Interference Between Gestures and Words

Stephen R. H. Langton, B.Sc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, February 1996

Contents

Abstract

This thesis explores the idea that a speaker's gestural and verbal behaviours are mutually influential in the comprehension process. A Stroop-type interference paradigm was adopted as a tool for investigating whether or not listeners process to-be-ignored gestural information and how this information influences the processing of spoken words.

In Experiments 1-4, static pointing (deictic) gestures and corresponding spoken and written words showed symmetrical interference. Incongruent words slowed responses to gestures, and incongruent gestures slowed responses to words, compared with congruent arrangements. These findings support the idea that both pointing gestures and words are processed in comprehension. Furthermore, the results of Experiments 5-11 suggest that the mutual influence of the two dimensions is largely independent of specific stimulus-response compatibilities. Collectively, these findings are difficult to reconcile with models of Stroop interference which place the locus of the effect at response selection. Instead, they are more consistent with the position that the two sources of information interact at a semantic stage of processing.

Arrows (Experiment 12) and spatially positioned dots (Experiment 13) also produced symmetrical interference effects when paired with spoken words, raising the possibility that it is the spatial nature of the pointing gestures which is important in influencing the comprehension of spoken words. In support of this suggestion, other non-spatial gestures such as emblems (Experiment 14), iconics (Experiment 15) and facial gestures (Experiment 16) did not interfere with responses to verbal material. However, symmetrical effects did return when subjects were asked to make affective judgements to either emotional words or schematic facial gestures (Experiment 17).

The results are discussed with reference to research on the orienting of social attention, the stimulus-driven "capture" of attention, models of integration, and a processing framework which incorporates the notion of informational integration at "semantic" levels of processing.

Acknowledgements

There are many people to thank for their support, advice and in some cases, money. I thank my supervisors, Claire O'Malley and Vicki Bruce. I am particularly indebted to Vicki who stands as one of the few people in this list to whom I owe nothing but thanks. As tutor, boss and supervisor she has provided unrelenting and inspirational enthusiasm, time, at least three research jobs, and along with Claire O'Malley, the opportunity to undertake this thesis.

Thanks are also due to members of the communications group at the Universities of Nottingham, Glasgow and Stirling. Along with Claire and Vicki, Anne Anderson, Gwyneth Doherty-Sneddon, Simon Garrod and Stephanie Walker provided the friendly, enjoyable and supportive academic environment which stimulated this work.

I thank David Wood and other members of the ESRC Centre for Research in Development, Instruction and Training at Nottingham for financial support and time to pursue my own work between July 1994 and September 1995.

I would also like to thank my friends and colleagues at Nottingham, in particular Andy Waters for his interest, insight and longer lunch breaks than might otherwise be expected. Thanks to James Bright for good humour, cricket, and a keen understanding of office politics, to Julian Pine for leading the renaissance enjoyed by the department's hitherto unsuccessful five-a-side football team, to David Gilmore for his wisdom, and to Shaaron Ainsworth and Sharon Ding for space in their office and an introduction to political correctness. Thanks also to the faces group at Stirling for making writing-up a good deal less painful than it might have been.

Finally I cannot offer enough thanks to both Helen Murgatroyd and my parents for their love, understanding and support. They are last in the list only because of the recency effect.

This thesis represents my own work, completed while registered as a part-time candidate for a higher degree at the University of Nottingham during the academic years 1992-1995. Throughout this period I held the appointment of Research Assistant in the Department of Psychology.

Experiments 2-4, 11 and 12 of this thesis also appear in Langton, S. R. H, O'Malley, C., & Bruce, V. (in press). Actions speak no louder than words: Symmetrical cross-modal interference effects in the processing of verbal and gestural information. *Journal of Experimental Psychology: Human Perception and Performance.*

Chapter One
Gesture and Speech

To what extent does a speaker's nonverbal behaviour actually influence the way we perceive and understand what he or she is attempting to say to us? The present thesis aims to address this question by adopting the so-called "Stroop-type" interference paradigm as an investigative tool (Stroop, 1935). In this respect, the work represents an attempt to marry literature from two differing perspectives; the vast quantity of information on non-verbal behaviour, largely amassed by social psychologists; and the equally massive body of knowledge concerned with the Stroop effect, which, despite over 60 years worth of effort on the part of experimental psychologists, still awaits cogent explanation. In attempting this alliance, through the experiments reported in Chapters 3-7, the hope was first to demonstrate that gestures are influential in the comprehension of verbal material, and in doing so, to begin to describe the mechanisms which might underpin any gesture-speech interaction. However, as a result, it is possible that more insights have been gained into the nature of interference in general rather than specifically into the processing of gestural and verbal information per se. Having said this, the results of the present experiments, in particular those described in Chapters 4 and 5, furnish us with an answer to the question posed above. These studies suggest that, at the very least, listeners process the information contained in a speaker's pointing gestures. Furthermore, it is argued that this information is integrated with the verbal content of the utterance to yield a complete understanding of the speaker's intended meaning.

The notion that information from different modalities is combined or integrated in comprehension reflects the ideas of researchers interested in the production and comprehension of gesture and speech, and those of some scholars of Stroop interference and dimensional interaction. In the following two chapters these ideas are elaborated. In the remainder of this chapter some questions concerning the comprehension of gesture and speech are formulated following a review of some of the relevant literature. In Chapter 2 the Stroop interference paradigm is introduced, and described as an appropriate tool with which we can study the interaction between various dimensions, including those of gesture and words.

Why Study Gesture?

It has long been realised that spoken language is not the exclusive communication medium available to the human species. In addition to the voice, we make use of facial expression, gaze, gestures, posture, spatial behaviour and certain non-verbal vocalisations (e.g. prosody and paralinguistic features) in any social situation in which we find ourselves. Unlike facial expressions, which can be categorised into seven or so emotional states, an individual's repertoire of possible gestures is vast. They change and adapt from moment to moment, maintaining synchrony with both the physical properties of the speaker's voice (e.g. amplitude peaks) as well as the semantic and syntactic content of the vocalisation. Gestures can be used in the absence of speech to express affect or meaning and indeed systems of sign language have evolved with all the combinatorial and syntactic complexity of spoken language. It is perhaps the close relationship of gestures to language which arouses most interest. Researchers such as McNeill (1985) and Kendon (1986) stress this intimate alliance, choosing to view gestures as essential components of language itself. They suggest that by studying gesture we can gain new insights into the nature of language, how we produce and understand it, how thought and language are related and even how knowledge is represented. The gestural behaviour of young children can shed some light on their conceptual development (Goldin-Meadow, Alibali & Church, 1993) whilst an understanding of how children integrate speech and gestural information might be important in the study of language acquisition (Thompson & Massaro, 1986, 1994). The integration of gestural and verbal information is really the primary interest of the present work. If we endorse the view that gestures and spoken words are components of a language system then, just as we are able to use prosodic information along with the meaning of the words of a sentence in the understanding of irony or sarcasm, we might expect listeners to combine information obtained from the gestural and verbal performances of a speaker in order to reach an understanding of the intended meaning.

Historically gestures and other aspects of non-verbal behaviour have been treated as a "body language". These signals were thought to provide privileged access to speakers' emotional states, interpersonal attitudes and personality traits in addition to any information carried in their verbal "channels". The body language approach perhaps fostered the conviction that "actions speak louder than words", the idea that in the processing of emotional information, for instance, we show a preference for non-verbal as opposed to verbal performances. However, the evidence for this so-called "video primacy" hypothesis is shown to be rather equivocal. Moreover, much of the evidence from experimental cognitive psychology and the neuropsychological study of brain injured individuals points towards a much closer relationship between gesture and speech in *production*. Gestures are not seen as providing a privileged window on the soul, but are thought to be intimately linked with the production of spoken language. For instance, researchers such as McNeill and Kendon (e.g. McNeill, 1985, 1987b, 1992; Kendon, 1983, 1986) believe that gestural and verbal behaviours serve to represent different but *complementary* aspects of the underlying meaning that a speaker is striving to express. Both types of expression might refer to the same event but each offers a different "view" of it. Certain gestures might emphasise the shape of the referent, spatial arrangements or actions, whilst others might stress intonation, pauses or the logical structure of an utterance. An important point for the listener then, is that neither the gestural nor the verbal performance alone is necessarily sufficient to specify the speaker's underlying cognitive representation. The complete picture can only be fully appreciated by considering jointly both gestural and verbal behaviours. Under this view a listener might be expected to process both verbal *and* gestural components of an utterance, combining this information at some point in processing to provide an integrated representation of the speaker's intended meaning.

The hypothesis that gestures are *processed* and then *integrated* with speech in comprehension forms the theoretical motivation behind the experiments described in the present thesis. Just how we arrive at this hypothesis is hopefully illustrated in the remainder of this chapter. After describing the types of gestural behaviours which have aroused researchers' interest, the discussion turns to "body language" and the evidence in support of this approach. The evidence and theoretical positions on gesture/speech production are then introduced, followed by a discussion of the role that gestures might play in the comprehension process.

Gesture Classification

That gesture is far from a unitary phenomenon is made quite clear by a glance at the variety of classification systems proposed by various authors (e.g. McNeill, 1985; Ekman & Friesen, 1969; Efron, 1941/1972; Rimé & Schiaratura, 1991). Rather than providing a detailed account of any single system, it is perhaps more useful to highlight some of the commonalities. Most draw a distinction between gestures which are either speech related or which can occur independently of speech. In the latter category we might include: *pantomimes* where the gesture depicts the action which might be performed on an object; *sign-languages* (e.g. American Sign Language); and so-called *emblems* which often occur in the absence of speech and can function as language substitutes (e.g. the "thumbs-up" gesture or the "okay" sign). On the other hand much spontaneous speech is accompanied by movements of the arms and hands which Kendon (1986) refers to as *gesticulation*. Much gesticulation consists of the types of hand movements which often occur in close co-ordination with the rhythmic nature of speech. These are variously described as *batonlike* (Efron, 1972), *beats* (McNeill, 1987a) or *batons* (Ekman & Friesen, 1972). These gestures provide stress or clarity and perhaps "chunk" the sentence according to the underlying reasoning. Other gesticulations display some concrete aspect of the accompanying speech. These *iconic* gestures might depict the shape of an object being referred to (e.g. a spiralling movement of the finger to describe a spiral staircase), its spatial position (e.g. a palms together gesture indicating that the object is "sandwiched" between two others), or perhaps an action performed by the object (e.g. a descending gesture which might parallel someone or something "falling down the stairs"). *Metaphoric* gestures are similar to iconics in that they relate to the concurrent speech, however, these gestures depict images of more abstract concepts and relationships occurring in the speech. For instance in western culture a common metaphor for a mental product (e.g. an "idea") is a bounded physical object. A speaker might refer to this idea whilst simultaneously holding or grasping an imaginary object.

As a summary it is useful to consider Kendon's continuum of gestural behaviour cited by McNeill (1987). This ranges from gesticulation (beats, iconic and metaphoric gestures) through emblems to fully lexicalised sign languages. Emblems and sign languages are historically long-lived; culturally specific (see Morris, Collett & O'Shaughnessy, 1979); do not refer to objects actions or events and often appear without concurrent speech. In contrast, gesticulations are more ephemeral; are used similarly in different cultural groups; refer to objects, actions and events; and only occur with concurrent speech. (McNeill, 1987).

Bavelas (1994) argues that taxonomic categorisations are retrogressive in that they imply mutually exclusive categories of gesture. Indeed the boundaries of the categories are often blurred. For instance certain symbolic gestures (e.g. the gesture for "money") are often used both autonomously (i.e. silently) and at other times can be used to "illustrate" a spoken utterance. An iconic gesture might also serve to emphasise a point, or provide a cue to the structure of the sentence etc. Bavelas recommends that the function of the gesture should also be taken into account, both in terms of providing information regarding the topic of conversation, and in providing a means of regulating the interaction, perhaps eliciting help or agreement from the addressee.

Despite this criticism, the system is included here to give a flavour of the diversity of human gestural activity, both in form and in conditions of use. A second reason for its inclusion is to illustrate the possibility that the various types of gesture might be processed in different ways, both in comprehension and production. McNeill (1987) suggests that pantomimes, emblems and lexicalised sign languages have a temporal relationship with speech which is different from that of iconic type gestures. As a consequence, McNeill ventures that these types of bodily movement should not be indiscriminately lumped together in discussions of the relationship between the underlying processes mediating gesture and speech production. In comprehension it is possible that emblems, which often occur without speech, might be processed independently of verbal information. On the other hand, speech-related gestures such as iconics and beats might enjoy no such autonomy, perhaps being constrained by, and/or constraining, the processes involved in decoding speech signals.

Having briefly reviewed the various forms of gestural behaviour we now go on to discuss the development of ideas on their relationship with concurrently produced verbal material. The emphasis has gradually shifted from the notion of gesture and speech as autonomous communication channels to one where both verbal and visual behaviour are considered linguistically, comprising integral parts of the communication process.

Gestures as Body Language

Many of us are seduced by the idea that with the purchase of a text on "body language" and a few hours of study one can read the romantic intentions (or otherwise) of someone we have just met at a party, that the body movements of a friend might reveal whether they are speaking truths or untruths, or whether an opponent across the card table is attempting to bluff or double bluff us. Under this view gestures form part of a system of body movements which might offer a privileged means of knowing and perceiving one another, a system thought to follow its own laws and transmit affective, cognitive and regulating mechanisms distinct from those carried by the accompanying speech.

The Secret Code: What Can We Infer From Body Language?

This popular view appears to have fuelled a great deal of research which seeks to examine what Sapir (1927) referred to as "an elaborate and secret code that is written nowhere, known by none, and understood by all". The idea is to discover exactly what we are able to infer from one another simply by virtue of the movements of our hands and bodies in isolation from any verbal information. For instance Wolff (1945) observed that certain movements of the hands and body were associated with the expression of emotional states such as depression, elation and anxiety. However, Ekman & Friesen (1967) have suggested that the face is most important in this regard but that the body conveys the intensity of the expressed emotion. Certain behavioural attitudes are also expressed gesturally. Maxwell, Cook & Burr, (1985) found that when experimental subjects liked another subject whom they met in the laboratory, they engaged in more active body movements and gesture, higher levels of gaze, and adopted a livelier tone to their voices. Finally, despite the unimpressive correlations between results obtained from personality questionnaires and measures of non-verbal behaviour (Feyereisen & de Lannoy, 1991), the perception of various non-verbal behaviours yield distinct and consistent attributions about the personality of the sender (e.g. Riggio & Friedman, 1986).

Thus there is evidence to suggest that subjective emotions, attitudes and perhaps personality traits can be communicated and perceived by gestural means. People seem able to make these various attributions despite the fact that many of the gestures were not intentional acts of communication. This type of approach, where gestures are studied in isolation, resulted in the idea of separate verbal and non-verbal communication channels and to the associated conviction that gestural and verbal communication are subserved by entirely separate processing systems. The idea is that the operations governing the production and comprehension of gestures are essentially independent of the cognitive processes associated with the production and comprehension of speech. Thus, having demonstrated that we can learn much about one another simply by examining the contents of the non-verbal channel, many sought to examine the relative importance or competence of this information compared with that gained from the analysis of information in the verbal channel. The common perception is that gestures and other aspects of non-verbal behaviour are somewhat more effective than language in the communication of attitudes and emotions. In the next section some of the evidence for this assumption is considered.

Video Primacy: Do Actions Speak Louder Than Words?

Several lines of evidence have been used to investigate the relative contribution of verbal and non-verbal behaviour in the communicative process. Emotional and attitudinal judgements can be measured in conditions where either verbal or non-verbal signals are made unavailable. Interactions in face-to-face conditions can be contrasted with situations where interlocutors are unable to see each other. Finally conditions can be contrived where verbal and non-verbal messages are contradictory.

It has already been noted that body movements carry information regarding the emotional state of the gesturer. The interest here is to what extent gestures and other body movements are more important than the verbal "channel" in emotional recognition and social evaluation. Mehrabian (1972) found that facial expression carried more information regarding positive attitudes than tone of voice, and that both carried more weight than the actual verbal content. O'Sullivan. Ekman, Friesen & Scherer (1985) noted that information from the face and body movements correlated highly with judgements based on full audiovisual input, whereas the correlations were lower with both tone of voice and verbal content. Finally Rosenthal & DePaulo (1979) used the PON's (Profile of Nonverbal Sensitivity) which indicated that facial expression is judged more accurately than information from the body followed by that from the tone of voice. In summary these findings suggest that the face carries most information (regarding emotion at least), followed by the body, followed by tone of voice and verbal content. This preference for visual, as opposed to verbal information has been termed the *video primacy* effect.

A number of other studies have departed from this video primacy point of view. Berman, Shulman, & Marvitt (1976) had actors express warmth or coldness in relationships. The evaluation of these messages was more accurate when based on verbal information than when subjects relied on bodily or facial signals. Krauss et al. (1981) extracted judgements from observers who watched a political debate under different conditions; audiovisual, video only, and on the basis of a written transcript. For ratings on the positive-negative dimension, responses to the written transcript correlated significantly with those made in the audiovisual condition but responses under the video only condition did not. They concluded that there was "no support for the widespread assumption that non-verbal channels....form the primary basis for the communication of affect" (Krauss et al., 1981, p.312). Others have noted that the type of judgement task, and the conditions in which the behaviour occurred, affected the relative weights given to facial expression, speech and body cues (Ekman, Friesen, O'Sullivan, & Scherer, 1980). Thus, at the very least, the video primacy effect is dependent on context.

In summary, the evidence for the common assumption that gestures and other non-verbal cues are more effective in the communication of emotional information is equivocal. The video primacy effect appears to be qualified by a number of variables, the type of judgement task, context and perhaps individual preferences. In their review, Feyereisen & de Lannoy (1991) make the observation that correlations are often noted between judgements from isolated cues (e.g. face only) and from cues based on full information, suggesting redundancy of the different communication channels rather than a preference for the visual channel. On the basis of this type of evidence the most likely conclusion is that actions do not necessarily speak louder than words.

The Removal of Visual Cues

Another way of examining the importance of visual cues is to remove them and examine the consequences for normal social interactions. It has been suggested that non-verbal cues are critical in the transmission of affective, attitudinal and attributional information (e.g. Ekman & Friesen, 1967) as well as playing an important role in the turn-taking mechanism (e.g. Duncan, 1972; Duncan & Fiske, 1977; Dittman & Llewellyn, 1968; Kendon, 1967). With this in mind, one might expect that the removal of visual cues in a communicative context will have a number of effects, disruption of the turn-taking mechanism, reduced intimacy and perhaps a change in the content and style of the interaction.

The evidence from studies concerning the effects of the removal of visual signals on the turn-taking mechanism is inconsistent. Some studies have found increases in overlapping speech and interruptions in audio-only conditions as might be predicted from a disruption in turn-taking (e.g. Argyle, Lalljee & Cook, 1968; Boyle, Anderson & Newlands, 1994) whilst others found higher rates of interruptions (Cook & Lalljee, 1972, Rutter & Stevenson, 1977; Butterworth, Hine & Brady, 1977) and simultaneous speech (Beattie & Barnard, 1979) in face-to-face dialogues. The inconsistency of these findings is probably due, at least in part, to the variety of interactional contexts adopted in the studies. Some look at dialogues produced in more "social" situations whilst others examine task-oriented dialogues.

As to reduced intimacy, Argyle, Lalljee & Cook (1968) varied the amount one interactor could see another. As visibility was reduced, from normal vision to dark glasses, mask showing eyes only and one-way screen, they rated the interaction as increasingly more uncomfortable.

Rutter (1987) has suggested that the lack of visual cues in telephone conversations affected the style and content of discourse, exchanges became less spontaneous with more filled pauses and fewer interruptions. Rutter attributed this increase in the formality of the interaction to the larger "psychological distance" experienced when speaking on the telephone. In their work, Boyle, Anderson & Newlands (1994) and Doherty-Sneddon et al. (submitted) have suggested that communication style is affected by the communicative context within which the interactions take place. The removal of visual cues has marked effects on the dialogue structure during a co-operative problem solving task. Subjects compensated for the elimination of visual information by requiring more exchanges (turns) and words to complete the task successfully in unseen compared to copresent contexts. Specifically, subjects used more backchannel responses (mhm's etc.) and increased their elicitation of feedback (so-called ALIGN games) in unseen contexts. These results suggest that the visual channel is normally used to provide feedback regarding the current status of the interaction, supporting Boyle et al's (1994) proposal that the visual channel carries a communicative function rather than merely acting as a turn-taking regulator.

In summary the removal of visual cues in interactive contexts appears to reduce the feeling of intimacy and decrease the feeling of "being there". The more rigorous studies on the effects of "eye contact" on dialogue structure and content (Boyle et al, 1994 etc.) suggest that the visual channel assists in the turn-taking mechanism but also carries communicative information, to the extent that verbal information is used to compensate for the elimination of the visual cues. Thus speakers are adept at using visual *and* verbal information to express both meaning and pragmatic/discourse related information. Again it seems that whilst speakers routinely make use of non-verbal information, there is little support for the notion that the visual channel is special. The hypothesised regulating functions are equally well accomplished by verbal as well as visual means (e.g. in Boyle et al's studies, subjects performed equally well at the problem solving task whether or not they could see one another).

Discrepant Messages

The video primacy hypothesis has also been examined in the decoding of incongruent messages. This has particular relevance for the present work as many of the experiments to be described involve placing verbal and non-verbal stimuli into conflict.

Video primacy would suggest that, when faced with discordant information in the verbal and non-verbal channels, listeners should give priority to the non-verbal signals. In a frequently cited study, DePaulo, Rosenthal, Eisenstat, Rogers & Finkelstein (1978) asked their subjects to make affective judgements (positive/negative, dominant/submissive) based on either audio-only messages, video-only messages or discordant audio-video messages. The ratings in the audio-video condition resembled those in the video-only condition more so than the audio-only ratings. DePaulo et al. claim that these results support the video primacy hypothesis. However, there are at least two problems with this study. First, the conclusion rests on the absence of a difference between audio-video and video only ratings. Secondly, it may be that task difficulty was not equated across the conditions. More specifically it seems as though the audio-only judgements were likely to be much harder than the video-only judgements. In the former condition, subjects were required to make their decisions on the basis of *incomprehensible* verbal messages. That is, the voice signal was either "muffled" by removing "critical frequencies", or rearranged by cutting the audiotape and splicing it back together. Judgements in this condition would have to be based on pitch, intensity and rhythm information, rather than the actual content of the message. On the other hand, no equivalent manipulations were made to the visual information. Judgements in this condition could therefore be based on the intensity, frequency and rhythm, as well as the actual form or identity of the gestures etc. This would seem to bias judgements in favour of the visual channel and hence a video-primacy effect.

In a later study, using a similar procedure, Zuckerman, Blanck, DePaulo & Rosenthal (1980) noted complex developmental changes in decoding discrepant audio and visual cues. Video primacy interacted with age, rated dimension (positive/negative, dominant/submissive) and degree of discrepancy. Video primacy increased with age, but only for facial, as opposed to body cues, and only for cues of positivity, as opposed to cues of dominance. Finally Trimboli & Walker (1987) demonstrated that experimental results depended on the proportion of inconsistent messages in the stimulus materials. When discrepant cues are frequent, subjects give priority to non-verbal cues (video primacy), otherwise they give priority to verbal information.

Thus the evidence for a video-primacy effect in the processing of discrepant messages is not all that compelling. The effect has been noted, but appears to be contingent on certain experimental conditions and developmental factors.

Summary and Conclusions

Gestures and other bodily movements do indeed appear to convey information relating to emotions and attitudes as conjectured by the common approach to non-verbal communication and "body language" in the popular sense. These findings have probably reinforced the idea of a separate non-verbal communication system. However the evidence for the video primacy effect is rather equivocal, or at least qualified by numerous variables (Feyereisen & de Lannoy, 1991). On the whole subjects do not show a preference for the non-verbal channel. Information regarding affect and attitude is probably carried equally in both the verbal and non-verbal channels.

Removal of visual cues does seem to affect discourse, probably due to disruption of the turn-taking mechanism and because of the elimination of non-verbal cues which are critical in providing feedback in certain discourse situations. Nevertheless we are still able to communicate effectively without the visual channel, suggesting that we are able to use the verbal signals to compensate for the absence of non-verbal cues. Turn taking can be signalled equally well by verbal factors such as sentence well formedness, certain other linguistic cues (e.g. expressions like "well" or "you know") and prosodic signals such as falling pitch contour (e.g. Duncan, 1972). Backchannel responses may replace head nods, facial expressions or gestures which would otherwise provide visual feedback to the speaker. Thus whilst gestures and other non-verbal signals are almost certainly used in discourse and impression formation, there is little evidence that they perform functions which the verbal channel is ill-equipped to deal with. In this sense, actions speak no louder than words.

We have seen that approaching gestural behaviour as a form of body language appears to be inappropriate. Furthermore the notion that gestures and other non-verbal cues offer a *privileged* access to a speaker's innermost thoughts and feelings, has not received much empirical support. Whilst gestures and speech do seem able convey emotional and dispositional information regarding the state of the sender, it is not clear whether these signals are simply *expressions* or whether they should be considered as intentional *communicative* acts. Other researchers have concentrated on the role that gestures play in more cognitive, as opposed to emotional activity. Many of these authors consider hand gestures and other bodily movements as linguistic entities insofar as they are performed and processed as integral components of language. Under this view, gestures are seen as more than simply regulators of turn-taking in discourse, but are thought to complement speech in providing important communicative information which can assist in establishing mutual understanding between speakers. Thus, rather than viewing gesture and speech as functionally separate behaviours, perhaps we should be mindful of the notion that they are, in fact, interacting processes.

Interactions and Separations

In the mid 1980's a number of authors published work which criticised the widely held view of gestures and other non-verbal behaviours as "body-language" along with the associated belief that gestural information is carried in a separate non-verbal channel of communication. McNeill (1985) suggested that gestures and speech, far from being psychologically distinct, "share a computational stage; they are, accordingly, parts of the same psychological structure" (p. 350). This prompted rebuttals from Feyereisen (1987) and Butterworth & Hadar (1989) with accompanying replies from McNeill (McNeill, 1987b, 1989). Most seem to agree that gesture production depends, to some extent, on the mechanisms responsible for speech production (see also Rimé, 1983; Kendon, 1983). The arguments centred around specifying the locus of the interaction, elaborating McNeill's conception of *inner speech* as the shared computational stage. These ideas will be elaborated in a later section, for now we simply note that this work represented a shift in emphasis from the social impact of non-verbal behaviour, to an approach which sought to examine the processes underlying the performance of body movements and, in particular, the relationships between these processes and the structures mediating vocal behaviour.

Information Processing Approach

Information processing models undertake to explain a simple form of overt behaviour, say word reading, in terms of a set of processing stages. In its simplest form, the results of one stage of processing forms the input to the next stage. In this way a set of input-output relations can be built up to account for the performance on an experimental task. For instance a simple model of word recognition might proceed as follows: visual analysis of the input word leads to an orthographic description of the item. This input code then causes activation of a unit in memory (e.g. a logogen) corresponding to that particular pattern. The word is now recognised as familiar. Having identified the word, its pronunciation and/or its meaning can be accessed resulting in word naming or some form of semantic categorisation depending on the experimental task (e.g. Morton & Patterson, 1980). This approach has also been applied to tasks such as face (e.g. Bruce & Young, 1986) and object recognition (e.g. Ellis & Young, 1988). In more recent years the adherence to a serial, stage-by-stage approach has been relaxed. Contemporary models of picture naming (Humphreys, Riddoch & Quinlan, 1988) and word recognition (McClelland & Rumelhart, 1981), for instance, allow the stages to operate in "cascade", with operations at a later stage beginning before those at a previous stage have been completed.

Similarly a cognitive approach to the study of gesture and speech would attempt to identify the processes underlying the performance and perception of body movements and concomitant speech. The relationship between these two actions can also be examined from such a perspective. One can use evidence from at least two sources: experiments performed on the normal population examining the relationships between gestural and verbal behaviour, and from observations of brain injured individuals who demonstrate functional associations and dissociations of verbal and gestural abilities (e.g. Shallice, 1988). Using these methods we can begin to sketch a cognitive architecture of the multiple components involved in both the comprehension and production of gesture and speech, identifying both modality specific processes (i.e. independent processing stages for gesture and speech) and those which might be shared by the two behaviours.

Providing a functional account of gesture and speech *production* will quite possibly reveal clues regarding the relative contribution of speech and gesture in the *comprehension* process. For instance, it may be that gesture and speech are linked at some deep, conceptual stage of processing, so that the complete meaning of an utterance is expressed partly in gesture and partly in speech. If this is the case then in order to provide an adequate representation of the "encoder's" intended message, the "decoder's" cognitive system would need to integrate the two sources of information at some stage of processing. Alternatively, gesture and speech might be linked, but express the same information via different modalities. In this case "decoders" could well exploit this redundancy in the comprehension process, ignoring one or other of the signals. These alternatives are discussed in more detail below, for now they serve to illustrate how research on the production of gesture and speech can inform those of us interested in their comprehension.

In the remainder of this chapter the research and evidence for the integration of gesture and speech in *production* is reviewed and a number of theoretical positions are presented. Finally, the discussion turns to the *comprehension* of gesture and speech from the same cognitive perspective.

Gesture and Speech Production

A glance at someone holding a conversation over the telephone will convince you that people continue to use gestures and facial expressions despite the fact that these non-verbal activites cannot possibly influence the "decoding" processes of the listener. Indeed a number of empirical studies have demonstrated this simple fact. For instance Rimé (1982) noted no decreases in aspects of non-verbal behaviour (e.g. trunk movements, rhythmic nodding of the head, and eyebrow movements) in pairs of subjects who were unable to see one another, as compared to pairs interacting in face-to-face conditions. This observation supports the intuition that certain body movements (e.g. hand gestures, trunk movements etc.) are linked to the mechanisms responsible for speech production. In what follows both experimental and neuropsychological evidence for this position is reviewed.

Experimental Studies

Rimé & Gaussin (1982) examined the form of non-verbal behaviour under two conditions of communication content. When repeating a series of digits subjects engaged in a certain amount of body movement. These actions, however, were far more frequent and appeared in a flow of motor activity when subjects were asked to describe noteworthy events that had occurred in their personal lives over the last week. Thus it appears that merely producing words gives rise to some degree of body movement and secondly, that the content of the verbal, and the form of the non-verbal gestural activity is linked in some way.

Further evidence is consistent with the suggestion that gestures aid in the shaping of the speaker's verbal activity, these studies demonstrate that speech is affected when conversants are unable to gesture. Rimé et al. (reported in Rimé, 1983) restricted the movements of the head, arms and legs of subjects during interviews and noted increased non-verbal activity of eyebrows, eyes, mouth and fingers as if in compensation for the restriction of the critical gestural effectors. In addition they noted that the vividness of the imagery contained in the accompanying speech was reduced during movement restriction. However, this effect was only "marginally significant ($p<.10$)" (Rimé, 1983, p.99) and furthermore the actual content of the speech was unaffected by the manipulation. Graham & Heywood (1975) also noted that speakers used more pauses, more use of words or phrases describing spatial relations and fewer demonstratives (e.g. "there", "like this"), when asked to keep their arms folded whilst describing line drawings of two-dimensional shapes.

Another line of research which suggests that the gestural performances of speakers are inextricably linked to their verbal activity comes from microanalytic studies of the temporal relationships between gesture and speech. The work of William Condon and his colleagues (e.g. Condon, 1970; Condon & Sander, 1974) has indicated that the organisation of body movements is synchronised with the organisation of speech articulation at very fine grained levels, such as the boundaries between phonemes. Astonishingly this research has also revealed that not only are our body movements synchronised with our own speech but that these actions also tend to coordinate with the verbal utterances to whomever we may be listening.

Several studies conducted in the laboratory environment have also revealed temporal links between gesture and speech in production. Levelt, Richardson & La Heij (1985) had subjects point to lights which were illuminated near or far from a centreline. In some conditions subjects were asked to use the deictic expressions "this light" or "that light" whilst pointing. Movements of gesture initiation and apex were recorded and related to the speech onset times. The results indicated that the initiation of the gesture and the time to reach the apex were influenced by the distance of the light from the subject. Longer movements took more time to be planned and executed. The voice onset time was also influenced by this factor suggesting an interaction of the two systems. Moreover the distance variable had no effect on speech onset time when no pointing movement was required. It appears that the two systems are linked at the planning phase where motor responses are prepared. However, the unexpected application of a load on the arm after the initiation of a pointing movement delayed the apex time but did not influence the voice onset time. Levelt et al. concluded that the interaction between gesture and speech was restricted to the planning phase prior to the beginning of a movement but that in execution the movements were "ballistic".

In a similar study, Kelso, Tuller & Harris (1983) noted that finger movement amplitude was related to stress in the utterance of lists of words. Again it seems as though motor programs controlling manual and vocal movements were exchanging information or interacting in some way.

More recently Morrel-Samuels & Krauss (1992) noted that referential gestures precede the speech onset of their lexical affiliate by an amount which is inversely related to the affiliate's rated familiarity. Thus gestures appear longer in advance of unfamiliar lexical items. Morrel-Samuels & Krauss (1992) claim that their data is not compatible with the notion that gesture and speech are produced by independent modules. Instead they maintain that their findings are consistent with the idea that gestures facilitate lexical access by serving as cross-modal primes.

Neuropsychological Studies

McNeill appeals to neuropsychological evidence to support his assertion that gesture and speech share a common conceptual stage (McNeill, 1985). He suggests that gesture and speech dissolve together in aphasia pointing toward a shared neural mechanism for the production of these activities. Furthermore the gestural performances of patients with sub-types of aphasia appear to be consistent with their language problems.

Patients with Broca's aphasia have great difficulty in combining sets of referential terms into grammatical wholes. Whilst maintaining the ability to produce iconic or referential gestures, their use of beats (batonic gestures) is impaired. Conversely Wernicke's aphasics seem unable to form coherent semantic plans whilst maintaining the ability to construct sequences of words. In terms of their gestural performances, they make few iconic gestures but retain beat-like or batonic gestures. Simply put, patients with problems producing semantic speech fail to produce iconic (or meaningful) gestures, whilst patients with problems producing grammatical speech also fail to produce the kinds of speech marking gestures which have been related to the syntactical and rhythmic nature of speech. McNeill (1985) uses this data to suggest that gesture and speech production are mediated by the same parts of the dominant cerebral hemisphere, consistent with the idea of a common computational stage.

However McNeill's use of this type of evidence has been criticised by Feyereisen (1987) in particular. He notes that associations of verbal and gestural deficits are not sufficient to indicate the existence of a common underlying mechanism. Instead the associations might arise from simultaneous damage to anatomically neighbouring, but none-the-less distinct, neural control centres. Moreover Feyereisen introduces a number of *dissociations* between gesture and speech production. For example aphasics can often pantomime objects which they cannot name (Davis, Artes & Hoops, 1979) and some cases have been observed of the reverse dissociation, apraxia without aphasia (e.g. De Renzi, Motti & Nichelli, 1980). This type of double dissociation is logically consistent with separable, rather than common, underlying mechanisms (e.g. Shallice, 1988).

The argument between McNeill and Feyereisen is probably based around a confusion of the gestural types which McNeill (1985) suggests are linked with speech in production (see McNeill, 1987). Briefly, Feyereisen's dissociations are associated with emblematic type gestures which "with a diabolical accuracy....are exactly the kind of gesture that would show independence from speech" (McNeill, 1987, p.500). McNeill's associations are realised in speech related gestures such as *beats* and *iconics* (what Kendon terms *gesticulations*). Thus it may be that only these types of gestures associate with speech in production. A lesson should be learnt here, it would appear to be a mistake to lump together all forms of gestural behaviour when discussing possible interactions with speech and common computational stages etc. Gestures form a somewhat heterogeneous class (e.g. see above taxonomy) and as such one should make explicit the particular gestural type under investigation.

Research from rather diverse sources points toward the fact that certain gestures and speech interact in some way in production. People persist in gesticulating even when they cannot be seen whilst there is evidence to suggest that the fluidity of speech and the associated gestures are linked in some way. Indeed restricting these body movements appears to affect the content and articulacy of the concurrently produced speech. Microanalytic observational studies have noted a certain synchrony in gestural and speech production which has also been demonstrated in more experimental paradigms. Finally both associations and dissociations have been observed in neuropsychological research suggesting that the picture is complicated insofar as certain gestural types may associate (e.g. *gesticulations*), whilst others (e.g. *emblems*) appear to dissociate from speech.

Theoretical Positions

i) McNeill

David McNeill (1985) used a certain amount of the evidence presented above in support of his argument for a close association between the systems controlling gesture and speech output. For McNeill (e.g. McNeill, 1985, 1987b, 1989), gestures and speech arise from a single cognitive structure where meanings are stored independently of language formats. When speaking, meanings are transformed directly into either linguistic or gestural form by a single underlying process. This process transforms "global and imagistic" representations to the "segmented and syntactic" representations underlying speech. Gestures reveal the early, imagistic stages of this thinking process whilst verbal speech is the manifestation of the final stages. The final utterance then, represents a synthesis of imagistic and syntactic "thinking". McNeill suggests that the balance between the two changes until a form of "thought" is reached that can be spoken. This balancing act is at the whim of the speaker who is able to choose how to divide meaning between the gesture and speech channels. An important point is that neither outcome is necessarily sufficient to specify the underlying cognitive representation. Thus gestures and speech combine together to express the meaning of a given utterance. "To get the full cognitive representation that the speaker had in mind, both the sentence and the gesture must be taken into account." (McNeill, 1985, p.353). This would seem to have important implications for an approach to gesture and speech comprehension. If meaning is embodied in both the gestural and speech forms, then presumably efficient processing of this information would involve some sort of combination or integration of information from the two sources. At the very least it suggests that the two forms of expression should receive some degree of processing.

This The studies of the above mentioned researchers on speech production would seem to have important implications for an approach to gesture and speech comprehension *ii) Kendon*

Similarly Kendon (1983, 1986) views gestures and speech as arising from a single conceptual structure. Like McNeill's, Kendon's theory suggests that gestures manifest a surplus of meaning beyond what is permitted in the verbal channel i.e. gestures represent components of the utterance not represented in words. In Kendon's view, gestures are not subject to the same syntactic and lexical constraints as spoken language and as such possess more "degrees of freedom" for the expression of meaning. Whilst McNeill's theory suggests that the tactical decision of the speaker determines the distribution of meaning between the gestural and speech channels, Kendon emphasises situational constraints, such as noisy conditions, which contribute to the speaker using the more flexible gestural channel as a means of expressing the underlying representations. All forms of speech related gestures are seen by Kendon as integral in the communication process. Speech marking gestures or *beats* serve as a visual analogue of phonological "chunking" (e.g. pauses or falling pitch contour indicating the end of a sentence), they are able to represent visually stress, intonation, pauses and the logical structure of an utterance, assisting the listener in structuring sections of the discourse. Iconic gestures, for Kendon, depict aspects of the utterance not referred to in the verbal component such as the shape of the referent, spatial arrangements or actions, whilst emblems are used by utterers as an alternative to speech.

Again if we accept Kendon's view that in gesticulation "we observe...components of the utterance content that are *not* represented in words" (Kendon, 1986, p.12) and, as Kendon stresses, that gestures serve a particular communicative function, we might be forgiven for assuming that perceivers of utterances use *all* the information available in order to process the speaker's intended meaning. The processing of gestures would be of obvious benefit in both highlighting the structure of the speech and in decoding the meaning of the utterance.

iii) Freedman

Freedman (1972) has a rather different view of the relationship between gesture and speech. Unlike McNeill and Kendon, Freedman sees gesticulation not as an explicitly communicative process but one of facilitating the verbal encoding mechanism. Again the idea is that both gesture and speech originate from a single "imagistic" representation. *Speech-primacy* gestures (*beats*) occur when there is a trouble-free translation of the central representation into a verbalisable form. These movements, according to Freedman, serve in a self-monitoring and clarifying function. In contrast *motor-primacy* gestures (*iconics*) arise when there is a failure to translate the central imagistic structures into linguistic representations. This gestural output can be seen as the manifestation of the internal representations in a kind of interim format, i.e. one which has not yet been organised for linguistic encoding but that seeks immediate expression. Freedman has suggested that these gestures also serve to reactivate decaying images and perhaps activate connections between the image and the searched-for word or phrase (see Morrel-Samuels & Krauss (1992), described earlier, for some recent empirical support for this theory). In this way hand movements are critical in maintaining the fluent articulation of speech.

In Freedman's view, speech related gestures are used by a speaker to aid in the process of utterance construction and not explicitly as a medium of communication. However the gestures produced under these circumstances express meanings that will not be simultaneously present in the vocal channel. For a listener then, processing information in the gestural channel will obviously be of some use.

iv) Rimé

In many respects Rimé's cognitive-motor view of gestural behaviour (e.g. Rimé, 1983; Rimé & Schiaratura, 1991) is similar to Freedman's theory in that gestural activity helps to retrieve linguistic elements of the underlying representational structure, i.e. we gesture to help us speak.

Rimé stresses the importance of motor processes in perception. He suggests that our representations of reality include a motoric component either by virtue of the movements we make when encoding a particular stimuli, or by the movements of the stimuli themselves. Thus a representation of a spiral staircase contains some coded spiralling movement. A mouse might be motorically coded by its speedy movement, twitching nose etc. Motoric elements, along with verbal attributes and images, form components of the *raw structure*. Through continuous activity these representations are refined and integrated into a conceptual form or *expressive structure* which is ready for verbal expression.

The raw elements of the complex representational structure are often reactivated as the speaker attempts to transform a poorly articulated expressive structure into the logical, syntactic and lexical elements of speech. This reactivation materialises in iconic type gesturing. The reactivated motoric representations "prime" linguistic and conceptual forms related in meaning. The result of this motor activity is then the retrieval of the appropriate linguistic forms for verbal expression. Under these circumstances speech will be slightly hesitant and poorly constructed but accompanied by a rich variety of iconic gestures.

Speech-marking gestures (*beats* or *batons*), on the other hand, are likely to occur when the speaker is attempting to verbalise what is represented in a highly articulated expressive structure. In these circumstances the deep, raw structure of the representational network will only be weakly activated thus the accompanying gestures will be small and rhythmical.

Thus in Rimé's view both gestures and speech arise from complex representational structures, iconic gestures materialising as a result of the reactivation of primitive motor representations. For this reason there is much redundancy between the verbal and non-verbal channels, gestures and speech are simply different manifestations of the same underlying representational structure. Because of this redundancy, a listener need only attend to one or other of the channels and so might not be expected to integrate the information contained therein. In fact Rimé suggests that hand movements are largely ignored when interactions occur in normal contexts. Instead listeners devote the lion's share of their attentional resources to the processing of the verbal information contained in speech. This view is elaborated in the discussion of Rimé's figure-ground model of the relationship between verbal and non-verbal materials in comprehension (see Chapter 4).

Summary and Conclusions

To recapitulate, evidence has been reviewed which suggests that gestures and speech are closely associated in the production process. The various theoretical positions outlined above all agree that gestures originate somewhere in the process through which abstract conceptual structures are translated into speech. What they disagree about is the reason for the emergence of gestures in the first place. For Freedman and Rimé, gestures do not occur as communicative acts, rather they play an instrumental role in restoring fluent speech by retrieving or priming linguistic structures. On the contrary, Kendon sees gesture as an alternative medium of communication used when verbal resources are insufficient for global representation, or when situational or social contingencies constrain the use of the verbal channel. In any case gestures represent meanings not contained in the verbal component of the utterance. McNeill also suggests that meaning is underspecified by information in a single channel but views gestures as the manifestation of imagistic or global thinking. Like Kendon, McNeill suggests that a speaker can choose how to distribute meaning between the two channels however the reasons for this choice are not clearly specified by McNeill's theory. The empirical evidence that speakers are able to compensate for the removal of visual cues by increasing verbal feedback (Boyle et al., 1994 but see Rimé, 1982) is in line with this way of thinking.

In terms of the benefit to listeners or decoders, McNeill and Kendon's theories suggest that the complete meaning of an utterance is embodied in both the gestural and verbal channels and not in either one alone. Thus the integration of information contained in both verbal and non-verbal acts would be critical for the complete understanding of a speaker's intended meaning. For Freedman as well, the gesture manifests some form of meaning not currently available in the verbal channel, again we can see the benefit to the decoder of processing and integrating both types of information. Only Rimé's ideas seem to suggest complete redundancy between the two channels. Integration of information by a listener would seem to be unneccessary in this case. These ideas are expanded in the following section.

Gesture and Speech Comprehension

Despite the large volume of work dedicated to the study of gesture and speech *production,* the field of gesture *comprehension* remains a "neglected field in cognitive psychology" (Feyereisen, 1991, p.57). Basically there are two questions which must be addressed. First, are gestures processed or attended by listeners? In previous sections we have seen how emblematic gestures can express symbolic meanings for a particular cultural or social group and how emotional and dispositional states can be attributed to speakers on the basis of their bodily movements. It seems as though listeners do process certain aspects of a speaker's non-verbal behaviour. Secondly, given that they are attended, do gestures play a *significant* role in the communication process? Gestures are linked to speech in production, are they also integrated in comprehension?

Theoretical Positions: Do Gestures Communicate?

