14 The production of gesture and speech

Jan Peter de Ruiter
Max Planck Institute for Psycholinguistics, Nijmegen

1 Introduction

Research topics in the field of speech-related gesture that have received considerable attention are the function of gesture, its synchronization with speech, and its semiotic properties. While the findings of these studies often have interesting implications for theories about the processing of gesture in the human brain, few studies have addressed this issue in the framework of information processing.

In this chapter, I will present a general processing architecture for gesture production. It can be used as a starting point for investigating the processes and representations involved in gesture and speech. For convenience, I will use the term 'model' when referring to 'processing architecture' throughout this chapter.

Since the use of information-processing models is not believed by every gesture researcher to be an appropriate way of investigating gesture (see, e.g., McNeill 1992), I will first argue that information-processing models are essential theoretical tools for understanding the processing involved in gesture and speech. I will then proceed to formulate a new model for the production of gesture and speech, called the Sketch Model. It is an extension of Levelt's (1989) model for speech production. The modifications and additions to Levelt's model are discussed in detail. At the end of the section, the working of the Sketch Model is demonstrated, using a number of illustrative gesture/speech fragments as examples.

Subsequently, I will compare the Sketch Model with both McNeill's (1992) growth-point theory and with the information-processing model by Krauss, Chen & Gottesman (this volume). While the Sketch Model and the model by Krauss et al. are formulated within the same framework, they are based on fundamentally different assumptions. A comparison between the Sketch Model and growth-point theory is hard to make, since growth-point theory is not an information-processing theory. Nevertheless, the Sketch Model and growth-point theory share a number of fundamental assumptions.

An important conclusion is that information-processing theories are essential theoretical tools for exploring the processing involved in gesture and speech. The presented Sketch Model is an example of such a tool. It accommodates a broad range of gesture phenomena and can be used to explain or predict these phenomena within a consistent formal framework.

1.1 Terminology

In this chapter, the word 'gesture' is used in the restricted sense of spontaneous body movements that occur during speech and often appear to represent aspects of the topic of the accompanying speech. Although most gestures are hand gestures, other body parts, such as the head, are also often used for gesture. The typology I will use for distinguishing different kinds of hand gestures is similar to McNeill's (1992) typology, but because it is not based on semiotic properties, but rather on the representations underlying the processing of gesture, it is different in important ways. I distinguish the following types of gestures:

Iconic gestures: Depicting aspects of the accompanying speech topic. This category includes what McNeill calls *metaphoric* gestures, because from the perspective of gesture production it is of no relevance whether the imagery underlying the gesture is related to abstract or to real entities.

Pantomimes: Gestures that are imitations of functional motor activities.

Deictic gestures: Pointing gestures.

Beat gestures: Biphasic movements of the hands or fingers that do not represent anything.

Emblems: Gestures whose form–meaning relation is lexicalized.

2 Information processing

A common way to formulate theories in cognitive psychology is to use the *information-processing approach*. In this approach, theories are often specified by using highly suggestive box and arrow drawings. These 'boxologies' may be helpful visual tools, but they do not reveal the underlying assumptions of the information-processing approach. I will therefore clarify and justify these assumptions before presenting a model for gesture and speech.

The term information-processing itself is precise: the core assumption of the approach is that the brain does its job by processing information. This is a weak, but highly plausible, assumption supported by a wealth of neurobiological data. In the formal definition of information by Shannon &

Weaver (1949), information is defined to be anything that reduces the uncertainty about a number of possible alternatives. The general nature of this definition is the main reason why the Information Processing assumption is relatively weak. Neural networks, Artificial Intelligence models, and spreading activation models, to name just a few, are all information-processing models, even though they differ wildly in the way they represent and process information.

In a stronger version of this approach, the word 'information' is taken to mean 'representation'. A representation is some information that is stored as a retrievable entity. The fact that the sun shines is information in Shannon & Weaver's sense of the word, but if I note down in my diary "sun shines today," the entry in my diary will be a representation. The difference between the fact that the sun is shining and my note of that fact is the possibility of retrieving my note at a later time, even when at that time weather conditions have changed. I will use the abbreviation RP (Representations and Processes) for this approach.

The extra assumption implicit in the RP approach is that we can view representations and the processes operating on those representations as *functionally* distinct entities. From this assumption it does not follow necessarily that processes and representations have different spatial locations in the brain, or that the processes involved operate in a certain order. All the assumption states is that once we have a description of the processes and the representations involved in a certain cognitive activity, we know what computations are performed in order to perform this cognitive activity. Often the term 'symbolic' or 'classic' is used to refer to this approach. However, 'classical' theorists usually make stronger assumptions about representations and processes than those of the RP approach (see Fodor & Pylyshyn 1988).

Even when an RP theory is correct, we have no complete knowledge about the cognitive domain of interest. For example, we do not know how the processes are carried out by our neural hardware, or how the representations are stored in the brain. From the RP perspective, to answer those questions it is necessary to know *what* the brain does before trying to figure out *how* it does it.

Needless to say, it is possible to make mistakes in developing an RP theory. If such a mistake is made, researchers investigating lower levels of processing (e.g., neuroscientists) can be wrong-footed in their research, because they have an incorrect view of the computation involved. In that case we have no choice but to hypothesize *another* RP theory, and try again. It is impossible to find out how the brain works only by looking 'under the hood'. If there is no understanding of the computations the brain has to

perform, even detailed knowledge about the anatomical structure of the brain will hardly be interpretable.

However, it is possible (and desirable) for neuroscientific knowledge to guide and constrain the development of a functional description. For instance, knowledge of the human retina influences our ideas about the kinds of representations involved in vision, and neuropsychological knowledge about human motor control could constrain and inspire theories about gesture. As Churchland (1986) has argued persuasively, cognitive scientists and neuroscientists can and should cooperate in order to arrive at a full understanding of the workings of the brain.

Assuming that cognitive faculties can be described by specifying representations and processes at some level of abstraction has many advantages. First, the formal properties of processes operating on representations have been studied extensively by mathematicians and information scientists (e.g., in formal languages and automata). It is possible to use this knowledge in proving certain properties of an information-processing model. Second, as with all information-processing models, it is often possible to use computer simulations to explore an RP model or parts of it. Simulations are an effective way of checking the coherence of an RP theory. Something important could be missing from an RP theory, or there could be inconsistencies in it. Simulation will reveal such faults, simply because the simulation will either not run or will produce the wrong kind of output. This forces the researcher to track down and address the problem. Another advantage of using computer simulations is that the processing assumptions are fully specified in the computer program. Verbal theories or processing accounts often have multiple interpretations, which tends to make them immune to potential falsification.

Many RP theorists make the additional assumption that one or more subprocesses of their model are "informationally encapsulated" (Fodor 1983). Models of this kind are often called *modular*. This means, roughly, that computations performed by the subprocess are not affected by computations that take place elsewhere in the system. It should be emphasized that the assumption of informational encapsulation, although it is adopted in the model presented below, does not follow automatically from the assumptions of RP processing.

Modular models are highly vulnerable to falsification, because they prohibit certain interactions between subprocesses. Any data showing an interaction between two encapsulated processes will be sufficient to falsify the model, or parts of it. Without any modularity, every computation can potentially influence every other computation. This makes both the formulation and experimental falsification of predictions considerably more

prone to multiple interpretations. Thus, for any specific case, it makes sense to carry on with a modular RP model until it has been proven beyond reasonable doubt that the modularity assumption is false.

