preverbal Message is a semantic representation (Levelt, 1989, p. 71, see also Guhe, 2003), adapted to language-specific requirements (Levelt, 1989, pp. 103–105). Every language has its own semantic expressivity, the conflation of manner and path or the availability of arc-like motion trajectories being part of that expressivity. The difference between the way the Interface and the Sketch Model allow for adaptation of the speech to the gesture lies in whether this happens in the Conceptualiser or in the Formulator. This in turn is related to the rather confusing and somewhat implicit re-interpretation of Levelt’s original model by Levelt, Meyer, and Roelofs (1999); in Levelt (1989), the model specified, as do Kita and Özyürek (2003), that the selection of lemmas takes place in the Formulator on the basis of semantic criteria provided in the Preverbal Message. In the reformulation of the theory by Levelt et al. (1999), which deviates clearly (but implicitly) from the assumptions in Levelt (1989), it is assumed that the Conceptualiser has access to the words of a language, implying that the process of lexical access now changes into a trivial 1:1 mapping exercise (Roelofs, 1992). The Sketch Model relies on that additional assumption. But either way, the difference between these two accounts of how speech can adapt to gesture is cosmetic, and it is hard to see which empirical evidence could inform us about which of these two nomenclatures is more correct.

The similarity of the Interface Model and the Sketch Model can also be illustrated by looking at the respective diagrams for the Interface Model (Kita & Özyürek, p. 28) and the Sketch Model (De Ruiter, 2000, p. 298). If one would simply draw a box called “Conceptualiser” around the three processes in the Interface Model labelled “Communication Planner”, “Action Generator”, and “Message Generator”, and also change “Action Generator” into “Sketch Generator”, one ends up with near-identical models. One residual difference is that in the Interface Model, the “Action Generator” projects straight into the motor control process. The Sketch Model cannot assume this, as it also has to account for other gesture patterns, such as emblems and deictics, which are not based on action patterns, but on parameterised “gesture templates”. To model the production of these noniconic gestures the Sketch Model needs to specify an additional level of representation.

The central and important contribution of Kita and Özyürek is therefore not primarily their proposed model, which, in its central assumptions and architecture does
not differ substantially from the Sketch Model, but rather their empirically well-motivated elaboration of processes assumed to take place in the conceptualiser in the Sketch Model to accommodate their cross-linguistic results. So this means that the accommodation of the aphasia results by the Sketch Model discussed above do also apply to the Interface Model.

DISCUSSION AND COMPARISON OF THE MODELS

Now that we have analysed these four models/theories in detail, we are in a position to simplify the discussion considerably. As it turns out, we do not have to compare four different models, but only two. One is the model by Krauss et al. (2000), which assumes that (what they call lexical, and what the other authors call iconic) gestures indeed are generated purely from imagery (the “Free Imagery Hypothesis” as identified by Kita & Özyürek, 2003), and that these gestures function to facilitate word form retrieval. The second one is the model that has been inspired by GP theory, implemented by the Sketch Model, and then later refined by the findings by Kita and Özyürek (2003). GP theory originally hypothesised the shared representational origin of gesture and speech, the Sketch Model took this original hypothesis and expanded it into a general computational architecture, and the work by Kita and Özyürek (2003) provided crucial information about how the conceptualiser splits a communicative intention into the two modalities. Fundamentally, however, these three approaches share the same architectural assumptions and therefore lead to the same explanations for the aphasia findings.