Opinion seems to be divided on the subject of whether or not gestures play a significant role in the communication process. Authors such as Rimé (1983, Rimé & Schiaratura, 1991), Feyereisen (Feyereisen et al., 1988) and Krauss (Krauss et al., 1991) are sceptical of the idea that gestures are produced for the benefit of others as an act of communication, whilst in line with their ideas on the production of gestures and speech, McNeill and Kendon view gestures as integral to the communication process (see above). The position of the sceptics is reviewed below, with a critical assessment of the main aspects of Rimé's theory of gesture/speech comprehension. Thereafter some evidence is presented suggesting that gestures are influential in comprehension

Number of Rimé's observations that in the theory

Rimé's ideas on the relationship between gesture and speech in comprehension arise from his belief that gestures play only a minor role in the communication process, serving instead to assist the *speaker* in formulating thought into words (see also Krauss et al., 1991 for a similar position). Because of the resulting redundancy between the gestural and verbal forms of expression, a listener need not concern herself with the task of "decoding" any information provided by the speaker's hand movements. Instead she can devote her attentional resources to the more demanding task of decoding the verbal message and encoding potential verbal answers. Thus, according to Rimé (e.g. Rimé, 1983; Rimé & Schiaratura, 1991), under normal circumstances we are able to attend selectively to a speaker's voice with no effects from his or her non-verbal behaviour.

As evidence for this theory, Rimé presents a number of observations. First, he appeals to intuition, "ask a person who has just been speaking to someone to describe that other person's gestures and movements. People can rarely answer this question" (Rimé, 1983 p.132). Secondly, he suggests that communication contexts in which the interactants cannot see each other are no different from face-to-face situations from the point of view of message reception and understanding (e.g. Rimé, 1982). Thirdly, Kendon's (1967) study is cited as demonstrating that in an interaction, gaze oscillates between the interactants' faces and points of the environment removed from the speaker. Thus according to Rimé (1983) body activity does not play an important role and is often "overlooked", particularly in interactions where subject's involvement in the exchange is low or moderate. Finally in a more recent paper Rimé & Schiaratura (1991) suggest that "subjects are unable to guess the speech content to which gestures relate, which supports the view that hand gestures do not provide independent access to a meaning expressed in words"(p.271).

There are several points concerning this evidence worthy of mention. First, if you ask someone who has just been involved in a conversation to repeat the words used by their conversant they might have some difficulty recalling this information, instead they are likely to recall the "gist" of the conversation (e.g. Bransford, Barclay & Franks, 1972). Secondly, verbal report has been shown, in a number of demonstrations, to be an inadequate indication of knowledge. Studies on implicit learning have shown that subjects can manifest learning of artificial grammars (e.g. Reber, 1969) or certain invariant features (e.g. McGeorge & Burton, 1990; Bright & Burton, 1994) from a list of test items with no explicit verbalisable knowledge of these features. Thirdly, the history of experimental psychology is littered with examples of the influence on subject's performance of unattended items, or items presented below the threshold of conscious identification (for a review see Holender, 1986). Clearly then, asking subjects if they can remember the gestural content of an interaction is not sufficient evidence to suggest that they do not process this type of information. A more sophisticated methodology must be established to test this claim. Indeed there are a number of studies which suggest that gestural behaviour *can* influence the recognition and recall of verbal material (see below).

With regard to contexts where face-to-face communication is prevented, Boyle, Anderson & Newlands (1994) and Doherty-Sneddon et al. (submitted) have recently demonstrated that whilst performance in a collaborative problem solving task remains consistent across face-to-face and unseen communicative contexts, interlocutors are able to compensate for the removal of visual cues and, as a consequence, produce more dialogue in these "unseen" conditions. Whilst it is not clear exactly which non-verbal cues are of importance, these studies clearly demonstrate that non-verbal actions carry a communicative function.

There is also plenty of evidence to suggest that visual spatial attention can be shifted "covertly" with no corresponding change in eye or head position (e.g. Posner, 1980). Thus the fact that subjects' gaze is not directed toward their partner's gestural movements does not necessarily entail an absence of attention.

Turning to Rimé & Schiaratura's (1991) suggestion that observers are unable to guess the speech content to which gestures relate, at least two studies suggest that subjects are able to match words or expressions to gestures seen in the absence of speech. Rather ironically, the authors of these studies remain sceptical of a communicative view of gestures. Feyereisen, van de Wiele & Dubois (1988) presented subjects with a sample of gestures which had been previously judged as iconic. In a forced choice matching task, subjects were asked to attribute one of three verbal expressions to the gestures, either correct, plausible, or implausible words. Participants were more likely to match either correct or plausible expressions to the gestures than implausible words. In a similar study, Krauss, Morell-Samuels & Colasante (1991), again using a forced choice matching task, found that subjects were more likely to choose the "lexical affiliate" of a gesture presented without speech, than a lexical affiliate of a gesture chosen at random. Furthermore, subjects asked to guess the verbal expression associated with the gesture were more likely to write an interpretation similar to the actual affiliate than not. Subjects were also fairly accurate at assigning gestures to semantic categories in the absence of speech, although those who were presented with either speech alone or speech plus gesture were significantly more accurate.

These studies appear to suggest that subjects are at least able to derive some semantic information from the gestures, and in some cases even the appropriate lexical affiliate. However, despite these findings, both sets of authors are rather sceptical about the role played by gestures in communication. Feyereisen et al. (1988) concluded that gestures conveyed some semantic information but that this was only of a very general nature. At best, certain gestures suggested a range of possibilities rather than any precise meanings. Similarly Krauss et al. (1991) concluded that whilst gestures do convey some semantic information, they do not communicate as articulately as speech. They submit that gestures are unlikely to add any information to speech when that speech is intelligible and as such do not usually play a significant role in communication.

Rimé's ideas seem to be based primarily on observations which, it is argued, are somewhat flawed; subjects are unlikely to remember specific gestures in an interaction, but are more likely to remember the gist, or as in a study by McNeill et al. (1994), memory is determined jointly by verbal and gestural performances (see below); Face-to-face interactions have been shown to be different from interactions in unseen contexts; and finally there is some evidence to suggest that subjects are reasonably consistent in their judgements of the lexical affiliates and semantic content of gestures seen in the absence of speech.

As discussed above, researchers such as McNeill and Kendon offer a rather different view of the role of gestures in comprehension. For them gestural and verbal behaviours serve to represent different aspects of the underlying meaning that a speaker is striving to express. Gesture and speech might both refer to the same event but each offers a different view of it. One form of expression is thought to complement the other. An important point for the listener is that neither the gestural nor the verbal outcome alone is necessarily sufficient to specify the speaker's underlying cognitive representation. The complete picture can only fully be appreciated by considering jointly both gestural and verbal behaviours. Under this view a listener might be expected to process both the gestural and verbal components of an utterance, combining this information at some point in processing to provide an integrated representation of the speaker's intended meaning. Thus, in McNeill's terms, gesture and speech may well share a common computational stage in comprehension as well as in production.

There are a small number of empirical studies which suggest an interaction between speech and gesture in the comprehension process and a still smaller number of neuropsychological investigations which have been somewhat inconclusive.

Experimental Evidence

Empirical studies have examined subjects' comprehension of utterances under conditions where the accompanying gestures were made either visible or invisible (or not included in the utterance), whilst others have investigated the effects of gestural context on verbal memory.

Rogers (1978) compared speech comprehension in three different conditions. The same text was presented in full audiovisual, audio only and audiovisual but with facial cues used in lipreading masked. Various amounts of noise were added to the speech stimuli in order to reduce ceiling effects. Subjects' understanding of different grammatical aspects of the text was assessed. These included factors such as the agent (who or what acts), action (what is done), location and recipient of action. Rogers reported that there was better understanding of the text in both the visual conditions (which did not differ) as compared to the audio-only condition. Moreover, the facilitatory effects of visual cues on comprehension were inversely related to the signal-to-noise ratio, i.e. the noisier the speech the more subjects relied on visual cues. Thus subjects were clearly using both verbal and non-verbal information in the comprehension process.

Riseborough (1981) noted faster responses to a defined but unnamed object when the speaker accompanied the definition with an illustrative gesture. However this was only the case for one of the three items used as stimuli in the experiment. The remaining objects received responses at ceiling. Clearly this experiment was limited in that task difficulty was not equated across the (small) range of gesture and speech stimuli. Furthermore the temporal distribution of the input verbal and gestural stimuli was not controlled. This is critical if reaction time is used as a dependent variable.

Graham & Argyle (1975) had speakers describe a set of drawings under conditions where a folded arm posture was maintained and a second where gestures were allowed. The task for the audience was to reproduce these drawings based on the verbal description of the speakers. Drawings resulting from the gesture condition were rated as more similar to the original than those from the no-gesture condition. Moreover an analysis of the discourses produced in the two conditions showed very few differences, suggesting that the advantage to the listeners in the gesture condition resulted from the speakers use of hand movements and not because of a reduction in the fluidity in speech that has been noted when movements are restricted (e.g. Graham & Heywood, 1976).

Thompson & Massaro (1986, 1994) demonstrated that both adults and children relied heavily on referential gestures when making decisions concerning the identity of an object simultaneously referred to by synthesised speech syllables. Moreover these pointing gestures were found to have greater influence when the synthesised syllables were ambiguous.

Other studies have demonstrated an influence of gestural context on the recall of verbal material. Riseborough (1981, Experiments 2 & 3) noted that free recall of a list of verbs and cued recall of words from a short story were improved when the presentation of the verbal material was accompanied by illustrative gestures (*iconics*) as compared to conditions where the words were paired with "vague gestures" or no gestures at all. Similarly Woodall & Folger (1981, 1985) found superior recall of the verbal content of a conversation when representational gestures (again, *iconics*) were included. McNeill et al. (1994) presented subjects with a videotape of a person telling the story of an animated cartoon using a number of gestures which did not match the accompanying speech. In later recall of the story, subjects were found to be sensitive to these mismatches, producing changes in the story which were revealed in either their gestures, their speech, or in both their speech and gesture together. McNeill et al. (1994) claim that these results are consistent with the idea that speech and gesture combine into a single unit of meaning in comprehension. When faced with mismatches, subjects integrate information from both sources "without the necessity of conscious attention". Thus the integration process is deemed to be involuntary or automatic in character.

Neuropsychological Evidence

Taken together the above studies offer some evidence that listeners process the gestural information provided by a speaker. This information appears to influence the comprehension and recall of the verbal component of the utterance. This being the case, one might expect gestural cues to facilitate auditory comprehension in aphasia. However the few studies which have examined this hypothesis have yielded inconsistent and inconclusive results. Venus & Canter (1987) studied a group of 16 subjects with communication disorders. They found no differences in performance from matching spoken words to pictures when the words were either presented alone or with a gestural cue. On the other hand, Feyereisen & Hazan (unpublished) noted improved performance from two aphasic subjects in a similar word-to-picture matching task when a gestural cue was added to the word. However a third subject with similar verbal impairments showed no improvements in the gesture condition. The heterogeneity of the aphasic population makes it difficult to draw any firm conclusions from these studies. The verbal comprehension problems of the aphasic subject might stem from defective phonological processing, from impaired semantic access from verbal forms, or perhaps a general semantic disorder. In the latter case, gestural cues are unlikely to be of any assistance in the comprehension process.

Finally, support for the specificity of gesture processing comes from the observation of an agnosia for gesture. Rothi et al. (1986) described a 65 year old man who had sustained damage to the left occipital region. This man performed well below controls in tasks designed to assess gesture naming, discrimination and comprehension. He was only able to name 63% of the gestures performed by an examiner as compared to the 95.7% named by controls. In the discrimination task, the patient was asked to choose from among three gestures presented on a video, which best depicted the use of a target object. He performed at 68% accuracy on this task whereas control subjects performed flawlessly. Finally in the gesture comprehension task, the patient saw 20 trials of an actress performing "pantomimes". He was only able to indicate 50% of the acts being performed compared with 98% accuracy of control subjects. Whilst his ability to process gestural information was clearly impaired, the patient's ability to recognise pictures of objects remained intact, as was his auditory comprehension. In addition he was able to imitate gestures he could not name, ruling out any visual defects. Thus the patient's agnosia appears to be restricted to certain gestures. However, it should be noted that the patient was tested only on symbolic, pantomimic gestures. It remains to be seen whether a similar dissociation will be evident in the comprehension of those gestures which are closely associated with speech in production.

In summary the evidence reviewed suggests that gestures may indeed influence the comprehension and memory of verbal items in both normal subjects and perhaps aphasics, whilst the observation of a gestural agnosia indicates the possibility of a functionally specific "module" concerned with the identification of certain gestures.

Summary and Conclusions

Two questions were posed at the beginning of this section: Do listeners process gestural information and do they serve any communicative function? The evidence reviewed above suggests that both questions should be answered in the affirmative. A number of empirical findings suggest that the processing of verbal information may well be influenced by the associated gestural context. This implies that not only are gestures attended by listeners, but that this information is somehow combined with information in the verbal "channel". This position is at odds with Bernard Rimé's model of the comprehension of verbal and non-verbal information. Rimé suggests that because of the redundancy assumed to exist between gestural and verbal components of an utterance, listeners need only attend to the message in the verbal channel. In the interests of processing economy much of the non-verbal material is ignored under normal conditions. The evidence is more consistent with the positions of McNeill and Kendon who emphasise what they view as the non-redundant relationship between gesture and speech, the importance of gestures in communication, and consequently the need for listeners to encode and combine both sources of information.

The Present Research

The above discussion has stressed that the idea of gestures as part of a "body language" isolated from speech is mistaken. Instead the evidence points towards much closer relationships between the two sources of information. Most researchers agree that gestures arise as part of the speaking process but offer differing opinions concerning their communicative significance. Some see gestures as arising for the benefit of the speaker, functioning to maintain fluent speech and lexical access (e.g. Rimé & Schiaratura, 1991) and play only a minor role in communication (Krauss et al., 1991), whilst others stress that gestures and speech are equally important components of an individual's communicative effort (e.g. McNeill, 1985; Kendon, 1986).

The primary aim of the present work was to explore further the communicative status of the gesture, initially by demonstrating that certain *deictic* gestures are able to influence the comprehension of verbal material.

The understanding of deictic expressions seems to represent a fine example of the mutual role that gesture and speech have in fostering understanding. Deictic gestures are simply movements of the hand, head or fingers directed towards some actual or symbolically present object that is simultaneously referred to in the accompanying speech. Efron (1941/1972) suggested that deictics belonged to the class of gestures which could convey meaning independently of verbally expressed utterances, whilst others have suggested that they belong in a class of speech-related iconic gestures which carry no such independent meaning (e.g. Ekman & Friesen, 1972; McNeill, 1985; for a review of gesture classification see Rimé & Schiaratura, 1991). This type of gesture is clearly important in the comprehension of ambiguous deictic expressions such as "here", "there", "that one" or "this one". In addition, an appreciation of these kinds of gestures may also be important to the language learner. For a young child, understanding the referent of an unfamiliar word may well involve the processing of the speaker's "social attention". For example, if a father says to his young daughter "look at the horse", her understanding of the concept "horse" can only be facilitated if she understands what her father is referring to. This information may be gleaned from cues such as the direction in which his head, hand or body is pointing, or the location to which his eye-gaze is directed. In this case, complete understanding requires that both verbal and non-verbal information be taken into account. In this respect, pointing gestures and speech may well be mutually influential in comprehension.

In order to examine this proposition, an interference paradigm was adopted wherein gestural and verbal attributes are placed into conflict. This method has proved particularly useful as a diagnostic of the processing of an irrelevant or to-be-ignored aspect of a stimulus (see Chapter 2). The intention is to use this paradigm to demonstrate that a to-be-ignored deictic gesture can influence the comprehension of an auditorily presented word and reciprocally that an irrelevant spoken word can influence the response to a pointing gesture. This would constitute evidence for the mutual influence of the two sources of information in comprehension, a view consistent with those of researchers such as McNeill et al. (1994) who argue that listeners are impressed by both the verbal *and* gestural performances of the speaker.

As will become apparent in Chapter 2, interference effects may be viewed from the perspective of dimensional interaction, the study of how we are influenced by multiple sources of information in our environment. Under this view, interference effects arise because of a "crosstalk" of information between processing "channels" at some level of analysis. By considering the information stages involved in the processing of gestural and verbal information we can examine which stage or stages might be implicated in any interference effects that are obtained (see Chapter 4).

Finally, given some of the evidence reviewed above, we might ask whether gestures are "special" particularly with regards to their proposed interaction with verbal information in comprehension. This question is explored in Chapter 5 where again an interference paradigm is used to examine the extent to which non-linguistic visual stimuli are able to influence the processing of spoken words, and in Chapter 6 where the same question is asked of emblematic and iconic gestures.

Thus, the experiments reported in this work are aimed at addressing three principal questions. First, do subjects attend to or process speech related gestures in the comprehension of an utterance? Secondly, if gestures do interact with verbal material, where does the "crosstalk" of information occur? Finally, are gestures special in their interaction with verbal information? These issues are addressed in Chapters 3-7. In the following chapter the interference paradigm is introduced along with its use in the study of selective attention and dimensional interactions.

Chapter Two

Stroop Interference and Dimensional Interactions

The first of the three questions posed at the end of the last chapter asked whether or not people process or attend to the gestural information available in a speaker's utterance. It is suggested that this question is best addressed with the use of a Stroop-type interference paradigm (Stroop, 1935). In this chapter some of the work on the Stroop effect is reviewed along with a number of related experiments adopting Stroop-like procedures. The intention is to illustrate how this type of procedure can be used as a diagnostic of the processing of a to-be-ignored or irrelevant aspect of a stimulus, and to suggest that such a methodology is appropriate for the study of gesture and speech. By considering Stroop interference in the light of literature on dimensional interactions it becomes apparent that interference effects arise as the result of an interaction, or "crosstalk" of information between processing channels at some level of analysis. In the second half of this chapter this notion is introduced and the alternative loci of interference effects are discussed. It becomes clear that in addition to revealing the extent to which irrelevant gestural stimuli might be processed, the interference paradigm also provides a useful method for approaching the second issue of interest, namely the locus of any interference effects which are obtained.

The Stroop Effect

The Stroop effect is one of the more persistent findings in experimental psychology. Despite over 700 articles spanning the 60 years since Stroop's seminal paper, the effect has proved rather resistant to coherent explanation. In the original version of the Stroop colour word interference task (Stroop, 1935) subjects were slower to name the colour of the ink in which an incongruent colour word was printed (e.g. RED in blue ink) relative to a control condition of naming solid colour squares. However, reading the colour word was largely unaffected by the ink colour in which it was printed. The effect seems to demonstrate how an ignored or irrelevant aspect of a stimulus can "capture" attention. This fact has been exploited by a number of authors with a wide range of research interests, from the study of attentional bias to threat-related material (e.g. Mathews & MacLeod, 1985) to cognitive functioning of bilinguals (e.g. Gerhand, Deregowski & McAllister, 1995). At a more fundamental level, the Stroop effect and its variants (see below) have been used as evidence for late selection theories of selective attention. Briefly, proponents of this idea (e.g. Deutsch & Deutsch, 1963) argue that attention operates *after* stimulus identification to select material which might demand some response, influence cognitive activity, or perhaps be stored in long term memory. In contrast, early selection theories, originally propounded by Broadbent (1958), hold that attention serves to select the representations of rudimentary features from the total input (e.g. colours, edge orientations, locations in the case of vision). Such features can then be integrated or bound together in some way so that stimulus identification can occur (e.g. Treisman & Gelade, 1980). Thus the main bone of contention between the two theories is the extent to which unattended stimuli achieve identification. The Stroop effect, according to the late selection theorists, establishes that irrelevant or unattended words are identified as the effect depends on the word's meaning. The important point is that the observation of Stroop interference can be regarded as diagnostic of the processing of an irrelevant dimension of the stimulus. Indeed, this is how the task is to be used in the present studies.

The Stroop Asymmetry

An important aspect of the Stroop effect to bear in mind is its asymmetry. The typical finding with colour-word stimuli is that colour naming is affected by the identity of the to-be-ignored word, but word reading is unaffected by the colour of the ink in which the word is printed (e.g. Stroop, 1935). This asymmetry is a robust and almost defining characteristic of Stroop interference and many of its derivatives (see below). Its importance is illustrated in the efforts of researchers seeking to produce a general account of interference. These models all include assumptions or mechanisms which create a measure of imbalance in the processing of the component stimuli (e.g. Posner & Snyder, 1975; Logan, 1980; Virzi & Egeth, 1985; Glaser & Glaser, 1989; Cohen et al, 1990). Several of these models are elaborated and critically evaluated in Chapters 4 and 7. Symmetrical or bi-directional interference effects are seldom obtained in the Stroop-type interference task although there have been a few isolated examples (e.g. Shor, 1971; O'Leary & Barber, 1993; see Chapter 3). As will become apparent, observations of this kind have important ramifications for interpretations of Stroop interference.

Development of The Stroop task

The last 60 years has also seen a large number of adaptations to Stroop's original task both in terms of methodology and the choice and presentation of the interfering stimuli. In the remainder of this chapter some of these variants are reviewed. Since the list is long and the diversity of studies almost overwhelming, the present review is limited to manipulations relevant to the proposed application of the task to the study of gesture and speech comprehension. It examines extensions of the original colour-word task to experiments involving spatial and picture-word stimuli along with auditory and cross-modal analogues of the effect.

Methodological Developments

The most prominent variation from Stroop's original experimental procedure was the change from the timing of *lists* of words etc. to the presentation and timing of responses to individual stimuli, the method currently used in most studies of Stroop interference. This procedural modification allows for randomisation of the stimulus materials, however the asymmetry characteristic of the original Stroop task remains prominent in this manipulation (e.g. Dunbar & MacLeod, 1984; Glaser &

Glaser, 1982). That is, colour naming is affected by the irrelevant word, but word reading is unaffected by the colour of the ink in which the word is printed. This asymmetry is a robust and almost defining characteristic of Stroop interference and many of its derivatives (see below). The asymmetry is seldom reversed and there are very few examples of symmetrical interference where each dimension exerts effects on the processing of the other. It is important to bear these points in mind for the purposes of the present research. Along with the use of individual items, many have switched the task from a verbal to a manual response. Under these conditions the original Stroop effect persists, although perhaps slightly attenuated (e.g. Logan et al., 1984; Keele, 1972; Roe et al., 1980). This result is important when we come to consider the role of stimulus-response compatibility in producing interference effects (Chapter 4). Briefly the idea is that Stroop interference occurs because the verbal nature of the irrelevant word stimulus matches that of the required response. To summarise a vast literature, the evidence seems to suggest that whilst stimulus-response compatibility matters, it is not in itself fully sufficient to explain the Stroop effect.

Dalrymple-Alford & Budayr (1966) were also the first to introduce congruent items among incongruent ones with Stroop stimuli, a condition that Stroop himself did not include. This allows for the measurement of facilitation (i.e. a processing "gain" over a control condition) as well as the normal interference (often referred to as inhibition). As MacLeod (1991) points out, the choice of the control condition appears to be critical in measuring the relative contributions of inhibition and/or facilitation. Originally Stroop used colour patches as the colour-naming control but these were replaced by a nonalphanumeric character as Stroop considered this more akin to a letter in appearance (Stroop, 1935, Experiment 2). His choice of the swastika as this control character might now be regarded as somewhat unfortunate. Subsequent studies have used rather less provocative controls; unrelated words (e.g. Dalrymple-Alford, 1972; Redding & Gerjets, 1977); nonwords (e.g. Hintzman et al., 1972; Redding & Gerjets, 1977); and rows of X's (e.g. Dalrymple-Alford, 1972;

Glaser & Glaser, 1982; Dunbar & MacLeod, 1984). On the whole, congruence between the irrelevant word and the to-be-named ink colour results in facilitation although the size of this effect is generally much less than the corresponding interference in the incongruent condition (MacLeod, 1991). However, Lindsay & Jacoby (1994) have recently suggested that there is "no type of control item that can reliably provide a factor pure baseline measure of colour naming." (p.219). They claim that the existence of interference and facilitation can be established using a baseline measure but that this technique cannot be used accurately to measure the magnitudes of these effects.

In summary, the main developments in methodology from Stroop's original article have been the switch from lists to individual trials, the switch from a verbal to a manual response, and the inclusion of congruent colour word stimuli. The latter has resulted in a debate concerning the choice of an appropriate baseline measure required to assess the relative contribution of facilitation and inhibition. In general it seems that the size of the inhibitory effect outweighs that of facilitation in the Stroop task. The discussion of interference versus facilitation is elaborated in later chapters.

Variants of the Stroop Task

Picture-Word Interference

The basic Stroop asymmetry has been obtained with picture-word stimuli. In this variant, words are printed inside congruent or incongruent pictures. Rosinski et al. (1975) obtained classic Stroop interference with combinations of picture-word stimuli, that is, incongruent words caused a strong inhibition in a picture naming task whereas pictures had only weak effects on word reading. Others have established that this asymmetry is at least as consistent as that obtained with Stroop colour-word stimuli (e.g. Glaser & Düngelhoff, 1984; Smith & Magee, 1980). Moreover, congruent words appear to facilitate picture naming whilst incongruent words cause inhibition (e.g. Rayner & Posnansky, 1978; Underwood, 1976).

Clearly the picture-word effects are similar to those evident with the traditional Stroop stimuli. Both result in asymmetrical interference effects and, as Glaser & Düngelhoff (1984) point out, both have similar time courses.* In fact the similarities are such that Glaser & Glaser (1989) consider that both effects are based on the same underlying cognitive process, specifically that the normal Stroop effect should be considered as a special case of a general reading-naming interference. Glaser & Glaser's (1989) model of Stroop-like effects is critically discussed in Chapter 7.

Spatial Stimuli

In a further variation of the Stroop task, Seymour (1973a) presented the words ABOVE, BELOW, LEFT and RIGHT to his subjects and demonstrated both inhibition and facilitation when subjects were required to name the position of the word relative to a dot. Indeed the use of conflicting spatial attributes of various stimuli has been popular in the Stroop literature. For instance Shor (1970, 1971) embedded the words LEFT, RIGHT, UP and DOWN in arrows pointing in directions incongruent with the embedded word and obtained interference in naming the direction of the arrow. Dyer (1972) obtained inhibition and facilitation effects from the same four directional words which he moved in either the congruent or incongruent direction. White (1969) reported interference from the words NORTH, SOUTH, EAST and WEST when subjects were asked to name the word's position inside a rectangle, whilst Clark & Brownell (1975) obtained interference from the height of an arrow printed inside a rectangle when the criterial attribute was the direction of the arrow.

---

* The time courses of Stroop and picture-word effects are determined by varying the temporal separation between distractor and target onset (SOA). Normally both are presented simultaneously (SOA = 0 ms), however distractors can precede targets (negative SOA's) and vice-versa (positive SOA's). Typically interference in both colour naming (Glaser & Glaser, 1982) and picture naming (Glaser & Düngelhoff, 1984) is maximal with SOA's between -100 and +100 ms.

Virzi & Egeth (1985) managed to reverse the usual Stroop-type asymmetry, obtaining an interference effect from the spatial location (to the left or right of fixation) of a directional word (LEFT or RIGHT) when the task was to make a *manual* response to the meaning of the word. On the other hand the response to the position of the word was unaffected by its meaning. Virzi & Egeth (1985) suggested that when interference occurs, values on the relevant dimension (e.g. verbal) require a translation into the cognitive domain of the response (e.g. spatial keypress) but that the irrelevant stimuli require no such translation (e.g. spatial location to a spatial keypress response). This explanation has much in common with the notion of stimulus-response compatibility and is considered further in Chapter 4.

Virzi & Egeth's findings have much in common with the so-called "Simon effect" (after Hedge & Marsh, 1975) first noted by Simon & Rudell (1967). In this paradigm subjects might typically make a left/right keypress response contingent on the *identity* of a stimulus (e.g. colour of light, direction of arrow, the words "left" or "right" etc.) presented randomly to the left or right of some central point. The results indicate that the location of the stimulus provides an irrelevant cue that interferes with the processing of the target stimulus. Thus in Simon & Rudell's original paper subjects were slower to respond with a right keypress to the word "right" presented in the *left* ear than to an identical stimulus presented to the *right* ear. Thus the reverse Stroop-type effect observed by Virzi & Egeth (1985) might also be treated as an example of the Simon effect, the irrelevant spatial location of the directional words produced an interference effect when making a manual response to the meaning of the word.

One point of note is that in both the colour-word and picture-word tasks the size of the effect often exceeds 100 ms whilst the magnitude of the effect is smaller for spatial stimuli (e.g. around 20 ms in the Dyer (1972) study).

Cross-Modal Effects

There is also some evidence that both Stroop and picture-word interference occur when opposing dimensions are presented cross-modally. Cowan & Barron (1987; see also Cowan, 1989a, 1989b) had subjects perform a colour naming task to lists comprising incongruent colour words (as in the original Stroop task) and lists comprising strings of the letter *x*. Whilst performing these tasks, subjects were presented with each of 5 types of auditory distractor. These consisted of audiotapes playing spoken colour words, letters of the alphabet, repetitions of the word "the", music (Bartok's "Sonata No.2 for Violin and Piano") and silence. Subjects' colour naming performance was significantly slower on the colour-word card than on the control card, the usual Stroop effect. In addition, spoken colour words produced slower colour naming performance on both colour words and *x*'s relative to the other auditory conditions. However, Miles, Madden & Jones (1989) and Miles & Jones (1989) failed to replicate this type of cross-modal Stroop effect arguing that "irrelevant speech" should only produce interference effects on tasks involving the rehearsal of items prior to ordered recall.

Clearly Cowan & Barron's (1989) procedure is rather different to those used in usual Stroop-interference experiments. In their studies, Cowan & Barron's irrelevant auditory stimuli were presented at a continuous rate, regardless of the rate at which the colour-word lists were read. Normally a distracting stimulus in a Stroop-type study is presented at the same time as the target stimulus so that each target stimulus is paired with a single irrelevant word. Perhaps it is this procedural modification which results in the Cowan's seemingly ephemeral effects. However, the arguments in favour of the cross-modal Stroop effect are bolstered by Shimada (1990) who noted significant interference effects of auditorily presented colour words on the naming of colour patches in a more typical Stroop-type procedure.

Turning to the picture naming task, Schriefers & Meyer (1990) have demonstrated cross-modal, visual-auditory picture-word interference. Longer picture naming latencies were observed when the picture was paired with an associated or unrelated word than in a silence condition. A similar result has also been reported by Stuart & Carrasco (1993). Taken together the available evidence points toward the existence of cross-modal Stroop interference.

Summary

The Stroop effect provides an excellent illustration of how a stimulus which is irrelevant to the current goal of the subject is nevertheless still able to influence performance. In this respect the Stroop effect and related interference paradigms serve as useful diagnostics of the processing of to-be-ignored information. Stroop's original procedure has been refined and extended over the years from presenting lists of colour words for subjects to read, to the presentation of single items, including picture-word and spatial stimuli, which command either a speeded naming or a manual keypress response. Moreover, both the Stroop and picture-word effects have been obtained with cross-modal presentations of the stimuli. The usual finding is that the verbal attribute influences processing of the pictoral/spatial attribute but not vice-versa. However, this usual asymmetry has been reversed where subjects are asked to make manual judgements based on the spatial attributes of the various stimuli.

Dimensional Interaction

Rather than considering interference effects as simply reflecting the processing of an unattended dimension, the effects might be explained in terms of "breakdowns" of selective attention. If subjects show interference effects in a focused attention experiment such as the Stroop task, then selective attention has faltered to some degree. More specifically, selective attention fails because information from the separate sources *interacts* at some level of processing. Thus, under this view attentional effects emerge because of the nature of the processing of the component dimensions. Stroop interference can therefore be considered as indicative of some kind of dimensional interaction, as well as diagnostic of the processing of the to-be-ignored or distracting information.

Garner Interference

The recent work of Robert Melara and his colleagues has been particularly influential in the study of dimensional interaction and selective attention (e.g. Melara & O'Brien, 1987; Melara, 1989; Melara & Marks, 1990a, 1990b, 1990c; Melara & Mounts, 1993). In these studies congruity effects are measured within the "Garner Interference" paradigm (e.g. Pomerantz, 1983, 1986) which essentially examines the effect which orthogonal variations in the irrelevant dimension have on selective attention to the relevant dimension of the stimulus. For instance in their work with the classic Stroop dimensions (e.g. Melara & Mounts, 1993), subjects might respond to the identity of a colour word in one block of trials whilst the irrelevant colour of these words remains unchanged. This forms the "baseline" condition. In the "filtering" condition, subjects continue to respond to the identity of the word but this time the colour of the ink varies orthogonally (i.e. black on one trial, white on another etc.). Garner interference is defined as the difference between baseline and filtering RT's and is a measure of the ability of the variation in ink colour to disrupt selective attention to word identity. Pairs of dimensions which produce significant Garner interference are said to be *interacting* dimensions whereas pairs which produce little or no Garner interference are considered to be *separable*.

Levels of Crosstalk

One of Melara & Marks' goals was not only to identify pairs of interacting dimensions but to determine at what level(s) of processing and in which direction(s) the interaction or *crosstalk* occurs. Students of the Stroop effect and its analogues have been involved in a similar endeavour. For a classification response to be made to any stimulus, it is assumed that the stimulus must undergo a number of distinct levels of analysis. For instance in order to respond to a picture of an object, some perceptual representation of that object must first be constructed. This representation is then compared with some description of that object stored in long term memory. The object's semantic representations can then be accessed and finally the appropriate object name can be retrieved from a stored "output lexicon" (e.g. Ellis & Young, 1988). Alternatively, if a manual response is required, the encoded object information must be mapped onto the appropriate response code by some decision process. In general we can consider three stages; perceptual analysis leading to stimulus identification*; semantic processing; and response selection. Work within the Garner interference paradigm has mainly considered crosstalk at the perceptual and semantic levels whereas research into Stroop, Simon and Picture-word interference has predominantly concentrated on semantic and response level interactions.

Perceptual Crosstalk

So-called *integral* dimensions such as the saturation and brightness of colours (Garner & Felfoldy, 1970), or the pitch and loudness of vibrations (e.g. Melara & Marks, 1990a) are thought to interact at an early, perceptual stage of processing. High pitched notes "sound" louder than low pitched tones and yellow hues appear to us as brighter than greens or blues. On one view, stimuli formed by combining attributes from these dimensions (e.g. a "bright" blue or a loud, low pitched tone) are initially perceived as dimensionless because of the perceptual "fusion" or interaction between the dimensions (e.g. Garner, 1981 Lockhead, 1979). These stimuli are perceived as "blobs" or "wholes" and can only be broken down into their constituent "parts" by applying some sort of cognitive effort. An alternative view (Melara & Marks, 1990a, 1990b) suggests that perceivers do in fact extract information from each of the dimensions comprising a multidimensional stimulus. So, when presented

---

* In the case of verbal stimuli, the process of identification will involve access to either stored visual (graphemic), or auditory (phonemic) descriptions of those words, depending on the input modality.

with a set of sounds, subjects perceive each sound as comprising attributes from psychologically real dimensions such as loudness, pitch, timbre etc. However, the context established by the presence of one attribute (e.g. high pitch) acts to constrain the perceptual experience of the other attribute (e.g. loud). This is what Melara & Marks regard as perceptual crosstalk.

Returning to the area of communication, McGurk & McDonald (1976) have demonstrated how visual information can influence the perception of speech sounds. In a typical "McGurk" experiment, subjects might see the shape of a mouth uttering the syllable /ga/ whilst being simultaneously presented with the sound /ba/. In this instance subjects often report hearing a blend of the two pieces of information, the syllable /da/. This is a striking demonstration of a perceptual "fusion" between the verbal and visual dimensions. Outside the rather contrived laboratory situation, we all use information from the shape of the mouth and lips to constrain our perception of phonemic information. Consider the ease with which we are able to distinguish the similar sounding letters "F" and "S" in a face-to-face conversation. We are able to use the differing mouth shapes formed when these letters are uttered to "weight" the perception of the speech sound. In the absence of this visual information similar sounding phonemes become hard to distinguish.

There are a number of dimensional interactions which are clearly not located at a perceptual level. The Stroop effect, for example does not appear to be a perceptual phenomenon. According to Melara & Marks (1990b) certain dimensions, including the Stroop stimuli, *interact* but are not *integral.* For these dimensions, crosstalk occurs at more central levels of processing.

Semantic Crosstalk

The observation of Garner interference suggests that subjects have trouble classifying one dimension in the presence of variation in the irrelevant dimension. However, with certain pairs of dimensions subjects are also sensitive to the *kind* of dimensional change. That is they appreciate that the value or meaning of the irrelevant dimension conflicts with that on the relevant dimension. Melara & Marks suggest that this type of interaction produces a significant congruity or Stroop-like effect, measured by subtracting reaction times to congruent pairs from those to incongruent stimulus pairs (both congruent and incongruent stimulus pairs appear in baseline and filtering blocks).

By measuring congruity effects, Melara and his colleagues (Melara, 1989; Melara & O'Brien, 1987; Melara & Marks, 1990b) have suggested that a number of *corresponding* dimensions interact at a semantic level of processing. Attributes from certain pairs of dimensions are considered to correspond well with one another both within and between sensory modalities. For instance most people report that high pitched sounds are associated with bright colours. In fact, certain individuals, known as synesthetes, actually report *experiencing* bright colours on hearing a high pitched tones. For these reasons, pitch and colour are considered to be corresponding dimensions. Melara & O'Brien (1987) noted both symmetrical Garner interference and congruity effects between the corresponding dimensions of vertical position and pitch, whilst Melara (1989) later replicated these findings in experiments involving the dimensions of colour (black/white) and pitch (high/low). These bi-directional effects have also been obtained between corresponding linguistic and nonlinguistic dimensions. Melara & Marks (1990b) obtained symmetry between the visual words "HI" and "LO" and low or high pitched tones, between the same visual words and the vertical position from which they were presented (Experiment 2), and a similar interaction between the spoken words "high" and "low" and their vertical position (Experiment 3). Melara & Marks reject the thesis that these dimensions interact at a perceptual level as the qualitatively different nature of the dimensions make it unlikely that subjects fail to recognise the individual constituents of the multidimensional stimuli, perceiving them holistically as a single "blob" or point in psychological space. Rather, they suggest that the correspondences between these dimensions should be considered as arising from bi-directional crosstalk at a semantic level of analysis, with the extraction of meaning units from one attribute affected by the processing operations in the second channel.

A number of authors have suggested that Stroop interference has a semantic locus (e.g. Seymour, 1977; Stirling, 1979). The fact that a semantic "gradient" has been noted in the original Stroop task seems to confirm this point. Klein (1964) noted large interference (37.5 s) in the normal incongruent colour-word task with the words RED, GREEN, YELLOW and BLUE and smaller interference (15.5 s) for colour-related words (FIRE, GRASS, LEMON and SKY). The size of the effect dropped to 12 s with unassociated words (PUT, HEART, TAKE and FRIEND), to 7.5 s for rare words (SOL, HELOT, EFT and ABJURE) and to 5 s for nonsense syllables (HJH, EVGCJ, BHDR and GSXRQ). Dalrymple-Alford (1972) replicated Klein's findings but also noted that both congruent colour-words and colour-related words caused reliable facilitation. In general it appears that more interference is obtained as the semantic similarity between the colour and the word increases.

In addition to the colour naming task, the picture-word and spatial variants of the Stroop effect have been shown to be sensitive to the relationship between the target and distractor stimuli. Several researchers (Rosinski, 1977; Lupker, 1979; Glaser & Düngelhoff, 1984) have observed more inhibition from distractor words belonging to the same semantic category as the picture (e.g. the word DOG on the picture of a mouse) than from words from another category (e.g. the word CAR on the picture of a mouse). Fox et al (1971) replicated Klein's (1964) semantic gradient effects with colour-word stimuli and extended their investigation to spatial stimuli. These consisted of words printed at various locations inside a square. Subjects suffered interference effects when their task was to name the location of the word. Moreover the "same" direction names (UP, DOWN, LEFT and RIGHT) caused more interference than "different" direction names (NORTH, SOUTH, EAST and WEST).

Response Stage Crosstalk

Although firmly rejected by Melara (e.g. Melara, 1989; Melara & O'Brien, 1987; Melara & Marks, 1990b) a number of dimensional interactions might well be rooted in response competition. Indeed, traditional accounts of both the Simon and Stroop effects suggest that the stage where the response is selected is the most likely source of the interference effect (e.g. Morton, 1969; Posner & Snyder, 1975; Kornblum et al., 1990). According to this view, the two dimensions are processed in parallel culminating in the determination of two separate response codes, one of which must be selected in order for the correct response to be programmed and executed. Inhibition occurs when the various dimensions evoke conflicting response tendencies leading to slowed overall classification time. Conversely, congruent stimuli produce identical response tendencies which improves overall classification performance. A number of models of Stroop interference which place the locus of the effect at this response stage are further discussed in Chapter 4.

Summary

From one perspective, interference effects can be viewed as resulting from the crosstalk of information from different dimensions. From this standpoint, attentional effects are seen as rather more "stimulus driven" or bottom-up, emerging from the interaction of different sources of information. The challenge for the researcher is not only to identify which dimensions interact, but to determine under what circumstances and at what level of processing crosstalk occurs.

The Present Experiments

To return to the area of gesture and speech comprehension, recall that the ideas of researchers such as McNeill and Kendon lead to the prediction that both gestural and verbal material should be processed in comprehension, and that this information should be combined somehow in order to furnish the listener with a complete representation of the speaker's intended meaning. By placing gestural and verbal information into conflict in an interference paradigm, both these points can be addressed. For instance, the observation of an interference effect clearly indicates that the irrelevant stimulus (e.g. a gesture) has been processed. Furthermore, as the above discussion has stressed, this kind of effect can be thought of as arising from the crosstalk of the component sources of information. Thus, the objectives of the present experiments are threefold: first to ascertain whether or not gestures are processed along with spoken words in comprehension; secondly, given any interaction, to determine the locus of the crosstalk or interaction which precipitates the interference; and thirdly to explore whether or not the peculiarly linguistic, or language related status of gestures contributes to any influence which they exert in the comprehension process.

In the experiments reported in Chapter 3, pointing gestures are paired with either congruent or incongruent verbal attributes and subjects are asked to respond to either the verbal or the gestural stimuli in separate blocks of trials. By comparing the speeds at which subjects are able to identify words in the face of congruent versus incongruent gestural information, or gestures in the face of congruent versus incongruent verbal information, the extent to which subjects process *both* gestural and verbal information can be explored. The symmetrical or bidirectional interference effects obtained in these experiments offer support for the notion that gestures and words are mutually influential in comprehension.

Chapter 4 addresses the second of the objectives, to locate the source of the interaction. The idea is to determine where in the sequence of processing operations any "crosstalk" of information occurs. As described earlier, the observation of an interference effect is consistent with the position that information encoded from the interfering dimensions interacts at either a perceptual, semantic or response stage of processing. Chapter 4 addresses the notion that deictic gestures and spoken words interact at the latter response selection stage of processing. More specifically, the experiments reported in this chapter are designed to evaluate the hypothesis that deictic gestures exert their influence by virtue of their compatible relationship with the type of spatial response task used in the experimental procedure. The results of these studies offer little support for this "stimulus-response compatibility" (SRC) account, nor indeed for other response selection models of Stroop-type interference. It is concluded that the gesture-word interference effects are likely to be produced by the interaction of information at a logically earlier semantic stage of processing.

The experiments reported in Chapters 5 and 6 together attempt to address the possibility that gestures are "special" with regard to their interaction with verbal information. In Chapter 5 it is suggested that deictic gestures are processed *spatially* rather than as gestures per se. The experiments reported in this chapter offer some support for this suggestion. In these studies, non-linguistic directional cues (arrows and spatially located dots) and spoken words produce qualitatively similar interference effects to those obtained with deictic gestures and verbal materials in Chapter 3. The suggestion is perhaps further bolstered by the results of the experiments reported in Chapter 6. Here the technique is extended to the use of non-spatial emblematic and iconic gestures, neither of which appear to interact with spoken words.

The final set of experiments described in Chapter 7 are an attempt to extend the investigation from bodily to facial gestures. A mutual interaction of facial expressions and spoken words is obtained when the task is to make an affective judgement to either stimulus. However the usual Stroop asymmetry is obtained when using an identification judgement as the experimental task.

Chapter Three

Gesture-Word Interference Effects

In Chapter 1 current views on the production of gesture and speech were introduced. These theoretical positions on *production* provided a springboard for the development of theories of gesture and speech *comprehension*. In Chapter 2 it was suggested that a Stroop interference paradigm might provide a useful tool for investigating the extent to which gestural and verbal dimensions influence one another. In this chapter, four experiments are presented which adopt an interference paradigm, placing pointing gestures and verbal information into conflict. The results provide support for a model where listeners process and combine the two sources of information in the comprehension process.

Gesture and Speech Revisited

As discussed in Chapter 1, current theoretical ideas on the production of gesture and speech indicate a rather closer relationship between the two dimensions than was considered in the "body language" approach. However, views differ on the function of gestural behaviour. Some suggest that gestures are merely a by-product of verbal performance (e.g. Rimé & Schiaratura, 1991; Freedman, 1972) whilst others emphasise their communicative function, viewing them as an essential "part" of the utterance as a "whole" (e.g. McNeill, 1985; Kendon, 1986). The implications for comprehension are clear. Under the former view, gestures are not expected to influence a listener's decoding activity, whereas under the latter, a speaker's intended meaning can only be fully appreciated by a consideration of both verbal *and* gestural behaviour.