Some researchers believe that making the assumption of modularity is dangerous, for if this assumption is wrong, the knowledge accumulated by means of experimentation can be misleading. For instance, in Levelt's (1989) model of speech production, the phonological representations used by the process of word–form encoding are stored in a lexicon. If lexical retrieval were not a relatively independent process, the knowledge obtained from picture-naming experiments could not be generalized to the process of spontaneous speech (S. Duncan, pers. com.). However, there is empirical evidence that the results obtained using experimental tasks are comparable to results found under more naturalistic conditions. For instance, the well-known effects of semantic priming in reaction-time research (see Neely 1991) have been replicated using the Event Related Potential methodology (Hagoort, Brown, & Swaab, in press), even though subjects in ERP experiments typically perform no explicit task – they just passively listen to or read language fragments. Another source of support for modularity is the amazing speed and fluency of speech, which makes it likely that there are specialized subprocesses that operate in highly automated, reflex-like fashion, enabling important subprocesses to operate in parallel (Levelt 1989: 2).

McNeill (1987) argues that information processing has certain built-in limitations that make it impossible to apply it to language behavior:

> The most basic [limitation] is that information-processing operations are carried out on signifiers alone, on *contentless* symbols.[1] Given this limitation the only way to take account of 'meaning' and 'context' is to treat them as inputs that are needed as triggers to get the machinery moving, but that are not modeled by the information processor itself. (133; emphasis in original)

However, the fact that the elements of a computation (symbols) do not have inherent content is not a limitation of information-processing theories. As Pylyshyn puts it:

> [Turing's notion of *computation*] provided a reference point for the scientific ideal of a mechanistic process which could be understood without raising the specter of vital forces or elusive homunculi, but which at the same time was sufficiently rich to cover every conceivable informal notion of mechanism. (1979: 42)

In other words, it is necessary to define computation as operations on form rather than on meaning (or content), for if symbols have inherent meaning, there also needs to be an entity to whom those symbols mean something. This would introduce a homunculus in the theory.

Context information should obviously be incorporated in theories of

language processing. The information-processing framework allows for that, provided there is sufficient knowledge about the role context plays in speech production. For instance, in Levelt's model, the conceptualizer keeps track of anaphoric references and previously produced speech (the 'newness' of information) in the form of a discourse record. The information stored in the discourse record can be used by the conceptualizer to take contextual factors into account while producing speech and gesture (see Levelt 1989 for details).

The only limitation on information processing in general is that it does not allow 'vital forces' or 'homunculi' to be used as explanatory devices. This limitation is in fact one of the main virtues of the approach.

3 A model for gesture and speech

The gestures of interest in this chapter usually occur during speaking and are meaningfully related to the content of the speech. It is therefore plausible that these gestures are initiated by a process that is in some way linked to the speaking process. Sometimes people do gesture without speaking, for instance when speech is not possible (e.g., in a noisy factory), but for the moment I will ignore this phenomenon. The fact that gesturing and speaking are in many ways related to each other led to the choice of extending an existing model for speaking to incorporate gesture processing. Another reason for doing this is to make use of, and be compatible with, existing knowledge about the speaking process. The model of the speaking process that is extended to incorporate gesture processing is Levelt's (1989) model of speech production. This model consists of a number of subprocesses (or 'modules') that each have a specific type of input and output. Given a communicative intention, the *conceptualizer* collects and orders the information needed to realize this intention. It retrieves this information from a general knowledge base. The output of the conceptualizer is a representation called the *preverbal message* (or 'message', for short), which contains a propositional representation of the content of the speech. The message is the input for the *formulator*. The formulator will produce an articulatory plan. In order to do that, the first subprocess of the formulator, *grammatical encoding*, will build a (syntactic) *surface structure* that corresponds to the message. It will access a *lexicon* in which the semantic and syntactic properties of lexical items are stored. The second subprocess of the formulator is *phonological encoding*. During phonological encoding, the surface structure built by the grammatical encoder will be transformed into an *articulatory plan* by accessing phonological and morphological representations in the lexicon. The resulting articulatory plan will be sent to the *articulator*, which is responsible for the generation of overt speech. Overt speech is

available to the comprehension system because the speaker can hear it. Internal speech is also fed back to the speech-comprehension system, allowing the conceptualizer to monitor internal speech as well,[2] and possibly correct it before the overt speech has been fully realized.

In order to extend Levelt's model to incorporate gesture, it is important to make an assumption about the *function* of gesture. Some authors (most notably Kendon 1994) argue that gesture is a communicative device, whereas others (Krauss, Morrel-Samuels & Colasante 1991; Rimé & Schiaratura 1991) believe that it is not. There are several arguments for either view. The fact that people gesture when there is no visual contact between speaker and listener (e.g., on the telephone), while this does not present any problems for the listener, is often used as an argument for the non-communicative view. Furthermore, Krauss et al. (1991) argue that iconic gestures (lexical gestures, in their terminology) are hardly interpretable without the accompanying speech. As I have argued in de Ruiter (1995a), there is no real conflict between the two views. Gestures may well be intended by the speaker to communicate and yet fail to do so in some or even most cases. However, one cannot draw the conclusion that gestures do not communicate from the fact that iconic gestures are hard to interpret without the accompanying speech, as Krauss et al. (this volume) do. Gestures are normally perceived *together with* the accompanying speech, and then seem to be an interpretable and non-redundant source of information (see for instance McNeill 1992).

The fact that people gesture on the telephone is also not necessarily in conflict with the view that gestures are generally intended to be communicative. It is conceivable that people gesture on the telephone because they always gesture while they speak spontaneously – they simply cannot suppress it. Speakers could adapt to the lack of visual contact by producing more explicit spatial information in the verbal channel, but they need not suppress their gesturing. Finally, there is evidence for the view that gesturing facilitates the speaking process (Morrel-Samuels & Krauss 1992; Rimé & Schiaratura 1991; de Ruiter 1995b, 1998), implying that communication could indirectly benefit from the gesturing of the speaker. This could be the reason why speakers do not suppress their gesturing in situations without visual contact.

To conclude, I will assume that gesture is a communicative device from the speaker's point of view. The *effectiveness* of gestural communication is another issue that will not be addressed here.

To extend Levelt's model for speaking to incorporate gesture, the first question to be answered is where gestures originate from. In information-processing terminology, what process is responsible for the initiation of a gesture? People do not gesture all the time, nor do they gesture about every-

thing that is being said, so some process must 'decide' whether or not to gesture and what to gesture about.

Considering the formulation of Levelt's model for speaking, the main candidates in that model for the initiation of gesture are the conceptualizer and the grammatical encoder (the first subprocess of the formulator). Butterworth & Hadar (1989) have suggested that iconic gestures are generated from lexical items. If they are correct, the process of lemma retrieval (a subprocess of grammatical encoding) would be responsible for the initiation of gesture. However, there is ample evidence that most gestures cannot be associated with single lexical items. In a corpus of gestures by Dutch subjects (de Ruiter 1998), many subjects drew a horizontal ellipse in the air with the index finger, while saying, "liggend ei" [ENG 'lying egg']. The gesture did not represent the concept of 'lying', nor did it express the concept 'egg'. Rather it represented simultaneously the concepts of 'lying' and 'egg' together. Similarly, in data from the McNeill Gesture Laboratory in Chicago, a subject speaks about a small bird in a cartoon throwing a bowling ball down into a drainpipe (see McNeill & Duncan, this volume; McNeill, this volume). During the utterance, the subject performs a 'pantomimic' gesture of this action. The gesture reveals aspects of the bowling ball itself, of holding it, of throwing it, and of throwing it in a downwards direction. If gestures were associated with lexical items, one would expect that a gesture would reveal only information that is semantically equivalent to the meaning of the 'lexical affiliate'.