Having reduced the number of models to two still leaves us with the question of which model is more adequate. This is difficult, as even the most specific and challenging of the aphasia findings (Findings III and IV, the lower gesture rate per time unit, and the higher gesture rate per word) can be explained by both the Lexical Facilitation Model and the GP/Sketch/Interface Model. The Lexical Facilitation Model explains this straightforwardly by assuming that more gestures are made to facilitate lexical access, while, in the GP/Sketch/Interface Model, these findings can be accommodated by assuming that the production system has trouble in the Formulator, and therefore the conceptualiser adapts by producing smaller speech units. But since the gestures are more holistic, they cannot be broken into smaller parts, and consequently fewer words go with the same number of gestures, hence the higher gesture/word rate. So both accounts adequately accommodate the central findings. However, there are a number of reasons to prefer the account based on the GP/Sketch/Interface Model. First, the experimental evidence from Beattie and Coughlan (1999) and Frick-Horbury and Guttentag (1998) shows that gesturing does not help to avoid or resolve TOT states. To date this is the only real experimental evidence on this issue, and it casts serious doubt on the lexical facilitation hypothesis. Second, the GP/Sketch/Interface account is more compatible with the notion that speakers with nonfluent aphasia use gesture to compensate for problems in the speech formulation processes. In this case, the GP/Sketch/Interface account can also explain why gestures are produced instead of speech, without necessarily being connected to a lexical entry (De Ruiter, 2000; Goodwin, 1995, 2000; McNeill, 1992). Finally, the idea put forth here in the GP/Sketch/Interface account that the conceptualiser is forced to generate smaller speech units is compatible with and supported by Kolk and Heeschen’s (1990) influential theory about the cause of elliptic speech in Broca’s type aphasia.
In order to understand the role of gesture in nonfluent aphasia (in our restricted definition of the phenomenon stated in the introduction), it is very helpful to have a computational model that gives a comprehensive overview of the different representations and processes that are assumed to underlie the production of gesture and speech (Rose, 2006). However, models need to be empirically validated. For that, it is necessary that the models can accommodate the most important findings about gesture in aphasia. In this article, we have identified some solid findings about gesture in aphasia, concentrating on iconic gestures and nonfluent aphasia. Then we have discussed four models (or rather, three models and one model-inspiring theory) about the production of gesture and speech in healthy speakers, and evaluated how these models would accommodate these findings. The models are GP Theory (McNeill, 1992), the Lexical Access Model (Krauss et al., 2000), the Sketch Model (De Ruiter, 2000) and the Interface Model (Kita & Özyürek, 2003). A detailed analysis of the underlying assumptions of these models revealed that the initial set of four models could be reduced to two meaningfully different modelling accounts. First, GP Theory has inspired the Sketch Model, which fleshes out the implementational aspects left unspecified in Growth Point Theory. Moreover, the Interface Model does not differ conceptually from the Sketch Model, but rather specifies in more detail certain processes that new empirical findings suggest to take place in the conceptualiser of that model. So in the end, we have two fundamental models: the Lexical Access Model, and the GP/Sketch/Interface Model. Both of these models give an account of the gesture/aphasia findings under discussion, but these accounts are different. The Lexical Access Model would have to assume that the higher ratio of gestures per word is caused by speakers trying to internally facilitate word-form retrieval by making more (lexical, or iconic) gestures, while the GP/Sketch/Interface Model assumes that speakers adapt to difficulties in grammatical and phonological word form retrieval by generating smaller units, which nevertheless are accompanied by the same (holistic) gestures.

Of course, more research is necessary to evaluate the relative merits of these two computational accounts. The models also need to be tested against the symptomatic profiles of fluent types of aphasia, which was not taken into account here. A differential set of predictions that is perhaps testable can be found in more qualitative data, rather than in the quantitative data that have been discussed in this article. For instance, the Lexical Access Model would predict that most of the “surplus” gestures (per word) could be linked to certain words (that speakers are trying to retrieve), whereas the GP/Sketch/Interface Model would predict that the additional gestures would semantically correspond not necessarily to words, but to the clause, or the communicative intention as a whole. To test this differential prediction, the communicative context and the intentions of the speakers should be taken into account which is a good idea anyway in psycholinguistics, irrespective of the object of study.

REFERENCES


Lanyon, L., & Rose, M. (2009). Do the hands have it? The facilitation effects of arm and hand gesture on word retrieval in aphasia. Aphasiology, 23(7–8), 809–822.