The existence of an interference effect can be interpreted as evidence that the to-be-ignored attribute of the stimulus has received some degree of processing. Of particular interest in the study of the interaction of gesture and speech is whether an irrelevant gesture will interfere with the processing of a spoken word. If movements of the hands and/or body provide only redundant information we might not expect them to be attended and as such they should not interfere with the processing of verbal information (cf. Rimé & Schiaratura, 1991). On the other hand, McNeill's ideas on the complementarity of gesture and speech suggest that both dimensions should be attended. In this case one might expect a mutual influence of the two sources of information. Gestures should interfere with the processing of information in the verbal channel and reciprocally, verbal information should interfere with the processing of gestural information. This might materialise as a symmetrical pattern of interference.

Symmetrical Interference Effects

As discussed in Chapter 2, asymmetry is the norm in Stroop and picture-word interference. Typically, the verbal dimension exerts interfering effects on target visual stimuli (e.g. colours or pictures), whilst relevant verbal stimuli generally appear to be immune from interference from to-be-ignored visual items. In fact this reverse Stroop effect (interference in word reading caused by an incongruent colour) has proved to be notoriously difficult to obtain (e.g. Glaser & Glaser, 1982; Glaser & Düngelhoff, 1984; Dunbar & MacLeod, 1984; MacLeod & Dunbar, 1988). The few cross-modal studies that exist did not investigate the influence of the colour or picture on the processing of the auditorily presented word (Cowan & Barron, 1987; Shimada, 1990; Schriefers & Meyer, 1990).

In recent years a number of studies claim to have demonstrated symmetrical interference effects with Stroop-like stimuli. The work of Robert Melara and his colleagues, using the Garner interference paradigm (see Chapter 2), has been particularly influential in this regard (e.g. Melara & O'Brien, 1987; Melara, 1989; Melara & Marks, 1990a, 1990b; Melara & Mounts, 1993). Under conditions where baseline discriminabilities of colour and word dimensions were matched (see Chapter 8), Melara & Mounts obtained small but symmetrical congruity effects in a

Stroop task. Within the same paradigm, symmetrical congruity effects were noted with the dimensions of position and pitch, e.g. a high or low pitched tone presented at either high or low spatial positions (Melara & O'Brien, 1987); Colour and pitch, e.g. white or black dots paired with either high or low pitched tones (Melara, 1989); visual words and position, e.g. the words "HI" and "LO" presented either above or below the midline of a computer screen (Melara & Marks, 1990a; Experiment 2); and spoken words and position, e.g. the words "HI" and "LO" auditorily presented from high or low spatial positions (Melara & Marks, 1990a; Experiment 3).

Symmetrical interference patterns have also been reported in spatial variants of the Stroop task itself. Shor (1971) reported such effects using directional words embedded in arrows employing a naming task, whilst more recently O'Leary & Barber (1993) managed to obtain symmetrical interference using the words LEFT and RIGHT to the left and right of the screen with verbal naming responses to either the location or the meaning of the word. Symmetrical effects were also obtained using a left/right keypress response again to the meaning or location of the stimulus word (the Simon effect and reverse Simon effect respectively), but only when the salience of the irrelevant word in the location judgement task was marked.

In summary, over the past 60 years since Stroop's original article, an overwhelming number of studies (Macleod, 1991 reckons on over 700) have appeared in the literature. A great many of these have reported Stroop-like asymmetries. However, despite being relatively uncommon, particularly within the Stroop, as opposed to the Garner interference paradigm, symmetrical interference effects have recently materialised in the experimental psychology literature.

In the present experiments it was decided to investigate any mutual interaction of deictic (pointing) gestures and verbal information using an interference paradigm. In Chapter 1 it was suggested that these are exactly the kind of movements which a listener might be expected to process in comprehension, as the complete understanding of certain deictic expressions requires that both verbal and gestural information be taken into account. The beauty of these gestures, at least for the present purposes, is that they can be adequately represented as static images and that their meaning can be expressed as a single word. Moreover, by using gestural and verbal representations of spatial directions it was relatively easy to construct congruent and incongruent gesture-word pairs analogous to the pairings used in the picture-word and spatial variants of the Stroop task reviewed in Chapter 2.

In Experiments 1-4 verbal attributes were paired with either congruent or incongruent gestural attributes with a manual response required to both dimensions presented in separate blocks. Experiments 1, 2 and 4 made use of these static deictic gestures with their spoken verbal equivalents. Experiment 3 investigated the within-modality interaction by replacing the auditory stimuli with written words.

Experiment 1

In this experiment static directional gestures (up, down, left and right) were paired with their congruent or incongruent verbal equivalents and subjects' reaction times were measured in a *manual** response to both verbal and non-verbal dimensions in separate blocks. The "complementarity" hypothesis suggests that listeners process and combine information in both the gestural and verbal channels. In this case a mutual interaction of the two dimensions is predicted. When responding to spoken words, subjects should be slowed by the presence of an incongruent, relative to a congruent gesture, indicating that the gestural information has been processed. Similarly, as in the usual Stroop effect, responses to the visual gesture stimuli should be slower when paired with incongruent, compared to congruent spoken words. Thus a symmetrical pattern of interference effects is predicted. On the other hand, the "redundancy" hypothesis suggests that gestures and speech express essentially identical information. Under this view, gestures serve

---

* When the task is switched from a verbal to a manual naming response the original Stroop effect persists although perhaps slightly attenuated (e.g. Logan et al., 1984; Keele, 1972; Roe et al., 1980).

very little communicative function and are therefore not expected to be processed in the comprehension of an utterance. If listeners largely ignore gestural information, a normal Stroop asymmetry would be the expected result. Responses to the verbal dimension will be unaffected by the presence of irrelevant deictic gestures, whereas irrelevant spoken words might be expected to influence responses to the gestural stimuli.

To recapitulate, the complementarity hypothesis predicts a symmetrical pattern of interference effects, congruity effects of equal sizes should be observed when responding to gestural and verbal stimuli respectively. This will manifest as a main effect of congruity (i.e. a significant difference between congruent and incongruent RT's) with no interaction with the type of response. The redundancy hypothesis predicts that a congruity effect should only be evident when responses are made to the gestural dimension. In this case an interaction between congruity and response factors is expected.

Method

Subjects: Twelve undergraduate and postgraduate volunteers participated in this experiment, all were between 18 and 30 years of age with normal or corrected to normal vision.

Materials and Apparatus: Frame grabbed images of a person gesturing to the left, right, up and down were obtained using a Sun workstation. Examples of the images are presented in Figure 1. These stimuli subtended approximately 13° of vertical visual angle. The verbal stimuli were recorded using audio software on hypercard and edited using "Soundedit" software on a Macintosh IIci. Four spoken words (left, right, up and down) were recorded and edited to be approximately the same length (0.8 secs). The visual and auditory stimuli were presented together using the "SuperLab" software on the Mac IIci , this enables the two types of stimuli to be presented simultaneously and claims to record reaction times to four millisecond accuracy.* The onset of the presentation of the speech stimuli coincided with the presentation of the visual gesture stimulus.

Two types of test stimuli were prepared from the frame grabbed images and the auditory speech stimuli. Congruent trials consisted of the spoken word "up", "down", "left" and "right" paired with their respective gestures, whilst incongruent trials consisted of the spoken words "up" and "down" paired with *down* and *up* gestures respectively along with "left" and "right" paired with *right* and *left* gestures respectively.

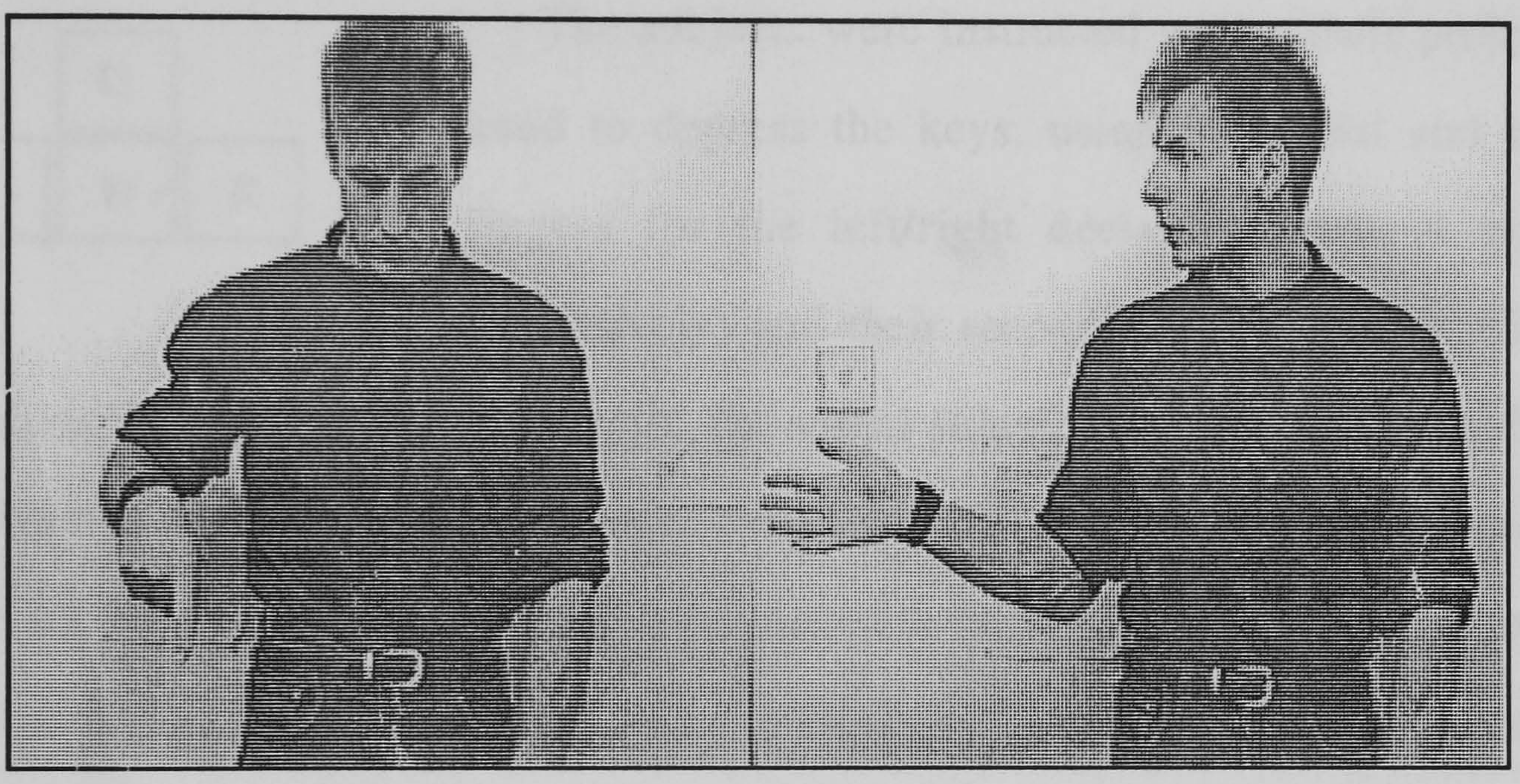


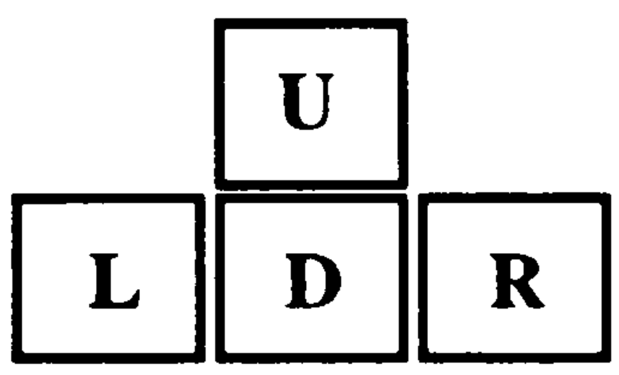
Figure 1. Examples of the "down" (left panel) and "left" (right panel) pointing gestures used as stimuli in Experiment 1.

Design: The experiment took the form of a 2 x 2 x 2 repeated measures design. The three factors were: Response (either to a gesture or to a voice), Congruity (congruent or incongruent pairs of stimuli) and Decision (either left/right or up/down decisions). The stimuli were blocked by response, either to gesture or to voice. Half of the subjects responded to the voice first and half to the gesture. With ten trials per cell, subjects responded to 40 trials in each response block.

---

* In this and all other experiments, manual responses were recorded using the Apple ADB keyboard using the Time Manager 1 timer. Cedrus, the software manufacturers, *claim* four millisecond accuracy for this configuration using a Macintosh IIci processor.

Procedure: The stimuli were presented in a random order using the "SuperLab" software with the visual and auditory stimuli being presented simultaneously. The subjects were required to make a response by pressing one of four buttons on the keyboard. These keys were selected to give a correspondence between stimulus attributes and response. Thus the keys used were 4, 5, 6 and 8 on the keypad area of the keyboard. These keys possess the same spatial relationships as the responses required in the experiment i.e. "left", "right", "up" and "down" and were assigned these values accordingly. The arrangement of the response keys is illustrated below.



The subjects were instructed to use their preferred hand to depress the keys, using their first and third fingers for the left/right decisions (keys 4 and 6 respectively) and their second finger for the up/down decisions (keys 8 and 5 respectively). The visual stimuli persisted until the subject had made a response which was then followed by a 500 ms interval and then the next trial. In order to ensure that subjects continued to watch the screen throughout the experiment they were occasionally asked to respond to a question mark by indicating the direction of the previous gesture, again with a manual directional response. Eight such question mark trials appeared in each response block.

A set of ten practice trials were presented before each block, these included two question mark trials and a cross section of the other trials, ensuring that all the experimental conditions were represented at least once. Both reaction times and percentage of errors were recorded as dependent variables in the experiment.

Results

In this, and all other experiments reported, outliers were removed from individual subject's scores by removing those reaction times greater or less than two standard deviations from the mean.

The mean correct reaction time and error scores are summarised in Table 1a* . Subjects were generally slower, and made slightly more errors in responding to incongruent compared to congruent stimuli. Moreover, overall RT's to the voice dimension were rather slower than those to gestures.

Table 1a.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent and Incongruent Conditions of Experiment 1.*

	Up/Down Decisions		Left/Right Decisions		Overall Mean	
Congruity	*M* RT	%E	*M* RT	%E	*M* RT	%E
			Responses to Voice			
Cong	788	0	816	0.83	802	0.42
Incong	880	1.67	995	5.83	938	3.75
Mean	834	0.84	906	3.33	870	2.09
			Responses to Gesture			
Cong	696	0	645	0	671	0
Incong	790	4.17	699	1.67	745	2.92
Mean	743	2.09	672	0.84	708	1.46

A 2 (response) x 2 (congruity) x 2 (decision) ANOVA applied to the reaction time data largely confirmed the above observations. This analysis yielded a significant main effect of response ($F(1,11)=12.47$, $p<0.01$) such that RT's were faster to gesture than to voice stimuli (707 ms vs. 869 ms). The main effect of congruity also reached significance ($F(1,11)=27.68$, $p<0.001$) with congruent stimuli receiving 105 ms faster RT's than incongruent pairs (i.e. a congruity effect of 105 ms). The response by decision interaction was also significant ($F(1,11)=9.51$, $p<0.05$). Simple main effects analysis qualified the main effect of response (that responses were faster to gesture stimuli) but indicate that the switch in response

* Standard deviations and complete ANOVA summary tables for this, and all other experiments, can be found in Appendix B.

dimension from gesture to voice had larger detrimental effects on left/right compared with up/down decisions. Left/right decisions were slowed by some 234 ms ($F(1,22)=453.70$, $p<0.001$) whereas up/down decisions were slowed by 91 ms ($F(1,22)=68.70$, $p<0.001$). This interaction together with the effects of congruity are illustrated in Figure 2. No other effects reached significance (all p's >0.1).

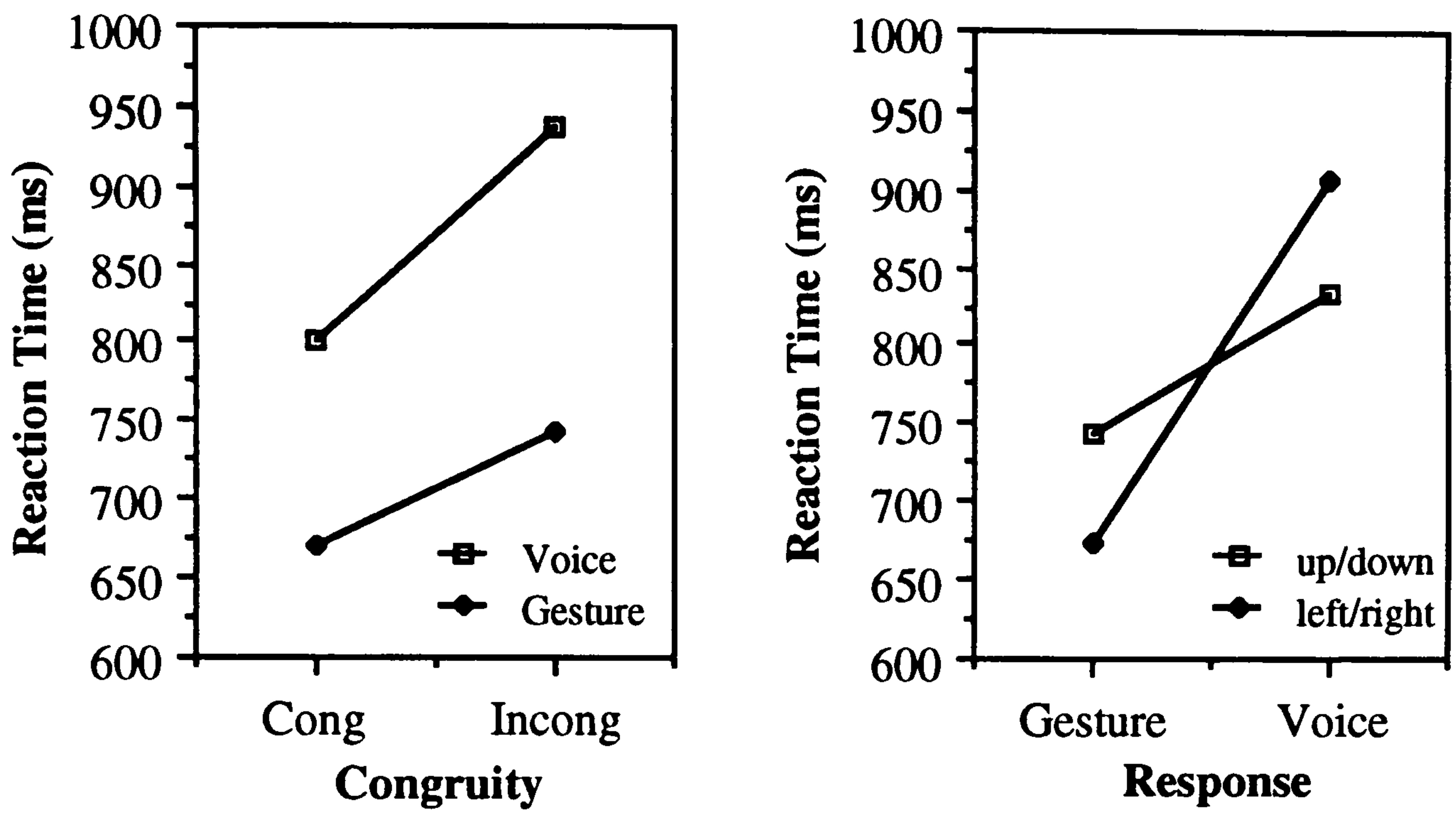


Figure 2. Mean Reaction Times (ms) obtained in Experiment 1 as a function of congruity and response factors (left panel), and response and decision factors (right panel).

From Table 1a error scores closely mirror the reaction time data. The overall mean percentage error was 1.77%. The correlation between RT's and errors was 0.73 suggesting no evidence of a tradeoff between speed and accuracy. Because of the relatively low rate of errors in the cells of the design (including scores of 0% in some cases) made by subjects in this and other experiments, no further analysis was conducted on the error data.

Table 1b summarises the overall size of the interference effects (calculated by subtracting congruent from incongruent RT's) for each order of response task. Clearly the magnitude of the effect is substantial regardless of the order in which subjects completed the experimental tasks. In particular, incongruent gestures slowed RT's to target voice stimuli by 111 ms even when this was the first task attempted. Thus it is unlikely that the overall effect of irrelevant gestures was due to an inability to suppress a response to a gesture after extended practice at responding to gestures (cf. Schneider & Shiffrin, 1977).

Table 1b.
Magnitude of the Interference Effect (in milliseconds) as a Function of Order of Response Task..

	Response Task	
Order	Responses to Voice	Responses to Gesture
Voice First	+111	+84
Gesture First	+159	+63

Discussion

The non-interactive effects of response and congruity obtained in this experiment indicate a symmetrical Stroop-like interference effect consistent with McNeill's ideas on the complementary nature of gesture and speech. This pattern of congruity effects is compatible with the notion that deictic gestures and verbal information are mutually influential in the comprehension process i.e. gestures appear to influence the processing of verbal information and reciprocally verbal information influences the processing of gestural information. This was perhaps a rather striking result in view of the overwhelming number of studies involving Stroop-type methodologies which tend to produce asymmetric effects. The reverse Stroop effect is only obtained by either manipulating the SOA between the two dimensions (e.g. Glaser & Glaser, 1982), dramatically slowing down the reading process (e.g. Dunbar & MacLeod, 1984), or by practising colour naming (e.g. Stroop, 1935; MacLeod & Dunbar, 1988). Here a reverse, as well as a normal Stroop-type effect has been demonstrated merely by manipulating the task instructions. The fact that an interference effect has been obtained at all is interesting given the exchanges between Cowan (e.g. Cowan & Barron, 1987; Cowan, 1989a, 1989b) and Miles (Miles, Maddon & Jones, 1989;

Miles & Jones, 1989) concerning the existence of cross modal Stroop-type interference. Clearly the results of this experiment suggest that a visual-auditory interference effect is at least possible, supporting the conclusions of Cowan and his colleagues (see also Schriefers & Meyer, 1990; Shimada, 1990 for evidence of cross modal effects).

In general the switch from gesture to voice as the relevant response dimension resulted in a significant increase in RT. This may well be due to the relative temporal parameters of the verbal and gestural stimulus presentations. Because visual stimuli are presented almost instantaneously whereas, by their nature, verbal stimuli are temporally extended, the complete identity of a relevant gesture will become available to a subject before the identity of a corresponding verbal stimulus in the voice response condition. Regardless of the relative processing speeds of visual and verbal information, the head start given to the gestural stimuli is likely to produce faster overall RT's to this dimension. However, the effect of response was found to interact with that of decision. Left/right decisions were slowed by a larger magnitude than up/down RT's when the relevant response dimension was switched from gesture to voice. This differential effect on left/right decisions was probably due to a left/right confusion effect. A number of studies have shown that it takes longer to make locational discriminations when the relevant spatial dimensions are given the *verbal* labels "right" and "left" than when they are described by terms such as "above" and "below" (e.g. Farrell, 1979; Maki, Grandy & Hauge, 1979; Sholl & Egeth, 1981). This effect is often referred to as the right/left confusion, and is thought to result from the lack of a natural referent in the horizontal dimension caused by the symmetrical right-to-left axis of the human body (Corballis & Beale, 1976). In view of this, left/right decisions should be slower when subjects respond to voices than when they respond to gestures. In the latter case, the presence of the gesture removes any left/right ambiguity resulting in a speedier decision.

To summarise, a normal Stroop-type effect has been demonstrated in this experiment, i.e. interference caused by an irrelevant, or to-be-ignored verbal stimulus in responding to a gestural stimulus. Notably, in this case, the dimensions in question were presented cross modally and required a manual response. Significantly a reverse effect has also been observed, i.e. interference caused by an irrelevant gesture when responding to a verbal stimulus. This suggests that subjects were able to monitor the visual stimulus and extract information from the gesture even though this was not required by the task demands. This symmetrical pattern of effects is consistent with the complementarity hypothesis derived from McNeill's ideas on the production of gesture and speech. Furthermore, the results of this experiment are suggestive of the existence of cross-modal, visual-auditory interference effects in a picture-word interference task (cf. Schriefers & Meyer, 1990).

Experiment 2

As already mentioned, the observation of a reverse Stroop-type effect, that is the interference caused by an irrelevant visual, as opposed to a verbal stimulus, is rather infrequent. It has been suggested that this effect is evidence for the processing of the to-be-ignored gestural stimuli. However the ancillary "question mark" task employed in Experiment 1 necessitates the explicit recognition of each of the gestures. Recall that the appearance of a question mark prompted subjects to respond to the direction indicated by the previously presented gesture. In order to successfully perform this task, subjects must presumably extract the relevant information from the gesture (i.e. its direction) and deposit the results of this processing in working memory. Should a question mark appear in the next trial, this information could be retrieved and an appropriate response programmed and effected. It is possible that in the absence of this type of secondary task, subjects might not extract any spatial information from an irrelevant gesture. In these circumstances a to-be-ignored gesture might not be expected to interfere with the processing of a target verbal stimulus*. For these reasons the question mark task employed in Experiment 1 was replaced with a task which, although still requiring subjects to attend to the screen, no longer necessitated the explicit recognition of a gesture.

It is again hypothesised that gestural and verbal information will interact in a symmetrical fashion yielding an effect of congruity with no interaction with response type.

Method

Subjects: Fourteen undergraduate and postgraduate volunteers participated in this experiment, all were between 18 and 30 years of age with normal or corrected to normal vision.

Materials , Design and Procedure: These were identical to Experiment 1 with the exception of the question mark task. Upon the presentation of the question mark, subjects were now simply asked to press the space bar. This modified secondary task should no-longer require any processing of the gestural information and is not expected to place any demands on working memory capacity.

Results

The mean correct reaction time scores and percentage of errors recorded in each condition are reported in Table 2a.

A 2 (response) x 2 (congruity) x 2 (decision) ANOVA conducted on the reaction time scores revealed a main effect of congruity ($F(1,13)=21.44$, $p<0.001$) of some 57 ms in the predicted direction. There was no evidence of an interaction between

---

* Of course this explanation assumes that gestures will interfere with verbal stimuli at a post-perceptual stage of processing. If the locus of the interference was at a perceptual level of analysis, merely directing visual attention to the gesture might be sufficient to cause interference. This issue is discussed at some length in Chapter 6.

congruity and response ($F(1,13)=1.52$, $p=0.24$) indicating a symmetrical pattern of interference or congruity effects. A main effect of response was also evident ($F(1,13)=6.11$, $p<0.05$) indicating that subjects were generally faster to respond to the gesture information. However the interaction of response and decision reached significance ($F(1,13)=5.49$, $p<0.05$). Left/right decisions were slowed by some 123 ms when the relevant response dimension was switched from gesture to voice, a difference which proved to be significant ($p<0.01$). Up/down decisions were also slowed but the difference of 50 ms did not reach significance ($p=0.2$). The effects of congruity and the response by decision interaction are displayed graphically in Figure 3.

Table 2a.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent and Incongruent Conditions of Experiment 2.*

	Up/Down Decisions		Left/Right Decisions		Overall Mean	
Congruity	*M* RT	%E	*M* RT	%E	*M* RT	%E
			Responses to Voice			
Cong	687	2.14	739	2.86	713	2.50
Incong	773	1.43	798	7.14	786	8.57
Mean	730	1.79	769	5.00	750	5.54
			Responses to Gesture			
Cong	652	1.43	633	0	643	0.72
Incong	707	7.14	658	2.86	682	5.00
Mean	680	4.29	646	1.43	663	2.86

As in Experiment 1, the congruity effect appears to be reasonably consistent across the two different task orders (see Table 2b). However, there is a relatively small 18 ms congruity effect caused by irrelevant verbal stimuli when subjects first responded to gesture information.

Table 2b.

Magnitude of the Interference Effect (in milliseconds) as a Function of Order of Response Task..

Order	Response Task: Responses to Voice	Response Task: Responses to Gesture
Voice First	+74	+63
Gesture First	+71	+18

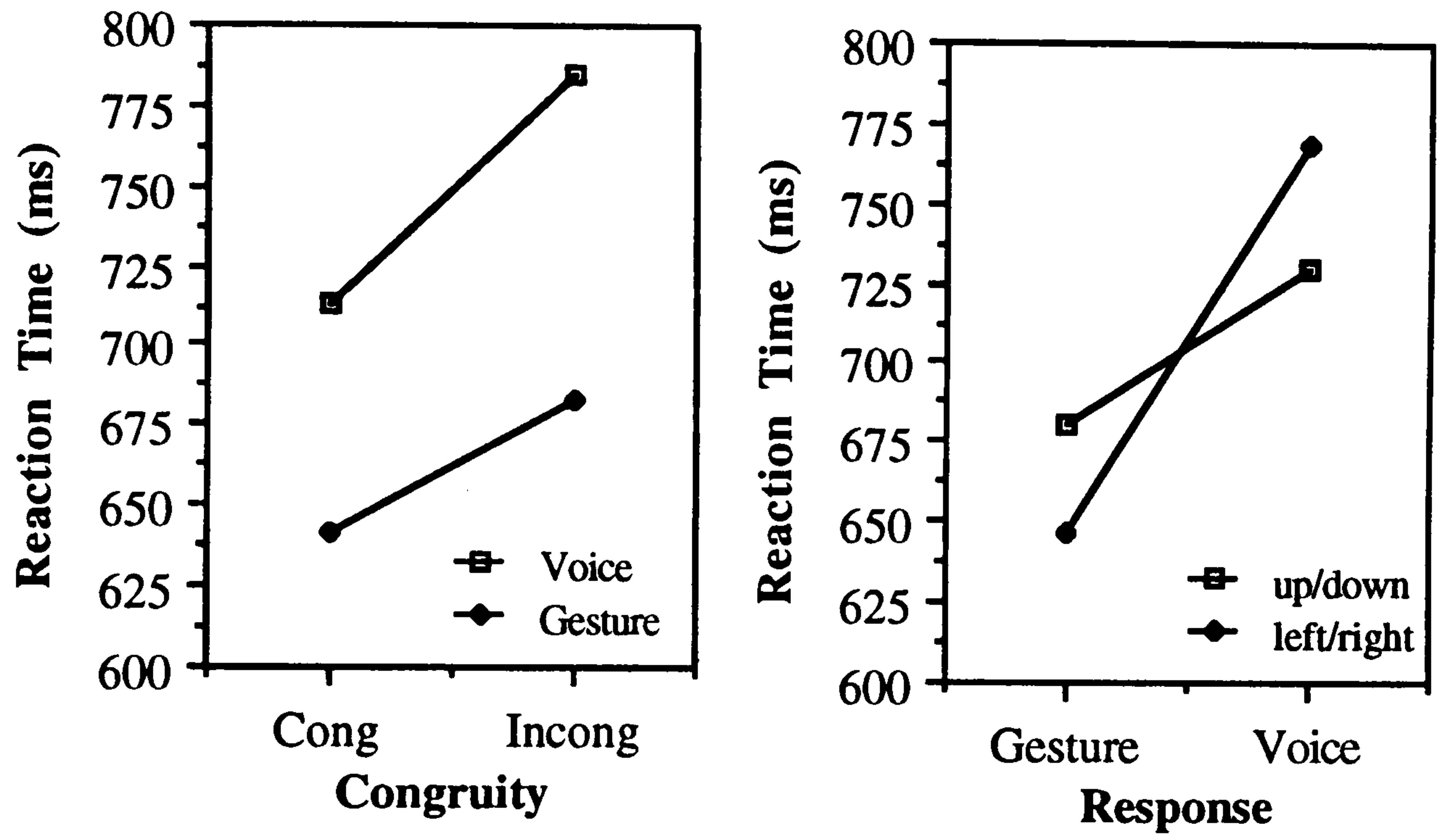


Figure 3. Mean Reaction Times (ms) obtained in Experiment 2 as a function of congruity and response factors (left panel), and response and decision factors (right panel).

From Table 2a the error scores generally appear to mirror the reaction time data, the overall mean percentage error was 3.13%. The correlation between RT's and errors averaged across all conditions was 0.55 suggesting no evidence of a trade-off between speed and accuracy.

Discussion

A comparison of Tables 1a and 2a suggests that the overall RT's in Experiment 2 were rather lower than in Experiment 1 indicating that the switch in secondary task has indeed reduced the task demands of the experiment. Despite this, the findings were essentially identical to those of Experiment 1. The non-interactive effects of response and congruity indicate a symmetrical pattern of interference whilst the interaction between response and decision can again be attributed to the left/right confusion and the inevitably slower responses to auditorily presented verbal stimuli. It appears that the change in the secondary task has not eliminated the interference caused by irrelevant gestures, despite the fact that the demands of the experiment no longer require subjects to explicitly process the information contained in the gestures.

Experiment 3

It might be the case that the symmetrical effects obtained in Experiments 1 and 2 were due to the *cross-modal* nature of these stimuli rather than specifically to the combination of gestural and auditory materials. For instance, it is the cross-modal presentation of the stimuli in Experiments 1 and 2 which largely contributes to the "head start" in the processing of the gestural dimension. This processing advantage may well be the cause of the interference effects exerted by the irrelevant gestures in Experiments 1 and 2. Relative speed of processing accounts of the Stroop effect rely on just such processing mismatches (e.g. Morton, 1969; Morton & Chambers, 1973; Posner & Snyder, 1975). These "horse race" models are discussed more fully in Chapter 4. Briefly, they assume that the faster dimension will interfere with a slower dimension but not vice-versa.

To investigate whether or not the cross modal nature of the stimuli were responsible for the pattern of effects obtained, Experiment 2 was repeated intra-modally. Presenting both dimensions within the visual modality ensures that both gestural and verbal information is available to the subject at the same time. Any processing advantage enjoyed by the gestural stimuli due to the temporal extent of the auditorily presented verbal material is therefore eliminated. Thus in Experiment 2 the words "up", "down", "left" and "right" were printed across the chest of the gesturer rather than being presented auditorily. This arrangement is more like the classic picture-word interference paradigm with less confusing predictions. Rosinski et al. (1975) obtained classic Stroop interference with combinations of picture-word stimuli, that is, incongruent words caused a strong inhibition in the picture naming task, whereas pictures had only weak effects on word reading. These effects were also demonstrated by Glaser & Düngelhoff (1984) who, in addition, largely failed to induce reverse effects of comparable magnitude to the normal interference pattern by pre-exposing the distracting picture. Indeed when both stimuli were presented together incongruent pictures had no effect on word reading.

Thus in this experiment, assuming the stimuli to be analogous to a picture-word paradigm, a normal asymmetric pattern of interference effects was expected. Incongruent words should slow responses to the gestures (pictures), but gestures might be expected to have no effect on responding to the words.

Method

Subjects: Fourteen undergraduate and postgraduate volunteers participated in this experiment, none of whom had taken part in Experiment 2.

Materials, Design and Procedure: These were identical to Experiment 2 with two exceptions. The auditory stimuli were replaced by the equivalent words which were printed across the chest of the gesturer (see Figure 4). The height of each letter subtended approximately 1° of visual angle (the horizontal angle depended on the length of each word). This process was completed using image processing software on the Macintosh. Secondly, as completion of the task required subjects to attend to the visual stimuli presented on the screen, the question mark task was not included in the design of this experiment.

As before, the experiment took the form of a 2 x 2 x 2 within subjects design. The three independent variables were: response (to gesture or word), congruity (congruent or incongruent stimulus pairs) and decision (up/down or left/right decisions). Again both reaction time and percentage of errors were recorded as dependent variables. Subjects responded to words and gestures in separate blocks of trials. The order of presentation of these blocks was alternated between subjects so that half responded to words first, and half responded to gestures first.

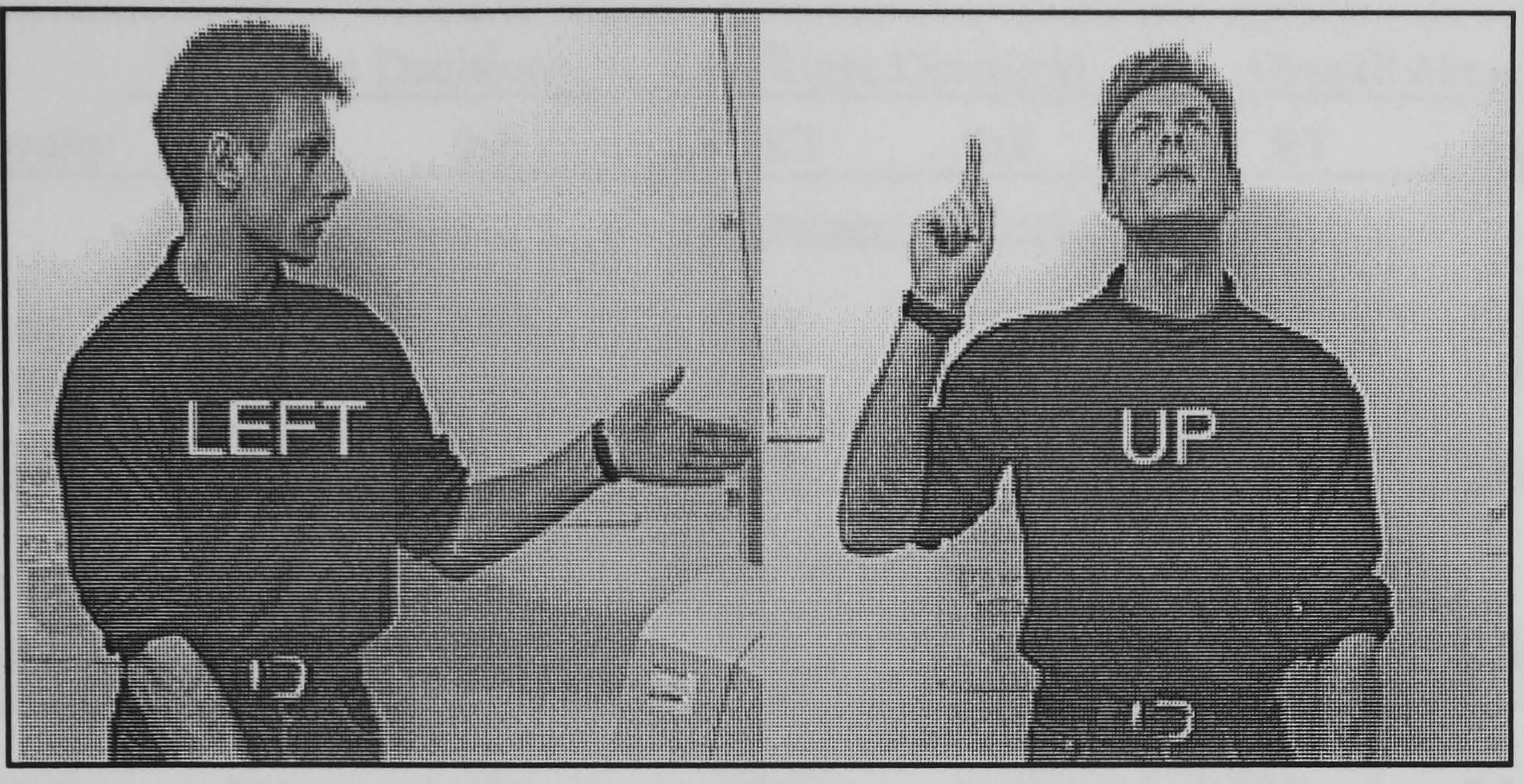


Figure 4. Examples of incongruent (left panel) and congruent (right panel) stimuli used in Experiment 3.

Results

The average reaction time and error scores obtained under all the experimental conditions are presented in Table 3a.

A 2 (response) x 2 (congruity) x 2 (decision) ANOVA was conducted on the reaction time data. This revealed a main effect of congruity, $F(1,13)=41.09$, $p<0.001$, such that congruent responses were made 75 ms faster than incongruent responses. The main effect of decision also approached significance, $F(1,13)=4.19$, $p=0.06$, indicating a trend toward faster reaction times to left/right decisions. No other effects reached significance (all p's > 0.1). Notably, the interaction between response and congruity was not reliable ($p=0.33$) indicating a symmetrical pattern of interference effects.

Table 3a.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent and Incongruent Conditions of Experiment 3.*

	Up/Down Decisions		Left/Right Decisions		Overall Mean	
Congruity	*M* RT	%E	*M* RT	%E	*M* RT	%E
			Responses to Word			
Cong	675	2.86	633	0	654	1.43
Incong	755	2.14	722	5.71	739	3.93
Mean	715	2.50	678	2.86	697	2.68
			Responses to Gesture			
Cong	626	3.57	561	2.14	5.93	2.86
Incong	686	2.14	628	0.71	657	1.43
Mean	656	2.86	595	1.43	625	2.15

Again, the magnitude of the effect remains consistent across both response task orders (see Table 3b).

Table 3b.

Magnitude of the Interference Effect (in milliseconds) as a Function of Order of Response Task..

	Response Task	
Order	Responses to Word	Responses to Gesture
Word First	+84	+44
Gesture First	+86	+83

The overall mean error score was only 2.42%. The correlation between RT's and errors was 0.38 again suggesting no evidence of a trade-off between speed and accuracy. Because subjects made relatively few errors including an overall mean error rate of 0% in one of the cells in the design, no further analysis was conducted on these data.

Discussion

Here again symmetrical Stroop-type interference has been demonstrated despite the predictions arising from the picture-word interference literature. It seems that the pattern of effects found in Experiments 1 and 2 cannot be solely attributed to the cross modal relationship between the stimuli, but rather more to the nature of the stimuli per se. It appears that deictic gestures at least, do indeed influence the processing of verbal information at the same time as processing of verbal information influences that of non-verbal gestural information. These findings support the idea that listeners attend to both verbal and gestural information in the comprehension process, combining the two sources of information at some point in processing.

The switch from a cross-modal to a within modality presentation of the stimuli also eliminated the response by decision interaction found in Experiments 1 and 2. Recall that the main cause of this interaction was the slowing of responses to the verbal stimulus dimension, particularly to left/right decisions, caused by the temporal extent of the voice stimuli and by a left/right confusion effect. In Experiment 3, however, there was no significant effect of response, and no evidence of a left/right confusion effect. The absence of a response effect might be expected as this was originally thought to be due to the temporal extent of the voice stimuli in Experiments 1 and 2. In the present experiment, all the stimulus information required by a subject to identify a printed word is available at the same time as that required to identify the gesture. Thus, in contrast to the cross-modal case, we might expect written words to show no RT advantage over gestures. Nevertheless, although non-significant, gesture responses were performed 72 ms faster than average word responses in Experiment 3. Thus, there appears to be an advantage for keypress responses to gestural, over verbal stimuli, regardless of the modality of presentation. It is suggested that this might be due to a spatial stimulus-response compatibility

(SRC) effect (e.g. Fitts & Seeger, 1953; Simon et al., 1981). The issue of SRC is discussed at some length in Chapter 4.

The reason for the absence of a left/right confusion effect when the relevant response is switched from the gestural to the verbal dimension is unclear. Instead, left/right decisions showed a consistent trend towards an advantage over up/down decisions for both gestural and verbal response dimensions. This may be due to a "fingering" advantage for left/right over up/down decisions caused by the arrangement of the response keys. This required subjects to respond to up/down stimuli by moving a single finger to one of two keys. In contrast two fingers were used for the left/right decisions so that no initial movement to the location of the correct key was necessary. This presumably results in a natural advantage for left/right decisions.

In summary, Experiment 3 has again yielded a symmetrical pattern of interference effects when both stimulus dimensions were presented in the visual modality and consequently both available for identification at the same time. Again this is suggestive of a mutual influence of gestural and verbal information in comprehension.

Experiment 4

A possible problem with the interpretation of the deictic gestures used in Experiments 1-3 was that subjects may have been somewhat confused as to whose left or right the gestures referred. The responses "left" and "right" referred to the side of the *image* rather than to the side of the *gesturer* on the image. It is possible that a tendency existed to interpret the left/right gestures from the point of view of the gesturer and not with respect to the image. In order to investigate this possibility Experiment 2 was repeated but with the gesturer turned around to face away from the subject. The ambiguity as to whose left or right the gesture referred was therefore no longer present.

Method

Subjects: Fourteen undergraduate and post-graduate volunteers participated in this experiment, all had normal or corrected to normal vision and hearing.

Materials, Design and Procedure: The auditory stimuli were the same as those used in Experiments 1 and 2, however the gesture stimuli were replaced by similar images but with the gesturer facing away from the subject (see Figure 5). Both the design and procedure were identical to Experiment 2.