Given the fact that many iconic gestures reveal properties that can, at best, only be represented by phrases, the conclusion is that gestures do not have lexical affiliates but rather 'conceptual affiliates'. While it is possible to argue that gestures do have a lexical affiliate, such a proposal would be neither parsimonious nor empirically supported even when there is more information in the gesture than in the affiliate. There is much evidence, most notably from McNeill (1992), that gestures are synchronized with and meaningfully related to higher-level discourse information. The notion of a conceptual affiliate can also explain the occurrence of the occasional gesture that seems to be related to a single word (such as pointing upwards while saying "up"). All content words have an underlying conceptual representation, but not all conceptual representations have a corresponding content word.

Another reason why the grammatical encoder is an unlikely candidate for initiating gestures is the fact that the formulator's input is a *preverbal message* which is a propositional representation. In other words, it does not have access to imagistic information in working memory.

Gestures often represent spatial information that cannot be part of the preverbal message. While it is possible to grant the formulator access to

non-propositional information, that would imply a radical change in Levelt's speaking model. My goal is to leave the core assumptions of the speaking model unchanged as much as possible, for a number of reasons. First, there is an extensive body of literature devoted to testing and investigating Levelt's model and its underlying assumptions. This literature will, for a large part, still be relevant to the speech/gesture model if the speech/gesture model is compatible with it. Second, developing a new speaking model is not only beyond the scope of the present chapter, but also unnecessary, as I hope to demonstrate below.

Aside from these considerations, the conceptualizer is the most natural candidate for the initiation of gesture. Many of the problems the conceptualizer has to solve for speech (e.g., perspective-taking) also apply to the gesture modality. Furthermore, selecting which information should be expressed in which modality is a task that is very similar to Levelt's notion of *Macroplanning*, which "... consist[s] in the elaborating of the communicative intention as a sequence of subgoals and the selection of information to be expressed (asserted, questioned, etc.) in order to realize these communicative goals" (1989: 107).

Because the conceptualizer has access to *working memory*, it can access both propositional knowledge for the generation of preverbal messages and imagistic (or spatio-temporal) information for the generation of gestures. The conceptualizer will be extended to send a representation called a *sketch* to subsequent processing modules. Because of the central role the sketch plays in the model, I will call it the Sketch Model.

3.1 The conceptualizer

3.1.1 When to gesture. People do not gesture all the time, nor do they gesture about everything they speak about. McNeill (1997) has proposed that gestures occur when new elements are introduced in the discourse. While I have no reason to disagree with this analysis, there are other factors involved as well. In some cases, it might be necessary to express certain information in gesture, as is most notably the case in pointing gestures that accompany deictic expressions. If someone says, "John is over there," without pointing to a location, or when someone says, "It is shaped like this," without in some way providing shape information by gesture, the utterance as a whole is semantically incomplete.

Even when gesturing is not obligatory, the conceptualizer can generate a gesture which would serve the function of enhancing the quality of the communication. 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.

People are often found to gesture when their speech is failing in some way. Krauss et al. (1991) and Butterworth & Hadar (1989) claim that gesturing in case of a speech failure helps the speech system resolve the problem, for instance by providing the lexical search process with a cross-modal accessing cue. Another interpretation, which I prefer, is that the temporary speech failure is recognized in the conceptualizer (e.g., by means of internal or external feedback, as in Levelt's model). This recognized speech failure could then be compensated for by the transmission of a larger part of the communicative intention to the gesture modality. Similarly, when circumstances are such that it is difficult to express a communicative intention in speech (e.g. in a noisy environment or when one does not speak the language), gestures can be generated to compensate for the lack of communicative efficiency in the verbal modality.

The assumption that information that is hard to encode as a preverbal message is encoded in a sketch would also explain why narratives involving salient imagery such as motion events will usually evoke many iconic gestures: the conceptualizer applies the principle that a gesture can be worth a thousand words. As mentioned above, the question whether transmitting information by gesture is always *effective*, i.e., whether the listener will detect and process the information represented in gesture, is another issue.

3.1.2 What to gesture. 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. It is the conceptualizer's task to extract the relevant information from a spatio-temporal representation and create a representation that can be transformed into a motor program. The result of this extraction process will be one or more spatio-temporal representations that will be stored in the sketch. Because these representations can involve spatial elements combined with motion, I will call these 'trajectories', for lack of a better term.

Emblems have a lexicalized, hence conventional, shape, so they cannot be generated from imagery. I will assume that the conceptualizer has access to a knowledge store that I will call the *gestuary*. In the gestuary, a number of emblematic gesture shapes are stored, indexed by the concept they represent. If a certain propositional concept is to be expressed, or a certain rhetorical effect intended, the conceptualizer can access the gestuary to see if there is an emblematic gesture available that will do the job. If there is, a reference (e.g., a 'pointer') to this emblematic gesture will be put into the sketch.

A third possibility is that the conceptualizer generates a *pantomimic* gesture. A pantomimic gesture is an enactment of a certain movement performed by a person or animate object in the imagistic representation. A

good example from the McNeill Gesture Laboratory in Chicago is a Tweety Bird narration in which a subject says, "and Tweety drops the bowling ball into the drainpipe" while moving both hands as if the subject herself is throwing a bowling ball down. This kind of gesture cannot be generated from imagery alone, but has to be generated from (procedural) motoric knowledge. The example makes clear why this is the case; Tweety is about half a bowling ball high, and therefore the movement that Tweety makes when throwing the bowling ball is quite different from the movement the (much larger) subject in the study makes in enacting the throwing. For encoding pantomimic gestures, a reference to a motor program (e.g., an action schema; Schmidt 1975) will be encoded in the sketch.

Finally, it is also possible to encode a *pointing* gesture into the sketch. This is done by encoding a vector in the direction of the location of the referent. Since the handshape of the pointing gesture is conventionalized (Wilkins 1995), and for some languages different types of pointing handshapes indicate the level of proximity of the referent, the conceptualizer will have to encode a reference to the appropriate pointing template in the gestuary.

As with the production of speech, another important issue that has to be resolved by the conceptualizer is that of *perspective*. If spatio-temporal information is stored in four dimensions (three for space and one for time), different perspectives can be used to represent the information using gesture. For instance, if a gesture is accompanying a route description (assuming that speaker and listener are facing each other), the gesture that might accompany the speech "taking a right turn" might be made to the speaker's right (speaker-centered perspective) or to the speaker's left (listener-centered perspective).

In some cultures, iconic gestures preserve *absolute* orientation (Haviland 1993, this volume; Levinson 1996), so in that case the conceptualizer has to specify the orientation of the gesture within an absolute-coordinate system. For details about gestural perspectives, see McNeill (1992). A convenient way of encoding perspective in the sketch is by specifying in the sketch the position of the speaker's body relative to the encoded trajectories.

To summarize, the final output of the conceptualizer, called a *sketch*, contains the information in Table 14.1. The sketch, which will be sent to the *gesture planner*, might contain more than one of the representations above. How multiple sketch entries are processed will be described below.