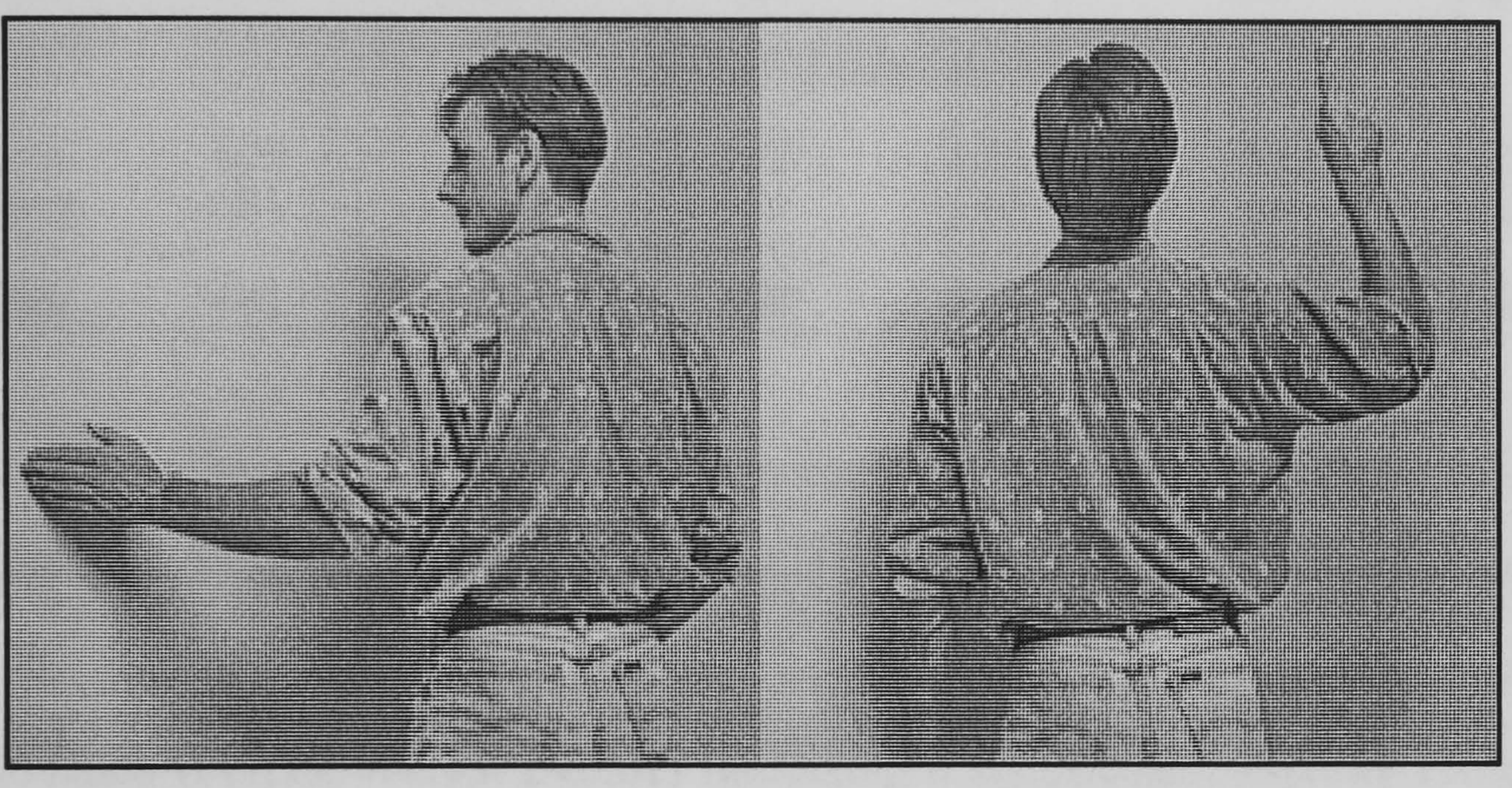


Figure 5. Examples of the "left" (left panel) and "up" (right panel) pointing gestures used as stimuli in Experiment 4.

Results and Discussion

The mean correct reaction time scores and percentage of errors recorded in each condition are reported in Table 4a. Again there is a consistent congruity effect, both in terms of RT's and errors. Subjects' responses are both faster and more accurate to congruent, as opposed to incongruent stimuli.

Table 4a.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent and Incongruent Conditions of Experiment 4.*

	Up/Down Decisions		Left/Right Decisions		Overall Mean	
Congruity	*M* RT	%E	*M* RT	%E	*M* RT	%E
			Responses to Voice			
Cong	651	1.43	679	2.86	665	2.15
Incong	748	5.00	752	7.14	750	6.07
Mean	700	3.22	716	5.00	708	4.11
			Responses to Gesture			
Cong	600	1.43	562	1.43	581	1.43
Incong	681	5.00	651	1.43	666	3.22
Mean	641	3.22	607	1.43	624	2.33

Statistical support for the above observations comes from 2 (response) x 2 (congruity) x 2 (decision) ANOVA conducted on the RT data. This analysis revealed a similar pattern of results to Experiments 1 and 2. The effect of congruity reached significance ($F(1,13)=18.13$, $p<0.01$) indicating a significant congruity effect of some 85 ms. The interaction of congruity with response failed to reach significance ($p=0.98$) showing a symmetrical pattern of interference effects. A main effect of response ($F(1,13)=7.70$, $p<0.05$) indicated faster responses to gesture (624 ms) as opposed to voice stimuli (708 ms), presumably because of the temporal extent of the voice stimuli as in Experiments 1 and 2. Finally, the interaction between response and decision was also significant ($F(1,13)=10.82$, $p<0.01$). Further analysis of this interaction largely qualified the advantage of gesture over voice responses for both decision dimensions (although failing to reach significance for up/down decisions, $p=0.07$). As in Experiment 1, the reason for the interaction appeared to be that the response manipulation had a much larger effect on left/right as opposed to up/down decisions (109 versus 59 ms), probably because of the left/right confusion effect discussed above. Analysis of the simple main effects also confirmed that left/right decisions were performed significantly faster (34 ms) than up/down decisions to the gesture dimension ($p<0.01$). This may well be due to the left/right fingering advantage discussed in Experiment 3, an advantage which is presumably negated by the left/right confusion effect when the relevant response is switched from the gestural to the verbal dimension. This interaction is illustrated, together with the effects of congruity, in Figure 6.

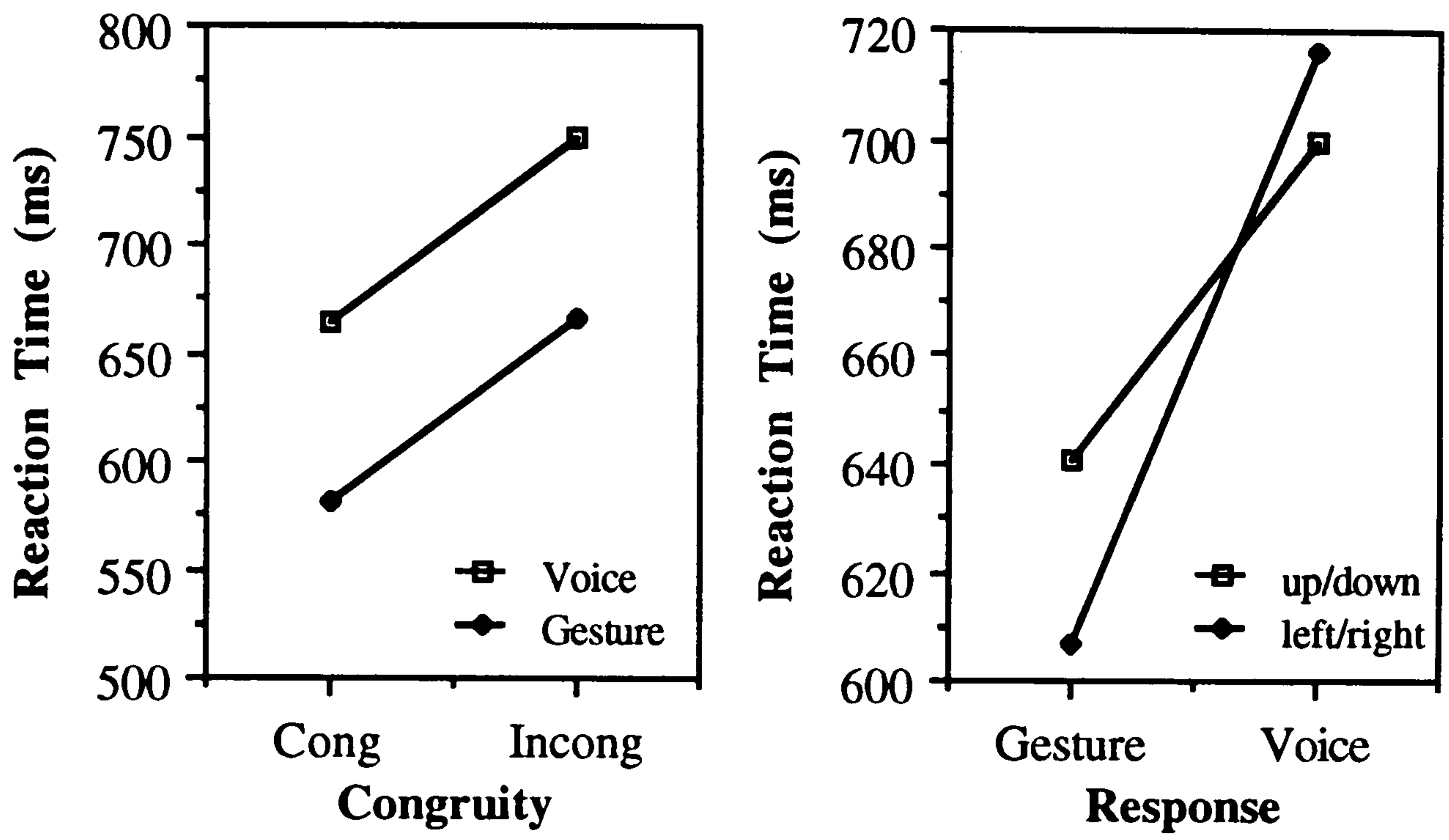


Figure 6. Mean Reaction Times (ms) obtained in Experiment 4 as a function of congruity and response factors (left panel), and response and decision factors (right panel).

Table 4b indicates that the interfering effects of both gesture and verbal information persist regardless of the order of presentation of the response tasks.

Table 4b.

Magnitude of the Interference Effect (in milliseconds) as a Function of Order of Response Task..

	Response Task	
Order	Responses to Voice	Responses to Gesture
Voice First	+75	+84
Gesture First	+96	+84

As can be seen from Table 4a the error rates in this experiment were fairly low, indeed the overall percentage of errors was 3.22%. The correlation between RT's and percentage of errors across all conditions was 0.84 offering no evidence of a speed/accuracy trade-off.

In summary, the results of this experiment were essentially identical to those of Experiment 2, both in terms of the symmetry of the interference and with regard to the interactive effects of response and decision factors. This suggests that subjects did not have a problem interpreting the various spatial attributes of the gestures and also provides more evidence for the mutual interaction of deictic gestures and verbal information in comprehension.

General Discussion

The primary aim of these experiments was to investigate the possibility that gestural and verbal information are mutually influential in the comprehension process. A Stroop interference paradigm was adopted wherein pairs of congruent and incongruent deictic gestures and verbal stimuli were simultaneously presented to subjects. Experiments 1 and 2 revealed symmetrical patterns of interference effects, that is auditorily presented verbal information interfered with the processing and response to concurrently displayed static deictic gestures and vice-versa. Furthermore, these symmetrical effects were obtained when presentations of both dimensions were restricted to the visual modality (Experiment 3), suggesting that the effects reported in Experiments 1 and 2 were not due to the cross-modal presentation of the stimuli, but to the particular combination of gestural and verbal materials per se. Experiment 4 was designed to examine the effects of any confusion arising from the interpretation of the directional gestures. The results of this experiment confirmed the symmetrical nature of the interference effects.

The observation of a symmetrical interference effect indicates that subjects were unable to attend selectively to either the gesture or the voice dimension when conflicting information was present in the to-be-ignored channel. In particular the fact that the identity of an irrelevant gesture influenced the RT's to respond to verbal items clearly indicates that these gestures receive some kind of analysis. This finding adds to the literature which suggests that recipients are indeed influenced by speech related gestures (e.g. Rogers, 1978; Riseborough, 1981; Graham & Argyle, 1975, Thompson & Massaro, 1986;1989; McNeill, Cassell & McCullough, 1994) and contradicts the claims made by Rimé & Schiaratura (1991) who argue that speech related gestures are not produced for the benefit of the addressee but facilitate the verbal encoding process of the speaker (see also Freedman, 1972). Whether or not deictic gestures are intentional acts of communication, the fact remains that they are quite capable of influencing the comprehension and recall of verbal material. Moreover, the symmetry of the interference effects concurs nicely with the idea that listeners process both gestural *and* verbal sources of information in the comprehension of an utterance (e.g. McNeill, 1992; McNeill, Cassell & McCullough, 1994).

The Locus of the Interaction

Breakdowns in selective attention, as we have observed in Experiments 1-4, may be thought of as being caused by a crosstalk or interaction of information between processing channels (see Chapter 2). Symmetrical interference between gesture and verbal information, for instance, may reflect *bidirectional* crosstalk as irrelevant gestures cause as much interference in responses to words as do words on responses to gestures. The question remains, however, as to where in the sequence of processing operations this interaction actually occurs. In Chapter 2 we considered how, broadly speaking, dimensional interactions might be caused by crosstalk at any of three levels of analysis: perceptual; semantic; or response stages. In what follows a perceptual encoding account of the gesture-word interaction is considered.

Perceptual Encoding

First it may be possible for certain gestures to influence the perception of verbal information. Although it is unlikely that hand gestures could facilitate the phonemic discrimination of speech sounds as in the McGurk effect, certain "beat" like gestures (McNeill, 1987a) might act as an aid to the processing of intonation patterns, or prosody, in speech. These types of hand movements frequently occur in close coordination with the rhythmic nature of speech. Moreover the maximal amplitude of a such a gesture will often correspond to a loudness peak in the concurrent speech stream (e.g. Chang & Hammond, 1987). These gestures might provide a visual cue to prosody in the same way that the shape of the mouth provides a cue to the identity of a phoneme.

Thus one can see how a perceptual interaction between gestures and speech might be possible. However, of current interest is whether or not the symmetrical interaction between pointing gestures and verbal information achieved in Experiments 1-4 is realised at a perceptual level of analysis. The qualitatively different nature of these stimuli and the relationship between them make this possibility unlikely. The perceptual experience of one dimension (e.g. verbal) does not seem to be affected by the presence of attributes in the second dimension (e.g. gestural). Of course it might be argued that the presence of a distracting gesture simply *slows* the perceptual encoding time for the relevant word and vice-versa. In fact this is essentially the explanation Hock & Egeth (1970) offered for the interference between the classic Stroop dimensions. In their view, the Stroop effect occurs because word meanings affect the perception of colour. Thus, the encoding of the ink colour is slowed by the presence of an incongruent colour word rather than there being any actual perceptual "fusion" as in the case of colour/brightness, pitch/loudness, or the dimensions used in the McGurk illusion. That is, certain dimensions such as the Stroop stimuli, or perhaps those of gesture and speech, may *interact* but are not *integral* (Melara & Marks, 1990b).

The data obtained here might be used to shed some light on the suggestion that the effects of congruity are due to some form of perceptual "crosstalk". Recall that in Experiments 1, 2 & 4 the congruity and response variables produced independent or additive effects on RT's. In the discussion to Experiment 1 the effect of response was attributed to the "head start" afforded to the visual, over the auditory stimuli. In terms of a stage-like information processing model this effect would presumably be located at an early perceptual level of analysis. According to Sternberg's (1969) additive factor method, non-interacting variables are assumed to affect different stages of information processing. By this logic one would expect the effect of congruity to be located at a different stage to the response variable, presumably at a later, post-perceptual level of processing.

Notwithstanding the additive factors treatment of the present findings, it is difficult to see how a perceptual encoding account could explain why congruent and incongruent stimulus pairs produce different RT's. According to Hock & Egeth's (1970) theory, subjects are "sensitised" to stimuli which are related to the semantic domain of the task. As a consequence, these items will be recognised sooner and, when acting as distractors, further disrupt the encoding of the target stimuli. Thus in the Stroop task, colour words will be encoded sooner than neutral words hence causing more interference. In the present experiments, all the distracting stimuli (whether gestures or words) are semantically related to the spatial response domain. Thus one might expect *all* of these distracting stimuli to produce roughly equal amounts of interference relative to some appropriate control condition. Although no such controls were included in Experiments 1-4 (e.g. a neutral word or gesture, or perhaps a word-only, or gesture-only condition), incongruent RT's were consistently slower than congruent RT's. It therefore seems reasonable to assume that congruent and incongruent distractors cause differing amounts of interference. In fact congruent distractors actually facilitate the response in many Stroop and Picture-word interference tasks. Hock & Egeth's theory cannot accommodate such observations. Indeed, the perceptual encoding account of the Stroop effect (Hock & Egeth, 1970) has been rejected by Dyer (1973) whilst more recently Simon & Berbaum (1990) have discounted this as an explanation of both the Stroop and Simon effects. Furthermore, Jacoby has recently used a "process dissociation procedure" (see Lindsay & Jacoby, 1994) to suggest that colour naming and word reading can operate *independently* to determine responses in the Stroop task. This evidence seems to challenge the notion that the Stroop effect occurs at the perceptual stage of processing.

In summary, the lack of support for Hock & Egeth's (1970) theory of a perceptual locus for the Stroop effect, coupled with the arguments arising from the results of the present experiments, make it unlikely that the interference between gestural and verbal information arises at a perceptual level of processing. This logically implies a role for higher level cognitive processes. For instance, the ideas of McNeill (e.g. McNeill, 1992; McNeill et al., 1994) suggest that gesture and speech should interact at a *semantic* level of analysis so that "gesture and speech...combine into a single system of meaning" (McNeill et al., 1994, p.235). Indeed, as mentioned in Chapter 2, a semantic account of the Stroop effect has been proposed (e.g. Seymour, 1977). However, a response-based locus of the effects also seems worth pursuing, particularly given the ease with which the attributes of the gestural and verbal dimensions can be mapped onto the spatial responses adopted in the present experiments. In particular, the interfering effect of the to-be-ignored pointing gestures may be due, at least in part, to this stimulus-response mapping. This hypothesis is examined in next chapter where response selection accounts are discussed in greater detail.

Concluding Comments

To the best of my knowledge the application of a Stroop-type interference procedure to the study of gesture and speech is novel whilst the use of cross-modal stimuli within such a procedure is also rare. The observation of Stroop-type interference with such cross-modal stimuli is also interesting given that the existence of such an effect has been challenged (e.g. Miles, Maddon & Jones, 1989). Finally, the results of the present experiments make an addition to a number of recent studies which have yielded symmetrical interference effects within both the Stroop and

Garner interference paradigms and with a variety of dimensional combinations (e.g. Melara & O'Brien, 1987; Melara, 1989; Melara & Marks, 1990; Melara & Mounts, 1993; O'Leary & Barber, 1993). As we shall see in later chapters, such observations represent a challenge to many models of Stroop and picture-word interference.

Overview

The aim of this chapter was to make use of the Stroop paradigm to explore the various ideas on the comprehension of gestural and verbal information. In all four experiments reported, spatial gestures exerted large and consistent interference effects on responses to the verbal dimension of the cross-modal stimuli. This in itself indicates that spatial gestures received some degree of processing. Similarly, incongruent spoken and written words slowed responses to spatial gesture stimuli compared with congruent arrangements. It seems that subjects are unable to attend selectively to either the verbal or the gestural dimension when faced with contradictory information in the other channel. It was suggested that these results are consistent with the view that listeners process both gestural and verbal information in the comprehension of an utterance. A discussion of the recent work on dimensional interaction suggested that the effects reported here are likely to located at a post-perceptual level of processing, perhaps reflecting interchannel crosstalk at either a semantic or response stage of processing. A number of response selection accounts are discussed in the next chapter. The set of experiments reported therein are primarily aimed at testing one such model, based on the notion of stimulus-response compatibility.

Chapter Four

Stimulus-Response Compatibility and the Locus of the Interaction

In the previous chapter we have seen that pointing gestures and spoken words interact symmetrically when placed into conflict in a Stroop-like interference task. Of particular interest, and somewhat at odds with the picture-word and Stroop interference literature, was the interference caused by an irrelevant gesture. This result provides us with an answer to the first of our questions; subjects do seem to process the gestural component of an utterance, at least when these gestures consist of pointing actions. It was also suggested that this type of interference effect might reflect some kind of post-perceptual interaction or crosstalk of information between processing channels. However, the possibility remains that the influence of irrelevant gestures is caused by the compatible mapping between them and the spatial response. This kind of stimulus-response compatibility account places the locus of the effect at the response selection stage of processing and critically, does not assume that information from the two channels is actually integrated or combined in producing the effect. In what follows the notion of SRC is introduced and a model of the gesture-word interference based on this conception is developed and tested in Experiments 5-11. The results suggest that SRC does play a role in gesture-word interference but is not critical as the effects seems to persist in its absence. The chapter concludes by considering two other response selection accounts of Stroop interference, the relative speed of processing and the translational models, neither of which can provide an adequate explanation of the symmetrical nature of the gesture-word interference. On the basis of this evidence, a response stage account of the effect is rejected in favour of a semantic-based locus.

Stimulus-Response Compatibility

One of the most potent influences on the speed of response in a choice reaction time task is the relationship between the properties of the stimulus and those of the response. Manipulation of this relationship has been shown to produce powerful effects on performance (e.g. Fitts & Seeger, 1953). For instance, response latencies are shorter if a stimulus appearing on the left hand side of the screen commands a response with the left, than when it requires a response with a right hand key (e.g. Simon, 1969). The former arrangement is said to be spatially S-R compatible and the latter, S-R incompatible. The difference in RT's between incompatible and compatible arrangements may be thought of as a *stimulus-response compatibility* (SRC) effect. The most widely accepted explanation for this effect maintains that the poorer performance of an incompatible arrangement arises because a larger number of coding transformations are required to translate the stimulus code into a response code compared with the compatible case. In general, if the task requires few transformations the compatibility is high and the RT fast. On the other hand, if many transformations occur, the compatibility is low and the RT slower (e.g. Nicoletti & Umiltà, 1984, 1994).

The Simon Effect

Many have considered the Simon effect to be a particular instance of a S-R compatibility manipulation (e.g. Simon, Sly & Vilapakkam, 1981; Kornblum et al., 1990; Kornblum, 1994; Nicoletti & Umiltà, 1994). Recall that in this paradigm subjects might typically make a left/right keypress response contingent on the *identity* of a stimulus (e.g. colour of light, direction of arrow, the words "left" or "right" etc.) presented randomly to the left or right of some central point. The results indicate that the location of the stimulus provides an irrelevant cue that interferes with the processing of the target stimulus. Thus in Simon & Rudell's (1967) original paper subjects were slower to respond with a right keypress to the word "right" presented in the *left* ear than to an identical stimulus presented to the *right* ear. So in contrast to normal SRC effects, RT's in the Simon effect are affected by a spatial S-R relationship which is not actually relevant to the task in hand. It is assumed that both the relevant and irrelevant aspects of the stimulus undergo parallel S-R transformations, thus subjects perform a word-to-position transformation in order to make the correct response whilst simultaneously making a position-to-position transformation in processing the irrelevant location of the stimulus. This irrelevant S-R transformation will proceed automatically if a compatible mapping exists between the possible irrelevant locations of the stimuli and the position of the response keys. For example the directional words LEFT and RIGHT might appear on either the left or right of the screen whilst the response to the directional word must be made by making a left or right manual response. Having encoded both stimuli and their associated responses, some arbitration process operates to select the correct response. The Simon effect arises because of the longer arbitration process which occurs if the responses do not correspond (e.g. Craft & Simon, 1970; Simon, 1982).

Gesture-Word Interference

The spatial code consistent with the direction of a pointing gesture and the spatial code required to make a response to this gesture are essentially identical. Thus very few transformations are required to mediate between stimulus and response. Thus, it is suggested that the relationship between the pointing gestures and the spatially located response keys is one of spatial stimulus response compatibility (Simon, Sly & Vilapakkam, 1981). In the previous section we have seen that an S-R relationship need not be relevant in a task for it to influence RT's. Therefore it is possible that irrelevant deictic gestures exert their effects on directional words in the same way that the irrelevant location of a stimulus exerts its influence in the Simon effect.

If the simple model of the Simon effect described above is applied to the gesture-word situation, a response code consistent with an irrelevant gesture is automatically activated by virtue of the compatible spatial S-R relationship. In addition, a spatial response code contingent on the identity of the spoken word, must also be activated in order to comply with the demands of the experimental task. Thus two response codes are present at the decision stage, one of which must be selected, programmed and executed. As in the Simon effect, interference occurs when a conflict between non-corresponding (incongruent) codes must be resolved (see Figure 6). An important point to note is that this type of explanation side-steps the issue of informational integration. In this type of account the information encoded from the two dimensions is kept separate and some kind of executive process selects the appropriate response. At no point is information actually combined. A processing architecture along these lines is clearly not commensurate with McNeill et al's (1994) suggestion that gesture and speech channels "smoothly combine into a single idea unit" (p.235), a proposal which brings to mind an interaction of the two sources of information. Clearly response selection accounts, such as the one described above, need to be explored if we are to claim that the interference obtained in Experiments 1-4 reflects a genuine integration of information.

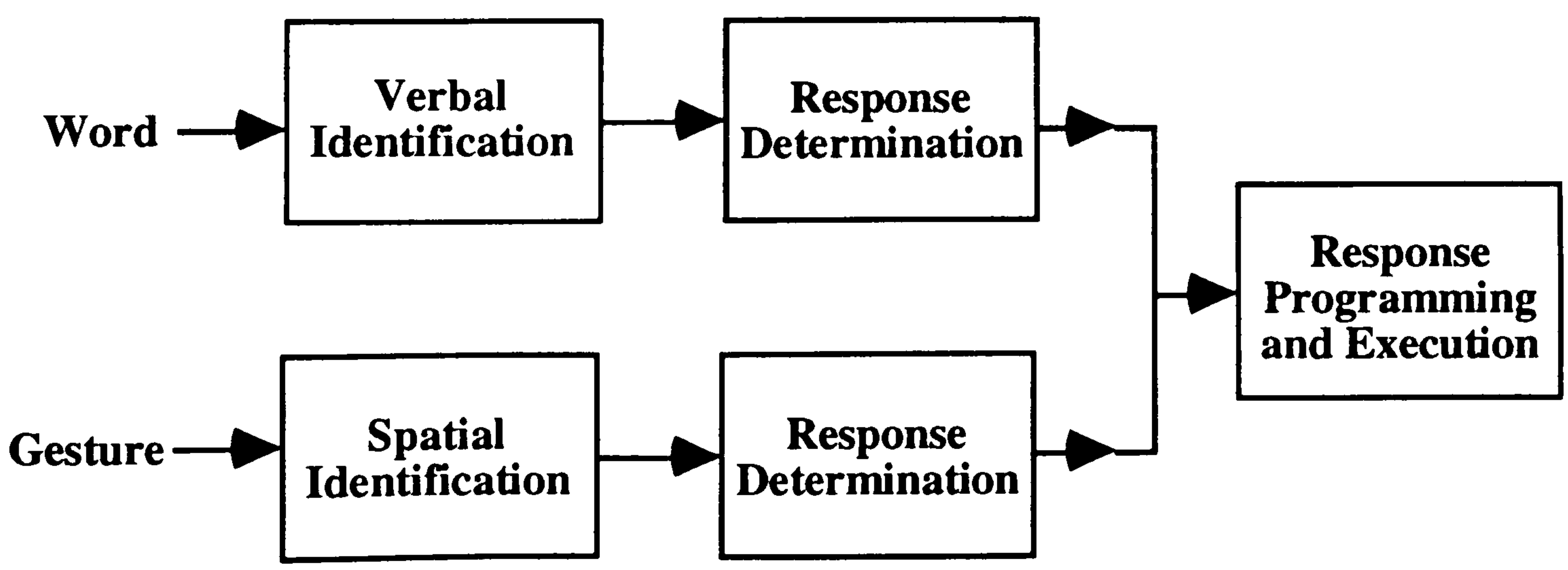


Figure 6. A possible 3-Stage representation of the information processing of gestural and verbal dimensions locating the source of the interference at the response selection stage of processing.

The Present Studies

In this chapter we attempt to investigate the contribution of SRC to the interference effects noted in the previous experiments. If the S-R transformation model is correct, that is the effects of irrelevant gestures are caused by the unintentional encoding of the inappropriate response as a result of SRC, then manipulating this S-R relationship so that the compatibility no longer exists should eliminate the interference effects.

Matching and Mapping

Broadly speaking there are two methods of experimentally manipulating the S-R relationship (Kornblum et al., 1990). First, one can vary the *mapping* of the stimulus set onto the response set simply by changing the task instructions. For instance, suppose the stimuli in a simple experiment consist of lights appearing either to the left or right of fixation. The subjects' task in this example is to press a left hand key when the light appears to the left, and a right hand key when the light appears to the right. In this case the transformation from stimulus to response is ipsilateral. However, another set of subjects might be asked to do the reverse, that is make a contralateral S-R transformation by pressing the left key to a light appearing to the right and vice-versa. Reaction times in this condition will almost certainly be slower than those in the former ipsilateral case (e.g. Schwartz, Pomerantz & Egeth, 1977). Neither the stimuli nor the responses have been changed, only the relationship between them.

The second method of manipulating the S-R relationship is to vary what is called the *match* between stimulus and response variables independently of the mapping. Matching S-R variables come from the same conceptual domain, thus an S-R association between colour stimuli (green-red) and responses with colour labelled keys would be between matching variables. In this case subjects make a colour-colour S-R transformation. In contrast, an association between the colour stimuli and spatial left-right keypress responses would be between non-matching variables. Here subjects make a colour-position transformation. A number of studies have suggested that stimulus and response sets which overlap conceptually receive faster RT's than non-matching sets (e.g. Greenwald, 1970; Stanovich & Pachella, 1977).

In Experiments 5-8 the mapping between the gesture stimuli and the manual responses is manipulated. Subjects continue to make a spatial response to a spatial stimulus but the relationship is made incompatible. In Experiment 9-11 the match between gesture and response is manipulated independently of the mapping. By using a spoken, naming response, subjects are forced to make a spatial-verbal transformation between the gesture stimuli and the response. The logic for both sets of experiments is as follows. If gestures exert their effects on verbal information solely by virtue of their spatial compatibility with the response, then removing this spatial SRC, either by manipulating the mapping or the match, should have the effect of eliminating any interference caused by these irrelevant gestures. Returning to the model presented in Figure 6, a response to an irrelevant gesture will no longer be encoded if the S-R relationship is not compatible. This will leave the response based on the relevant verbal stimulus unopposed at the determination stage, allowing unhindered selection, programming and execution.

Experiment 5

In this experiment the positioning of the left/right response keys was altered from a side by side to a vertical orientation whilst maintaining the positioning of the up/down response keys. Thus, by manipulating the mapping between the left/right gestures and the left/right keypress decisions the spatial S-R compatibility has been removed.

If the interference caused by irrelevant left/right gestures was due to the automatic tendency to encode a left/right response based on these irrelevant stimuli, changing the orientation of the left/right response keys might be expected to disrupt the ease or efficiency of this irrelevant S-R translation process. Under this hypothesis little or no interference is expected from left/right gesture distractors when the voice forms the target dimension.

Method

Subjects: Eighteen undergraduate and post-graduate volunteers participated in this experiment, all had normal or corrected to normal vision and hearing.

Materials, Design and Procedure: These were identical to Experiments 1 and 2, however this time the left/right responses were shifted to the vertical orientation. Keys 7 and 4 on the keypad now represented the left and right decisions respectively whilst keys 8 and 5 represented the up/down decisions. The keys were labelled as in the previous experiment and arranged as illustrated above.

L	U
R	D

Results

The mean RT's and percentage of errors from this experiment are summarised in Table 5. Looking first at the voice responses, there is a 51 ms congruity effect for up/down decisions but a smaller 15 ms effect for left/right decisions. Turning to the gesture responses, there are large congruity effects for both up/down (121 ms) and left right (94 ms) decisions.

Analysis of the obtained sample variances for each cell of the design indicated heterogeneity and so a single analysis of the data was impossible using the ANOVA technique. However the variances within the four cells of each of the response conditions were homogeneous and so separate 2 (congruity) x 2 (decision) ANOVA's were performed on the data, one on each of the response conditions across all subjects.

Table 5.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent and Incongruent Conditions of Experiment 5.*

Congruity	Up/Down Decisions *M* RT	Up/Down Decisions %E	Left/Right Decisions *M* RT	Left/Right Decisions %E	Overall Mean *M* RT	Overall Mean %E
			Responses to Voice			
Cong	672	0.56	765	1.67	719	1.12
Incong	723	0	780	0.56	752	0.28
Mean	698	0.28	773	1.12	736	0.70
			Responses to Gesture			
Cong	809	1.11	981	1.67	895	1.39
Incong	930	2.78	1075	4.44	1003	3.61
Mean	870	1.95	1028	3.06	949	2.50

The ANOVA performed on the "voice" data confirmed the observations made above revealing main effects of congruity ($F(1,17)=11.68$, $p<0.01$) and decision ($F(1,17)=18.97$, $p<0.001$). The crucial interaction between congruity and decision also reached significance ($F(1,17)=4.96$, $p<0.05$) and is illustrated in Figure 7 (left panel). Analysis of the simple main effects indicated a significant congruity effect of 51 ms for up/down decisions, $F(1,34)=16.39$, $p<0.001$ but a non-significant congruity effect of 15 ms for left/right decisions ($p>0.1$). The interaction also qualified the main effect of decision with left/right decisions receiving slower reaction times than up/down decisions.

The equivalent ANOVA performed on the "gesture" data indicated a congruity effect of 108 ms in the predicted direction, $F(1,17)=12.54$, $p<0.01$. The main effect of decision also reached significance, $F(1,17)=25.02$, $p<0.01$. Left/right decisions were 158 ms slower than up/down decisions. The interaction between decision and congruity failed to reach significance, $p>0.1$ (see Figure 7).

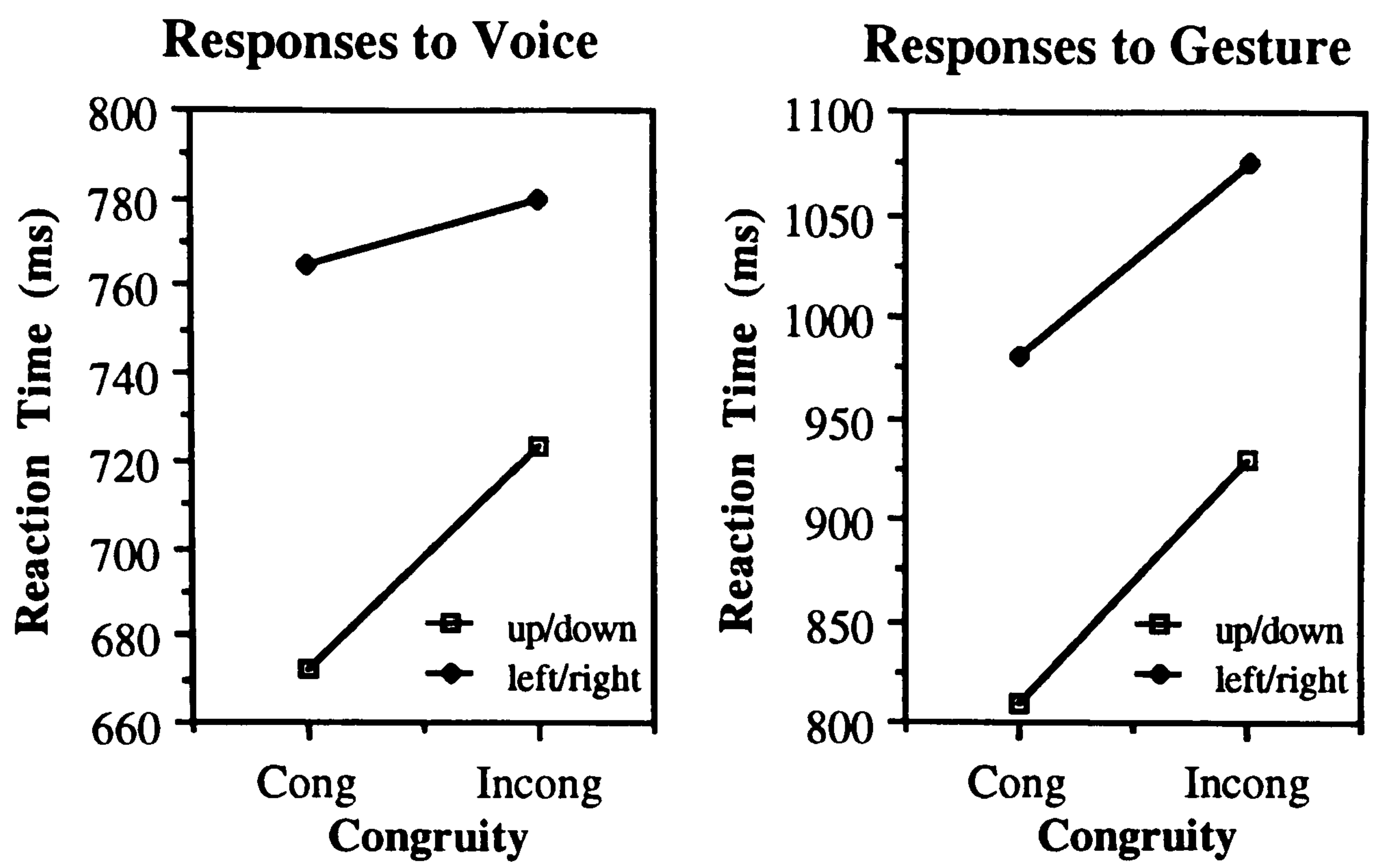


Figure 7. Mean reaction times (ms) as a function of congruity and decision for responses to the voice (left panel) and gesture (right panel) dimensions in Experiment 5.

The overall mean percentage error was 1.60%. The correlation between RT's and errors across both voice and gesture data was 0.87 suggesting no evidence of a trade-off between speed and accuracy. However, a glance at the voice data in Table 5 suggests some slight evidence of a trade-off despite the relatively low rate of errors (overall error for voice data was 0.70% which translates to less than one error per subject on the voice response trials). The RT/error correlation for these data was only 0.39.

Discussion

The results of this experiment provide some evidence for an S-R component to the interference effects observed in the previous chapter. Here the mapping between left/right stimuli and responses was made incompatible, whilst the mapping between up/down stimuli and responses was consistent throughout. When required to respond to the voice dimension, subjects showed no evidence of an interfering effect of the incompatibly mapped left/right gesture stimuli, but significant interference from up/down gestures which bore a compatible mapping relationship with the keypress responses. In contrast, when gestures formed the target stimuli, interference effects were observed for both the compatibly and incompatibly mapped voice stimuli.

Significant effects of decision were found for both gesture and voice response tasks, RT's to left/right, were slower than to up/down stimuli. This result might be expected given the manipulation to the left/right S-R mapping. However, in previous experiments we have seen that, for responses to *voice* stimuli, left/right decisions are usually made somewhat slower than up/down decisions. It is therefore difficult to determine whether the effect observed here is due to general problems with the left/right dimension, or to the compatibility manipulation per se. In contrast, the results of these same experiments have indicated that when *gestures* form the relevant stimuli, subjects are faster to make left/right as opposed to up/down decisions. Here this relationship was reversed implicating an effect of the incompatible mapping.

Thus the S-R mapping manipulation appears to have exclusively affected the process which translates a gesture stimulus into a spatial response. In contrast, the voice-to-response translation appears to be largely unaffected This might explain why left/right voice stimuli persisted in interfering with gesture responses despite the absence of any S-R mapping. O'Leary & Barber (1993) have recently suggested that the *mapping* variable can only be applied usefully to *matching* S-R combinations. Non-matching variables, they suggest, are already incompatible as they require a cognitive transformation between conceptual domains for an S-R association to be made. It is suggested that a spoken word and a manual spatial response constitute a non-matching association. Thus, by O'Leary & Barber's reasoning, the mapping manipulation used here is unlikely to have as profound an effect on the already incompatible voice-response relationship as on the S-R compatible gesture-response mechanism. If it is actually the case that the voice-response relationship is S-R incompatible, then the results of the present experiment certainly suggest that irrelevant voices are able to cause significant interference effects independently of any S-R factors.

To recapitulate, the manipulation to the spatial S-R mapping between left/right gestures and left/right spatial responses made in this experiment greatly attenuated the interference effect caused by these irrelevant gestures. Verbal stimuli persisted in interfering with responses to the gestural dimension despite the response key manipulation. Thus it can be concluded that SRC does indeed play a role in producing the influence of irrelevant gestures but perhaps not that of irrelevant voices. The investigation into the influence of spatial SRC is continued in the next experiment.

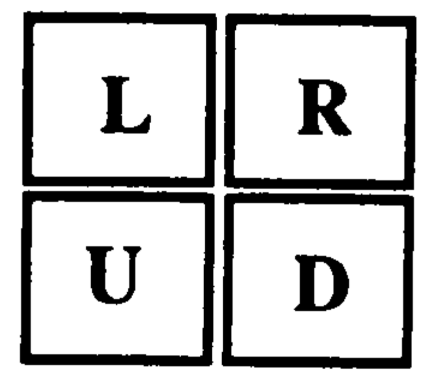
Experiment 6

The same logic was applied to this experiment as to Experiment 5, but this time the mapping from the up/down stimuli to the up/down decision keys was removed, but that for the left/right decisions was retained. This should prevent any automatic encoding of the response due to the irrelevant gesture stimuli. Thus a similar pattern of results as in the last experiment was predicted. As in Experiment 5, the manipulation is expected to have little effect on the interference caused by an irrelevant voice because the verbal stimuli and manual responses form non-matching S-R variables, but is likely to eliminate, or perhaps attenuate, the influence of up/down irrelevant gestures.

Method

Subjects: Eighteen undergraduate or post-graduate volunteers took part in this experiment, all had normal or corrected to normal vision and hearing.

Materials, Design and Procedure: These were identical to Experiment 1 with the exception of the change to the response keys. Keys 4 and 5 on the keypad now represented the up and down decisions respectively whilst keys 7 and 8 represented the left and right decisions. The keys were labelled "L", "R", "U" and "D" appropriately and arranged as illustrated.



Results

Table 6 summarises the error scores and average latencies for the correct responses to both voice and gesture stimuli. There are consistent congruity effects for both up/down *and* left/right decisions to voice stimuli (24 ms and 29 ms respectively) and rather larger effects for responses to gesture stimuli (91 ms for up/down and 68 ms for left/right decisions). These effects are also illustrated in Figure 8.

As in Experiment 5 the latency data from responses to the voice stimuli and those to the gesture stimuli were analysed separately.

The response latencies from the "voice" data were subjected to a 2 (decision) x 2 (congruity) ANOVA. The main effect of congruity just failed to reach significance although the 26 ms trend was in the predicted direction ($F(1,17)=3.59$, $p=0.075$). Similarly the main effect of decision also just failed to reach significance ($F(1,17)=4.39$, $p=0.051$). There was also no significant interaction between the congruity and decision factors ($F(1,17)=0.04$, $p=0.84$).

A similar 2 x 2 ANOVA performed on the response latencies of this "gesture" data revealed main effects of both congruity ($F(1,17)=6.91$, $p<0.05$) and decision ($F(1,17)=16.28$, $p<0.01$) in the predicted directions. The interaction term failed to reach significance ($F(1,17)=0.14$, $p=0.71$). This analysis therefore confirmed the above observation of consistent congruity effects throughout the "gesture" data.

The overall mean percentage error was 4.38%. Subjects made more errors to up/down stimuli than when making left/right decisions whilst fewer errors were made to congruent compared with incongruent pairs. The correlation between RT's and errors across both voice and gesture data was 0.81 suggesting no evidence of a trade-off between speed and accuracy.