3.2 The gesture planner

The gesture planner's task is to build a motor program out of the received sketch. The gesture planner has access to the gestuary, motor procedures

Table 14.1. *Gesture output of the conceptualizer*

Gesture type	Sketch content
Iconic	One or more spatio-temporal trajectories Location of speaker relative to trajectory
Deictic	Vector Reference to gestuary
Emblem	Reference to gestuary
Pantomime	Reference to motor-action schema

(schemata), and information about the environment. One of the reasons a separate module is specified for gesture planning is that the constraints that have to be satisfied in gesturing are radically different from those of manipulating the environment in 'standard' motor behavior.

A problem that the gesture planner has to solve, for all gestures, is that of *body-part allocation*. One hand may be occupied, in which case either the other hand must be used for the gesture, or the occupied hand must first be made available. If both hands are unavailable, a head gesture can sometimes be generated. For example Levelt, Richardson & La Heij (1985) had subjects point to one of a few lights while saying, "This light" or "That light." In one of their conditions, subjects were not allowed to point. The authors note that "When no hand gesture is made, the speaker will still direct his gaze or head toward the target LED; there will always be some form of pointing" (p. 143).

This illustrates one of the problems the gesture planner has to solve. The sketch specifies a location to be expressed in a deictic gesture, but the hands are not allowed to move. Therefore, the gesture planner selects another 'pointing device' to perform the (in this case obligatory) gesture. The fact that the same 'logical' gesture can often be realized overtly by different physical gestures provides support for the assumption that gesture sketch generation and the generation of a motor program are separate processes.

Another task of the gesture planner is to take into account restrictions that objects in the environment impose upon body movements. At a crowded party, too large a gesture could result in hitting another person. Although it probably happens, it seems reasonable to assume that people normally do not hit other people or objects during gesturing. If the environment imposes certain restrictions, the gesture will either be canceled or adapted to fit the circumstances.

For the generation of emblems and pointing gestures, the gestuary plays an important role in the creation of a motor program from the sketch. In

the gestuary, information is stored about gestural conventions. For instance, while it is possible to point to a location using the elbow or the little finger, in most Western European cultures people point with the index finger. There is a 'soft rule' that specifies that the preferred way to point is to use the index finger of the hand. However, when pointing to a location behind the back, Europeans will usually use the thumb (Calbris 1990). Speakers of Arrernte (a Central Australian language) will use the index finger for such backward pointing (Wilkins 1995).

It is not possible to have complete motor programs stored in the gestuary for any of these gesture types. Both pointing gestures and emblems have a number of degrees of freedom. In performing the emblematic OK gesture, there is freedom in the location of the hand and the duration of the gesture. The only aspect of this gesture that is fixed is the shape and orientation of the hand. The same holds for pointing: while the shape of the hand is subject to certain conventions, the location that the hand points to is dependent upon where the object of interest happens to be. It is therefore necessary to store gestures in the gestuary in the form of *templates*. A template is an abstract motor program that is specified only insofar as it needs to be. Taking the emblematic OK gesture as an example, the handshape is fully specified in the template, while duration, hand (left or right), and location (where the hand is held) are free parameters. The gesture planner will bind these free parameters to the desired values in order to obtain a fully specified motor program.

If the sketch contains one or more trajectories, the gesture planner has to convert them to motor programs that represent the information in these trajectories. In the simple case of one trajectory, the body part (usually the hand) can often be used to 'trace out' the trajectory, but things can be far more complicated. A more complex problem the gesture planner will have to solve is the generation of so-called *fusions* of different gestures. To take an example from data from Sotaro Kita (pers. com.), a subject is enacting a throwing movement, but adds a directional vector on top of the throwing enactment by 'throwing' the ball sideways, which was not the way the person in the stimulus film threw the ball. However, the movement sideways indicated that (from the speaker's viewpoint) the ball flew off in that particular direction. We must therefore conclude that there has been a fusion of a deictic component (indicating the direction of the ball) and an enactment. This type of fusion of different information in one gesture occurs frequently. If the fusion gesture involves a gestuary entry or an action schema, it is hypothesized that the unbound parameters of the template or action schema (i.e., its degrees of freedom) will be employed, if possible, to encode additional sketch elements. Example C below serves to illustrate this mechanism.

Further research is needed to find out which types of gesture can be fused together, and under which circumstances fusion of information is likely to occur. Of specific interest is also the question in what order information in the sketch will be encoded in case of a fusion gesture. If, for instance, the sketch contains both a pointing vector and an iconic trajectory, the degrees of freedom left over by the pointing template could be used to represent the iconic component; but the reverse is also possible: the degrees of freedom left over by the iconic gesture might be used to encode a deictic component into the gesture.

Once the gesture planner has finished building a motor program, this program can be sent to lower-level motor control units, resulting in overt movement.

To summarize the definition of the gesture planner: upon receiving a sketch, it will

- Generate a gesture from the sketch by retrieving a template from the gestuary (for pointing gestures and emblems), by retrieving an action schema from motoric memory (for pantomimes), or by generating a gesture from the trajectories (for iconic gestures) in the sketch.
- Allocate one or more body parts for execution of the gesture.
- Try to encode other sketch entries into the gesture as well.
- Assess potential physical constraints posed by objects in the environment.
- Send the motor program to the lower-level motor-control module(s).

The complete Sketch Model is graphically represented in Figure 14.1. Boxes represent processes, arrows represent representations that are sent from one process to another, ellipses represent knowledge stores, and dotted lines represent access to a particular knowledge store.

4 Synchronization

So far, nothing has been said about the temporal synchronization of gesture and speech. The Sketch Model provides the opportunity to hypothesize in detail how synchronization is achieved.

It should be pointed out that the issue of temporal synchronization is a nebulous one. It is problematic even to define synchronization. The conceptual representation (the 'state of affairs' [Levelt 1989] or the 'Idea Unit' [McNeill 1992]) from which the gesture is derived might be overtly realized in speech as a (possibly complex) phrase, and is not necessarily realized overtly as a single word. Therefore, it is by no means straightforward to unambiguously identify the affiliate of a given gesture. Even if a speech fragment has been identified as being the affiliate, it has a certain duration, and so does the gesture. The synchrony between gesture and speech is the