Table 6.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent and Incongruent Conditions of Experiment 6.*

	Up/Down Decisions		Left/Right Decisions		Overall Mean	
Congruity	*M* RT	%E	*M* RT	%E	*M* RT	%E
			Responses to Voice			
Cong	781	6.11	736	0.56	759	3.34
Incong	805	5.00	765	1.67	785	3.34
Mean	793	5.56	751	1.12	772	3.34
			Responses to Gesture			
Cong	923	6.67	809	0.56	866	3.62
Incong	1014	9.44	877	5.00	946	7.22
Mean	969	8.06	843	2.78	906	5.42

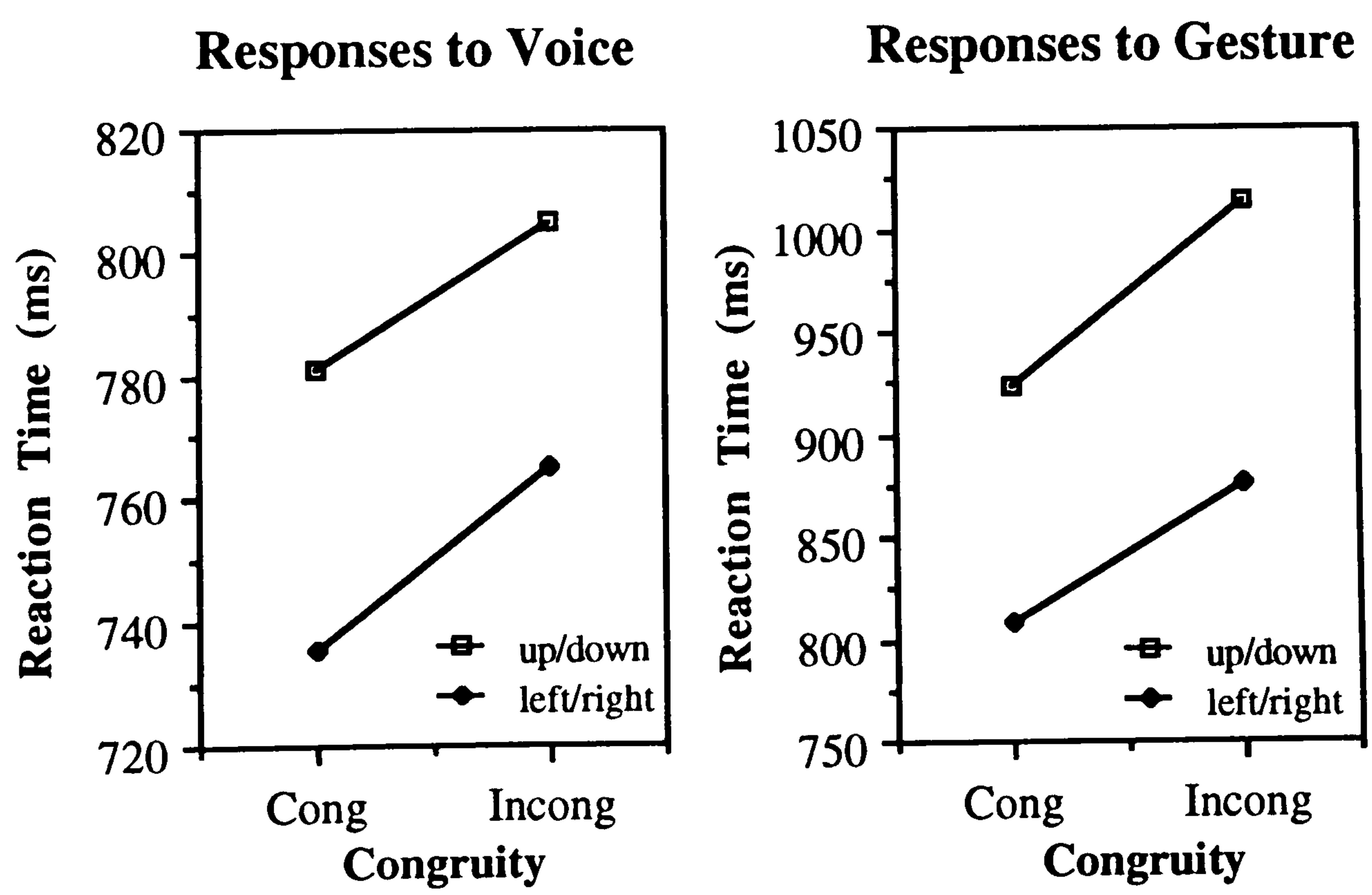


Figure 8. Mean reaction times (ms) as a function of congruity and decision for responses to the voice (left panel) and gesture (right panel) dimensions in Experiment 6.

Discussion

In the preceding experiment, up/down gestures enjoyed a compatible mapping relationship with the manual response and produced reliable interference effects when subjects were asked to respond to the voice stimuli. In contrast, the mapping between the left/right gestures and responses was made incompatible. As a result, these gestures no-longer interfered with responses to spoken words. This study was effectively a replication but with the incompatible mapping this time applied to the up/down S-R relationship.

When responding to gestures, the manipulation of the up/down response keys did not appear to remove the influence of the irrelevant verbal stimuli. The congruity effect persisted despite the removal of the SRC between the irrelevant spoken word and the up/down decision. This further suggests that the effects of irrelevant verbal stimuli are independent of SRC factors.

Similarly, and in contrast to Experiment 5, the trends in the data suggested that the congruity effect persisted when the required response was to the verbal stimuli. Here the manipulation of the up/down SRC appears not to have affected the interference effect caused by irrelevant up/down gesture stimuli. This result was rather at odds with the experimental hypothesis which suggested that removing the SRC between up/down gestures and decisions should eliminate the interference effect caused by these gestures. It should be stressed, however, that both the effects of decision and congruity for the voice response condition failed, albeit narrowly, to reach significance. This suggests that the experiment lacked the power to investigate the critical congruity by decision interaction. Nevertheless the magnitudes of the effects caused by both the compatible left/right gestures and incompatible up/down gestures were similar (29 ms and 24 ms respectively).

Significant effects of decision were found for the gesture response task and a near-significant effect for the voice data. It appears that the compatibility manipulation did indeed disrupt the gesture-to-response translation process. As to the voice-response translation, the incompatible up/down mapping reversed the normal advantage enjoyed by up/down over left/right decisions (Experiment 2). This time there is clear evidence that the disruption to the mapping affected the non-matching voice-manual response association (cf. O'Leary & Barber, 1993). Introducing an incompatible mapping, even to a pair of S-R variables which already require a translation between conceptual domains, can nevertheless make additional demands on the S-R encoding process.

The results of Experiments 5 and 6 are somewhat equivocal with respect to the contribution of SRC to the influence of distracting gestures on voice stimuli. Making left/right gestures incompatible with the left/right responses eliminated their ability to interfere with spoken words. On the other hand the same manipulation to the up/down dimension failed to remove or reduce the interference effect caused by to-be-ignored up/down gestures. Why should this be so? One explanation is that the effects of left/right and up/down gestures occur at different stages of processing with left/right gestures influencing response selection and up/down gestures disrupting voice classification prior to this stage of processing. However, an account which posits different levels of processing to such similar dimensions as left/right and up/down gestures clearly lacks parsimony. A second possibility is that the mapping manipulation had rather different effects on the two decision dimensions. For instance, it may be easier to translate an up/down stimulus into a left/right spatial code and map this onto an left/right response than to do the reverse, that is make a left/right-to-up/down translation and response. On this account a response to an irrelevant up/down gesture may still be encoded despite the inconsistent mapping, causing a response-based disruption to voice classification. In contrast, no response code will be activated from inconsistently mapped left/right gestures leaving voice classification free from interference. Neither of these explanations is particularly compelling.

Whilst the findings of Experiments 5 and 6 are somewhat ambiguous with regard to the effects of distracting gestures, none of the S-R mapping manipulations affected the interference caused by an irrelevant voice when making gesture classifications. This suggests that either the irrelevant S-R transformation takes place in spite of the manipulation of the S-R mapping, or rather that this particular congruity effect has its locus before response selection.

Because of the equivocal nature of the findings obtained in Experiments 5 and 6 it was decided to examine the two decision dimensions separately in two further experiments. In Experiments 5 and 6 one type of decision (e.g. up/down) was made S-R incompatible whilst the other remained S-R compatible (e.g. left/right). Thus, the comparison of compatible with incompatible S-R arrangements also involved comparing left/right with up/down decisions. The aim of the following two experiments was to eliminate this complication by manipulating the S-R compatibility of a single decision dimension.

Experiment 7

The aim of this experiment was to examine how the mapping between left-right gestures and manual responses influences the interference effects which these gestures have on responses to spoken words. In contrast to the preceding experiments, subjects are now only asked to respond to the spoken words “left” and “right” and to ignore the left/right gestures. The mapping between these gestures and the keypress response is manipulated as a between-subjects factor.

The results of Experiment 5 suggest that the congruity effect caused by an irrelevant left/right gesture will be eliminated when the S-R mapping between this irrelevant gesture and its appropriate response is removed. Thus in the incompatible condition no congruity effect should be evident, whilst in the compatible S-R mapping condition the interference effect should be reinstated. An interaction between compatibility and congruity is therefore predicted.

Method

Subjects: Thirty-two undergraduate and postgraduate volunteers acted as subjects in this experiment.

Materials: The left/right visual and auditory stimuli were selected from those used in Experiments 1, 3 and 4. These were combined in the same way to form 20 congruent and 20 incongruent trials.

Design: The subjects were tested in a 2 x 2 mixed design with one within-subjects factor, congruity (congruent or incongruent) and one between-subjects factor, S-R compatibility (compatible or incompatible).

Procedure: The stimuli were presented as before using the "SuperLab" software. Subjects were asked to respond only to the voice stimuli and to ignore the gestures. In the S-R compatible condition subjects responded using the keys 4 and 5 on the keypad, labelled "L" and "R" respectively. These keys are arranged horizontally and therefore have a consistent spatial relationship with the left/right gestures. In the incompatible condition subjects responded using the 8 and 5 keys, again labelled "L" and "R". These keys are arranged vertically and as such bear no direct spatial relationship with the gestures. As in Experiment 1, the experimental trials were preceded by ten practice trials representing a cross section of the stimuli. Eight question mark trials were also included to ensure that subjects did not close their eyes or look away from the monitor. Subjects were asked to use only their preferred hand in making the keypress responses.

Results

The mean correct latency scores and percentage of errors recorded in each condition are summarised in Table 7 and displayed graphically in Figure 9. There is a clear 48 ms effect of congruity for the compatible mapping trials but a -9 ms effect for incompatible trials. These observations were confirmed by a 2 (congruity) x 2 (compatibility) ANOVA conducted on the RT data. This revealed a main effect of congruity ($F(1,30)=6.11$, $p<0.05$) and a significant interaction between compatibility and congruity ($F(1,30)=12.93$, $p<0.01$). Further analysis of the simple effects indicated that the effect of congruity was present for the compatible condition ($F(1,30)=18.41$, $p<0.001$) but failed to reach significance for the incompatible trials ($F(1,30)=0.63$, $p=0.44$). The main effect of compatibility did not reach significance ($p=0.11$).

Table 7.

Mean RT's (in milliseconds) and Percentage of Errors (PE) *for Compatible and Incompatible Left/Right Decisions to Voices in Congruent and Incongruent Conditions of Experiment 7.*

	Compatible Mapping		Incompatible Mapping		Overall Mean	
	M RT	PE	*M* RT	PE	*M* RT	PE
Cong	559	0.94	658	1.56	609	1.25
Incong	607	3.43	649	0.94	628	2.19
Mean	583	2.19	654	1.25	619	1.72

On average subjects made only 1.72% errors. As with the RT data, there appears to be an effect of congruity for the compatible condition with subjects making fewer errors to congruent compared with incongruent stimuli. The reverse is true for the incompatible condition. Subjects made slightly more errors to the congruent compared with the incongruent trials. Again this difference reflected that of the RT scores. Indeed, the correlation between RT's and errors was only -0.02 offering no evidence of a trade-off between speed and accuracy. No further analysis was conducted on the error scores.

Discussion

The findings here replicate those of Experiment 5. When left/right gestures could be mapped simply onto left/right responses, a large 48 ms congruity effect was obtained. When the mapping was made incompatible, this effect was eliminated. The results of this experiment add weight to the suggestion that SRC factors play a role in producing the interfering effects of irrelevant gestures.

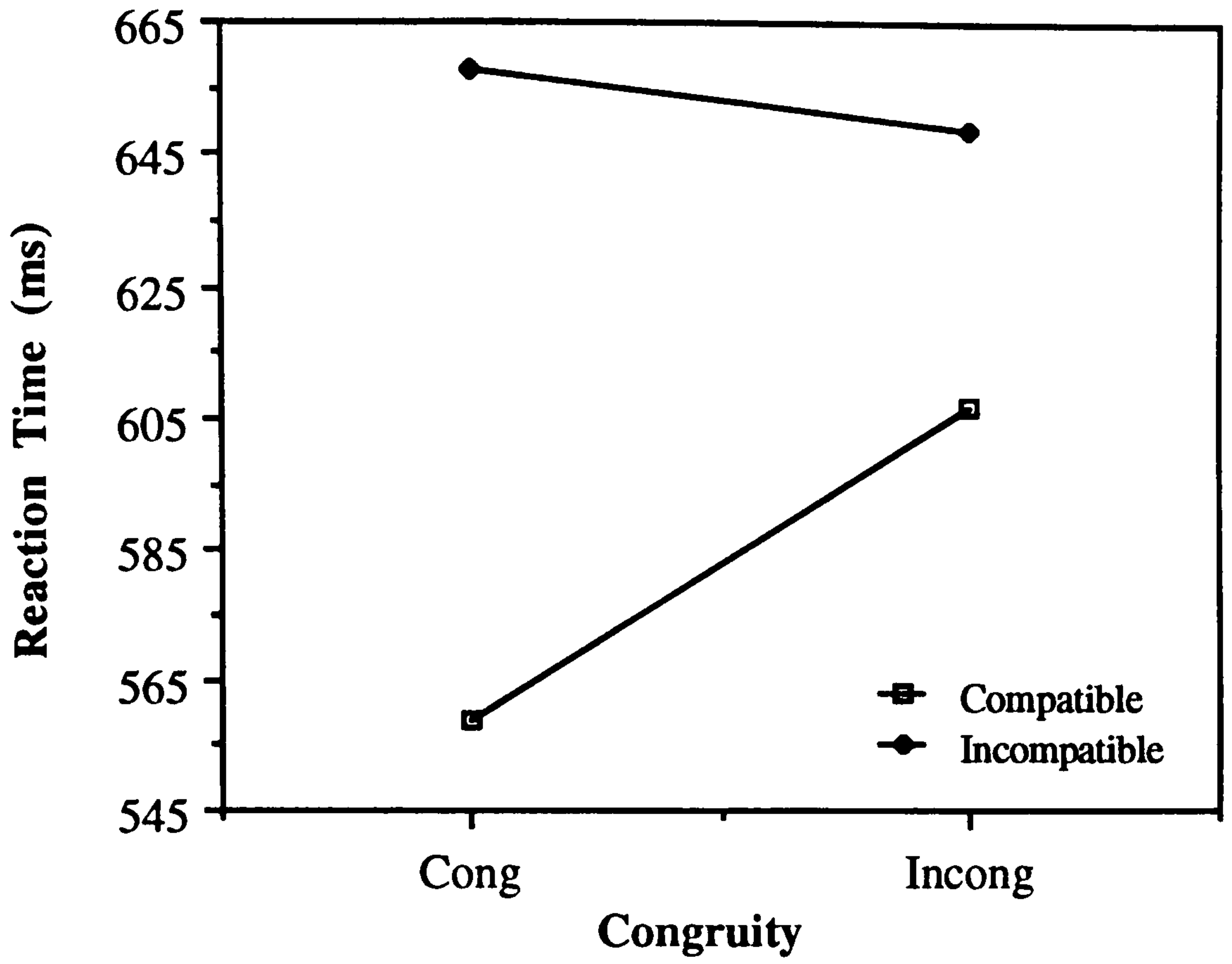


Figure 9. Mean reaction times as a function of congruity and compatibility factors of Experiment 7.

Experiment 8

This experiment was similar to Experiment 7 except here it was the up/down SRC that was manipulated as the between-subjects factor. Again, if the interference effect is caused by the encoding of a response to the irrelevant gestures by virtue of a compatible S-R mapping then eliminating this mapping should remove the interference effect.

Method

Subjects: Thirty-two undergraduate and postgraduate student volunteers acted as subjects in this experiment, none of whom had participated in any of the previous experiments. All had normal or corrected to normal vision and hearing.

Materials, Design and Procedure: These were identical to Experiment 5 with the exception that only the up/down visual and auditory stimuli were selected from those used in Experiment 1. These were combined to form 20 congruent and 20 incongruent trials as above. Subjects in the compatible S-R condition used keys 8 and 5 on the keypad labelled "U" and "D" respectively whilst those in the incompatible condition used keys 4 and 5, similarly labelled. Thus the response keys were changed from a vertical to a horizontal arrangement.

Results

The error rates and mean correct reaction times for all the conditions of this experiment are summarised in Table 8. Mean RT's are also illustrated graphically in Figure 10.

Table 8.
Mean RT's (in milliseconds) and Percentage of Errors (PE) *for Compatible and Incompatible Up/Down Decisions to Voices in Congruent and Incongruent Conditions of Experiment 8.*

	Compatible Mapping		Incompatible Mapping		Overall Mean	
	M RT	PE	*M* RT	PE	*M* RT	PE
Cong	524	1.88	534	1.25	529	1.57
Incong	567	2.19	546	0.63	557	1.41
Mean	546	2.04	540	0.94	543	1.49

There is a clear effect of congruity for compatible responses with congruent stimuli receiving faster responses by some 43 ms than incongruent responses. However there is a smaller 12 ms effect of congruity in the incompatible condition. These observations were confirmed by a 2 (congruity) x 2 (compatibility) ANOVA which indicated a main effect of congruity ($F(1,30)=17.70$, $p<0.001$) and a significant interaction between congruity and compatibility ($F(1,30)=6.02$, $p<0.05$). Further analysis of the interaction revealed a significant effect of congruity for compatible responses ($F(1,30)=22.19$, $p<0.001$) but no effect for incompatible responses (F(1,30)=1.54, p=0.22). The main effect of compatibility also failed to reach significance in this analysis (F(1,30)=0.04, p=0.84).

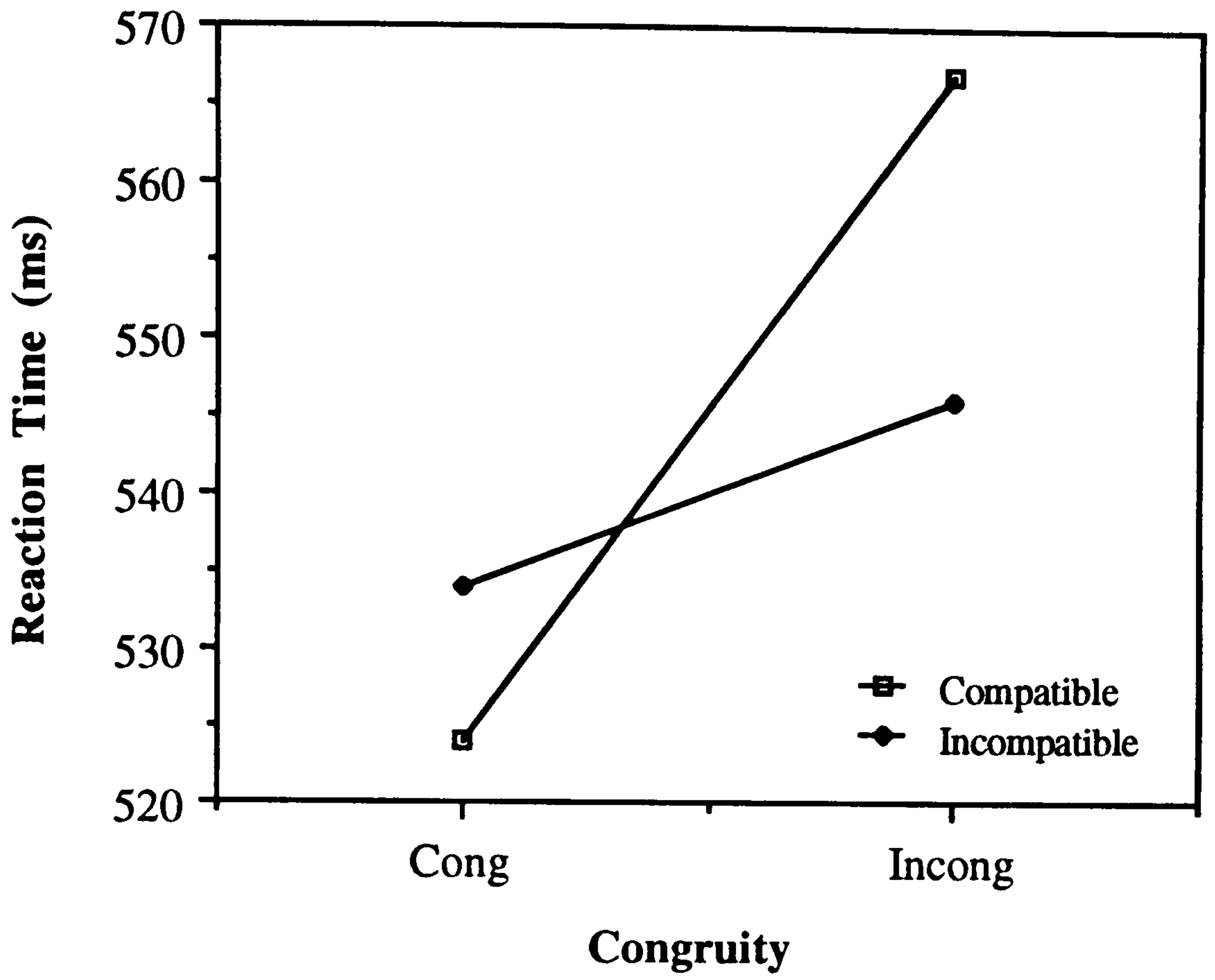


Figure 10. Mean reaction times as a function of congruity and compatibility factors of Experiment 8.

In the compatible mapping condition, subjects made slightly fewer errors on congruent than on incongruent trials whilst the reverse was true in the incompatible condition. Overall subjects completing the experiment with a compatible mapping arrangement made more errors than those with an incompatible mapping. The correlation between RT's and errors was 0.23 which was not suggestive of a trade-off between speed and accuracy. Because the overall error rate was low (1.49%), no further analysis was completed on the error data.

Discussion

In Experiment 6 the manipulation of the up/down S-R mapping did not affect the interference effect produced by to-be-ignored up/down gestures on voice classification. The results of Experiment 8, however, failed to replicate this finding.

Subjects showed no evidence of a significant interference effect when the mapping was inconsistent, but a significant effect when irrelevant up/down gestures could be easily mapped onto up/down responses. Again, this result adds further support to the notion that irrelevant gestures exert their effects at the response end of processing by means of an automatic S-R transformation process.

Experiments 6 and 8 offer rather contradictory findings with regard to the effects of manipulating the up/down S-R mapping. The absence of an interaction between decision and congruity factors in Experiment 6 indicated that the effects of up/down gestures are not mediated by SRC factors. However, these same gestures failed to cause any reliable interference in the incompatible condition in this experiment.

To summarise the results so far, in Experiments 5, 7 and 8 interfering effects of irrelevant gestures were eliminated under conditions where the gesture-response mapping was incompatible. However, in Experiment 6 there was some evidence of the influence of to-be-ignored gestures despite the mapping manipulation. In addition to these findings, in Experiments 5 and 6 the interfering effect of to-be-ignored verbal items was large and consistent despite manipulations likely to affect the response stage of processing. Thus it appears that SRC factors may well mediate the effects of to-be-ignored gestures whilst the effects of irrelevant voices appear independently of compatibility manipulations.

What implications do these observations have for the locus of the gestural-verbal interaction? In Experiments 5, 7 and 8 the effects of compatibility and congruity for voice responses were interactive. This suggests that the two factors affect the same stage of processing (Sternberg, 1969). Assuming that the mapping manipulation affects the latter response selection processes then, by this logic, so do the effects of congruity. In contrast, the additive effects of compatibility and congruity factors for responses to the gesture dimension (Experiments 5 and 6) indicate that these manipulations affect different stages of processing. The congruity effects caused by irrelevant voices therefore occur at a stage prior to response selection. The implication is that the symmetrical effects observed in Chapter 3 arise from an asymmetrical pattern of interactions, gestures exerting their effects at response selection by virtue of their compatible S-R relationship with the response, and voices exerting their effects at an earlier stage of processing. This kind of account seems somewhat unsatisfactory. Symmetrical interference effects seem to cry out for a "symmetrical" explanation, an account where interference occurs between the two dimensions at the same level of information processing, perhaps as the result of some kind of bidirectional "crosstalk" (e.g. Melara & Marks, 1990c). If we are to embrace a single locus of interference one of two things must be possible. Either (1) a response to an irrelevant verbal stimulus is automatically activated in a similar way to that encoded from an irrelevant gesture i.e. by way of an automatic S-R translation, or (2) despite the apparent evidence to the contrary, interference effects of irrelevant gestures actually persist in incompatible S-R conditions. A response selection locus is implicated by (1) whilst (2) would place the interference effects prior to this stage.

It seems unlikely that a spatial response code could be automatically activated from a verbal stimulus, particularly under incompatible mapping conditions. For this to occur a translation is required from the verbal code describing the stimulus to a spatial code describing the response. Moreover, with an incompatible mapping, a left/right spatial code would then require a translation to an up/down code to produce maximal interference. These extra processing stages make (1) unlikely, voices probably do not exert their effects at response selection.

Whilst the compatibility manipulations of Experiments 5, 6 and 8 resulted in statistically non-significant interference effects of gestures on voice classification, in only one case was the numerical value of the effect reduced beyond zero (the -9 ms effect in Experiment 7). In the two other cases a small trend in the predicted direction exists (15 and 12 ms in Experiments 5 and 8 respectively). Furthermore a near-significant 24 ms effect was found in Experiment 6. One suggestion is that the elimination of spatial SRC *reduced* but did not actually remove the influence of irrelevant gestures but that Experiments 5-8 lacked the power to indicate this. If this is the case (2) might be true i.e. gestures could exert an effect independently of response factors.

In summary, the majority of the experiments reported above provide some evidence for a spatial SRC component to the interfering effect of irrelevant pointing gestures. However, the persistence of the effects of to-be-ignored voices on gesture classification means that we are forced to posit separate loci for the interfering effects of gestures and voices, clearly an unsatisfactory state of affairs. Given this difficulty, and in view of the slight evidence of a residual interfering effect of irrelevant gestures in incompatible conditions, it was felt that further exploration of the contribution of SRC was warranted. This issue is pursued in the following three experiments.

The Naming Task

The results of Experiments 5-8 suggest that the interfering effects of to-be-ignored gestures on voice classification arise principally as a result of the compatibility between the spatial gestures and the manual responses. Gesture interference was eliminated (or reduced) when the S-R relationship between gestures and responses was disrupted. This manipulation was achieved by making the *mapping* between gestures and responses incompatible. It could be argued that an inconsistent mapping merely requires that an extra S-R translation be made between, say, a left/right gesture and an up/down response. Thus the SRC might simply be reduced rather than eliminated. In Experiments 9-11 the *match* between gestures and responses is manipulated whilst maintaining the mapping. This is achieved by switching from a manual to a verbal mode of response. According to O'Leary & Barber (1993) this will make the S-R relationship between gestures and responses completely incompatible and should therefore preclude any interfering effect of irrelevant gestures.

Spoken naming responses have been the norm in many studies adopting a Stroop interference paradigm. Indeed, in Stroop's original colour-word task, subjects were required either to *name* the colours of the colour-words or actually read the list of colour words. Similarly in picture-word studies, interference usually occurs when subjects are asked to *name* pictures with an irrelevant word printed either inside, or on the periphery of the picture (e.g. Glaser & Düngelhoff, 1984; Underwood, 1976). Typically Stroop interference is greater with oral as opposed to manual responses (e.g. White, 1969; McClain, 1983). This observation has led many to suggest that the influence of irrelevant verbal materials is largely due to stimulus-response factors as there is an obvious compatibility between a verbal stimulus and a verbal response. However, the fact that the interference effect persists with manual responses, suggests that the relationship between stimulus and response modes alone cannot completely account for the Stroop effect (see also MacLeod, 1991). Indeed, in Experiments 5 and 6 we have seen that the interfering effect of irrelevant voice stimuli is largely immune to SRC manipulations.

In the present experiments the response modality is manipulated to examine the role of response factors in producing gesture, rather than Stroop interference. Switching from a manual to a vocal response will introduce a compatibility between verbal stimuli and verbal responses. If "stimulus-response compatibility matters" (MacLeod, 1991, p.183) one might expect heightened interference effects of irrelevant voices i.e. a larger "Stroop" effect. However, if the interfering effect of irrelevant gestures is caused entirely by SRC factors, more specifically by an automatic activation of a response code based on an irrelevant, but spatially compatible gesture, then gestures will no longer interfere with verbal *naming* responses to voice stimuli. By switching from a manual to an oral response, gestures and responses now form non-matching S-R variables since they belong to different conceptual domains. Any S-R association between such non-matching variables is by definition incompatible (O'Leary & Barber, 1993) because of the need to translate stimulus codes from one dimension into another for the response. For instance the gestural stimuli in the "spatial" domain require a transformation into the "verbal" domain for a spoken response. This extra processing stage necessary to convert codes between domains might be expected to prevent any automatic encoding of a response to an irrelevant gesture, and so should eliminate the interference effect.

In summary, the likely result is an interaction between congruity and response factors. Large congruity effects are expected in responses to gesture where there is a strong SRC between the irrelevant voice and the response, but none in conditions where the voice is the relevant dimension and the spatial SRC between irrelevant gestures and responses is absent.

Experiment 9

In this experiment we reverted to the printed verbal stimuli printed across the chest of the gesturer as in Experiment 3 (i.e. intra-modal stimuli) as it was originally thought that any auditory stimuli would disrupt the operation of the voice-key apparatus used to record subjects responses (see below). However, the results of Experiment 2 and 3 showed that pointing gestures are capable of interfering with both auditorily *and* visually presented verbal materials. Thus it may be that the modality of the verbal input is of relatively little importance. Regardless of the input modality of the verbal stimuli, the question of interest here is to what extent gestures are able to exert their effects in the absence of spatial SRC.

To reiterate the predictions, in terms of a SRC account one would expect asymmetry, the irrelevant gestures will no longer interfere with responses to the voice stimuli as the removal of any spatial SRC will ensure that no response will be automatically encoded. On the other hand interference should persist from the to-be-ignored voice stimuli, either because of the SRC, or because the effects of verbal stimuli are mediated prior to the response selection stage of processing.

Method

Subjects: Thirty-two subject volunteers participated in this experiment, all had normal or corrected to normal vision.

Materials and Apparatus: The materials were identical to those used in Experiment 3. SuperLab 1.6, which has the capacity for auditory input from voice-key apparatus, was used as the software for this experiment. The voice-key consisted of a normal Macintosh microphone driven by the SuperLab software. Otherwise the specifications of the software were similar to the version used in the previous experiments. The microphone was attached to the subject's clothing as near as possible to the throat. Output from this microphone was detected by the computer and used to stop the timer.

Design and Procedure: The design of this experiment was identical to Experiment 3, except that in this experiment subjects were instructed to either name the direction of the gesture, or read aloud the direction word printed across the gesturer's chest, depending on the response block in question. Trials were presented in a random order within each block. The vocal response of the subject stopped the timer which, as before, measured the RT from the onset of the stimulus pair. The response terminated the stimulus display which was replaced by a blank screen. Subjects' responses were recorded on a response sheet by the experimenter. A 500 ms inter trial interval intervened between the subject's response and the presentation of the following trial. As in previous experiments question mark trials were included in each response block, upon which subjects were asked to press the space bar. Both reaction times and percentage of errors were recorded as dependent variables in this experiment.

Results

Trials were discarded where subjects' vocal responses failed to reach the threshold of the voice key microphone. The initial response of each subject was recorded, any self-corrections were marked as errors. Again individual RT scores under each condition were rejected if they deviated by more than two standard deviations from the cell mean. Table 9 summarises the mean correct reaction times and error rates recorded under each condition of Experiment 9. There are clear effects of congruity for responses to gesture, with performance in both decision conditions faster and more accurate for congruent than for incongruent trials. This pattern is repeated for responses to word stimuli but the overall size of the congruity effect appears to be much smaller. Overall, responses to word stimuli are made faster than those to gestures.

Correct reaction times were entered into a three-way analysis of variance (ANOVA) with three within subject factors: Response (to gesture or verbal stimuli), Congruity (congruent or incongruent trials), and Decision (up/down or left/right). This analysis largely confirmed the observations made above. There were significant main effects of congruity ($F(1,31)=21.56$, $p<0.001$) and response ($F(1,31)=117.03$, $p<0.001$) and a significant interaction between these two factors ($F(1,31)=26.68$, $p<0.001$). Further analysis of this interaction revealed a significant congruity effect (45 ms) for responses to gesture stimuli ($F(1,62)=40.98$, $p<0.001$) and a smaller congruity effect (14 ms) for responses to word stimuli ($F(1,62)=3.99$, $p<0.05$). The interaction also qualified the main effect of response with faster responses to word as opposed to gesture stimuli. Other significant sources of variance were the main effect of decision ($F(1,31)=16.34$, $p<0.001$) and the interaction between response and decision factors ($F(1,31)=34.11$, $p<0.001$) which again confirmed the RT advantage for word over gesture naming but also suggested a significant 44 ms advantage for up/down, over left/right responses to gesture stimuli ($F(1,31)=37.79$, $p<0.001$). As might be expected no such difference was evident for word reading. This larger effect of response for the left/right dimension might well reflect the ambiguity of left/right as opposed to up/down gestures. The interactions between response and congruity and response and decision are displayed graphically in Figure 11.

Table 9.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent and Incongruent Conditions of Experiment 9.*

	Up/Down Decisions		Left/Right Decisions		Overall Mean	
Congruity	*M* RT	%E	*M* RT	%E	*M* RT	%E
			Responses to Word			
Cong	491	0	494	0	493	0
Incong	515	0	499	0.63	507	0
Mean	503	0	497	0.32	500	0.16
			Responses to Gesture			
Cong	598	0	646	0.94	622	0.47
Incong	647	1.25	688	1.25	668	1.25
Mean	623	0.63	667	1.10	645	0.86

Error scores generally mirror the RT data (see Table 9). The correlation between RT's and errors was 0.9 suggesting no evidence of a trade-off between speed and accuracy. Subjects were slightly more error prone in incongruent compared with congruent conditions, whilst also making more errors in the incompatibly S-R matched gesture naming, as opposed to word naming response condition. Because of the low rate of errors (the overall error rate was 0.51%) no further analysis was conducted on these data.

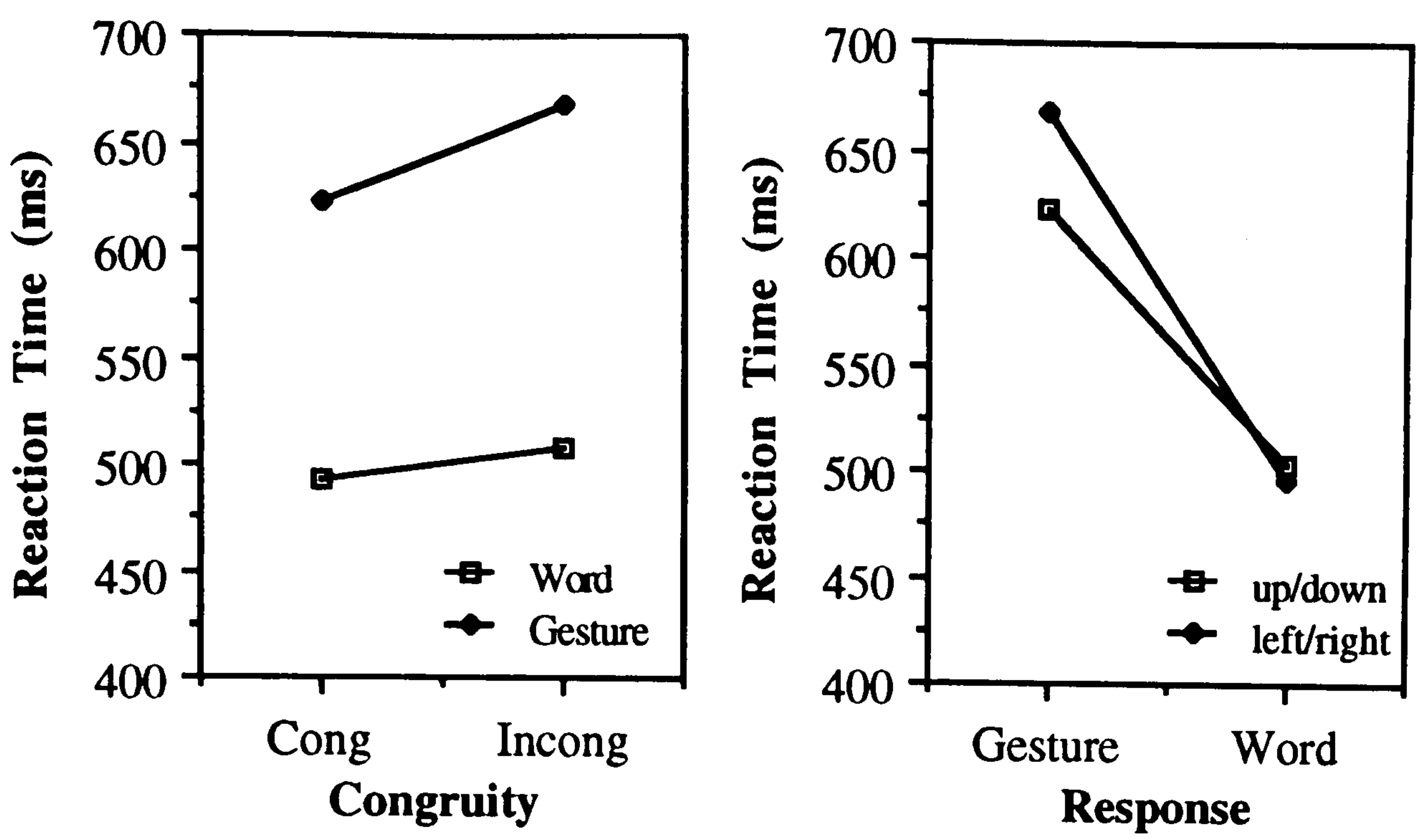


Figure 11. Mean Reaction Times (ms) from Experiment 9 as a function of congruity and response factors (left panel), and response and decision factors (right panel).

Discussion

The switch in response modality from manual to oral had the expected effect of eliminating the symmetry of the interference effects. Irrelevant verbal information caused a significant 45 ms interference effect on gesture classification whereas to-be-ignored gestures exerted a relatively smaller 14 ms effect on voice classification. Thus eliminating the S-R match between irrelevant gestures and responses reduced, but did not entirely eliminate their ability to interfere with concurrent verbal stimuli. This result implies that S-R compatibility does indeed "matter". Subjects may well automatically map a spatially compatible gesture stimuli onto a spatial response, however, the presence of the small residual interference effect observed here suggests that SRC is not sufficient to provide a complete account of gesture interference.

Left/right decisions were particularly slowed when the response was switched from word to gesture naming. It may be that with no spatial SRC to map gestures onto responses, subjects' naming responses were affected by the inherent ambiguity of the left/right gesture (e.g. to whose left does a "left" gesture refer to?). The difficulty of processing this type of gesture may also have contributed to the smaller overall influence of irrelevant gestures. Indeed, the magnitude of the interference effect caused by left/right gestures was only 5 ms, compared to the 24 ms effect of irrelevant up/down gestures. However it would seem that the experiment lacked the power to reveal this difference in a higher order, three-way interaction. Interestingly a similar pattern of results were obtained in Experiments 5-8. S-R incompatible left/right gestures exerted a 15 ms congruity effect in Experiment 5 and a -9 ms effect in Experiment 7. Contrast this with the 24 ms and 12 ms effects exerted by S-R incompatible up/down gestures in Experiments 6 and 8 respectively. It may well be that compatibility manipulations, whether caused by inconsistent mappings or matches, are more detrimental to the irrelevant left/right, than to the up/down gesture-to-response translation.

Overall, gesture naming responses were significantly slower than response to words. This is the reverse of the pattern observed in many of the earlier experiments and was presumably the result of the incompatible matching between gestures and oral responses, and the compatibly matched voice-response relationship. Indeed, comparing mean RT's here with those of Experiment 3, we find that overall gesture naming RT's were some 20 ms slower than manual responses to the gesture dimension. Thus the removal of spatial SRC has slowed responses to the gesture dimension as might be expected if subjects had difficulty mapping the spatial gestures onto verbal responses. However, mean word naming RT's in the present experiment were 197 ms *faster* than average manual responses to printed words in Experiment 3. In this instance, the relationship between the verbal stimuli and the verbal response is compatible. This form of SRC, like the spatial SRC existing in the previous experiments, will allow efficient encoding of a verbal response from a verbal stimulus. Thus the manipulation has slightly slowed gesture responses but massively speeded responses to verbal stimuli. It is possible that this mismatch in processing speeds is at the heart of the reduction in the effect of gestures. The processing of the verbal information may have been simply too fast for gestures to exert any larger effects. On the other hand, response effects of similar magnitude but in the reverse direction, occurred in Experiments 1, 2 and 4 and yet the slower verbal dimension still managed to interfere with the faster gesture classification responses. As will be discussed shortly, explanations of interference which rely on the relative speed of processing of the alternative dimensions fall some way short of explaining both the present findings and interference effects in general.

To summarise, the elimination of the SRC between gestural stimuli and responses caused by the switch from a manual to an oral response (an incompatible S-R "match") substantially reduced, but did not entirely eliminate, the interfering effect of to-be-ignored pointing gestures on word naming responses.

Experiment 10

In this experiment we reverted to the cross-modal stimuli used in Experiments 1 and 2 but employed the verbal response as in the previous experiment. The predictions are identical to Experiment 9. With the removal of spatial SRC, gestures might not be expected to exert any influence on the processing of verbal information.

A further addition in Experiment 10 was the inclusion of a neutral condition to examine any contribution of facilitation and/or inhibition to the interference effects. The S-R model and other response selection accounts of interference rely on the parallel processing of the two dimensions, resulting in the determination of two responses. Interference occurs when these responses are in conflict and some decision process must first identify the correct response, abort the interfering response and finally program and execute the correct response (e.g. Kornblum, Hasbroucq & Osman, 1990, Kornblum, 1994). As already outlined gestures are not expected to interfere with vocal responses to verbal stimuli. However to-be-ignored voices are expected to interfere with responses to gestures. A congruent voice/gesture pairing will result in a relatively speedy RT since the decision process will not require any abortion or re-programming of a response because both relevant and irrelevant responses are the same. A neutral word e.g. "car" will not activate a corresponding element in the response set as there is no corresponding element to activate (the response set consists of "up", "down" "left" and "right" response codes). The correct response is thus unopposed, resulting in a relatively fast RT. Therefore both congruent and neutral conditions should produce faster RT's than incongruent trials. A pattern of inhibition without facilitation might be expected.

To summarise, under the S-R model an asymmetric pattern of interference is predicted. Voices will interfere with gestures but not vice-versa. Moreover, in comparison to a neutral condition, incongruent words will produce an inhibitory effect on gesture responses but congruent words will not produce facilitation.

Method

Subjects: Sixteen subjects participated in this experiment, these were mainly undergraduates from the University of Nottingham. All had normal hearing and normal, or corrected to normal vision.

Materials and Apparatus: In this experiment a neutral condition was included. Along with the images of gestures used in Experiment 1, images of a car, a hat, a phone, a watch and a book were frame grabbed and used as neutral visual stimuli in this experiment. Similarly the words "car", "hat", "phone", "watch" and "book" were recorded using the audio and "Soundedit" software and used as the neutral verbal stimuli. As before all the verbal stimuli were edited to be approximately the same length (0.8 sec's).

Three types of test stimuli were prepared from the frame grabbed images and auditory speech stimuli. The congruent and incongruent trials were as before but now neutral trials were introduced. Neutral trials for the gesture response condition consisted of each directional gesture paired with one of the neutral words ("car", "hat" etc.) resulting in ten different neutral pairings. Similarly neutral "voice" trials consisted of each directional word paired with one of the neutral visual images again producing ten different neutral trials.

The voice key apparatus and software were identical to those used in Experiment 9. However, the sensitivity of the microphone was adjusted in order to prevent the auditory stimuli from triggering the voice key.

Design and Procedure: This experiment took the form of a within-subjects design with three factors: Response (to the gesture or to the voice), Decision (left/right or up/down), and Congruity (congruent, incongruent or neutral trials). Again the stimuli were blocked by response, either to the gesture or to voice stimuli. Half of the subjects responded to the voice block first and half to the gesture block. Ten trials were included in each cell of the design making a total of 120 trials. The procedure was identical to Experiment 9.

Results

Table 10 summarises both the reaction time and error data for each condition of the experiment. There is a 17 ms effect of congruity for responses to voice stimuli and a rather larger 29 ms effect for responses to gesture. For both responses there is little inhibition (i.e. the difference between incongruent and neutral conditions) but rather larger facilitation. Overall responses to voice were made faster than responses to gesture .

These observations were largely confirmed by a 3 (congruity) x 2 (response) x 2 (decision) ANOVA performed on the RT data. This revealed main effects of response ($F(1,15)=29.83$, $p<0.001$), decision ($F(1,15)=6.07$, $p<0.05$), and an effect of congruity ($F(2,30)=9.13$, $p<0.001$) of some 23 ms. The absence of a response by congruity interaction ($p=0.41$) confirmed the symmetrical nature of the congruity effects. Post hoc Tukey tests between pairs of means for the three congruity conditions revealed significant differences between congruent and incongruent conditions (i.e. a significant congruity effect) and between congruent and neutral conditions (p's<0.01). The comparison between the incongruent and neutral conditions failed to reach significance.

Table 10.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent, Incongruent and Neutral Conditions of Experiment 10.*

	Up/Down Decisions		Left/Right Decisions		Overall Mean	
	M RT	PE	*M* RT	PE	*M* RT	PE
			Responses to Voice			
Cong	586	0	579	0	583	0
Neutral	598	0	594	0	596	0
Incong	608	0.63	591	1.88	600	1.25
Mean	597	0.21	588	0.63	593	0.42
			Responses to Gesture			
Cong	646	0.63	672	0.63	659	0.63
Neutral	669	0	711	0.63	690	0.32
Incong	664	1.88	716	2.50	688	2.19
Mean	660	0.84	700	1.25	679	1.05

The interaction between response and decision also reached significance ($F(1,15)=12.71$, $p<0.01$). Analysis of the simple main effects qualified the main effect of response with generally faster RT's to the voice stimuli. However, the effect appears to be greater for left/right as compared with up/down decisions. This analysis also confirmed that up/down decisions were made significantly faster than left/right decisions when responding to gestures (40 ms, $F(1,30)=18.49$, $p<0.001$) but that no such difference existed when the required response was to the verbal stimuli. This pattern of effects was identical to Experiment 9. No other effects reached significance (all p's >0.1). The effects of congruity and the interaction between response and decision factors are illustrated in Figure 12.