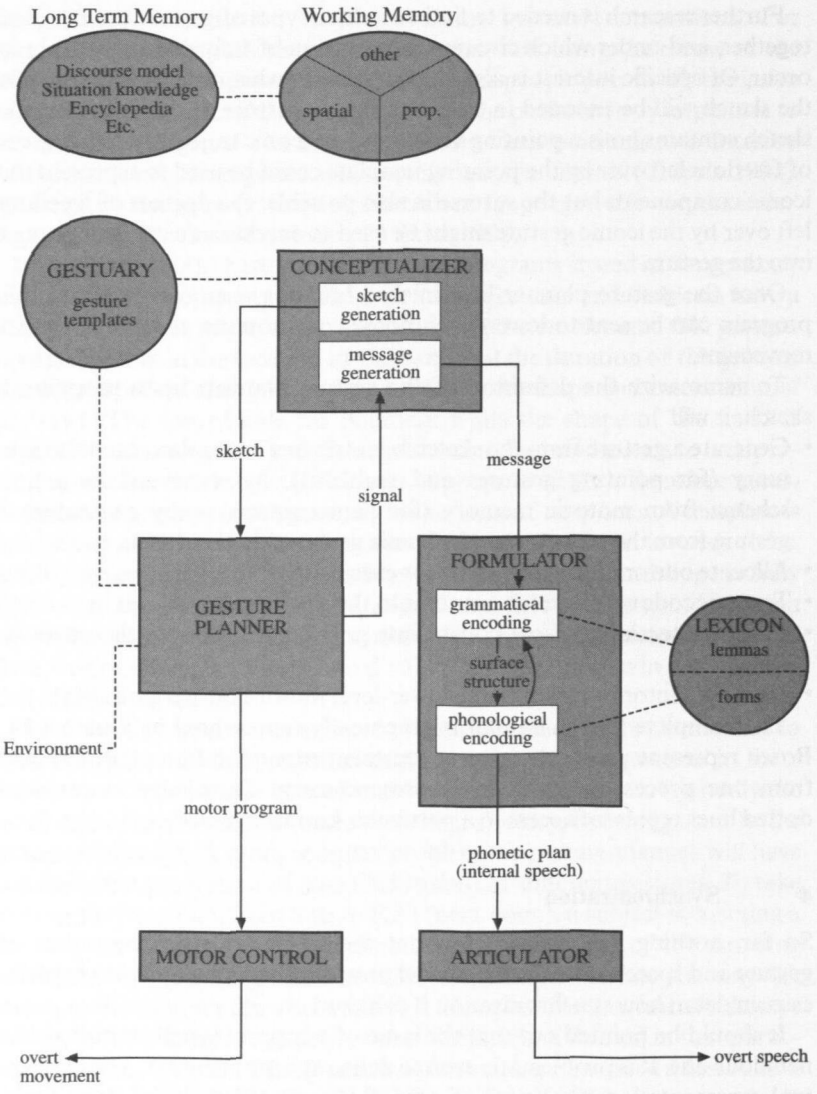

Figure 14.1. The Sketch Model.

synchrony between two time intervals that are often hard to define. However, there is evidence that the *onset* of gesture usually precedes the *onset* of the accompanying speech by a duration of less than a second (e.g., Morrel-Samuels & Krauss 1992; Butterworth & Beattie 1978; Nobe 1996, this volume). Butterworth & Hadar (1989) conclude that

> despite the fact that gestures may continue until after speech onset, the beginnings of movements must be considered as events potentially separable from speech onsets, and any processing model must take this fact into account. (P. 170)

The Sketch Model accounts for this phenomenon by assuming that although the preverbal message and the sketch are constructed simultaneously, the preverbal message is sent to the formulator only after the gesture planner has finished constructing a motor program and is about to send it to the motor-execution unit. Once the motor program for the gesture has been constructed, the gesture planner will send a message to the conceptualizer specifying when the generation of speech can be initiated.

This mechanism also accounts for other empirical findings, such as Levelt et al.'s (1985) finding that the onset of speech is adjusted to the duration of the gesture even though the gesture is not yet performed.

However, other synchronization phenomena need to be explained. The first to be addressed is the *gestural hold*, which comes in two varieties. In the *pre-stroke* hold, the gesturing hand moves towards its initial position and then waits for the accompanying speech to be produced before performing the stroke (meaningful part) of the gesture. In the *post-stroke* hold, the hand remains motionless after the stroke has been completed until the related speech has been fully produced. This phenomenon led Kita (1990) to claim that gesture is waiting for the accompanying speech to catch up.

The pre-stroke hold can be accounted for by assuming that the sketch can be sent to the gesture planner before the construction of the preverbal message has been finished. This allows the gesture planner to prepare the motor program for the sketch and put the hand(s) in the initial position for the stroke of the gesture.[3] When the preverbal message is finally sent to the formulator, the conceptualizer will send a 'resume' signal to the gesture planner, which will then send the rest of the motor program (the stroke of the gesture, corresponding to the sketch) to the motor units.

The post-stroke hold can be explained by assuming that the conceptualizer refrains from sending a retract signal to the gesture planner until the production of a preverbal message has been completed. The conceptualizer can detect the completion of a preverbal message fragment, owing to the internal and external feedback loop in the speech-production model. This mechanism does not only offer an account for the occurrence of the post-stroke hold, but also for another interesting finding by Kita (1990). He

found that repetitive gestures (e.g., a pantomimic gesture for sawing or hammering) are less likely to have post-stroke holds than are non-repetitive gestures. Instead of the hand stopping at the final position, the repetitive movement is repeated until the related speech fragment has been completed. This difference between repetitive and non-repetitive gestures could be a consequence of the motor programs that are constructed by the gesture planner. For a repetitive gesture, the motor program is specified as a 'loop' – when the stroke has been completed, it starts all over again. Therefore, it will continue to repeat until it receives a retract signal. If non-repetitive gestures are completed before the retract signal, there are no motor instructions left, so the hand stays in its stroke-final position.

Another phenomenon that could be seen as a kind of synchronization is the finding by Kita (1993) that if speakers interrupt their own speech because they detect an error in it, the gesture is often interrupted simultaneously with the speech. In the sketch model, the interruption of both the speech and the gesture is initiated by the conceptualizer. Upon detection of an error in the speech, the conceptualizer sends a stop signal to the formulator and to the gesture planner. These modules pass on this stop signal to lower processing modules.

If future studies reveal a tighter synchronization between iconic gestures and speech than the model accounts for at the moment, the Sketch Model will have to be adapted to incorporate these findings (and will be, in its present formulation, falsified). However, such studies will have to address a number of problems. First of all, synchronization should be defined in such a way that it is possible to locate the affiliate of any iconic gesture unambiguously. Second, synchronization should be defined carefully. Butterworth & Hadar (1989) pointed out that there are thirteen types of temporal relation between two time intervals. Since we can ignore the possibility that two points in time will be *exactly* synchronous, we are still left with six types of temporal relation from which to choose in defining the meaning of 'synchrony'. Finally, there is a measurement problem. Once the affiliate has been defined, the relevant speech interval can be measured to some degree of accuracy, but for gestures this is much harder. Locating the beginnings and ends of gestures (even if restricted to the stroke) is often problematic, especially when gestures follow each other rapidly.

It is tempting to interpret synchronization phenomena as evidence for interactive theories of speech/gesture production, in which lower-level speech and gesture processing continuously exchange information. To paraphrase Kita (1993), because interactive theories assume that speech processes and gesture processes always have access to each other's internal states, gesture can stop when speech stops. Plausible as this may seem, the explanation is incomplete as long as it does not specify what information

about what processes is shared in the course of processing. There are a multitude of ways in which speech and gesture processes could share information about their respective internal states. As I hope to have demonstrated in the formulation of the Sketch Model, the computational problems to be solved in the generation of gesture on the one hand and speech on the other are of an entirely different nature. Therefore, sharing all internal state information all the time is not necessary, and probably amounts to interactive 'overkill'. For the same reason, the assumption in growth-point theory that the generation of gesture and the generation of speech are the same process is suspect. It might be the same *general* process, but then this process must perform two rather different computations: one for gesture and one for speech. That again raises the question what information is shared between these two computations.

5 Some examples

A few examples will be helpful in illustrating the mechanics of the proposed model.[4] I will illustrate the Sketch Model by explaining what happens in case of (a) an iconic gesture, (b) a pantomimic gesture, and (c) an Arrernte pointing gesture.

Example A: an iconic gesture. A Dutch subject talks about a Sylvester and Tweety Bird cartoon she just saw. The fragment she is going to describe starts with a big sign saying, "Bird watchers society." She says:

> "Op den duur zie je zo'n eh, eh, vogelkijkersvereniging ofzo . . ."
> (After a while one sees a eh eh bird watchers' society or something)

Roughly during production of the compound "vogelkijkersvereniging" (ENG 'bird watchers' society') the subject uses both index fingers to draw a large rectangle in front of her body.

Computations in the conceptualizer result in the encoding of the introduction of the bird watchers' society in the speech channel. However, the fact that the bird watchers' society was introduced in the cartoon by showing a sign affixed to a building was encoded in the gesture channel. Therefore, a preverbal message corresponding to "vogelkijkersvereniging ofzo" (ENG 'bird watchers' society or something') is sent to the formulator, while a sketch containing the large rectangle is sent to the gesture planner. The hesitation before and after the described gesture/speech fragment was produced suggests that this part of the communicative intention was a separate fragment (or 'chunk', as Levelt 1989 calls it). Because the sketch and the preverbal message are sent at the same time, the gesture and the speech are synchronized roughly. Interestingly, the speech does contain enough

information to understand the cartoon fragment, but only by observing the gesture can the listener tell that there was a sign involved in the cartoon.

This example also illustrates how the conceptualizer can distribute information over different output modalities.

Example B: a pantomimic gesture. Another subject describing the Sylvester and Tweety Bird cartoon says:

> ". . . enne, da's dus Sylvester die zit met een verrekijker naar de overkant te kijken"
> (and eh so that is Sylvester who is watching the other side with binoculars)

During the production of "met een verrekijker naar" (ENG 'with binoculars at') the subject raises his hands in front of his eyes as if lifting binoculars to his eyes.

This example illustrates how difficult it can be to establish what the speech affiliate of the gesture is. In this fragment, it is hard to decide whether the gesture corresponds to "met een verrekijker" (ENG 'with binoculars') or to "zit met een verrekijker naar de overkant te kijken" (ENG 'who is watching the other side with binoculars'). In the latter case, the synchronization of gesture and speech violates the tendency formulated by Butterworth & Hadar (1989) that gesture onset precedes speech onset. We could therefore assume that the affiliate is the stretch of speech that is synchronized with the gesture. However, this leads inevitably to circular reasoning. Either we infer the affiliate from synchronization, or we infer synchronization from the affiliate. It can't be both at the same time, unless the meaning relation of a gesture and the accompanying speech can be established unambiguously, as is for instance the case with most deictic gestures. As example A illustrates, gesture and speech can communicate different aspects of the communicative intention, so the affiliate is not necessarily semantically related to the gesture. In case the meaning relation between gesture and speech is not clearcut, the logic of the Sketch Model requires one to infer the affiliate from the synchronized speech, because in the Sketch model the assumption is made that the sketch and the preverbal message are sent at the same time. In this example the conceptualizer, according to the Sketch Model, encodes a preverbal message corresponding to "met een verrekijker," or possibly "met een verrekijker naar de overkant." At the same time, a sketch is prepared with an entry to the motor schema for holding binoculars. The formulator then processes the preverbal message, and the gesture planner constructs a motor program from the specified motor schema.

Example C: an Arrernte pointing gesture. A (right-handed) speaker of Arrernte and Anmatyerre (Central Australia) is holding a baby in her right arm. (Prior to holding the baby she had made a pointing gesture with her right hand.) Now she says:

> Ilewerre, yanhe-thayte, Anmatyerre
> [place-name] [that/there(mid)-SIDE] [language/group name]
> "(The place called) Ilewerre, on the mid-distant side there, is Anmatyerre (people's country)."

Roughly during the utterance of "yanhe-thayte" she points to the south-east with her left arm. The hand is spread, and the arm is at an angle of approximately 90 degrees from the body.

Wilkins (pers. com.) has observed that the orientation of the palm of the Arrernte spread-hand pointing gesture matches the orientation of the surface that is being referred to. To paraphrase one of Wilkins' field notes, if the palm of the hand is facing out and vertical, it could indicate (for instance) paintings spread out over a cliff face.

Such a gesture can be interpreted as a fusion of an iconic gesture (representing the surface orientation) and a conventionalized deictic gesture. This provides support for the hypothesis that the way the gesture planner realizes fusions is to utilize degrees of freedom in gesture templates to represent additional iconic elements represented in the sketch.

Using the Sketch Model, this fragment can be described in the following way. On the basis of geographical knowledge, the conceptualizer chooses the mid-distant proximity for the deictic reference, and encodes it in the preverbal message. The indication of proximity is not sufficient to realize the communicative intention, so the conceptualizer will generate a sketch containing a vector in the direction of the location of the indicated place. The conceptualizer accesses the gestuary to find the appropriate pointing gesture. In Arrernte, both the shape of the hand and the angle of the arm in pointing are meaningful and conventionalized. The pointing gesture that is selected from the gestuary in this example is one with an arm angle of 90 degrees. Given that the indicated place is not visually available from that vantage point, and is at a significant distance away, the 90-degree angle corresponds with the mid-distant proximity. The spread hand is used to identify a region, something spread out over an area. The entry in the gestuary for this type of pointing gesture specifies hand shape and arm angle, but leaves the parameters specifying the planar angle of the gesture, the handedness of the gesture, and the orientation of the palm of the hand undefined. Finally, the sketch will contain a trajectory that represents a horizontal plane, indicating that the region is spread out over the ground.

The sketch containing the vector, the entry into the gestuary, and the horizontal plane trajectory is sent to the gesture planner. The gesture planner will retrieve the template for the pointing gesture. It will also notice that the right hand is occupied holding a baby, so it will bind the handedness parameter of the gesture template to 'left'. It will bind the parameter specifying the planar angle of the gesture to the angle of the vector that is specified in the sketch, in this case, the southeast. Finally, since the indicated region is an area (flat), the orientation of the hand will be specified as horizontal. Now all free parameters of the template are specified, yielding a complete motor program to be executed.

6 Discussion

To summarize, the most important assumptions of the Sketch Model are:

- The conceptualizer is responsible for the initiation of gesture.
- Iconic gestures are generated from imagistic representations in working memory.
- Gestures are produced in three stages:
  1. The selection of the information that has to be expressed in gesture (the sketch).
  2. The generation of a motor program for an overt gesture.
  3. The execution of the motor program.
- Different gestures can be 'fused' by an incremental utilization of degrees of freedom in the motor program.
- Apart from the conceptualizer, gesture and speech are processed independently and in parallel.

The model covers a large number of gesture types: iconic gestures (including what McNeill 1992 calls *metaphoric* gestures), pantomimes, emblems, and pointing gestures. The Sketch Model also accounts for a number of important empirical findings about gesture. The semantic synchrony of gesture and speech follows from the fact that both gesture and speech ultimately derive from the same communicative intention. Iconic gestures are derived from imagistic representations, while speech output is generated from propositional representations. The global temporal synchrony of iconic gestures and speech is a consequence of preverbal message (speech) and sketch (gesture) being created at approximately the same moment, while the tight temporal synchrony of (obligatory) deictic gestures is accomplished by a signal from the motor unit to the phonological encoder.

As has been mentioned before, it is important to note that in this model, iconic and metaphoric gestures as defined by McNeill (1992) are indistinguishable. Both types of gestures are generated from spatio-temporal representations in working memory. The fact that in the case of a metaphoric

gesture the spatio-temporal representation is about abstract entities has no consequence for the transformation of this representation into an overt gesture. On the other hand, pantomimic gestures, while being a subclass of iconic gestures in McNeill's taxonomy, have to be treated in a different way from other iconic gestures, owing to the fact that pantomimic gestures can't be generated from an imagistic representation alone, as explained in section 3.1.2.