Subjects committed an average of 0.74% errors. An inspection of Table 10 indicates that error scores closely mirror the RT data. Subjects made more errors in responding to incongruent compared to congruent trials. Subjects also made fewer errors when responding to the S-R compatible verbal stimuli than to the incompatible pointing gestures. The correlation between RT's and errors was 0.51 offering no evidence of a trade-off between speed and accuracy. No further analysis was conducted on the error scores.

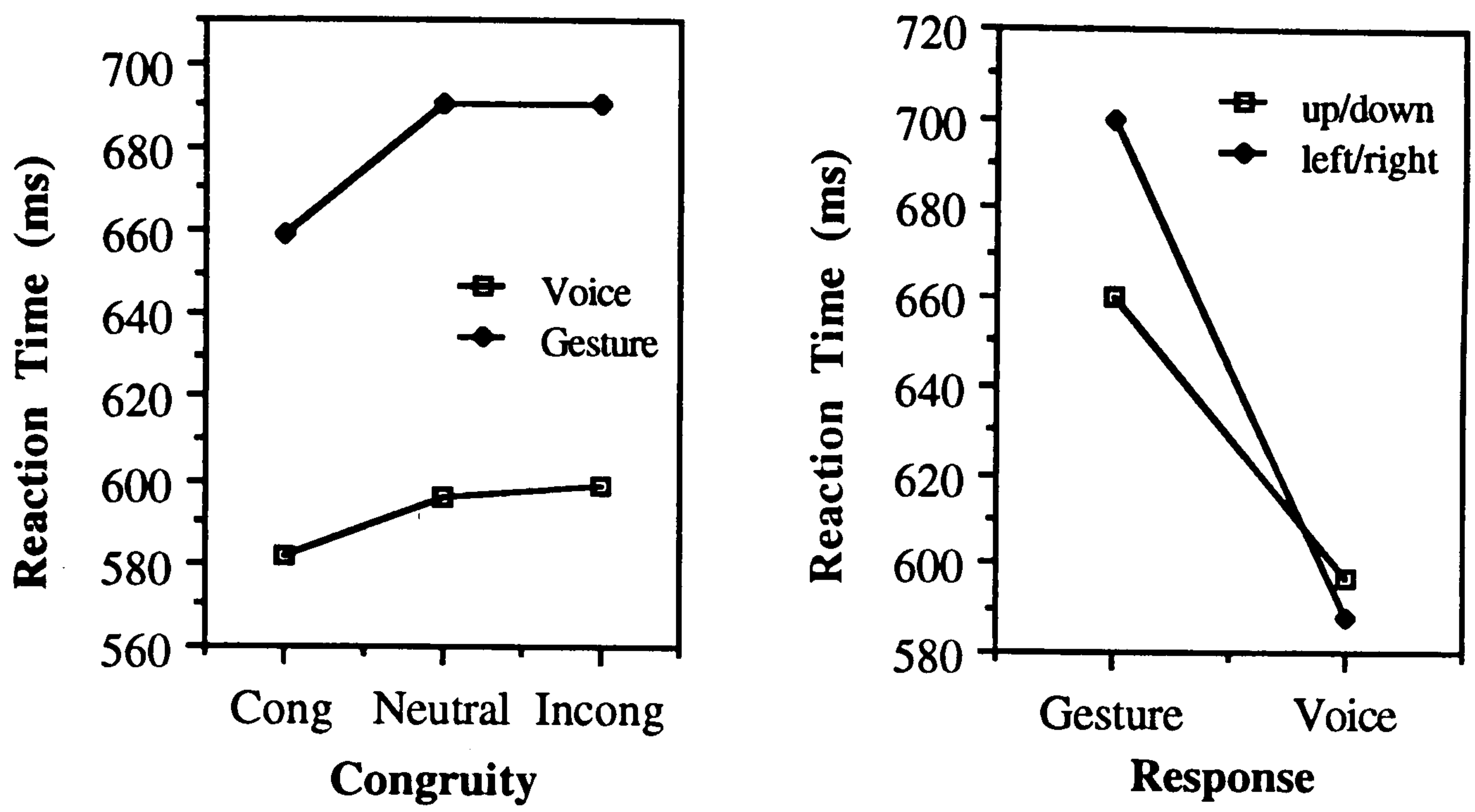


Figure 12. Mean Reaction Times (ms) from Experiment 10 as a function of congruity and response factors (left panel), and response and decision factors (right panel).

Discussion

The results of the previous experiment suggested that the interference effect caused by irrelevant gestures is reduced but not completely eliminated when the gesture-response S-R relationship is made incompatible by adopting a non-matching S-R arrangement. The results of this experiment also show significant effects of congruity for both voice and gesture naming responses. However, in contrast to Experiment 9, the effects of congruity and response are non-interactive, suggesting a symmetric pattern of interference. Furthermore the results suggest that this interference is best characterised as consisting of facilitation without inhibition.

These results are more consistent with the mutual interaction of gestural and verbal information than with the S-R hypothesis outlined above. Recall that under the S-R model, removal of the spatial SRC between a gesture and a response should prevent the encoding of the "gesture" response leaving the word-based response code unopposed. Despite the absence of spatial SRC, the results suggest that irrelevant gestures were still able to exert their effects on voice naming to the same degree as irrelevant voices influenced gesture naming. However the overall size of the congruity effect (23 ms) was somewhat reduced compared with Experiments 2 (57 ms) and 4 (85 ms) which were of similar design but employed the manual keypress response task. This finding is somewhat at odds with the literature on Stroop-type interference which suggests that the size of the effect should be *larger* when the response modality is switched from manual to oral (MacLeod, 1991). The reason for this may be that overall RT's in this experiment (636 ms) were slightly faster than in Experiments 2 and 4 (707 ms and 666 ms) resulting in a proportionately smaller congruity effect. Alternatively, as suggested in the previous experiment, it is possible that spatial SRC *contributes* to, rather than causes, the interference effects reported in this study. Removal of spatial SRC reduces, but does not completely eliminate the interference effect.

The change in response modality once again resulted in a reversal of the response effect observed in Experiment 2. Responses to gestures were made significantly slower than those to voices. Again, voice responses have been greatly speeded compared with Experiment 2 (593 ms versus 707 ms in Experiment 2) whilst gesture responses have been slightly slowed from 663 ms in Experiment 2 to 679 ms in the present experiment. Thus the reversal of the response effect appears to be caused by a speeding of responses to voice, and a slowing of gesture classification. However, despite the apparently profound effects of switching response modality, the mutual influence of the gestural and verbal dimensions persisted.

In a similar way to previous experiments, subjects appeared to have some problems processing left/right gestures. RT's to left/right gestures were again significantly slower than responses to up/down gestures. Whilst this left/right confusion might possibly reduce the overall effects of irrelevant gestures, the fact that symmetrical effects have been obtained suggests that it is not a critical factor.

Thus far then, the results of this experiment argue against a strong version of the S-R account of interference. In addition, the findings of symmetrical facilitation without inhibition are also problematic. Indeed, this pattern of results is uncommon in the Stroop literature. Facilitation is sometimes observed in the Stroop task but this is almost always accompanied by a larger inhibitory effect caused by incongruent items (e.g. Glaser & Glaser, 1982; Dalrymple-Alford, 1972). Facilitation is perhaps more common in the picture-word task but again these effects are found in parallel with much larger inhibition (e.g. Underwood, 1976; Glaser & Düngelhoff, 1984). The choice of the control items is crucial. Some use rows of X's or other characters (e.g. *#@?%) whilst others use non-words, or as in this experiment, unrelated words as neutral items in colour or picture naming tasks. However there is some evidence that unrelated words are able to cause a degree of interference and as such may not provide a factor-pure baseline measure of colour or picture naming. For instance Klein (1964) obtained small but reliable interference effects from unassociated words in a colour naming task whilst Fox et al. (1971) noted interference effects from unrelated common words in a direction naming version of the Stroop task. Thus it seems clear that unrelated words are able to exert some degree of interference in both colour and direction naming. Using such items as neutral words in an interference task would presumably produce an overestimation of facilitation and consequently an under-estimation of inhibition. Thus the present finding of facilitation without inhibition may well be the result of using inappropriate neutral items, at least for the gesture naming task. It is clear that we need to adopt more appropriate control items in order to gain better estimates of facilitation and/or inhibition.

In summary the results of this experiment provide some more evidence that the mutual interaction of gestural and verbal information can occur independently of the effects of SRC. However the relative contribution of facilitation and/or inhibition was not firmly established.

Experiment 11

Experiment 11 represented an attempt to confirm the findings of Experiments 9 and 10 where irrelevant gestures were found to cause a degree of interference in voice naming despite the absence of any obvious spatial S-R compatibility. However, in this experiment the neutral pictures used above were replaced with a single neutral gesture, whilst the word "blank" substituted for the neutral words. In the discussion of Experiment 10 it was suggested that the neutral stimuli used there acted as *irrelevant*, rather than as neutral items, and as such were likely to cause a degree of inhibition themselves, resulting in overestimates of facilitation and underestimates of inhibition. It was felt that single neutral items would cause less inhibition than the multiple items used in Experiment 10 and as such would provide better estimates of the contribution of facilitation and inhibition to the effects.

Method

Subjects: Fourteen subjects participated in this experiment. All had normal or corrected to normal vision and normal hearing.

Materials and Apparatus: In addition to the gesture stimuli used in Experiment 10, a neutral "gesture" was included. This consisted of the same gesturer standing with his arms by his side. The word "blank" was also recorded and edited as before and included as the neutral verbal stimulus. Neutral "gesture" trials consisted of each of the directional gestures paired with the word "blank" resulting in 4 different neutral pairings. Similarly neutral trials for the "voice" response condition consisted of each directional word paired with the neutral gesture again producing 4 different trials.

Design and Procedure: The design was identical to Experiment 10 i.e. a 2 (response) x 3 (congruity) x 2 (decision) repeated measures design. All subjects responded to ten trials in each of the twelve experimental conditions giving a total of 120 trials. As in previous experiments these trials were blocked by response with the order of presentation of the blocks alternated between subjects. The procedure for this experiment was identical to that of Experiment 10.

Results

Again both reaction times and error scores were recorded under each condition and these are presented in Table 11. Figure 13 illustrates mean RT's in all response and congruity conditions. Trials were discarded where subjects' vocal responses failed to reach the threshold of the voice key microphone. The initial response of each subject was recorded, any self-corrections were marked as errors. Again individual RT scores under each condition were rejected if they deviated by more than two standard deviations from the cell mean.

There are clear effects of congruity for both gesture and voice responses. Moreover, in contrast to Experiment 10, the pattern of effects is one of inhibition without facilitation for both responses to voice and gesture conditions.

The reaction time data were entered into a 2 (response) x 3 (congruity) x 2 (decision) ANOVA. This analysis yielded a main effect of decision ($F(1,13)=23.92$, $p<0.001$) with up/down responses being made 25 ms faster than left/right responses (676 versus 701 ms). The main effect of congruity was also significant ($F(2,26)=16.08$, $p<0.001$). Post hoc Tukey tests indicated reliable differences between congruent and incongruent conditions and between incongruent and neutral conditions (both p's<0.01) but no significant difference between congruent and neutral conditions. This translates to an overall congruity effect (i.e. the difference between incongruent and congruent conditions) of 23 ms (681 versus 704 ms) which consists of inhibition (24 ms) but no facilitation, confirming the observations made earlier.

Table 11.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent, Incongruent and Neutral Conditions of Experiment 11.*

	Up/Down Decisions		Left/Right Decisions		Overall Mean	
	M RT	%E	*M* RT	%E	*M* RT	%E
			Responses to Voice			
Cong	684	0	711	0	698	0
Neutral	690	0	703	0	697	0
Incong	722	0	731	0.71	727	0.36
Mean	699	0	715	0.24	707	0.12
			Responses to Gesture			
Cong	650	0.71	680	0.71	665	0.71
Neutral	650	0.71	675	0.71	663	0.71
Incong	657	2.14	705	2.14	681	2.14
Mean	652	1.78	687	1.19	670	1.19

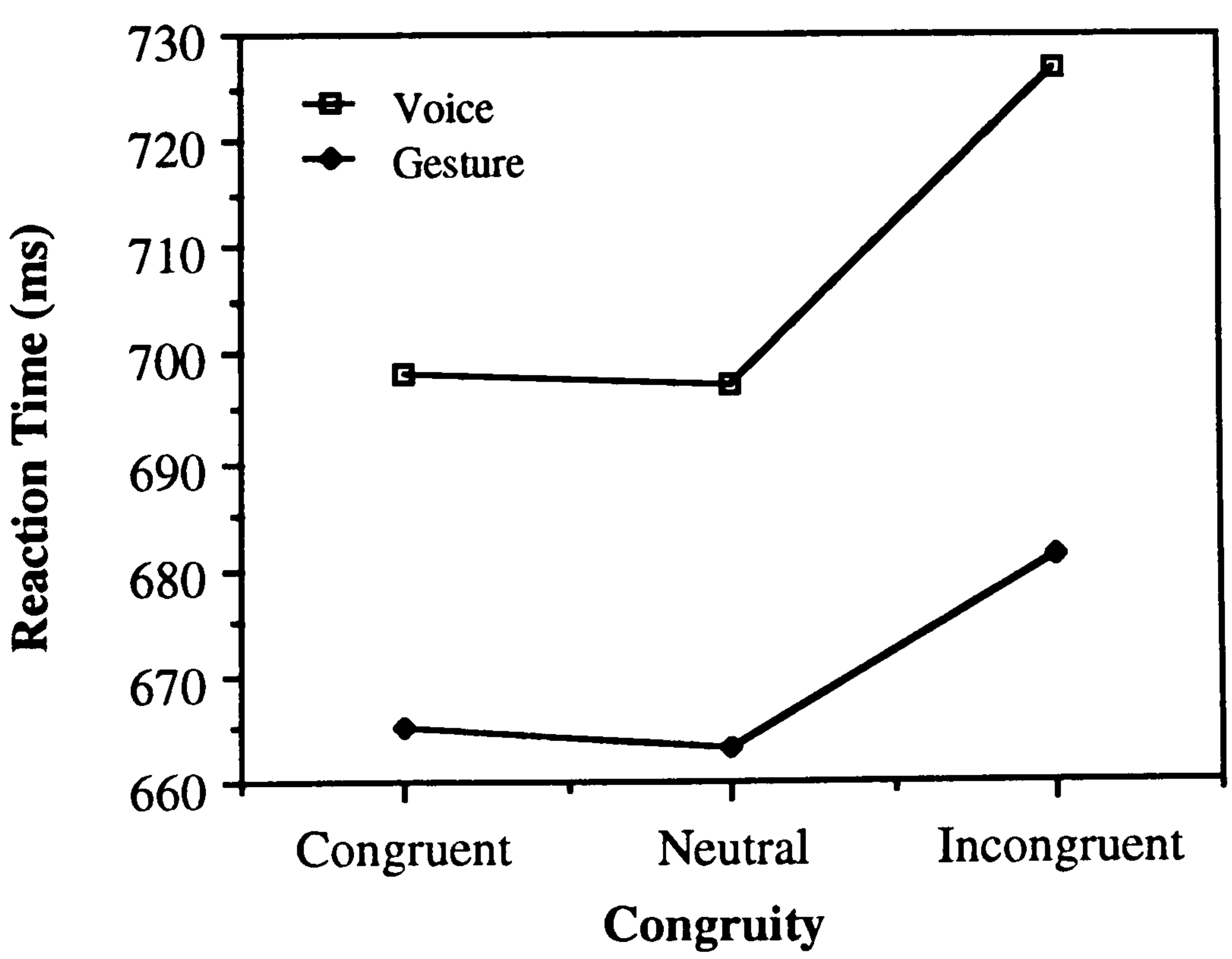


Figure 13. Mean Reaction Times (ms) from Experiment 11 as a function of congruity and response factors (left panel).

As in previous experiments, the error rate was very low (the overall mean percentage error was 0.66%). The correlation between RT's and errors was -0.3 suggesting little evidence of a trade-off between speed and accuracy.

Discussion

In terms of the symmetrical pattern of interference effects, the results of this experiment replicated those of Experiment 10. Once again gesture naming was affected by the presence of an irrelevant voice, and voice naming was influenced by a to-be-ignored gesture despite the fact that there was no direct spatial relationship between gestures and responses. These results are not consistent with the S-R account of the interfering effect of irrelevant gestures. However, in contrast to the pattern of facilitation without inhibition obtained in Experiment 10, the results of Experiment 11 indicate that the symmetrical interference consisted of inhibition without facilitation.

Once again the magnitude of the congruity effect (23 ms) in Experiment 11 was much smaller than the effects obtained in the manual response version of the task (57 ms and 85 ms in Experiments 2 and 3 respectively). However, in contrast to Experiment 10, mean RT's in this experiment (689 ms) were of similar magnitude to those of Experiments 2 and 4 (707 and 666 ms respectively). Therefore the reduction in the size of the congruity effect cannot be attributed to a proportional decrease in RT. Instead, it remains possible that spatial SRC contributes to the interference effects between deictic gestures and verbal information and that in its absence the magnitude of the effect is decreased. However, since the effect persists under these circumstances, spatial SRC is clearly not the only factor involved.

One other possibility that is consistent with an S-R model, is that a response code was indeed activated from an irrelevant gesture, not because of *spatial* SRC, but because of some sort of *symbolic* SRC between the verbal labels associated with the gestures and the verbal responses (e.g. Simon, Sly & Vilapakkam, 1981). That is, some form of compatibility might exist between the directional gestural stimuli and both the keypress *and* verbal responses. This ensures that a response is always encoded from the gestural input regardless of the response modality (keypress or verbal). An interference effect caused by an irrelevant gestural stimulus will, therefore, always result. However this explanation emphasises a potential problem with the SRC account of interference effects. The difficulty is in defining exactly what constitutes compatibility (e.g. faster RT's to compatible versus incompatible stimuli, past experience with the pairings involved etc.). With no *a priori* definition it becomes impossible to generate direct predictions about performance (McClain, 1983) which rather limits the application of SRC to models of interference. In summary we can conclude that the interference effects caused by to-be-ignored gesture stimuli, whilst probably influenced by the spatial relationship between the various stimuli and the responses, are not caused solely by *spatial* SRC. An explanation based on *symbolic* SRC is, however, possible.

Another point of note concerns the observation of inhibition without facilitation from both irrelevant gestural and verbal stimuli. This was the opposite pattern of effects to those obtained in Experiment 10 and more in keeping with Stroop and picture-word tasks where inhibition typically dwarfs facilitation (e.g. Glaser & Glaser, 1982; Glaser & Düngelhoff, 1984). In contrast to the symmetrical nature of the interference effects, this result is consistent with the S-R account. However, accepting this model hinges on the absence of the facilitation effect observed in Experiment 10. Facilitation may not have occurred for a number of reasons. First the experiment may not have been sensitive enough to reveal what would have to be a relatively small facilitation effect (the overall congruity effect was only 23 ms). Secondly, as in Experiment 10, it may be that the choice of neutral items was inappropriate. Another possibility is that both verbal and gestural identification in this experiment were at ceiling level so that no benefit of any congruent items could occur. Conflicting information could still be capable of slowing down target responses, producing an inhibitory effect. The issue of the inhibition and facilitation and the choice of neutral items is discussed in more detail in the general discussion to this chapter.

One final, and rather surprising finding was the absence of a response effect. As in Experiment 10, voice naming was expected to be performed faster than gesture naming, primarily because of the compatibility between the voice stimuli and the spoken responses. However, in this experiment voice naming was, if anything, slightly slower than gesture naming which remained relatively stable across throughout. The reason for the much slower voice naming RT's in the present study is unclear.

In summary Experiment 11 has demonstrated the persistence of an interfering effect of irrelevant gestures despite the removal of any spatial stimulus response compatibility between the gestures and their associated manual responses. This finding is not consistent with the S-R account of the interference effects outlined earlier.

General Discussion

The aim of these experiments was to explore the possibility that the interference effects obtained in the earlier experiments resulted from the close relationship between the directional stimuli and the spatial nature of the manual response. Of particular interest was the interference caused by irrelevant gestures, not least because visual stimuli rarely interfere with responses to target verbal materials. The suggestion was that irrelevant gestures might benefit from some kind of spatial SRC between themselves and the manual response. Under this hypothesis an irrelevant gesture, for example, will be perceptually encoded along with the relevant voice stimuli. However the spatial information encoded from the gesture will only be mapped onto a response by virtue of the compatible spatial S-R relationship. Interference occurs at the response selection stage. Responses to both relevant and irrelevant stimuli are encoded and some unspecified decision process must select the appropriate code in order to program the correct response. When non-corresponding codes are present, a conflict must be resolved. The delay in this resolution process is the source of the interference.

In the present experiments the spatial SRC between stimuli and responses was made incompatible by manipulating firstly the S-R *mapping* (Experiments 5-8) and secondly the *match* between gesture and response variables (Experiments 9-11). According to the model, removing the spatial SRC between an irrelevant item and its associated response should prevent the automatic encoding or activation of that response, leaving the relevant response unopposed. With no conflict to be resolved at the decision stage, no interference should occur. Manipulation of the S-R mapping between left/right gestures and manual responses (Experiment 5) removed the influence of these gestures on the processing of verbal information, a result consistent with the S-R model. The corresponding manipulation of the up/down SRC in Experiment 6, however, failed to influence the symmetric pattern of interference effects. In addition, neither of these experimental manipulations affected the interference effects caused by irrelevant voices on gesture classification. This suggests that voices exert their effects independently of any S-R manipulation. The implication is that the locus of the interference of voices is prior to response selection. Further investigation of the SRC effects of left/right and up/down gestures (Experiments 7 and 8 respectively) indicated that the interference effects caused by irrelevant gestures were eliminated under incompatible SRC conditions. However, in order to avoid the conclusion that gesture and voice interference effects occur at different stages of processing it was suggested that the disruption of the gesture-response mapping resulted in a reduction, as opposed to a complete elimination, of the congruity effect.

The results of Experiments 9-11 bolstered the claim that SRC factors contribute to, but are not the sole cause of the interfering effect of to-be-ignored gestures. In these experiments the relationship between gesture and response variables was made non-matching by asking subjects to make an oral as opposed to a keypress response to the gesture/verbal stimuli. Non-matching S-R variables, by definition are S-R incompatible (O'Leary & Barber, 1993). This manipulation introduced an asymmetry into the interference effects in Experiment 9. In this experiment written words continued to interfere with gesture classification but the effect of irrelevant gestures on word naming responses, although significant, was somewhat smaller. The symmetry of the interference effects were restored in Experiments 10 and 11 which used cross-modal gesture-word stimuli. In both these experiments the overall magnitude of the congruity effects were rather smaller than in the studies reported in Chapter 3, again suggesting that whilst spatial SRC makes a contribution to the interference effects, it is not the only factor involved. These results are not consistent with a model of interference which relies on the automatic activation of an irrelevant response by way of a compatible S-R relationship.

In the remainder of this discussion other response selection accounts of Stroop interference are addressed. It is concluded that, along with SRC accounts, these models fall short of providing a comprehensive account of the interference effects noted in Chapter 3.

Response Selection Accounts

An account based on SRC is not the only possible explanation of Stroop-type interference. At least two other models place the locus of the effect at a late, response selection stage of processing. These are the relative speed-of-processing model and the translational model.

Relative Speed of Processing Models

Speed-of-processing models (e.g. Morton, 1969; Morton & Chambers, 1973; Posner & Snyder, 1975) assume that the two codes in question (e.g. colour and word, or perhaps "gesture" and word) "race" to gain entry to a single channel response buffer (see Figure 14). Priority for entry into this buffer is determined by time of arrival. Interference occurs when an irrelevant response code wins the race to the response buffer and has to be overcome in order for the correct response to be selected. For instance, in the original Stroop task it is assumed that words are processed more quickly than colours, thus the word-based code must be overcome in favour of the colour-based response code. This mismatch in the processing speeds of colours and words is what produces the typical asymmetry of the Stroop effect.

The fact that gesture stimuli were processed faster than verbal stimuli in Experiments 1, 2 and 4 might explain the effects of irrelevant gestures. However, the slower dimension (verbal) was still able to exert effects on the dimension receiving faster processing (gesture) in both these experiments. Moreover, when the relative response times for the critical dimensions were matched (Experiment 2) symmetrical interference was obtained. Thus any model reliant upon assumptions of unequal speeds of processing will not support the findings of the experiments reported here. Indeed, investigations into the speed of processing account of the Stroop effect have not yielded much support. These studies have concentrated on attempting to reverse the pattern of interference by either previewing the slower dimension (Glaser & Glaser, 1982; Glaser & Düngelhoff, 1984), or by manipulating the speed of processing of a dimension by practice or reorientation (Dunbar & MacLeod, 1984; MacLeod & Dunbar, 1988). Neither of these manipulations succeeded in producing the "reverse Stroop effect" with any degree of stability.

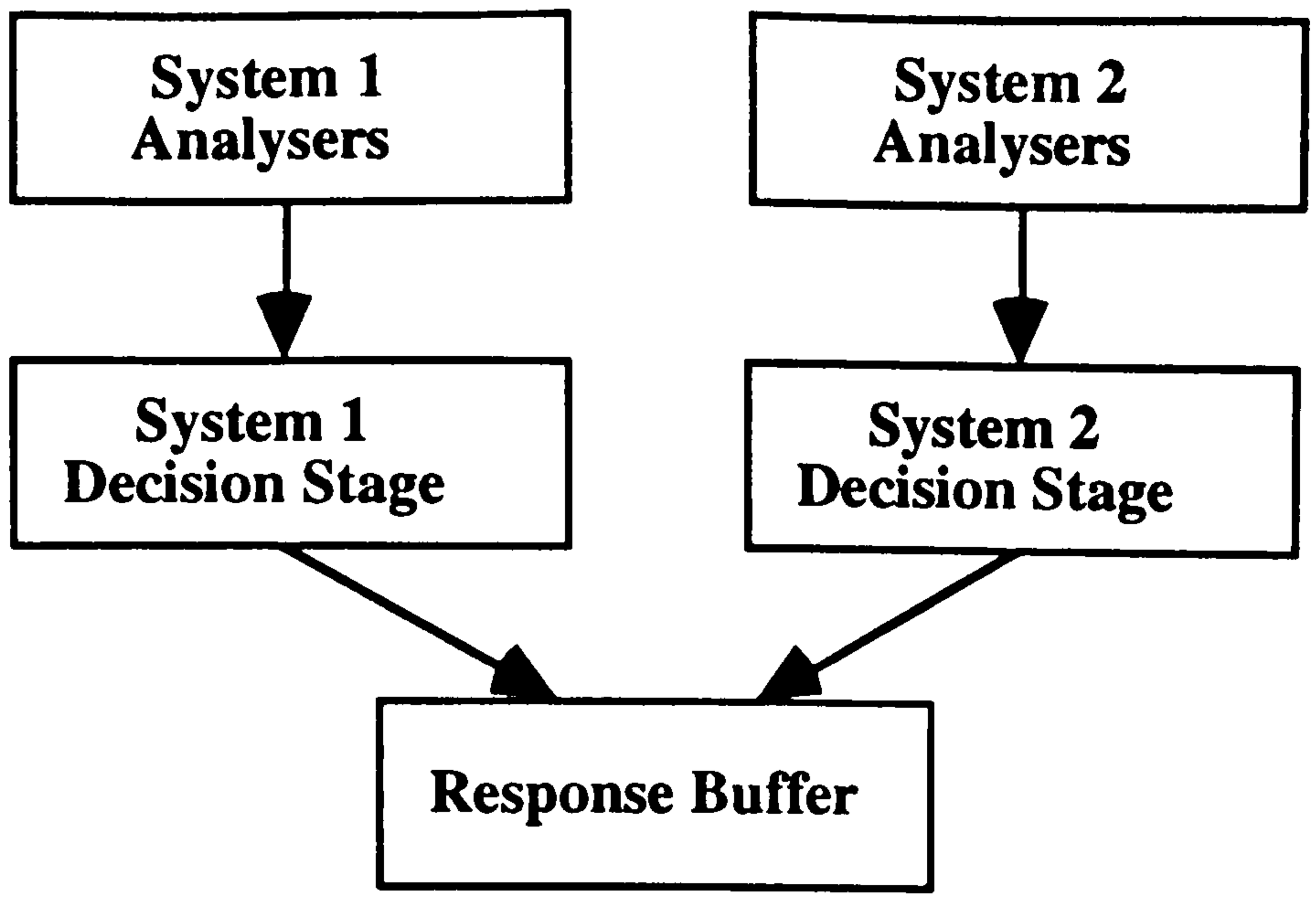


Figure 14. Schematic diagram of the speed of processing model of Stroop interference.

The Translational Model

Virzi & Egeth's (1985) translational model also contains multiple processing systems which operate in specific domains and handle specific types of information e.g. colour, pitch, verbal etc. (see Figure 15). Each system contains an input analyser, decision stage and in some cases a specific output or response stage. Responses are slowed if input to one system requires translation for output in a separate system. In the Stroop effect, for instance, the theory postulates one system devoted to the analysis of verbal input stimuli and verbal (speech) output, along with a second system concerned with colour analysis. If the colour of a stimulus has to be named, the encoded colour information must undergo translation into a verbal code for speech output in the verbal system. Competing codes, one translated and one untranslated, will both be present at the same verbal decision stage. The resolution of this conflict is what causes the interference effect. On the other hand when a translation is not needed, for instance when the task is to read the word of a colour-word stimulus, only one code will be present at each of the separate decision stages. Consequently there is no conflict to resolve and the correct response can be programmed and executed without interference.

In Virzi & Egeth's model of processing, we might assume that the spatial components of the gestural stimuli are extracted by some spatial processing system (see Chapter 5) producing spatial codes suited to the manual responses. When the required response is to the verbal stimuli, the verbal code must be translated into a spatial code for a manual response to be performed. Thus, in terms of the model, two spatial codes would be present at the decision stage of the spatial processing system, resulting in interference. However, the model predicts no such interference when gestures form the target stimuli. In this case no translation is necessary as subjects are required to make a spatial response to an essentially spatial stimulus. Interference was obtained from to-be-ignored verbal stimuli in all four of the experiments reported in this chapter.

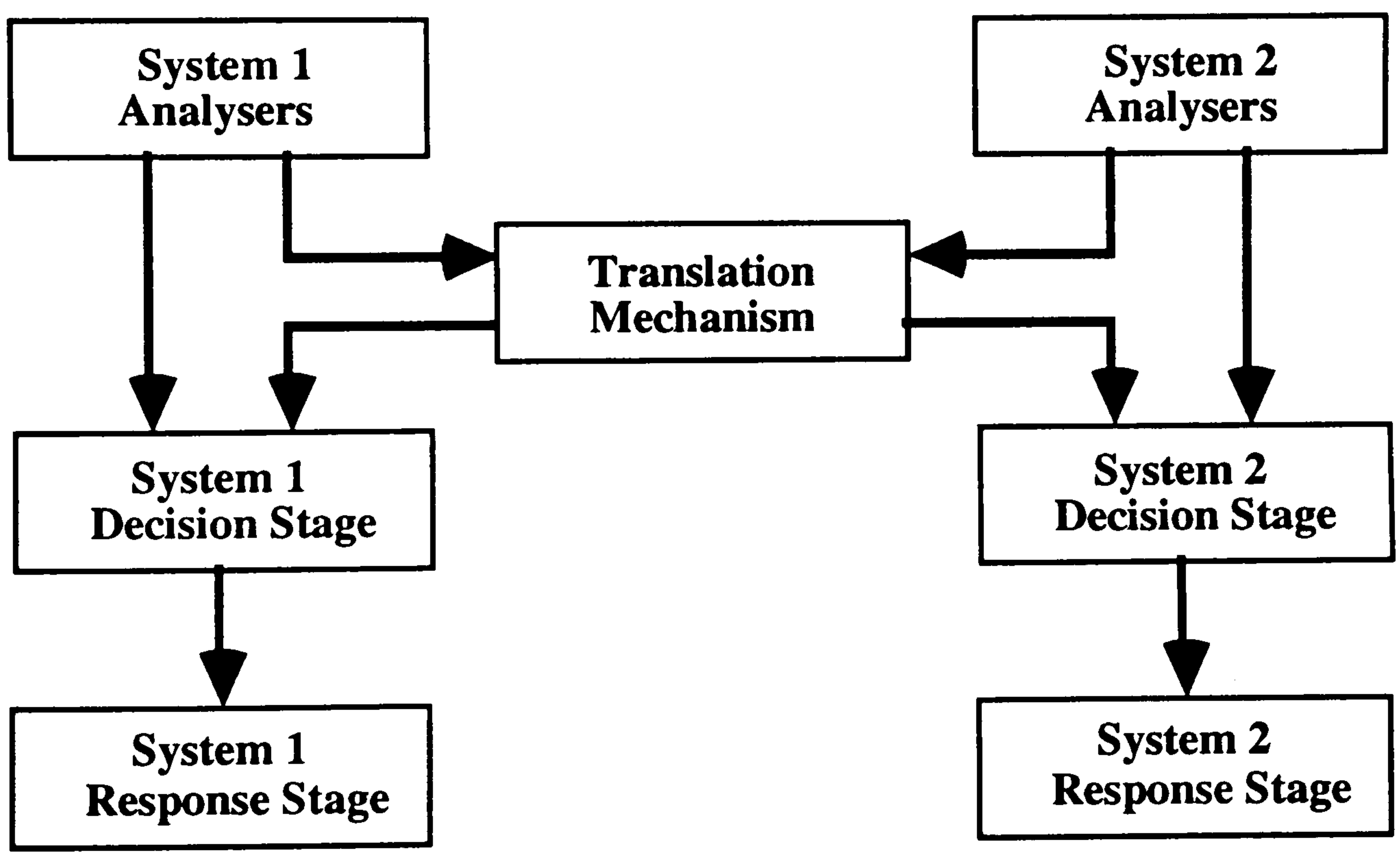


Figure 15. A translational model of Stroop interference adapted from Virzi & Egeth (1985). In the original Stroop task *System 1* might analyse colours whilst *System 2* analyses words.

In short, neither the speed-of-processing nor the translational account can cope with interference effects observed in the present experiments. These models were developed to account for the normal asymmetry of the Stroop effect and, consequently run into difficulty when faced with the symmetrical nature of the interference noted here. The implication is that the locus of the interference effect is prior to response selection, at either a perceptual or semantic level of analysis. However in Chapter 3 it was suggested that the interaction between gestural and verbal information is unlikely to be due to any "crosstalk" at a perceptual stage. The semantic stage of processing is therefore the more likely locus for the effects.

Conceptual Encoding Accounts

A number of authors have suggested that Stroop interference has a semantic locus (e.g. Seymour, 1977; Stirling, 1979). For instance, Seymour (1977) concluded that the Stroop effect is located at a stage between perceptual encoding and response activation (see below) where information contacts semantic memory, a stage he labelled as conceptual encoding. This view has been supported by Simon & Berbaum (1988) and endorsed as a plausible explanation of both the Stroop and Simon effects by Hasbroucq & Guiard (1991). In models of this kind, the separate response determination or decision stages of the speed of processing and translational models are replaced by a single response mechanism (see Figure 16) and interference in the Stroop task occurs prior to this stage. In Seymour's model, for example, Stroop stimuli give rise to two "perceptual" codes, a graphemic code produced by the distractor word and a pictoral code produced by the colour. These perceptual representations access "conceptual" codes in semantic memory via stored descriptions of words and pictures (in the "logogen" and "iconogen" systems respectively). In order to perform the task, a single conceptual code must be retrieved and converted into the appropriate articulatory response code. Problems occur when semantically overlapping codes are simultaneously accessed by an incongruent colour-word stimulus. In this case, some unspecified disambiguation process must select the appropriate code for conversion into a response. Stroop interference is the result of the increased processing necessary to resolve an ambiguity of this kind.

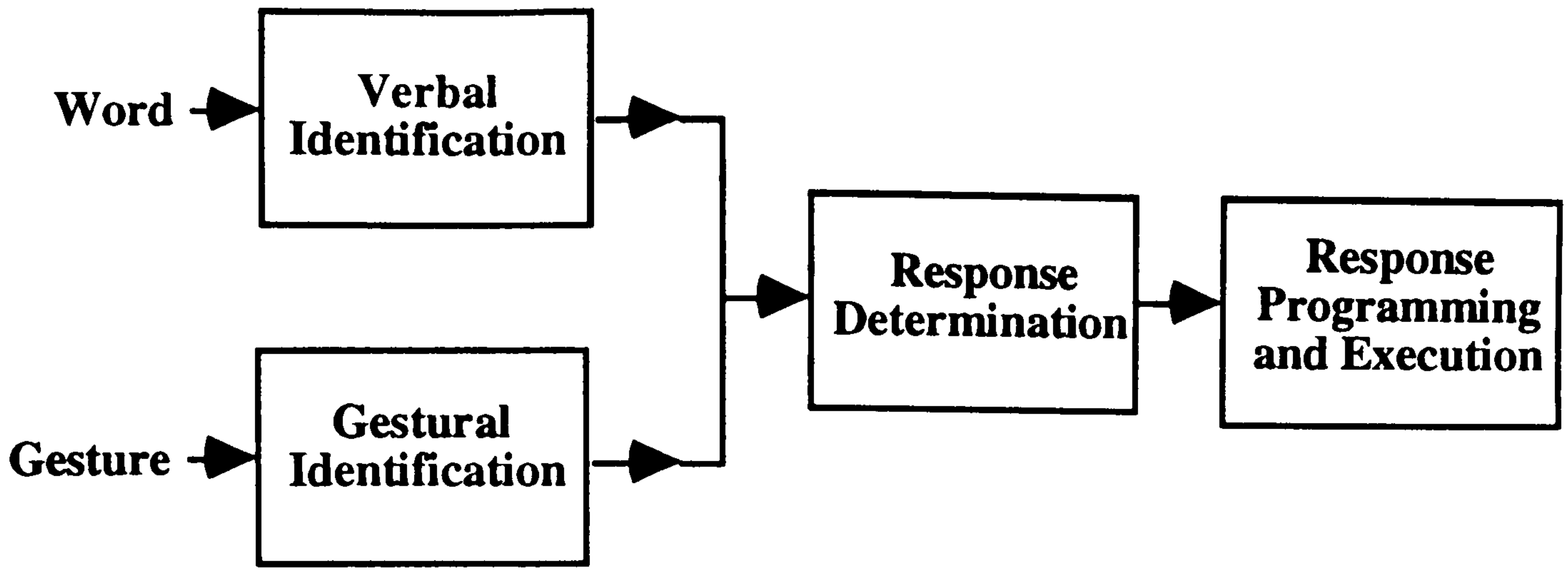


Figure 16. A possible 3-Stage representation of the information processing of gestural and verbal dimensions locating the source of the interference at the response selection stage of processing. The identification stages embrace perceptual encoding, retrieval of structural descriptions and the access of conceptual codes.

Results such as the semantic gradient effects mentioned in Chapter 2, are easily handled by this model by assuming first, that the conceptual codes of closely related items will "overlap" to a greater extent than less closely related stimuli, and second, that the time taken to disambiguate related stimuli is positively related to the degree of conceptual overlap.

Along these lines, a hypothetical account of gesture-word interference might see information from these multiple sources (i.e. words and gestures) processed in parallel. These independent processes contact an amodal semantic memory where meaning can be accessed from multiple sources. Information from the respective modules analysing gesture and verbal input is integrated, interacts or is pooled at this point. A single meaning unit is extracted leading to the selection and programming of the required response. The important point to note is that in this model, as opposed to response selection accounts, information from the separate sources is actually combined in some way in order to determine the response. This is presumably what McNeill et al (1994) have in mind when they explain that "gesture and speech combine into a single system of meaning" (p.235) and that a listener "attempts to form a single idea unit, an integrated combination of speech and gesture" (p.236).

This idea of a central amodal component finds echoes in both the work of McNeill on gesture and speech and Melara & Marks' ideas on Garner Interference and dimensional interaction. Melara & Marks (1990b) write,

> According to one hypothesis, then, codes for attribute labels are partially amodal, that is, lacking complete reference to the sensory modality from which they derive. (p.493)

Similar ideas are contained in McNeill et al's (1994) account of their speech-mismatched gestures experiment (see Chapter 1):

> What framework can we adopt to interpret the effects of speech-gesture mismatches? We take as a clue the evidence that information, once absorbed, is not indexed by the input channel and that subjects, after exposure to a mismatch, attempt to resolve the mismatch in their own retelling. Subjects, in other words, do not try to re-create the input to which they were exposed, but try to form a coherent mental model and introduce changes in memory, where it is necessary to make this possible. (p.234)

Thus it seems more plausible to suggest that the symmetrical relationship between pointing gestures and verbal information arises from a bidirectional crosstalk at a semantic level of analysis. However, this kind of account faces the same problem as response selection accounts, namely in providing a mechanism for symmetrical as well as asymmetrical interference effects. For the purpose of the present discussion it is sufficient to emphasise that the interaction between gesture and speech is likely to be more centrally located than in stimulus-response compatibility, speed of processing or translational accounts of interference effects.

Facilitation, Inhibition and Rimé's Figure-Ground Model

At a finer grained level of analysis, the observation of inhibition without facilitation (Experiment 11) does not seem consistent with the hypothesis that listeners routinely combine information from both gesture and speech in comprehension. If, as suggested, a complete representation of the intended "message" is only obtained from both gestural and verbal information, a congruent gesture, for example, should provide a performance *gain* compared with a neutral gesture. The presence of an incongruent gesture, on the other hand, might result in a detriment in performance over a neutral gesture as the listener integrates inconsistent information. Therefore, if listeners routinely integrate information from both gesture and speech in comprehension, we might expect facilitation from congruent stimuli as well as inhibition from incongruent stimuli. Instead, the current findings appear to be more in keeping with the ideas of Bernard Rimé (e.g. Rimé, 1983; Rimé & Schiaratura, 1991) who has produced a "figure-ground model" of the relationship between verbal and nonverbal materials in comprehension. Rimé suggests that hand gestures provide *redundant* information to speech and as such are largely ignored in the comprehension process. Under the figure-ground model a listener does not integrate information from both channels but, in the interests of processing economy, diverts the lion's share of his or her attentional resources to the speech channel, whilst the speaker's gestures remain in the periphery of the listener's attention. However Rimé adds that this normal relationship between the verbal and non-verbal channels may slip and reverse temporarily so that the non-verbal information becomes the figure and the verbal data the ground. This situation is likely to arise when either the intensity of the non-verbal channel rises e.g. the use of unfamiliar, bizarre or discordant gestures with regard to the context or situation, or the intensity of the verbal channel falls, possibly in noisy, confusing or complex situations. Therefore, under "concordant" conditions attention is allocated to the verbal channel whilst under "gesture neutral" conditions the same will be true. The attention shift only occurs when some discordant or bizarre gesture appears in the non-verbal channel. Thus there should be no processing "gain" for concordant gestures but a detriment in performance with incongruent gestures. This translates to a pattern of performance characterised by inhibition without facilitation as found in Experiment 11

However we should be cautious about making any strong claims on the basis of this pattern of results for at least two reasons. First, the absence of an effect (in this case facilitation) is not sufficient to reject any hypothesised model. As mentioned earlier it may be the case that the neutral conditions were inappropriate, or that the procedure was not sufficiently sensitive to reveal any such facilitative effects. Secondly, and relatedly, Lindsay & Jacoby (1994) have suggested that facilitation and inhibition in the Stroop task cannot be measured accurately using "neutral" control items, instead they suggest that the so-called "process dissociation procedure" (Jacoby, 1991) is more appropriate. In view of these points we should be cautious in making any claims based on the failure to demonstrate a facilitative effect.

Concluding Comments

One of the goals of the present research was to use an interference paradigm to investigate whether gestures are attended in the comprehension of an utterance. Regardless of the locus of interference, the results obtained in the studies reported in Chapter 3 have surely fulfilled this aim. However, the theoretical motivation driving this research is that gestures are in fact combined, or integrated with information in the verbal channel in order to facilitate the listener's understanding of a speaker's intended meaning. However, an account based on spatial SRC, which places the locus of the interfering effects of gestures at response selection, suggests that these gestures only influence the classification of verbal material as a rather artifactual consequence of the experimental task. It does not imply the integration of the two sources of information. This is because models locating the interference at the response stage side-step the integration issue. The respective codes remain separate throughout perceptual encoding, identification, decision and response programming stages. Thus, it is argued, information from the two sources is kept separate and is never actually combined. In this chapter I have argued and produced evidence to suggest that although stimulus-response compatibility plays a role in the interfering effect of irrelevant gestures, it falls some way short of providing a complete account. It has also been shown that other accounts which place the locus of the effects at response selection are similarly deficient, particularly with regard to the explanation of the symmetry of the effects. A semantic locus of the effects seems to be the likely default and one which best characterises the idea that gesture and speech interact in comprehension.

Overview

The aim of this chapter was to investigate the contribution of stimulus-response compatibility (SRC) factors to the symmetrical interference effects reported in Experiments 1-4 and more generally to examine accounts which place the locus of the interference effect at the response selection stage of processing. Reducing the SRC by manipulating the mapping between the directional stimuli and the spatial responses had the effect of eliminating or reducing the interfering effect of irrelevant gestures, suggesting a role for SRC factors. When the compatibility between gesture and response variables was eliminated by asking subjects to make an oral naming response, the original symmetry of the effects persisted although slightly attenuated. Taken together these findings suggest that SRC contributes to, but is not an essential component of, the influence of irrelevant gestures on voice classification. Furthermore, other response selection accounts of Stroop interference were discussed in relation to the present findings and were found to be unsuccessful in accounting for the symmetrical nature of the effects. It was concluded that the gesture-speech interaction is likely to be located at a semantic level of analysis.

Chapter Five
Are Pointing Gestures Special?

So far we have seen that subjects seem to process both the gestural and verbal dimensions of a gesture-word stimulus pair and that this interaction is unlikely to occur at either the perceptual or response stages of processing. In this chapter we ask whether or not these gestures enjoy a peculiarly linguistic or "special" status. For instance it is unclear whether the interference effects caused by the presence of incongruent gestures reflects the operation of a system processing *gestures* per se or one concerned with the manipulation of *visuo-spatial images*. Following a brief discussion of the linguistic status accorded to gestures, two experiments are reported which address this issue. In these Experiments both arrows and spatially positioned dots produce similar patterns of interference effects, although rather smaller in magnitude, to those found with pointing gestures in Experiments 1-4. These findings raise the possibility that it is the spatial nature of the pointing gestures, rather than their status as gestures per se, which gives rise to their influence on decisions to verbal material.

Gesture as Language

Parallels are frequently drawn between spoken languages and gestures as communication systems. Both could be said to embrace forms of arbitrarily derived symbolic tokens which are commonly understood by speakers of a particular language or culture. Some have even suggested that gestures have principles of organisation not unlike the "syntax" of spoken language. Variously the structure of speech, the control of the turn-taking mechanism and the ritualised forms of initiating and terminating social interactions have been suggested as providing a syntactic framework, or set of "rules" governing the organisation of gestural sequences (see Feyereisen & De Lannoy, 1991). Furthermore the study of "kinesics" (e.g. Birdwhistle, 1970) demonstrates how the hierarchical structure of linguistic forms (e.g. phonemes and morphemes) can be applied to gestural elements (e.g. kinemes and kinomorphemes). Does the linguistic status accorded to gestures suggest that they receive some form of linguistic processing? The presentation of a gesture may well result in a kind of lexical access. Indeed some have suggested the existence of a lexicon of symbolic gestures, a set of correspondence rules coded in long-term memory linking signals (gestures) to meanings (e.g. Ricci-Bitti & Poggi, 1991). If such a gestural input lexicon exists, it should be no surprise to discover the existence of a separate functional "module" dedicated to the processing of gestural material similar to the systems postulated for the analysis of spoken and written words, objects, facial identity, facial expression etc. Indeed there is some evidence for the specificity of gestural processing. Rothi et al.'s (1986) "gestural agnosic" patient, described in Chapter 2, could not comprehend or discriminate gestures but had normal auditory comprehension and no evidence of any object agnosia.

If gestures are processed by a specialised module it is possible that they are encoded somewhat differently from pictures or other visual information. Pictures, on the whole, are not intentional acts of communication, they do not form part of a large repertoire of communicative acts as has been suggested of gestures. Perhaps it is this peculiarly linguistic nature of the gesture which leads to its mutual interaction with verbal information in the comprehension process. If this is the case, one might expect no interaction between a spoken word and a non-linguistic, non-verbal symbol carrying the same informational content as a gesture. This hypothesis is examined in the two experiments which follow. The directional gestures of the previous experiments are replaced with arrows in Experiment 12 and simple "dots", positioned at equivalent spatial locations on the computer screen in Experiment 13. Arrows would seem to be universally recognised as symbolic of spatial direction and for this reason form adequate non-verbal, non-linguistic replacements of pointing gestures. The "dots" used in Experiment 13 are also obviously non-verbal and non-linguistic but do not form conventional symbols of direction.

Experiment 12

In Experiment 12 we ask whether gestures provide anything different from a non-linguistic, nonverbal spatial cue to direction such as an arrow. If different results are obtained with arrows and words as compared with gestures and words, there may well be good reason to follow up the possibility that certain non-verbal behaviours such as pointing gestures receive some specialised processing. On the other hand, the observation of a mutual interaction of arrows and words suggests that subjects may well be extracting spatial rather than linguistic information from the gestures used in the earlier experiments.

Shor (1971) obtained symmetrical interference effects with directional words embedded in arrows. That is, the direction of the arrow interfered with reading of the word, and word reading disrupted responses to the direction of the arrow. These effects were replicated in a study using a manual response. This result suggests that a symmetrical pattern of effects might well be found with a cross-modal presentation of words and arrows.

Method

Subjects: Twelve subjects participated in this experiment. All had normal or corrected to normal vision and normal hearing.

Materials: The pointing gestures used in the previous experiments were replaced by arrows pointing left, right , up and down. The arrows were black, presented on a white background. The non-pointed end of the arrow was positioned in the centre of the screen, thus a "left" arrow extended from the centre to the left of the screen, subtending a visual angle of some 2.9°. The spoken auditory stimuli were as in the previous experiments.

Design and Procedure: These were identical to Experiment 2 with the exception that subjects were now asked to respond to the direction of the arrow in one block and the meaning of the voiced word in a second block. Half of the subjects responded to the voice first, and half to the arrow. The order of presentation within the blocks was completely randomised.

Results

The mean reaction time scores and percentage of errors recorded in each condition of the experiment are presented in Table 12a. Congruity effects are apparent for both voice and arrow responses whilst overall RT's to voice stimuli are somewhat slower than for responses to arrows.

Table 12a.
Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Arrows in Congruent and Incongruent Conditions of Experiment 12.*

	Up/Down Decisions		Left/Right Decisions		Overall Mean	
Congruity	*M* RT	%E	*M* RT	%E	*M* RT	%E
			Responses to Voice			
Cong	618	4.17	644	0.83	631	2.50
Incong	635	2.50	661	3.33	648	2.92
Mean	627	3.34	653	2.08	640	2.71
			Responses to Arrow			
Cong	513	1.67	476	1.67	495	1.67
Incong	571	3.33	496	3.33	533	3.33
Mean	542	2.50	486	2.50	514	2.50

Consistent with the above observations, a 2 (response) x 2 (congruity) x 2 (decision) ANOVA conducted on the reaction time data yielded a 28 ms main effect of congruity ($F(1,11)=10.47$, $p<0.01$), a main effect of response ($F(1,11)=23.03$, $p<0.01$) and no interaction between response and congruity ($F<1$). The analysis also yielded a significant interaction between response and decision factors ($F(1,11)=19.29$, $p<0.01$). Again the cause of this interaction appeared to be that left/right decisions were slowed by 167 ms ($p<0.001$) whereas up/down decisions were slowed by 85 ms (p<0.01) when the relevant response was changed from arrow to voice. Left/right decisions were also made 56 ms faster than up/down decisions to arrow stimuli (p<0.01). The response by decision interaction and the effects of congruity are illustrated in Figure 17.

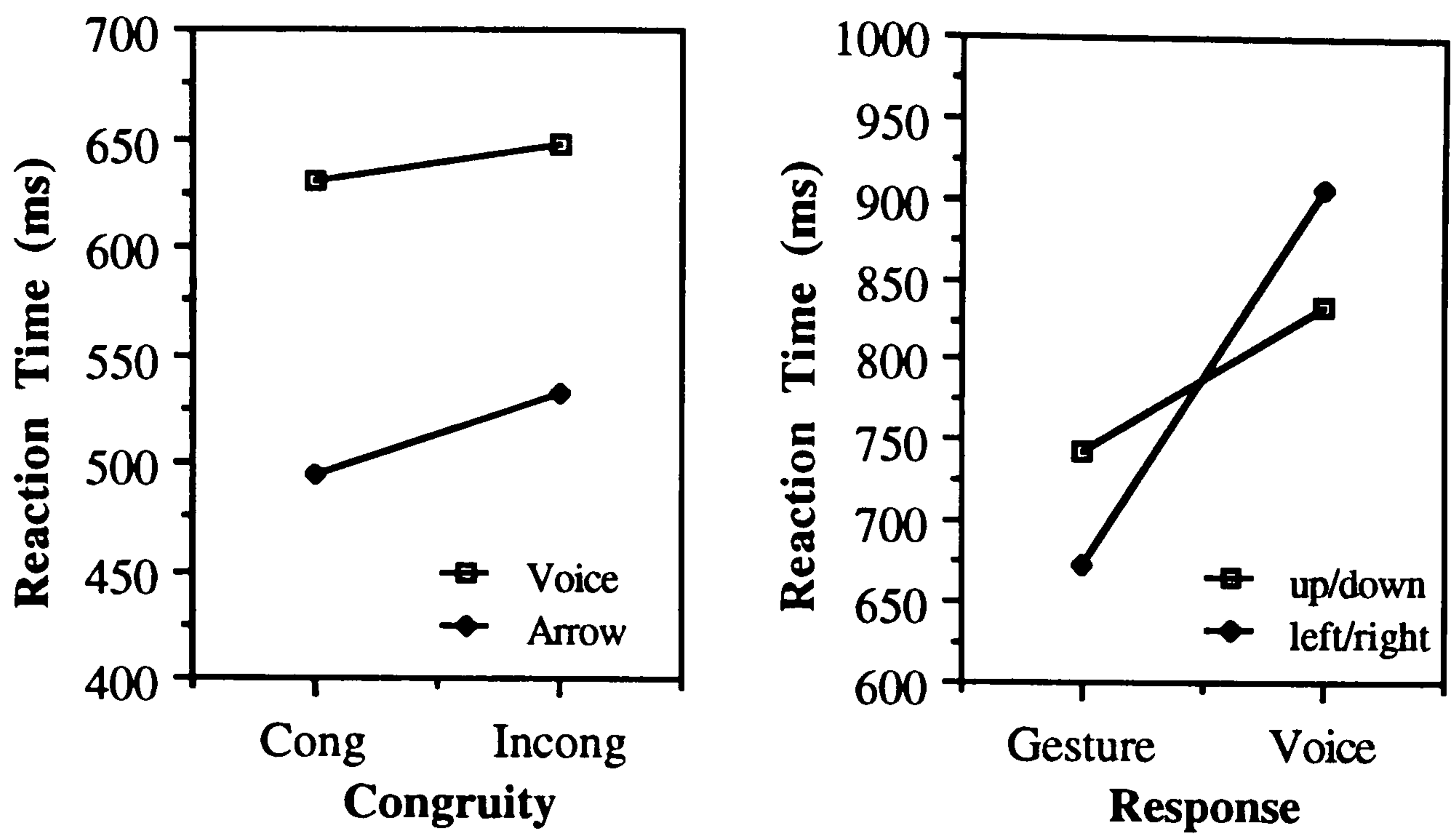


Figure 17. Mean Reaction Times (ms) for responses to voice and arrow stimuli as a function of congruity (left panel); and for up/down and left/right decisions as a function of response (right panel).

Table 12b summarises the magnitude of the interference effect for both response tasks as a function of the order of response task presentation. For those subjects who completed the voice block first, there appears to be a consistent interference effect. Notably, to-be-ignored arrows caused a 19 ms effect on the first block of experimental trials. The influence of irrelevant arrow stimuli is clearly not restricted to situations where subjects have first learnt to respond on the basis of the arrow stimuli and then are asked to switch tasks.

The overall error rate was reasonably low (2.61%). The correlation between RT's and error scores was 0.16 which was not suggestive of a trade-off between speed and accuracy.

Table 12b.

Magnitude of the Interference Effect (in milliseconds) as a Function of Order of Response Task..

Order	Response Task: Responses to Voice	Responses to Arrow
Voice First	+19	+14
Arrow First	+14	+64

Discussion

The results of this experiment again indicate a symmetrical pattern of interference effects. Responses to voice stimuli were slowed in the presence of an incongruent arrow relative to a congruent arrow, whilst a congruity effect of similar magnitude was evident for arrow classification.

The pattern of reaction times to the various decision and response dimensions also closely paralleled the earlier experiments. The left/right fingering advantage resulted in faster RT's to left/right, compared to up/down decisions when the arrow was the relevant dimension. RT's were slowed when the relevant response was switched from arrow to voice because of the temporal extent of the verbal stimuli. This effect was more marked for left/right as opposed to up/down decisions because of the ambiguity of the horizontal dimension when represented verbally.

On the face of it these observations attest to the similarity of the processing of the gestural and arrow stimuli. This may well suggest that the gestures used in Experiments 1-4 were processed by a system concerned with the manipulation of visuo-spatial material, rather than one processing gestures per se. However, the overall 28 ms interference effect observed in this experiment is rather smaller than the congruity effects obtained in Experiments 1-4 (105, 56, 75 and 85 ms respectively). Indeed a between-experiments ANOVA comparing the congruity effects of Experiments 2 and 12 yielded a near-significant interaction between experiment and congruity ($F(1, 24)=3.46$, $p=0.075$). In particular the arrows used in this experiment caused only a 17 ms interference effect on responses to voice stimuli compared with the 73 ms effect of irrelevant gestures in Experiment 2. Arrows appear to cause much less interference than gestures. There may be several reasons for this. First, the overall reaction times recorded here (the overall mean RT was 577 ms) were somewhat faster than in the previous experiments. The reduction in the magnitude of the interference effect might simply be a result of these faster responses. Secondly, the relative sizes of the arrows and gestures differ. The arrows of Experiment 12 subtended a visual angle of 2.9^{o} compared with the 13^{o} subtended by the gestures. The greater eccentricity enjoyed by the gestures may contribute to their relatively larger influence on voice responses.

The arrangement of the arrow stimuli ensured that the informative part of the arrow (the arrowhead) always appeared either above or below, or to the left or right of the centre of the screen. Moreover, the information carried by the arrowhead was always congruent with its spatial location. There is, therefore, a confound between the spatial location of the arrowhead and the symbolic information that it carries. As a result we cannot be sure whether the interfering effects of these arrows were caused by the processing of their symbolic content, or by the spatial encoding of the arrowhead's position. Indeed subjects may well encode the spatial location of a gesturer's hand in processing the pointing gestures used as stimuli in Experiments 1-4 and that this, rather than the gesture's meaning, is what causes their interfering effect on directional words. Thus it becomes important to tease apart the contribution of symbolic versus the purely spatial components of the effects. This issue pursued in the next study.

Experiment 13

Can a spatially located visual stimulus with no intrinsic symbolic meaning exert interfering effects on concurrently presented spoken directional words? In order to address this question the arrows used in the previous experiment were replaced by large dots which appeared at different locations on the screen. These dots have no intrinsic symbolic content so that any interference effects will presumably be caused by the encoding of the spatial location of these stimuli.

Method

Subjects: Twelve subjects participated in this experiment. All had normal or corrected to normal vision and normal hearing.

Materials: The gestures used in the previous experiments were replaced by large black dots which appeared on the left, right, top or bottom of the screen. These dots measured approximately 3 cm in diameter. The centres of the "left/right" dots subtended a visual angle of 8.1° and the "up/down" dots an angle of 5.7° from fixation in the centre of the screen. The spoken auditory stimuli were as in the previous experiments.

Design and Procedure: Again these were identical to Experiment 2 with the exception that in the visual response blocks subjects were asked to respond to the position of the dot which might appear on the left, right, top or bottom of the screen. The instructions to the subjects in the voice response blocks were as before. Again a manual keypress response was required with the response keys having the same spatial layout as in Experiment 1 (i.e. a compatible S-R arrangement).

Results

Data from one subject was excluded from further analysis as it was clear that this individual had misunderstood the task instructions and consequently made large numbers of errors in one experimental block. A summary of the mean reaction times and percentage of errors from the remaining eleven subjects appears in Table 13a. Again there are large and consistent interference effects for both responses to voice (59 ms) and dot (70 ms) stimuli. Overall responses to the dot stimuli were faster than those to the spoken words.

A 2 (response) x 2 (congruity) x 2 (decision) ANOVA confirmed the above observations, revealing main effects of response ($F(1,10)=12.76$, $p<0.01$) and congruity ($F(1,10)=13.31$, $p<0.01$) as well as significant interaction between response and decision ($F(1,10)=8.79$, $p<0.01$). No other effects reached significance. Further analysis of the interaction between response and decision qualified the main effect of response and also indicated that left/right decisions were made significantly slower than up/down decisions when responding to voice ($F(1,20)=5.97$, $p<0.05$) but that no such difference existed when responses were made to dot stimuli. This interaction together with the effects of congruity are illustrated in Figure 18.

Table 13a.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent and Incongruent Conditions of Experiment 13.*

	Up/Down Decisions		Left/Right Decisions		Overall Mean	
Congruity	*M* RT	%E	*M* RT	%E	*M* RT	%E
			Responses to Voice			
Cong	670	7.27	696	4.55	683	5.91
Incong	697	8.18	787	9.09	742	8.64
Mean	684	7.73	742	6.82	713	7.28
			Responses to Dot			
Cong	565	1.82	546	3.64	556	2.73
Incong	640	10.91	611	13.64	626	12.28
Mean	603	6.37	579	8.64	591	7.51

The interference caused by the dot stimuli is consistent across both orders of response task (see Table 13b). Error scores in Table 13b mirror the RT data. The overall error rate was 7.39%. The correlation between RT's and error scores was 0.33 which was not suggestive of a trade-off between speed and accuracy.

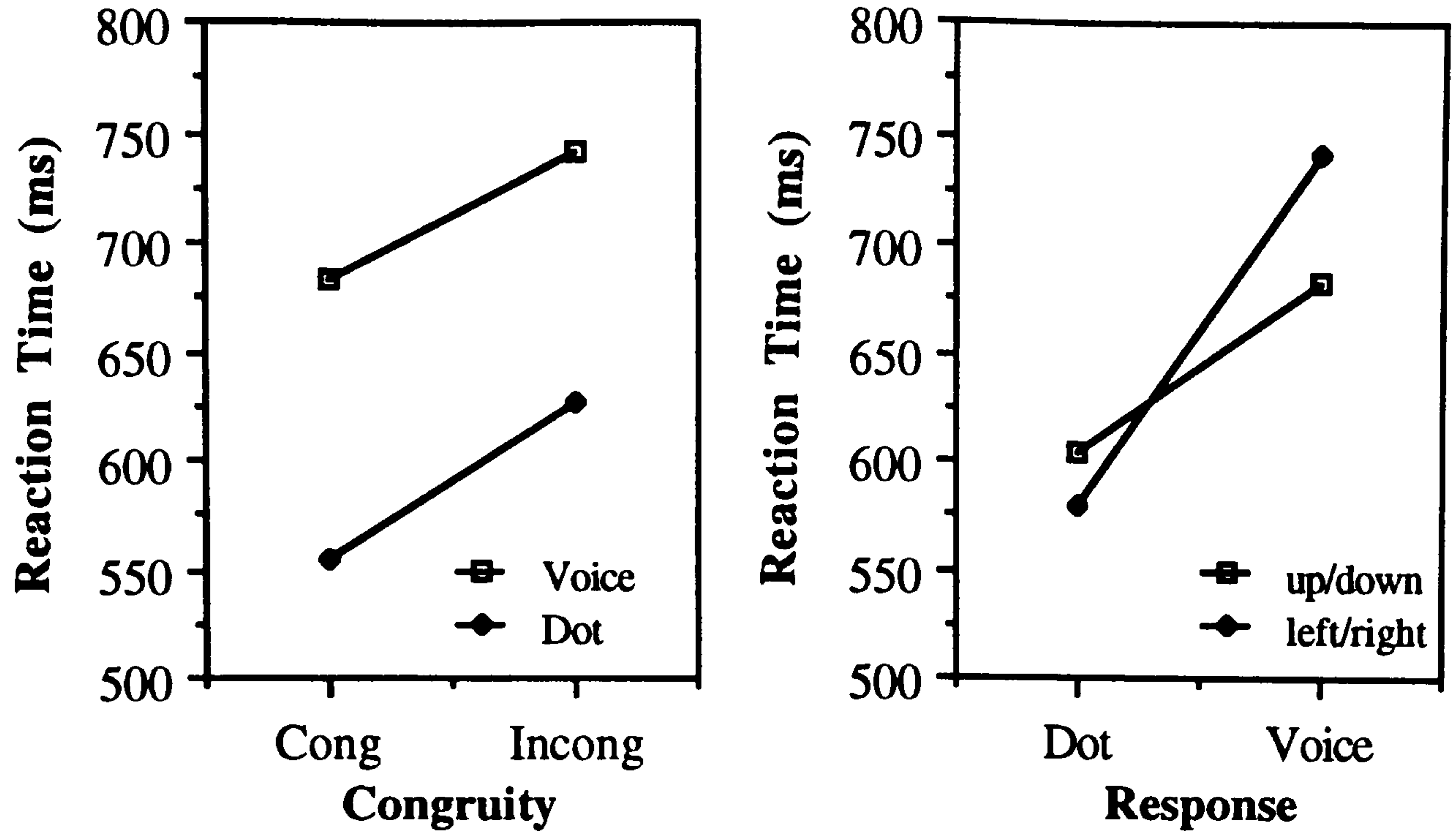


Figure 18. Mean Reaction Times (ms) from Experiment 13 as a function of congruity and response factors (left panel), and response and decision factors (right panel).

Table 13b.

Magnitude of the Interference Effect (in milliseconds) as a Function of Order of Response Task.

	Response Task	
Order	Responses to Voice	Responses to Dot
Voice First	+65	+65
Dot First	+84	+77

Discussion

The results of this experiment are rather similar to those of Experiment 12. The non-interactive effects of response and congruity indicate a symmetrical pattern of interference effects. Responses to voice stimuli were slowed by the presence of an incongruently positioned dot compared to a congruently located dot. Moreover, the magnitude of this effect (59 ms) was substantially larger than interfering effects of to-be-ignored arrows in Experiment 12 (17 ms). Reciprocally, RT's to dot stimuli were similarly affected by to-be-ignored congruent and incongruent voices. As in

Experiment 12 and Experiments 1, 2 and 4 (Chapter 3), left/right decisions were slowed to a greater extent than up/down decisions when the task was switched from responding to dot stimuli to responding to voice stimuli.

The results of this experiment clearly indicate that the location of a visual stimulus with no symbolic value is sufficient to evoke reliable interference effects on responses to verbal material. This result, coupled with the overall similarity between this and earlier experiments, raises the possibility that the gestures used in Experiments 1-4, and indeed the arrows used in Experiment 12, were processed spatially rather than symbolically. Subjects might encode the spatial location of the gesturer's hand yielding a spatial code which, in incongruent trials, conflicts with the spatial content of the directional word.

General Discussion

The experiments reported in this chapter aimed to explore the idea that the pointing gestures used as stimuli in earlier experiments were processed by some general visuo-spatial system, rather than a processor devoted to the analysis of gestural forms per se. Arrows containing the same symbolic information as the gestures, but lacking their linguistic status, interfered with concurrently presented verbal information and vice-versa (Experiment 12). This result implies that pointing gestures may well be processed in a similar way to other visual stimuli with spatial meaning. However, visual stimuli lacking any symbolic content (dots) also produced an interfering effect on responses to spoken directional words (Experiment 13) raising the possibility that subjects extract spatial, rather than semantic information from pointing gestures.

In this section the mechanisms are discussed which might be responsible for the interfering effects of both arrows and dots. Some of these ideas derive from research into both the Simon Effect (see Chapter 2) and the mechanisms of orienting visual attention. Finally, the discussion turns to the implications for the processing of the pointing gestures used in earlier experiments.

The Simon Effect

The results of the experiments reported in here and in Chapter 3 have much in common with the Simon effect. Recall that in the Simon paradigm, the location of the target stimulus (e.g. the word "left" printed on either the left or right of the screen) provides an irrelevant cue which interferes with a response contingent on the identity of that target. Thus in both the Simon effect and the experiments reported here it is the spatial position of a stimulus which apparently causes the interference effects. However, in the Simon effect it is the location of the *relevant* stimulus which causes the interference. In the present experiments the relevant verbal stimulus and the irrelevant spatial stimulus are entirely separate and it is the location of the *irrelevant* information which causes the effects.

An account of the Simon effect proposed by Umiltà and colleagues (e.g. Umiltà & Nicoletti, 1992; Nicoletti & Umiltà, 1994) provides a useful starting point in developing a theory which explains how the location of an irrelevant stimulus could produce an interfering effect on responses to a spatially distinct target. Their model comprises a synthesis of the coding and attentional hypotheses, the two main classes of explanation of the Simon effect. Briefly, the coding hypothesis (e.g. Wallace, 1971; Kornblum et al., 1990; Hasbroucq & Guiard, 1991, O'Leary & Barber, 1993) suggests that interference occurs between the various aspects of the stimulus which have all been encoded. Specifically the spatial dimension of the stimulus is encoded even though it is not relevant to the task. The Simon effect arises when the relevant stimulus code and the irrelevant spatial code differ and the conflict has to be resolved. As with the Stroop effect and dimensional interactions, there is a debate concerning the locus of the effect. Some argue that this interference occurs at response selection (e.g. Wallace, 1971; Kornblum et al., 1990) whilst others maintain that it is the degree of correspondence between the encoded stimuli, rather than response representations, which causes the effect (e.g. Hasbroucq & Guiard, 1991). However as Nicoletti & Umiltà (1994) argue, the main problem with the coding hypothesis is in providing a convincing explanation for why the irrelevant spatial location of the stimulus is actually encoded, never mind the actual locus of the effect. In order to provide such an explanation, Umiltà & Nicoletti draw on the attentional hypothesis.

Studies have suggested that if attention is directed to a particular spatial location, the ability to detect (Posner, 1980) or discriminate (Erikson & Collins, 1969) various stimuli presented at that location is improved. Thus, in the Simon effect subjects presumably direct their attention to the targets in an attempt to facilitate their identification. The attentional account suggests that when a stimulus is presented, spatial attention is shifted to the position of that stimulus, facilitating any response at that location (e.g. Simon, 1969; Verfaellie, Bowers & Heilman, 1988). For example a dot, arrow or gesture on the left hand side of the screen might cause a shift in visual attention to this location, "priming" a response on the left hand side. This tendency must be suppressed if the required response is a right hand keypress to the word "right". Rather than "priming" responses, Nicoletti & Umiltà (1994) considered that in order to accomplish the required shift of visual attention to the target's location, the spatial location of the target stimulus must first be encoded. According to the theory, this spatial code then causes a Stroop-like interference, as in the coding hypotheses, at either response selection, or an earlier stage of processing.

The account described above explains how an attentional shift to the target's location can cause an interfering effect in making a spatial response contingent on the target's identity. However, in the experiments reported above the target words are always centrally located, whether presented visually or auditorily. Thus if the interfering effects of the to-be-ignored visual stimuli are caused by any attentional orienting, we must have an account of how these stimuli are able to cause such a shift.

Orienting of Visual Attention

A number of authors (e.g. Jonides, 1981; Müller & Rabbitt, 1989) have suggested that shifts of attention can be guided by two mechanisms, one "top-down" or endogenous process under voluntary or strategic control and another "bottom-up" or exogenous mechanism which is stimulus-driven and more reflexive in nature. Exogenous orienting is considered to take place automatically and is caused by the appearance of certain salient visual (or auditory) stimuli presented in the periphery of the visual field. Under these circumstances attention is "pulled" to the location of the salient event. Endogenous orienting might be cued by the appearance of a central cue such as an arrow which might motivate a voluntary shift in attention from one part of the visual field to another. In this case, attention is "pushed" in the direction indicated by the cue. Jonides (1981) showed that subjects were able to suppress this endogenous mechanism when instructed to ignore the central cue. In contrast, similar instructions to ignore a peripheral cue did not prevent an exogenous shift to the cue's location.

These two mechanisms could explain why attention might be shifted by stimuli which are not relevant to the demands of the task. The dot stimuli presented in Experiment 13 may cause exogenous shifts of visual attention, producing a spatial code as described above. In incongruent trials this code will conflict with both the spatial meaning of the target word, and the spatial code created to allow the programming and execution of the correct response. In contrast, arrows might be expected to cue an endogenous shift in visual attention which subjects should be able to suppress. However, these stimuli persisted in causing a small, but reliable interference effect on responses to voice stimuli in Experiment 12. It may be that the spatial demands of the experimental task make it more difficult to suppress an endogenous shift in the direction of the arrow. On the other hand it has already been noted that the arrows used in Experiment 12 were not actually centrally located.

These stimuli may therefore behave more like peripheral than central cues, producing a spatial code on the basis of an automatic shift of attention.

Whilst this attentional account provides an appealing explanation for the spatial encoding of the irrelevant dot stimuli of Experiment 13 there are nonetheless a number of problems with its application to the present findings. First, in our paradigm we have no way of establishing whether any shifts of attention actually occurred on the appearance of a visual stimulus. In the location precuing paradigm which is frequently used in the study of covert orienting, subjects are first presented with a cueing stimulus such as a peripheral flash or a central arrow indicating the likely location of the target. The typical finding is that targets presented at these cued locations are detected more efficiently than those presented at uncued locations. This is taken as evidence that covert attention has been shifted to the cued location. In order to make a similar claim concerning the dots or arrows used in Experiments 12 and 13, we would need to demonstrate some facilitation of performance at the location indicated by the dot or arrow. Simply observing an interference effect is not sufficient to make such a claim. The second problem is that in covert attention studies, subjects are instructed to maintain their fixation at a central point. This ensures that any attentional shifts are "covert", i.e. not accompanied by overt eye or head movements. In the present studies, eye movements of the subjects were not controlled.

Secondly, whilst this account can provide an appealing explanation of the Simon effect and of the effects of irrelevant spatial stimuli, it is not able to account for the symmetry of the present findings. Like the Stroop effect, the Simon effect is typically asymmetrical. Responses to a relevant word are slowed if its spatial location is at odds with its meaning, but the meaning of a word does not affect responses to its location. In contrast, both auditory and written word stimuli interfere with responses to spatial gestures, arrows and dot stimuli in the experiments reported here.

Implications for the Processing of Spatial Gestures

The implications of the findings of Experiments 12 and 13 for the processing of spatial gestures are somewhat unclear. The fact that arrows and gestures both exert interfering effects on voice responses might suggest that pointing gestures are processed by mechanism more suited to the analysis of visuo-spatial information than to the processing of gestural material per se. We have also seen how the processing of this type of symbolic information might produce an endogenous shift of attention which, according to Umiltà and his colleagues, results in the formation of a potentially interfering spatial code. On the other hand, the interfering effects of non-symbolic "dot" stimuli suggests that it might be the spatial location of an irrelevant stimulus (e.g. the position of a gesturer's hand), rather than the nature of the stimuli itself, which is the source of the effect . We have also seen that the sudden appearance of such a stimulus in the periphery of visual field is capable of producing an exogenous or stimulus-driven shift of attention. Again, it is plausible that such a shift could yield an interfering spatial code.

The spatial processing of pointing gestures has indeed been implied in a number of studies by Mary Smyth and her colleagues (e.g. Smyth, Pearson & Pendleton, 1988; Smyth & Pendleton, 1989) within the field of working memory (e.g. Baddeley, 1986). Smyth & Pendleton (1989) measured subjects' memory spans for sequences of hand configurations and for a series of hand movements to spatial locations. These spans were measured normally or in conditions where a suppression task was attempted by the subjects during the presentation of the movement sequences. The spatial suppression task (tapping a repeated series of spatial targets) interfered with span for spatial locations but did not affect span for movement configuration. The opposite dissociation was observed in the movement suppression task (repeated squeezing of a tube) which affected memory span for movement pattern but did not interfere with span for spatial locations. Smyth & Pendleton interpreted these results as suggesting that the processing of movement configurations involves a different subsystem of working memory than does movement directed to spatial targets where a location in external space must be specified and maintained. This latter type of movement might be implicated in the functioning of a system designed to accommodate more general visuo-spatial information, such as the visuo-spatial scratch pad (e.g. Baddeley, 1986) or perhaps a more specific *spatial* component such as Logie's (1995) "inner scribe". The spatial information contained in pointing gestures, arrows and dots might all be maintained and processed by this type of system producing similar patterns of effects. Other types of movements may well gain access to a separate component of working memory, as suggested by Smyth & Pendleton (1989). This putative subsystem, coupled with the neuropsychological evidence of gesture agnosia (Rothi et al., 1986) leaves open the possibility that information from other non-spatial gestures might receive some specialised analysis.

Notwithstanding the problems with the attentional account of spatial encoding described earlier, there is a logical difficulty in assuming that a similarity in observed effects reflects a similarity in underlying processing mechanisms. Merely demonstrating that gestures, arrows and dots all have the capacity to interfere with voice responses does not necessarily entail that the mechanisms underlying the interactions are the same. The possibility remains that the quantitative differences in the effects of gestures and arrows, for instance, reflect a qualitative difference in their processing. Thus it may be premature to conclude, on the basis of the present results, that gestures receive the same kind of visuo-spatial analysis as arrows, or the same spatial analysis of "dots". Instead, it is still conceivable that deictic gestures are processed by a specialised system concerned with the identification of gestural material per se. Indeed the experiments conducted by Smyth & Pendleton (1989) coupled with the neuropsychological evidence of gesture agnosia (Rothi et al., 1986) leave open this possibility. Clearly more direct experimental comparisons are needed to fully explore this question. For now we should simply be mindful of the *possibility* that the effects of pointing gestures are mediated by some form of spatial encoding.

Overview

In the experiments reported in this chapter arrows and spatially positioned dots produced similar symmetrical patterns of interference to the deictic gestures in Experiments 1-4. This result raised the possibility that pointing gestures are processed spatially rather than by a system devoted to the analysis of gestures per se. Theoretical accounts of the Simon effect and the orienting of visual attention were used to derive a theory of how an irrelevant visual item such as a dot or an arrow could give rise to a spatial code by, respectively exogenous or endogenous shifts of attention. This spatial code could potentially interfere, either with a representation encoding the meaning of a target verbal stimulus, or with a spatial code specifying its appropriate manual response. A discussion of the work of Smyth & Pendleton (1989) introduced the possibility that pointing gestures are similarly processed, perhaps by a component of working memory which handles a broad range of spatial material. However, it was noted that merely demonstrating similar effects with spatial stimuli and pointing gestures is not sufficient evidence to claim that the underlying processing of the two types of stimuli are identical. The experiments in the following chapter examine the interaction of other non-spatial gestures with associated verbal information.

Chapter Six
Emblems and Iconics

In the preceding chapters we have seen that pointing gestures are able to influence responses to verbal material to the same extent that spoken words are able to influence gesture classification. Moreover in Chapter 4 both response, and possibly semantic factors were shown to influence the interference effects. Finally, in the last chapter the specificity or otherwise of gesture processing was addressed. The experiments reported there offered some tentative evidence that pointing gestures are processed by a system concerned with extracting spatial information from a visuo-spatial stimulus rather than by a system devoted to the analysis of gestures per se. The experiments reported in this chapter continue this investigation by examining whether or not other non-spatial gestures interact with verbal information in a similar way to the spatial gestures and visual stimuli explored in Chapters 3, 4 and 5. Neither emblematic nor iconic gestures interfered with responses to spoken words perhaps strengthening the proposition that pointing gestures exert their effects because of their *spatial* rather than their gestural properties.

Experiment 14
Emblematic Gestures and Spoken Words

Emblems form a class of gestures separate from the more illustrative deictic movements used in earlier experiments. Indeed, they are afforded special status in most gestural taxonomies (e.g. Efron, 1941/1972; Rimé & Schiaratura, 1991). According to Ekman & Friesen (1972) emblems are verbal acts with direct verbal translations of one or two words. McNeill (1985) suggests that in many cases emblems are in fact "unspoken words". Given that these highly lexicalised gestures often function as speech substitutes it should come as no surprise to find that

McNeill does not consider them to interact with speech in production: "With a diabolical accuracy, emblems are exactly the kind of gesture that would show independence from speech" (McNeill, 1987, p.500). Indeed there is some neuropsychological evidence that certain symbolic gestures and speech dissociate in production. Aphasics are often able to pantomime objects they are unable to name (e.g. Davis, Artes & Hoops, 1979). The reverse dissociation of apraxia (the inability to perform symbolic gestures on request) without aphasia is rare, but has been observed (e.g. De Renzi, Motti & Nichelli, 1980).

Whether or not emblematic gestures are likely to influence classification of verbal items is difficult to predict. Given that they are produced independently and often occur in the absence of speech then there may be little reason to suspect that they will be integrated with speech in the comprehension process. In addition the results of Experiments 12 and 13 have hinted that the influence of irrelevant deictic gestures is due to their spatial, rather than their communicative nature. An emblematic gesture will not necessarily produce an exogneous or endogenous shift in attention producing the interfering spatial code (Chapter 6). In this experiment these ideas are examined by pairing emblematic gestures with either related or unrelated words. Responses to words should not be influenced by the presence of an unrelated, compared with a related gesture. As to the effects of irrelevant verbal stimuli, there is evidence to suggest that unrelated words will cause a degree of interference in a picture naming task. Several studies have indicated a semantic "gradient" in picture-word and Stroop interference. Words sharing a semantic or categorical relationship with the picture cause more interference in picture naming than unrelated words (e.g. Glaser & Düngelhoff, 1984; Rosinski, 1977). However, and this is the crucial point, unrelated words also cause a degree of interference (e.g. Lupker, 1979; Glaser & Düngelhoff, 1984) compared with XXXX type controls. Moreover, in their study of cross-modal, visual auditory picture-word interference, Schriefers & Meyer (1990) obtained interference in picture naming from both associates and unrelated words relative to a silence control condition. Thus in the present study, the unrelated words were expected to cause some degree of interference relative to conditions where the gesture and word were related.

In summary, an asymmetrical pattern of effects was predicted in this experiment. Verbal information was expected to interfere with gesture classification but emblematic gestures were not expected to influence the processing of verbal information.

Method

Subjects: Twenty-four subjects participated in this experiment, all were undergraduate or post-graduate students who volunteered to take part in a series of experiments.

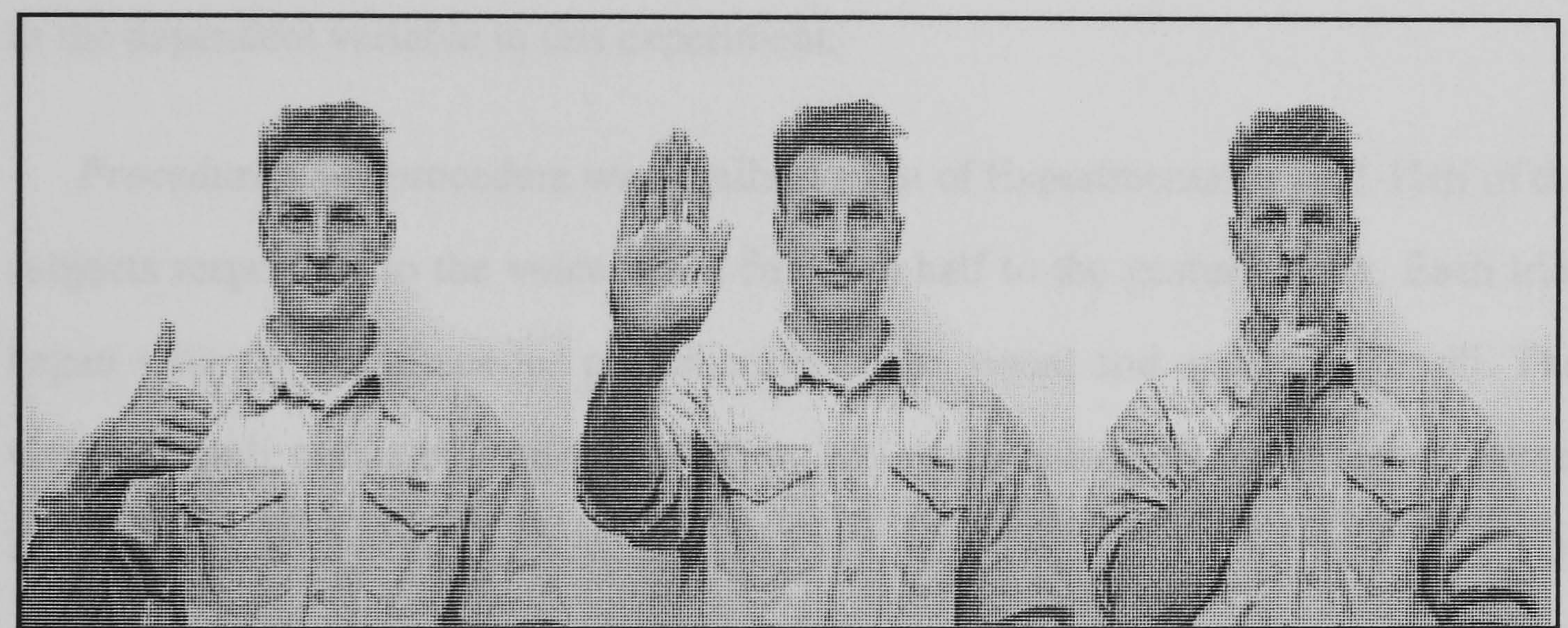


Figure 19. Examples of the "okay" (left panel), "stop" (centre panel) and "quiet" (right panel) emblematic gestures used as stimuli in Experiment 14.

Materials: The visual stimuli for this experiment consisted of three frame grabbed images of a male person gesturing: "Okay" (a thumbs up gesture); "Quiet" (index finger placed to the lips); and "Stop" (arm raised with palm facing the viewer). These gestures are illustrated in Figure 19. As before these images were frame grabbed on the "Sun" workstation. The complementary auditory stimuli consisted of the words "Okay", "Quiet" and "Stop" again recorded and edited on the Macintosh IIsi. The combined stimuli were presented using the SuperLab software on a Mac IIci. Unlike the previous experiment, the gesture and voice stimuli here could not be combined to form congruent and incongruent pairs so instead related and unrelated pairs were created. Related pairs consisted of all identical gesture and voice pairings whilst unrelated pairs consisted of all other combinations.

Design: The materials were presented in a repeated measures design with four conditions formed by all combinations of two response types (either to the voice or the gesture) and two levels of relatedness (related and unrelated pairings of gesture and voice stimuli). The stimuli were again blocked by response. Twelve trials were presented in each condition. Each related pair were presented four times in both the voice and gesture blocks whilst each unrelated pair were presented twice in each response block. Thus each subject made 48 responses. Reaction times were recorded as the dependent variable in this experiment.

Procedure: The procedure was similar to that of Experiments 1 and 2. Half of the subjects responded to the voice block first and half to the gesture block. Each trial began with the simultaneous presentation of the visual and auditory stimuli. The visual stimuli persisted until the subject had made a manual keypress response. Three adjacent keys on the keypad were used, labelled from left to right: "O", "Q" and "S" standing for okay, quiet and stop respectively. Having made a response there was a 500 ms interval before the next trial began. As before, the gesture orienting task was employed, subjects were asked to respond by pressing the space bar to a question mark which occasionally appeared on the screen. Five such trials were included in each block of the experiment. Before each block a series of 12 practice trials were presented to each subject. These consisted of a cross section of the experimental trials as well as two question mark trials.

Results and Discussion

Mean RT's and error scores from each condition of Experiment 14 are summarised in Table 14a. There are clear overall effects of response and relatedness.

Responses were made 176 ms slower to the gesture stimuli than to the voice dimension, whilst related pairs received 60 ms faster RT's than unrelated stimuli. However these effects appear to interact. The difference between related and unrelated pairs for responses to gesture is some 99 ms whilst the effect for responses to voice stimuli is only 20 ms.

A 2 (response) x 2 (relatedness) ANOVA conducted on the reaction time data largely confirmed the above observations. This analysis revealed main effects of relatedness, ($F(1,23)=9.30$, $p<0.01$) and response ($F(1,23)=13.69$, $p<0.01$), and a near significant interaction term ($F(1,23)=3.17$, $p=0.09$).

Table 14a.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Responses to Voice and Gesture Stimuli in Related and Unrelated Conditions of Experiment 14.*

	Response					
	Responses to Voice		Responses to Gesture		Overall Mean	
	M RT	%E	*M* RT	%E	*M* RT	%E
Related	788	6.46	924	5.38	856	5.92
Unrelated	808	7.58	1023	10.00	916	8.79
Mean	798	7.02	974	7.69	886	7.36

Statistically speaking, these results appear to represent a further observation of symmetrical interference. However, the imbalance in magnitude of the relatedness effect for the two response dimensions, coupled with the near-significant interaction between response and congruity factors, is more in keeping with the normal Stroop asymmetry. It may be that the experiment lacked sufficient power to investigate the interaction, indeed the large variability of the RT data attests to this fact. Nevertheless, a relatively small 20 ms effect of irrelevant gestures appears to exist. It is possible, however, that the order in which subjects completed the two experimental response tasks is of importance. For instance, any effect of an irrelevant gesture may only occur after subjects have been required to respond to those gestures. Thus we might expect no influence of an unrelated gesture on responding to a voice when the "voice" experimental block was presented first. Table 14b summarises the size of the relatedness effect as a function of task order.