There is one category of gestures that is not incorporated by the Sketch Model, namely, beats. The reason for this omission is that there is, at present, insufficient knowledge available about beats. McNeill (1992) has proposed that beat gestures serve metanarrative functions. Lacking detailed processing accounts for the metanarrative level of speech production, incorporating McNeill's proposal in the model is, at present, not possible. While it is sometimes possible to hypothesize the role of beats in a given speech/gesture fragment, it is still impossible to *predict* their occurrence using convenient landmarks: "Beats . . . cannot be predicted on the basis of stress, word class, or even vocalization itself" (McClave 1994: 65).

Since a major advantage of an information-processing model is its vulnerability to potential falsification, it is important to point out a number of predictions that can be derived from the model.

First, the model predicts that people sometimes do *not* gesture. When people read written material aloud, or when they are quoting (repeating) someone, they are predicted not to generate spontaneous gestures.[5] In quoting or reading, the conceptualizer is not constructing a new preverbal message and/or sketch. As the Sketch Model assumes that the generation of gestures is tightly coupled with the generation of preverbal messages, there will be no spontaneous gestures either.

With respect to the synchronization of iconic gestures with the related speech, the model makes the prediction that the onset of the iconic gesture is (roughly) synchronized with the onset of the overt realization of the conceptual affiliate in speech (usually a noun phrase or a verb phrase), independent of the syntax of the language. For example, if an English-speaker says, "He runs across the street," the onset of the iconic gesture that represents 'running across something' is predicted to be roughly co-occurring with the onset of the first word of the verb phrase, in this case "runs." If a Japanese-speaker says "miti o wata te" (street go-across), the onset of the iconic gesture is predicted to be synchronized with the onset of "miti" (street); although it has a different meaning from the first word in the English sentence, it is again the first word of the verb phrase.[6]

Another prediction concerning the synchronization of gesture and speech is that the model does not permit representations active in the formulator to influence the timing of the gesture. Lexical stress or pitch accent,

for instance, are therefore predicted to have no effect on gesture/speech synchronization.

A final, rather straightforward prediction of the Sketch Model is that gestures can only be interrupted (e.g., when there is a problem in the generation of speech) during the preparation phase of the gesture. Once the gesture stroke has started to execute, it can no longer be interrupted.

6.1 A comparison with growth-point theory

In comparing the Sketch Model with McNeill's (1992) growth-point (GP) theory, it is possible to point out both similarities and differences. The main similarity is that according to both the Sketch Model and GP theory, gesture and speech originate from the same representation. In the Sketch Model this is the communicative intention, while in GP theory it is the growth point. "The growth point is the speaker's minimal idea unit that can develop into a full utterance together with a gesture" (McNeill 1992: 220).

McNeill (1992) discusses and dismisses theoretical alternatives (pp. 30–35) for GP theory. His analysis applies equally well to the Sketch Model, because of two important assumptions that underlie both GP theory and the gesture model: gestures and speech are part of the same communicative intention, and are planned by the same process.

However, GP theory does not give any account of how (in terms of processing) growth points develop into overt gestures and speech. The growth point is an entity whose existence and properties are inferred from careful analyses of gestures, speech fragments, and their relation (McNeill 1992: 220). However, without a theory of how a GP develops into gesture and speech, gesture and speech data can neither support nor contradict GP theory. This also introduces the risk of circularity. If a fragment of speech that is accompanied by a gesture is interpreted as a growth point, and the growth point is also the entity responsible for the observed (semantic and temporal) synchronization, the observation and the explanation are identical. Therefore, GP theory in its present form does not explain how the speech/gesture system actually accomplishes the observed synchrony (cf. McNeill, this volume).

6.2 A comparison with the model of Krauss et al. (1996, this volume)

Krauss, Chen & Chawla (1996) and also Krauss, Chen & Gottesman (this volume) have also formulated a model that is based on Levelt's (1989) model. (For convenience, I will call their model the KCG model.) The most important difference from the Sketch Model is that in the KCG model the conceptualizer from Levelt's model is left unchanged, whereas in the Sketch

Model it is modified extensively. In the KCG model, gestures are not generated by the conceptualizer, but by a separate process called the spatial/dynamic feature selector. These features are transformed into a motor program which helps the grammatical encoder retrieve the correct lemma for the speech. Synchronization is accounted for in the KCG model by the assumption that the auditory monitor can terminate gestures upon perceiving the (spoken) lexical affiliate.[7] In the Sketch Model the assumption by Levelt (1989) is adopted that the monitor can use both overt speech and the output of the phonological encoder ('inner speech') to terminate gestures.

The main assumption of the KCG model is that iconic gestures (lexical gestures, in their terminology) are not part of the communicative intention, but serve to facilitate lemma selection. However, if the spatio-dynamic features in the generated gesture are to facilitate lemma selection, it is essential that the features that, taken together, single out a particular lemma are identifiably present in the motoric realization of the gesture. If the gesture contains features that are not associated with the lemma that is to be retrieved, these features will very likely *confuse* (hence slow down) lemma selection, because more than one lemma will be activated by the features in the gesture. The example given by Krauss et al. is a gesture about a vortex. A gesture representing a vortex might indeed contain features that, taken together, would help in retrieving the lemma for 'vortex'. Interestingly, a gesture containing these features will be very easy to identify as representing a vortex, even by a listener who does not have access to the accompanying speech. Especially when, as Krauss et al. assume, the features present in the gesture are to be apprehended proprioceptively, these features must be clearly present in the overt realization of the gesture.

This is in contradiction to the experimental findings reported in Krauss et al. that indicate that most iconic or lexical gestures are not easily recognizable at all without access to the accompanying speech. The paradox here is that those gestures that could facilitate lemma selection must be gestures whose meaning is largely unambiguous (such as the gesture for 'vortex'). Gestures that are hard to interpret without the accompanying speech will usually not contain enough information to facilitate lexical selection.

Another problem with the KCG model is that if gestures are indeed derived from spatio-dynamic working memory (an assumption their model shares with the Sketch Model), there is no reason to expect that there exists a single lemma that has the same meaning as the gesture. As mentioned above, in many cases *phrases* are needed to describe or approximate the meaning of a gesture. Therefore, if gesturing facilitates the speaking process, this is more likely to involve more complex (higher-level) representations than lemmas. For instance, de Ruiter (1995b, 1998) found evidence

suggesting that performing iconic gestures facilitates the retrieval of imagery from working memory.

Synchronization in the KCG model also differs from that of the Sketch Model. While the mechanism Krauss et al. propose is more parsimonious than that of the Sketch Model, it is doubtful whether it will explain all synchronization phenomena. The KCG model does not provide an account for the post- and pre-stroke hold phenomenon. In the KCG model, gestures are terminated once the corresponding lexical affiliate has been articulated, but especially the pre-stroke hold phenomenon indicates that gestures can also be *initiated* in synchronization with speech.

7 Conclusions

In investigating the processing involved in gesture and speech production, it is a great advantage to develop theories in the framework of information processing. IP theories are less prone to multiple interpretations, easier to falsify, and facilitate the generation of new research questions. These advantages become even more salient when the theory is in a stage of development that allows for computer simulations. The IP approach is a theoretically neutral formalism that enhances the resolution of our theories without restricting their content, apart from the fact that the IP approach does not allow the use of homunculi as explanatory devices.

The Sketch Model is an attempt to incorporate and explain many well-established findings in the area of gesture and speech with a largely modular IP model. It incorporates a large amount of knowledge about gesture and speech in a consistent framework. As has been shown, predictions can be generated from the Sketch Model that can be tested in future experiments. The Sketch Model allows for detailed specification of hypotheses about the synchronization between gesture and speech. It can also accommodate cross-cultural variation in gesture behavior, and proposes a new and detailed account for the fusion of information in gestures. Because the Sketch Model is an extension of Levelt's (1989) model, the model implicitly incorporates a large amount of accumulated knowledge about speech production.

While the formalism used to specify the Sketch Model is similar to the formalism used by Krauss et al. (1996), the underlying assumptions of the Sketch Model are more similar to growth-point theory. Most notably, the assumption that gestures and speech are planned together by the same process and ultimately derive from the same representation is made in both the Sketch Model and in growth-point theory. However, in contrast to the Sketch Model, growth-point theory does not specify how a growth point develops into overt speech and gesture.

NOTES

This research was supported by a grant from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, Munich. I wish to thank Asli Özyürek for initially suggesting to me that I write about this topic. The chapter owes a great deal to many interesting discussions with Susan Duncan about information processing and growth-point theory. Finally, I wish to thank (in alphabetical order) Martha Alibali, Susan Duncan, Sotaro Kita, Pim Levelt, Stephen Levinson, David McNeill, Asli Özyürek, Theo Vosse, and David Wilkins for their invaluable comments on earlier versions of the chapter.

1 Here McNeill cites J. A. Fodor (1980), 'Methodological solipsism considered as a research strategy in cognitive psychology', *Behavioral and Brain Sciences* 3: 63–73.
2 Wheeldon & Levelt (1995) found evidence suggesting that internal speech consists of a syllabified phonological representation.
3 The gesture planner knows which part of the motor program is the meaningful part (stroke) because only the stroke is constructed from the information in the sketch.
4 Examples A and B are taken from videotaped narrations collected at the Max Planck Institute for Psycholinguistics by Sotaro Kita. David Wilkins kindly provided a transcript from his Arrernte field data for example C. Any errors in representing or interpreting these examples are my responsibility.
5 Of course, it is conceivable that people will 'quote' the gestures made by the quotee, but these gestures are not spontaneous.
6 I took these example sentences from McNeill (1992).
7 This is a change from the assumption made in Krauss, Chen & Chawla (1996), where the phonological encoder terminates gestures by sending a signal to the motor planner unit once the affiliated lexical item has been encoded phonologically.

REFERENCES

Besner, D. & Humphreys, G. (eds.) 1991. *Basic Processes in Reading*. Hillsdale, NJ: Erlbaum.

Biemans, M. & Woutersen, M. (eds.) 1995. *Proceedings of the CLS Opening Academic Year 95–96*. University of Nijmegen.

Butterworth, B. & Beattie, G. W. 1978. Gesture and silence as indicators of planning in speech. In Campbell &. Smith (eds.), pp. 347–360.

Butterworth, B. & Hadar, U. 1989. Gesture, speech, and computational stages: a reply to McNeill. *Psychological Review* 96: 168–174.

Calbris, G. 1990. *The Semiotics of French Gesture*. Bloomington: Indiana University Press.

Campbell, R. N. & Smith, P. T. (eds.) 1978. *Recent Advances in the Psychology of Language*, vol. IV: *Formal and Experimental Approaches*. London: Plenum.

Churchland, P. 1986. *Neurophilosophy*. Cambridge, MA: MIT Press.

de Ruiter, J. P. A. 1995a. Classifying gestures by function and content. Paper presented at the Conference on 'Gestures Compared Cross-Lingustically', Linguistic Institute, Albuquerque, July.

de Ruiter, J. P. A. 1995b. Why do people gesture at the telephone? In Biemans & Woutersen (eds.), pp. 49–56.
de Ruiter, J. P. A. 1998. Gesture and speech production. Unpublished doctoral dissertation, University of Nijmegen.
Feldman, R. & Rimé, B. (eds.) 1991. *Fundamentals of Nonverbal Behavior*. New York: Cambridge University Press.
Fodor, J. A. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
Fodor, J. A. & Pylyshyn, Z. W. 1988. Connectionism and cognitive architecture: a critical analysis. *Cognition* 28: 3–71.
Hagoort, P., Brown, C. M. & Swaab, T. M., in press. Lexical-semantic event-related potential effects in left hemisphere patients with aphasia and right hemisphere patients without aphasia. *Brain*.
Haviland, J. B. 1993. Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology* 3: 3–45.
Kendon, A. 1994. Do gestures communicate? A review. *Research on Language and Social Interaction* 27: 175–200.
Kita, S. 1990. The temporal relationship between gesture and speech: a study of Japanese–English bilinguals. Unpublished master's thesis, Department of Psychology, University of Chicago.
Kita, S. 1993. Language and thought interface: a study of spontaneous gestures and Japanese mimetics. Unpublished Ph.D. dissertation, University of Chicago.
Krauss, R. M., Chen, Y. & Chawla, P. 1996. Nonverbal behavior and nonverbal communication: what do conversational hand gestures tell us? In Zanna (ed.), pp. 389–450.
Krauss, R. M., Morrel-Samuels, P. & Colasante, C. 1991. Do conversational hand gestures communicate? *Journal of Personality and Social Psychology* 61: 743–754.
Levelt, W. J. M. 1989. *Speaking*. Cambridge, MA: MIT Press.
Levelt, W. J. M., Richardson, G. & La Heij, W. 1985. Pointing and voicing in deictic expressions. *Journal of Memory and Language* 24: 133–164.
Levinson, S. C. 1996. The body in space: cultural differences in the use of body-schema for spatial thinking and gesture. Paper prepared for the Fyssen Colloquium: Culture and the Uses of the Body, Paris, December.
McClave, E. 1994. Gestural beats: the rhythm hypothesis. *Journal of Psycholinguistic Research* 23: 45–66.
McNeill, D. 1987. *Psycholinguistics: A New Approach*. New York: Harper & Row.
McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
McNeill, D. 1997. Growth points cross-linguistically. In Nuyts & Pederson (eds.), pp. 190–212.
Morrel-Samuels, P. & Krauss, R.M. 1992. Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory and Cognition* 18: 615–622.
Neely, J. H. 1991. Semantic priming effects in visual word recognition: a selective review of current findings and theories. In Besner & Humphreys (eds.), pp. 264–336.
Nobe, S. 1996. Cognitive rhythms, gestures, and acoustic aspects of speech. Unpublished Ph.D. dissertation, University of Chicago.

Nuyts, J. & Pederson, E. (eds.) 1997. *Language and Conceptualization.* Cambridge: Cambridge University Press.

Pylyshyn, Z. W. 1979. Complexity and the study of human and artificial intelligence. In Ringle (ed.), pp. 23–56.

Rimé, B. & Schiaratura, L. 1991. Gesture and speech. In Feldman & Rimé (eds.), pp. 239–281.

Ringle, M. (ed.) 1979. *Philosophical Perspectives in Artificial Intelligence.* Brighton, Sussex: Harvester.

Schmidt, R. 1975. A schema theory of discrete motor skill learning. *Psychological Review* 82: 225–260.

Shannon, C. E. & Weaver, W. 1949. *The Mathematical Theory of Communication.* Urbana: University of Illinois Press.

Wheeldon, L. & Levelt, W. 1995. Monitoring the time course of phonological encoding. *Journal of Memory and Language* 34: 311–334.

Wilkins, D. 1995. What's 'the point'? The significance of gestures of orientation in Arrernte. Paper presented at the Institute for Aboriginal Developments, Alice Springs, July.

Zanna, M. (ed.) 1996. *Advances in Experimental Social Psychology*, vol. XXVII. Tampa, FL: Academic Press.