An inspection of Table 14b clearly indicates that no interference effect of irrelevant gestures occurred with subjects who responded to voice in the first experimental block. Irrelevant gestures only seem able to cause interference after subjects have learnt their identities and/or mapped these onto particular responses. In contrast irrelevant voices exert interference effects regardless of the order of presentation of response blocks.

Table 14b.
Magnitude of the Interference Effect (in milliseconds) as a Function of Order of Response Task..

	Response Task	
Order	Responses to Voice	Responses to Gesture
Voice First	-9	+138
Gesture First	+50	+61

To summarise, the emblematic gestures used in this experiment exert no influence on the processing of verbal information aside from a small effect which appears to arise as an artefact of the experimental procedure. In contrast, unrelated words caused a large and consistent interference effect on the processing of gestures.

It has been suggested that the types of gesture likely to interact with speech in comprehension are the types that frequently occur with speech in production. Whilst emblems are excluded from this category, iconic gestures are not. Thus in the next experiment we examine the interaction of iconic gestures and verbal information.

Experiment 15
Iconic Gestures and Spoken Words

Iconic gestures (e.g. the type of hand movement a fisherman might make in exaggerating the size of "the one that got away") form a separate category of speech-related hand movements which are only performed with speech (e.g. McNeill, 1985), and cannot be understood completely without reference to the verbal component of the utterance (e.g. the words, "it was this big!"). These are exactly the types of gesture which, according to McNeill, share a common computational stage with speech (e.g. McNeill, 1985, 1987b, 1989) and for this reason ought to be the type of gesture most likely to be integrated with speech in comprehension. For instance, the mention of a "staircase" might be accompanied by a spiralling movement to indicate its shape, or a palms together gesture might indicate that the object of the speech is "sandwiched" between two others. In both these cases a complete representation of the speakers intended meaning is only achieved through the joint consideration of speech *and* gesture. Thus, in a similar way to deictic gestures, we might expect to observe a symmetrical interference effect between iconic gestures and verbal information, reflecting their mutual influence in comprehension. On the other hand, like the emblems used in Experiment 14, iconics might not be expected to provoke the shift of spatial attention which may well have produced the interference effects observed in earlier experiments.

In this experiment we used what Rimé & Schiaratura (1991) describe as pictographic gestures, a sub-class of iconic gesture which describes the shape of a referent. In this case the gestures used denoted size rather than shape, referring to big/small (e.g. "the one that got away") and tall/short. Given that these are speech-related gestures, a symmetrical pattern of interference was expected.

Method

Subjects: These were 24 undergraduate and postgraduates from the University of Nottingham. All had normal, or corrected to normal vision and normal hearing. Each was paid £1 for participating in the experiment which took around ten minutes.

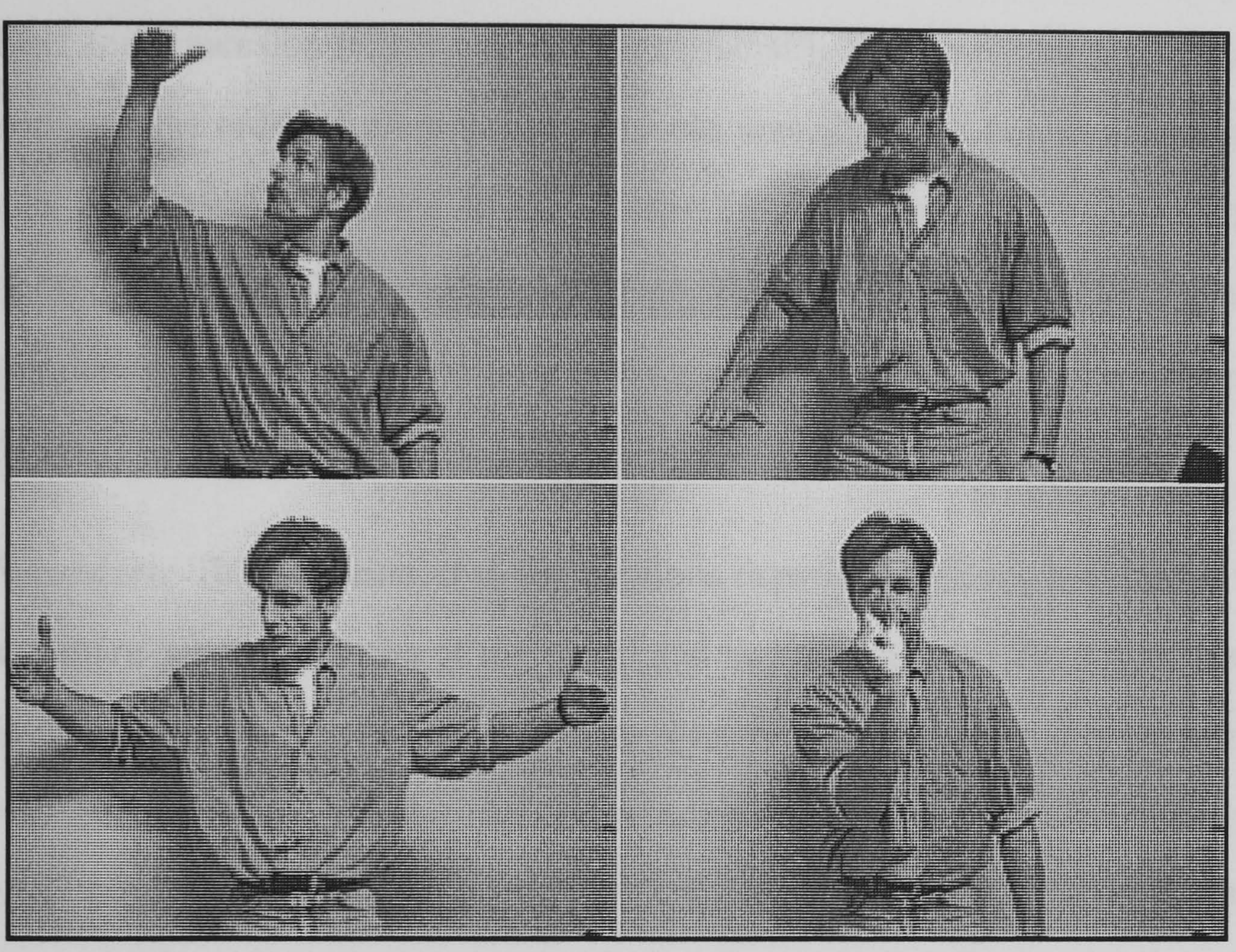


Figure 20. The "tall" (top left panel), "short" (top right), "large" (bottom left) and "small" iconic gestures used as stimuli in Experiment 15.

Materials: The visual stimuli for this experiment consisted of four digitised images of a male subject gesturing big/small and tall/short (see Figure 20). The auditory stimuli consisted of the words "tall", "short", "big" and "small" again recorded on the Macintosh IIsi. These were all edited to 570 ms, approximately the time taken to present each of the visual images. These stimuli were combined to form congruent and incongruent pairs. Congruent trials consisted of each of the four gestures paired with their associated word. Incongruent trials consisted of the words "tall" and "short" paired with a short and tall gesture respectively, as well as the words "big" and "small" paired with a small and big gesture respectively. The stimulus pairs, so formed, were presented to subjects using the SuperLab software on a Mac IIsi.

Design: The materials were tested in a within-subjects design with three factors: Response (to gesture or to voice), Congruity (congruent versus incongruent trials) and Decision dimension (either tall/short or big/small). The stimuli were blocked by response and the order of presentation of blocks was alternated between successive subjects. Each block contained 48 experimental trials consisting of twelve trials in each of the four congruity/decision conditions. Thus subjects responded to a total of 96 stimulus pairs. Both RT's and errors were recorded as dependent variables in this experiment.

Procedure: This was similar to the previous experiments. Each trial began with the simultaneous presentation of the visual and auditory stimuli. The visual stimuli persisted until the subject had made a manual keypress response. In this experiment subjects made a two choice decision by pressing the "z" key on the keyboard with their left forefinger to indicate a "big" or "tall" stimulus or the ">" key with their right forefinger to make a "small" or "short" response. These keys were labelled with the words "big/tall" and "small/short" respectively. Subjects were asked to respond as quickly and accurately as possible to either the identity of the gesture or that of the word, depending on the response block. Having made a response there was a 500 ms interval before the next trial began. As before, the gesture orienting task was employed; subjects were asked to respond by pressing the space bar to a question mark which occasionally appeared on the screen. Twelve such trials appeared in each experimental block. Before each block, a set of 20 practice trials were presented to each subject. These consisted of two examples of each of the eight different stimulus pairs along with four question mark trials.

Results and Discussion

Mean RT's and percentage of errors are summarised in Table 15a. Responses are generally faster and more accurate to congruent as opposed to incongruent stimuli across all conditions of the experiment. However the congruity effect, in terms of the RT data, is larger for the gesture responses. These observations were largely confirmed by a 2 (response) x 2 (congruity) x 2 (decision) ANOVA conducted on the RT data. This analysis revealed a main effect of congruity (46 ms, $F(1,23)=17.74$, $p<0.001$) and a main effect of decision ($F(1,23)=4.62$, $p<0.05$). Big/small decisions were made 12 ms faster than tall/short decisions. No other effects reached significance. However the response by congruity interaction term approached significance ($F(1,23)=3.64$, $p=0.07$). It seems as though the effects of irrelevant voices (66 ms) were rather larger than those of to-be-ignored gestures (26 ms). Despite this tendency toward asymmetry in the congruity effects, irrelevant iconic gestures still caused a degree of interference in subjects RT's to voice stimuli.

Table 15a.
Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Tall/Short and Big/Small Decisions to Voices and Gestures in Congruent and Incongruent Conditions of Experiment 15.*

	Decision					
	Tall/Short		Big/Small		Overall Mean	
	M RT	%E	*M* RT	%E	*M* RT	%E
			Responses to Voice			
Cong	648	1.67	644	0.67	646	1.17
Incong	676	4.13	667	2.04	672	3.09
Mean	662	2.90	656	1.36	659	2.13
			Responses to Gesture			
Cong	640	1.38	628	0.33	634	0.86
Incong	711	4.83	689	3.38	700	4.11
Mean	676	3.11	659	1.86	667	2.49

The overall error rate was low (2.30%) and these scores generally mirror the mean RT data. The correlation between mean errors and mean RT's was 0.95 suggesting no evidence of a tradeoff between speed and accuracy. Because of the low rate of errors (the overall error rate of 2.3% represents an average of only two incorrect responses per subject over the entire experiment), no further analysis was conducted on this data.

As in the previous experiments, the concern was that the order of presentation of the response blocks may have influenced the pattern of congruity effects. Table 15b summarises the congruity effects obtained for each response task as a function of block order. Subjects who were given the voice response task first showed only a small 5 ms interference effect from irrelevant gestures. In contrast, those who were initially given the gesture response task subsequently showed a large 45 ms interference effect when given the voice response task. Gestures only caused interference after subjects had had the opportunity to learn their identities and associate these with a particular response. Interference effects caused by to-be-ignored verbal stimuli, on the other hand, were large regardless of the order of block presentation.

Table 15b.
Magnitude of the Interference Effect (in milliseconds) as a Function of Order of Response Task. Mean RT's in parenthesis.

	Response Task	
Order	Responses to Voice	Responses to Gesture
Voice First	+5	+52
Gesture First	+45	+81

In conclusion, as in Experiment 14, to-be-ignored gestures caused very little interfering effect on the first block of experimental trials. As such, the overall effect of congruity on voice responses can be largely accounted for by an artefact of the experimental procedure, rather than to the interfering effect of gestures per se.

Iconic gestures of the kind used in this experiment do not seem to interact with verbal information in the Stroop-like task adopted in these studies. This is somewhat problematic for the hypothesis derived from McNeill's ideas on gesture/speech production. Iconic gestures are exactly the type of movement which might be expected to interact with speech in both production and comprehension. So why do they fail to exert any effects on the identification and response to verbal information? As in the previous experiment, it is possible that irrelevant iconic gestures were simply not identified without prior practice at the gesture response task. Again, this would leave response determination from a verbal stimulus unhindered by integration with irrelevant and incongruent gestural information, resulting in no interference effects.

General Discussion

The aim of the experiments reported in this chapter was to investigate the possibility that, along with deictic gestures, emblematic and iconic gestures affect the comprehension of verbal items in a Stroop-like interference task. Emblematic gestures were not expected to influence the processing of verbal information and this intuition was confirmed in Experiment 14. Here a normal Stroop-like asymmetry was obtained with spoken words interfering with responses to gesture stimuli, but no complementary interference from to-be-ignored gestures on responses to the verbal stimuli. Contrary to the experimental hypothesis, this same Stroop asymmetry was observed with iconic gestures and verbal information in Experiment 15. To-be-ignored spoken words exerted large congruity effects on gesture responses but irrelevant iconic gestures produced very little influence on responses to spoken words. However, the results of Experiments 14 and 15 did suggest that emblems and iconics were able to exert some effects on the processing of verbal information but only after first acting as relevant target stimuli themselves. That is, irrelevant gestures exerted no interfering effects unless subjects had "practised" responding to them in the first block of experimental trials.

The first issue to be considered in the remainder of this discussion is what the absence of an interference effect means for the non-interfering stimulus. The influence of "practice" is addressed in the light of this discussion, and thereafter the present findings are compared with the symmetrical effects produced by deictic gestures.

The most obvious conclusion to reach from the lack of an interfering effect of gestures in these experiments is that neither iconics nor emblems actually received any processing when they were acting as distracting stimuli. However, as pointed out by Driver & Tipper (1989), the conclusion that the absence of an interference effect entails the absence of identification of a distractor is subject to a logical flaw. The implicit assumption is that any identification of a distractor *necessarily* leads to interference with a response to a related stimulus. However, this need not always be the case. Both Driver & Tipper (1989) and more recently Fox (1995) have established that an ignored stimulus on trial *n* which does not produce Stroop-like interference, will nevertheless slow a response to a subsequent presentation of that stimulus on trial *n+1*. This phenomenon, negative priming in the absence of Stroop interference, suggests that the identity of the distracting stimulus on trial *n* must have been available, even though it was apparently unable to induce any interference effect. Thus it appears that non-interfering distractors *are* identified, at least under certain circumstances. This fact leads to a rather unfortunate weakness of the Stroop paradigm. In the absence of interference one cannot determine whether, or to what extent, the distracting information has been processed. The observation of an interference effect suggests that information from the distractor has somehow influenced the processing of the target. Further investigation can reveal the locus of the effect and perhaps the mode of combination of the two sources of information. However, the absence of interference either entails that the distracting information has not been identified, or that processing did indeed take place, but that no informational combination transpired.

In the present experiments, irrelevant gestures produced no interfering effects on the initial block of experimental trials. However, as discussed above, we cannot conclude that these gestures received no semantic processing. Indeed, all the accounts of Stroop interference discussed in previous chapters have assumed that the non-interfering stimulus (e.g. the colour of a colour word) is indeed identified, but fails to produce interference for any one of a number of reasons. In response competition accounts, for instance, the non-interfering dimension might be processed too slowly to exert any effects (i.e. speed of processing models), or would require some kind of translation into another cognitive subsystem in order to produce interference (i.e. Virzi & Egeth's (1985) translational model). Alternatively, in accounts where integration of information occurs (e.g. Logan, 1980), the non-interfering dimension makes little contribution to the response because of a relatively small and stable automatic "weight". What is certain, however, is that after practice at responding to both iconic and emblematic gestures, these stimuli were able to exert interfering effects on subsequent responses to verbal stimuli. However, we cannot be sure whether practice enables subjects to encode the identities of the gestures, or whether it acts post-semantically, perhaps speeding response determination, or increasing the contribution that the encoded gesture makes to the decision (e.g. by strengthening its automatic "weight"). On the other hand, both of these possibilities could well be true.

Whether or not the irrelevant gestures are identified, and regardless of the effects of practice, the initial absence of interference in both experiments suggests that neither emblems nor iconics are combined with verbal information in this Stroop-like interference task. This was the expected result for emblematic gestures, which often function as language substitutes, but somewhat of a surprise for iconics which are better characterised as language accompaniments (McNeill et al., 1994). In contrast to the bi-directional nature of the interaction between deictic gestures and verbal information, iconic gestures do not appear to influence the processing of auditorily presented verbal information. Why should this be so?

One obvious possibility is that if iconic gestures (and emblems) are encoded, and the evidence presented here does not rule out such a possibility, they are simply not processed in the same way as the deictic gestures used in earlier experiments. The results of previous experiments raised the possibility that pointing gestures exerted their effects because of their spatial nature. One suggestion is that the processing of these gestures results in a spatial code, perhaps because of the orienting of spatial attention or because of some general spatial analysis (Chapter 5). Either way this spatial code is able to influence the processing of verbal information with a spatial meaning. Interference occurs as the spatial information encoded from the gestures is integrated with that encoded from the verbal stimulus to jointly determine the response (Chapter 4). Returning to the present experiments, in most cases a spatial analysis of an iconic gesture will be insufficient to discriminate (or identify) that gesture. In addition, these gestures are unlikely to produce an orienting of spatial attention likely to result in a code related in meaning to either the concomitant verbal information or its associated response. Presumably effective discrimination of the many iconic and emblematic gestures is achieved primarily by visual processing in much the same way as a picture might be processed. This being the case, it comes as no surprise to discover that iconics and emblems behave in a similar way to pictures in a picture-word interference task, producing the hallmark Stroop asymmetry when in combination with verbal material.

Notwithstanding the above discussion, we must be careful in concluding that emblems and iconics do not exert any effects because they lack the spatial properties essential to produce interference. The involvement of spatial processing was only implicated by analogous findings with spatial (non-gestural) stimuli (Chapter 5), whilst the conclusion that it is the spatial properties of both verbal and non-verbal dimensions which largely cause the interference effects is based on an absence of exceptions. Both conclusions are thus grounded on rather weak logic. For instance the iconic and emblems in the present experiments may fail to exert any effects because movement is more important in the processing of these gestures compared with the pointing gestures used earlier. Alternatively these types of gesture may exert a relatively smaller influence on voice classification, an effect which the interference paradigm is not sufficiently sensitive enough to register.

Overview

The experiments reported in this chapter acted as an attempt to provide some generalisability to the results obtained with pointing gestures in earlier chapters. As before, the experimental hypotheses were generated from McNeill's ideas on the relationship between gestures and speech. Accordingly emblematic gestures were not expected to, and indeed did not, influence the comprehension of verbal items. Similar results were obtained with iconic gestures which *were* expected to exert some effects on spoken word classification. Whilst it is tempting to conclude that neither type of gesture was actually identified when acting as the distracting stimuli it was explained that the logic underlying this assumption is false. Instead it was concluded that the information which might have been encoded from iconics and emblems was not actually integrated with the encoded verbal information. Moreover, the fact that these *non-spatial* gestures fail to interfere with the processing of verbal information lends some weight to the proposition that the pointing gestures studied earlier exert their influence because of their peculiarly spatial nature.

Chapter Seven

Facial Gestures and Speech

It seems that deictic gestures and various other stimuli with spatial components interact symmetrically with verbal information in a Stroop-like interference task. In contrast, gestures without these spatial properties seem unable to influence the processing and response to auditorily presented words. In this chapter the relationship between certain *facial* gestures and speech is examined. Two experiments are described, the first of which demonstrates a familiar Stroop asymmetry using simple facial expressions and expression words. In the second experiment, however, the task was switched from the usual keypress discrimination to one which required semantic/affective processing of the stimuli. The results of this experiment indicate a symmetrical pattern of interference which cannot readily be explained by existing models of Stroop or picture-word interference (e.g. Glaser & Glaser, 1989). A brief review of some of the literature on the processing of facial expressions will help set the scene before these final experiments are tackled. First, we ask how people perceive and categorise facial expressions and secondly, how people use this information in conjunction with other aspects of communicative behaviour.

Expression Perception

The face is arguably the most important source of non-verbal information that the body has to offer. An accurate, moment-by-moment description of a facial pattern not only provides us with useful cues to its carrier's identity, but enables us to derive many other kinds of meaning. Perception of lip and tongue movements provides information during speech perception which can help disambiguate acoustically confusing consonants such as /f/ and /s/ (e.g. McGurk & MacDonald, 1976; Campbell, 1989). Others have emphasised the importance of eye-gaze in maintaining and managing social interactions (e.g. Kendon, 1967) and in communicating affective information (e.g. Argyle, 1983). Finally, and most relevant to this chapter, the perception of facial expression is hugely important in the "decoding" of underlying emotional states.

A number of authors have established that around seven different categories of facial expression are reliably identified from still photographs (e.g. Ekman, 1982), these include happiness, surprise, fear, sadness, anger and disgust/contempt. Furthermore, these expressions seem to be culturally universal in that they are interpreted and produced in similar ways by both literate and pre-literate societies (e.g. Ekman & Friesen, 1971). The perception of these expressions also appears to be categorical rather than continuous (e.g. Etcoff & Magee, 1992). For instance, a particular configuration of face shape is perceived as happy, another as sad. We do not appear to be sensitive to variation within a category, thus no configuration of a face is perceived as more-or-less happy, or reasonably sad. Given these constraints on processing, Ekman & Friesen have suggested that we might make use of relatively simple lists of features, or "action units" to encode different facial expressions. Moreover, there is converging evidence that the analysis of facial expressions proceeds independently of other forms of face processing. Single-cell recordings from monkeys have identified separate groups of neurons which are maximally sensitive to facial expressions and to face identity (Hasselmo, Rolls & Bayliss, 1989; Perrett et al., 1984). Brain injured patients have been identified who are unable to recognise faces but whose ability to recognise facial expressions remains intact (e.g. Bruyer et al., 1983), whilst the reverse dissociation, impaired expression recognition with spared face recognition, has also been observed (Kurucz & Feldmar, 1979). Experiments with normal subjects have indicated that subjects are able to match or make decisions to facial expressions irrespective of the familiarity of the faces (Bruce, 1986; Young et al., 1986). Taken together this represents good evidence that facial expression is computed separately from facial identity.

Dissociations have also been observed between expression identification and the analysis of facial speech. For instance, Campbell et al. (1986) describe two patients with complementary dissociations. The first, a prosopagnosic woman, was unable to identify faces or expressions but could judge what phonemes were being mouthed in still photographs, and was still susceptible to the McGurk illusion. Campbell et al.'s second patient had no trouble identifying facial expressions or identities but had difficulty in making phonemic judgements and was not susceptible to the McGurk effect. However, a recent study by Walker, Bruce & O'Malley (1995) casts doubt on the claim that the analyses of facial speech and facial identity are independent processes. Walker et al. created McGurk stimuli by pairing faces and voices from different individuals. Their results showed that when subjects were familiar with these faces they were less susceptible to the McGurk effect than those unfamiliar with the faces. If the processing of facial speech is independent of facial identification then we would expect the McGurk effect to appear irrespective of the familiarity of the speaker.

Whilst people are readily able to perceive and categorise various facial expressions they are also remarkably consistent at using this information to judge one another's emotional states (Ekman, 1982). Some evidence suggests that the cortical sites involved in this activity are different from those involved in processing the emotional content of verbal material. Kolb & Taylor (1981) describe a number of right hemisphere damaged patients who were impaired at matching emotional expressions of photographs of faces. Left hemisphere patients with similarly located lesions were relatively unimpaired at this task. The reverse pattern was noted in a task which involved matching a verbal category of emotion to a verbal description of an emotional situation. Right hemisphere patients performed in the normal range whilst the performance of patients with lesions to their left hemispheres was significantly poorer.

In summary, there is evidence to suggest that facial expressions are identified independently of other aspects of facial information such as facial identity and facial speech (but see Walker, Bruce & O'Malley, 1995). Furthermore the actual emotional content of this information may be derived separately from that of verbal materials.

As with the study of gesture and speech, the kinds of fractionation evident from experimental and neuropsychological approaches are also apparent in the social psychological literature. Thus authors have concentrated on exploring whether information derived from the face is more or less important than material encoded from other channels of communication. The *video primacy* standpoint arose from this perspective (see Chapter 1). Under this view the face is thought to convey most emotional information followed by the body, tone of voice and verbal content (e.g. Mehrabian, 1972). Implicit in this research is the view that, as perceivers, we either selectively attend to one or other source of information, or process both, but give priority to a specific component under particular circumstances. However, as recently pointed out by Ekman & O'Sullivan (1991, p.188), "there is no evidence that individuals in actual social interactions selectively attend to different channels....". Indeed, it seems we are remarkably adept at combining *relevant* sources of information to enable us to process effectively potentially ambiguous information (e.g. the perception of phonemes using both auditory and visual information). Thus, as with the comprehension of bodily gestures and speech, a more fruitful approach to the study of facial expression and voice might be to investigate if and how information from these sources is combined in the processing of emotional states.

The evidence in support of the video primacy hypothesis has come from experiments where verbal and non-verbal cues are placed into conflict. However, rather than examining the effects which one dimension might have on decisions to information in the other channel, researchers have basically asked which dimension has the most influence, or where subjects choose to allocate their attention in the comprehension of an utterance. For instance, De Paulo et al. (1978) found that in discordant conditions, listeners switched their allocation of attention from the verbal to the non-verbal signal. In the present experiments, the effects of to-be-ignored expression information on the comprehension of verbal items are examined. If a facial expression, which subjects have been explicitly asked to ignore, can influence the response to a verbal item and vice-versa, then this is very good evidence that both dimensions are routinely processed and possibly integrated in comprehension. This chapter seeks to determine whether expression information can influence; first the *identification* (Experiment 16); and second, judgements of the *affective content* (Experiment 17) of verbal material, and vice-versa.

Experiment 16

In this experiment the words happy, sad and angry were paired with happy, sad and angry facial expressions. These expressions were chosen as they are among the seven emotional expressions to be well judged by observers across different studies (Ekman 1982).

As with the experiments using emblems and iconic gestures, models of Stroop and picture-word interference would not generally predict any interference from to-be-ignored pictures. In contrast, the video-primacy hypothesis predicts that in incongruent conditions, visual information will take precedence producing an interference effect when subjects are asked to respond to the verbal stimuli. However, as before, the approach taken here is that information from the two sources will interact in processing, producing a symmetrical pattern of effects.

To summarise the predictions, picture-word studies predict a normal Stroop asymmetry, video-primacy predicts a reverse Stroop-like effect and the hypothesis developed here predicts a symmetrical pattern of interference.

Method

Subjects: Fourteen volunteers drawn from the same population as the previous studies were used as subjects in this experiment.

Materials: These consisted of frame grabbed images of a male face with happy, sad and angry expressions prepared as in the previous experiments (see Figure 21). The auditory stimuli consisted of the words "happy", "sad" and "angry" recorded and presented on the same equipment as before. The two types of stimuli were paired together to produce congruent (e.g. HAPPY-happy), incongruent (HAPPY-sad) and unrelated (e.g. HAPPY-angry) trials. The inclusion of the unrelated condition was thought necessary as happy-sad might not share the same diametrically opposed relationship as "up" and "down" which are clearly opposite. The inclusion of the unrelated condition was therefore as a comparison with the incongruent condition.

Figure 21. The "happy", "sad" and "angry" facial expressions used as stimuli in Experiment 16

Design: The experiment took the form of a repeated measures design with two factors: Congruity (congruent, incongruent or unrelated trials) and Response (to the voice or expression dimension). All combinations of auditory and visual stimuli were prepared but only responses to happy and sad stimuli were included in the design, responses to angry stimuli were used only as fillers. Thus 16 trials were included in each of the six cells of the design, half with "happy" decisions and half with "sad" decisions, with half of the subjects first responding to the voice dimension and half to the expression.

Procedure: This was largely the same as in the previous experiments . The subjects were instructed to depress one of three horizontally aligned keys (4, 5 and 6 on the keypad), labelled with the letters "H", "S" and "A" respectively, in response to either the expression or voice dimension depending on the response block. The stimuli were presented to the subjects in a random order using the SuperLab software. Again the question mark trials were included to ensure that the subjects continued to look at the screen when responding to the voice dimension. As before, these required the subject to press the space bar before the next trial was presented.

Results

Table 16 summarises the mean reaction times and error scores for each of the experimental conditions. These means are also illustrated graphically in Figure 22. Examining the RT data, the effects of congruity and response appear to be interactive. There is a clear 89 ms effect of congruity for expression responses but a -8 ms effect for voice responses. Indeed, a 3 (congruity) x 2 (response) ANOVA conducted on the reaction time scores revealed a main effect of congruity ($F(2,26)=5.53$, $p<0.05$) and a response by congruity interaction ($F(2,26)=7.15$, $p<0.01$). Further analysis of the interaction revealed a significant simple main effect of congruity for responses to the facial expressions ($F(2,52)=12.17$, $p<0.01$). Tukey HSD tests on the expression data indicated that RT's to congruent trials were faster than both incongruent and unrelated trials, whilst there was no reliable difference between RT's to unrelated and incongruent trials.

An inspection of Table 16 reveals that the error data closely mirror the pattern of RT data. In fact the correlation between RT's and errors was 0.92 suggesting no evidence of a tradeoff between speed and accuracy.

Table 16.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Responses to Voice and Expression Stimuli in Congruent, Unrelated and Incongruent Conditions of Experiment 16.*

	Responses to Voice		Responses to Expression		Overall Mean	
	M RT	%E	*M* RT	%E	*M* RT	%E
Cong	821	1.71	825	0.86	823	1.29
Unrelated	824	0.86	945	5.64	885	3.25
Incong	813	0.86	914	2.64	864	1.75
Mean	819	1.14	895	3.05	857	2.10

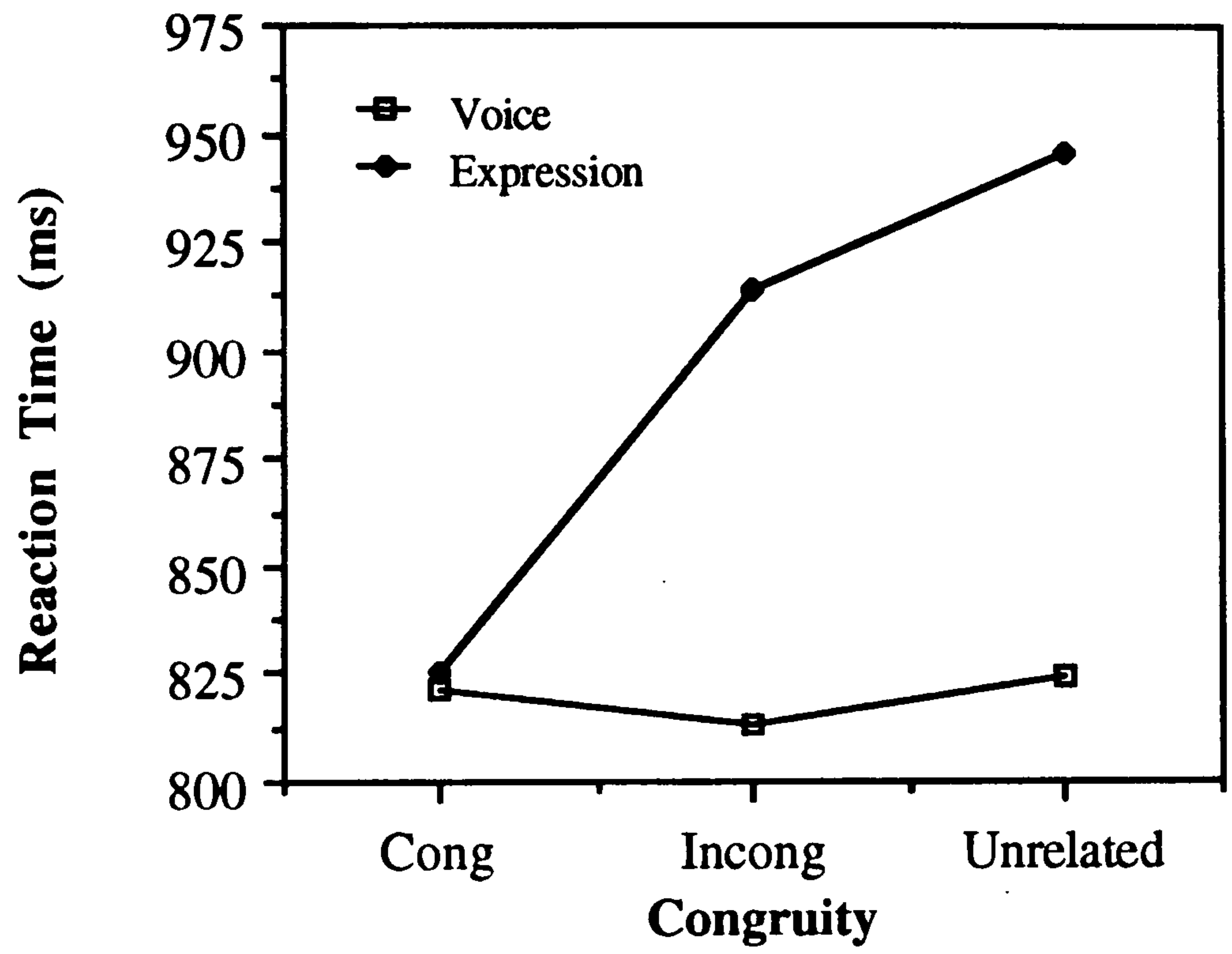


Figure 22. Mean Reaction Times (ms) to voice and expression stimuli as a function of congruity.

Discussion

The asymmetric interference effects found in this experiment suggest that, in contrast to the video-primacy hypothesis, to-be-ignored facial expressions do not influence the identification and response to auditorily presented verbal material. On the other hand, incongruent words exerted large interference effects in the expression response condition. This is the pattern of results predicted by most models of Stroop and picture-word interference and offers little support for the hypothesis that verbal and non-verbal information are mutually influential in comprehension.

An interesting finding from Experiment 16 was that both incongruent and unrelated speech stimuli appeared to interfere equally with the processing of facial expressions. Research on the Stroop effect has indicated a semantic gradient. For instance Klein (1964) demonstrated that incongruent colour words interfered more with naming than did unassociated words. Thus we might expect a similar gradient in this experiment i.e. more interference from incongruent as opposed to unrelated words. However, La Heij et al. (1990) has demonstrated that words from the same semantic category as the picture (e.g. Mouse/Dog) tend to produce interference in a picture-word task, whilst associates (e.g. Mouse/Cheese) produce facilitation. In the present experiment, happy/sad pairings are not only strong associates but are also members of the same category (emotional expressions). Thus, it is possible that the contribution of this associative facilitation produces an underestimate of the overall inhibitory effect in the incongruent condition. In contrast, "unrelated" pairs in this experiment (e.g. happy/angry) are likely to produce a categorical interference effect but much less associative facilitation (happy/angry, I would argue are not strongly associated). Thus, assuming that in general, inhibition is stronger than facilitation, La Heij's explanation would predict larger inhibition from "unrelated" compared with "incongruent" pairs. Indeed this was the pattern of results obtained, although the difference between incongruent and unrelated conditions failed to reach significance.

The earlier experiments in this thesis were influenced by the ideas of McNeill who views speech-related hand gestures as a linguistic phenomenon. Facial expressions, on the other hand, are not related to speech in quite the same way. They appear to be more like the symbolic gestures, or emblems, introduced in the last chapter. Like emblems they have culturally recognised "meanings" and are able to carry information in the absence of speech. However, these "meanings" are rather more *affective* than linguistic in nature. It is suggested that a mutual interaction between facial and verbal information is more likely to be found in a task involving affective processing rather than in simple Stroop-like identification. Thus in Experiment 17, rather than pressing a button contingent on the *identity* of either a visual or a verbal stimulus, subjects were asked to make a keypress response dependent on the *affective connotation* of either words or facial expressions in separate blocks of trials.

Changing the task from one of naming to one of semantic classification has important consequences for the predicted direction of Stroop interference. Both Smith & Magee (1980) and Glaser & Düngelhoff (1984) noted a reversal of the usual Stroop asymmetry when subjects were asked to categorise, rather than name, word and picture stimuli. In these studies, word categorisation was slowed in the presence of a distracting picture whereas picture categorisation remained relatively immune to interference from simultaneously presented incongruent words. Glaser & Glaser (1989) developed a model which sought to explain this, and many of the other data in the Stroop and picture-word interference literature. Recently De Houwer & Hermans (1994) have suggested that this model might be applied to the affective processing of pictures and words. For this reason we can derive predictions from the model concerning the affective processing of *facial expressions* and verbal information. At the same time we can examine its applicability to the range of Stroop-like interference effects presented in the present work. The discussion of this model is developed in the following section.

Glaser & Glaser's (1989) Model

Earlier it was suggested that models of the Stroop and picture-word effects which place the locus of interference at the response selection stage are neither sufficient to explain the full range of the Stroop phenomena, nor the effects reported in the present series of experiments. It was also ventured that accounts which locate the source of interference at an earlier "conceptual" stage of processing might prove to be more durable. The model formulated by W. R. Glaser & M. O. Glaser (1989) is one such account.

Glaser & Glaser's offering is broadly similar to Virzi & Egeth's (1985) model (see Chapter 4) but places the locus of the interference effect earlier in the processing system. In contrast to Virzi & Egeth's multiple systems, Glaser & Glaser (1989) suggested that only two functionally and locationally distinct processing modules exist, the lexicon and the semantic system. The semantic system consists of concept nodes linked by semantic relationships to other concepts, whilst the lexicon contains word nodes linked by non-semantic relationships to nodes sharing similar phonemic and orthographic properties. Each word node has a bidirectional link to the concept node containing its meaning. Colours and pictures are assumed to have direct access to the semantic system via an executive apparatus controlling the perception of objects and pictures, and the actions performed on physical objects including the keypress response. Lexical executive systems, on the other hand, control the perception of printed words and speech along with their associated written and spoken output responses.

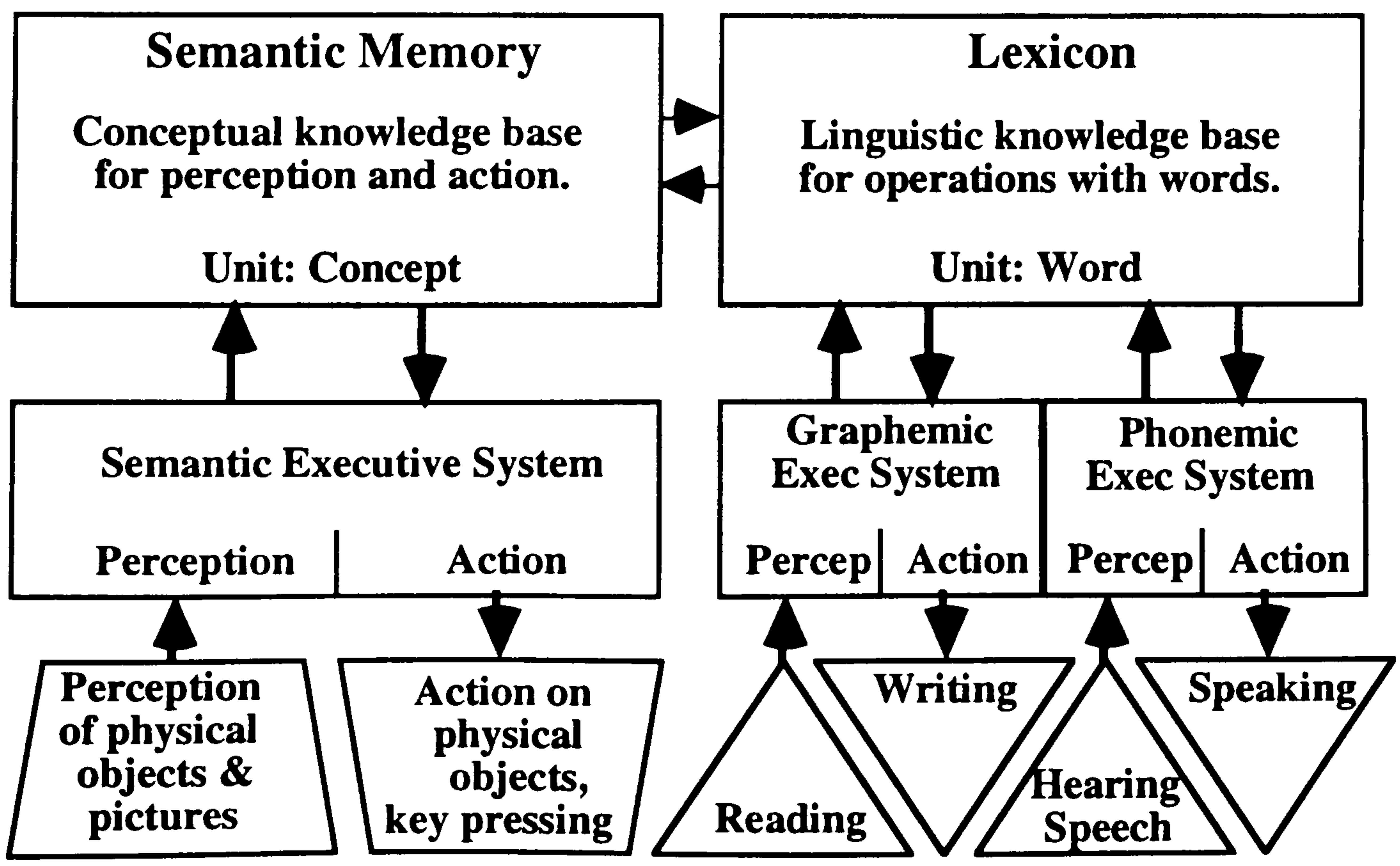


Figure 23. Glaser & Glaser's (1989) model of the cognitive structures underlying Stroop-like interference

Whether or not Stroop-like interference is obtained depends on a dominance rule based on the notion of *privileged loops*. McLeod & Posner (1984) introduced this idea to explain the relatively interference-free task of shadowing or repeating spoken words. For instance, Allport, Antonis & Reynolds (1972) described subjects who were able to shadow prose whilst playing the piano by sight reading and similarly, Shaffer's (1975) typist subject was able to copy type whilst performing a shadowing task. McLeod & Posner (1984) suggested that the underlying nervous system is structured in such a way that an articulatory program can be retrieved directly from perceptually encoded verbal information, and that this pathway from perception to action is isolated from general information processing. In this way the repetition of spoken words can be performed independently of other cognitive activities. Glaser & Glaser simply extended this idea by suggesting that in addition to the privileged loop from hearing speech to speaking (via the lexicon), there are loops from reading to writing (also via the lexical system) and from the perception of pictures to key pressing (via the semantic component of the model).

The presence or absence of these loops in a particular task is important in predicting whether or not interference will occur. If the two components of a Stroop-type stimulus are processed entirely within the same subsystem interference is always obtained. If, on the other hand, the relevant and irrelevant components are perceived by the separate semantic and lexical modules, interference will only occur if the irrelevant stimulus has privileged perceptual access to the system necessary for eliciting the appropriate response.

If a picture naming response is required from a picture-word stimulus, the lexicon will be relevant for the selection of the correct response. However, because the distracting word has privileged access to the lexicon it will influence the time taken to name the picture. On the other hand, an incongruent picture will not influence word reading time because it does not enjoy this privileged access. If a categorical decision is required, the model predicts a different pattern of results. A decision based on the semantic category of a stimulus will be controlled by the semantic system. Now pictures, but not words, have privileged access to this module. Thus the categorisation of a picture will not be affected by the presence of a word from an incongruent category. An incongruent picture, on the other hand, will delay the categorisation of a word because the picture has privileged access to the semantic system which controls the response. These are essentially the findings of Smith & Magee (1980) and Glaser & Düngelhoff (1984) which were introduced in the previous section.

Recently De Houwer & Hermans (1994) have applied Glaser & Glaser's model to the affective processing of words and pictures. In their Experiment 1, De Houwer & Hermans paired line drawings of animals which had either positive or negative affective connotations, with the names of these animals. Thus a drawing of a spider (negative connotation) might be paired with the word spider (congruent condition) or with a "positive" name such as "duck" (Incongruent condition). Half of the subjects were asked to make a verbal positive/negative judgement to the word, the other half made a similar judgement to the picture. The pattern of results obtained in this experiment were identical to those observed in picture word tasks involving categorical judgements (Smith & Magee, 1980; Glaser & Düngelhoff, 1984). Affective judgements to words were slowed by the presence of an affect-incongruent picture but affective categorisation of pictures was not influenced by an affect-incongruent word. This is the pattern of effects one would predict if pictures have privileged access to a network which contains affective as well as semantic information. Words, on the other hand, can only access affective information after they have passed the lexicon.

Returning to the present experiment, facial expressions should directly access the semantic system. Thus, according to Glaser & Glaser's (1989) model, any affective judgement of these stimuli will be immune from interference. On the other hand affective judgements to words will be slowed by the presence of affective-incongruent facial expressions. Thus the model, with De Houwer & Hermans' (1994) addition of affective information to the semantic system, predicts a "reverse" Stroop type effect. This is rather reminiscent of the "video primacy" hypothesis which suggests that when faced with discordant information, subjects should give priority to non-verbal cues. Alternatively, if subjects integrate information from the two channels, we might expect to observe symmetrical interference effects. These predictions were tested in Experiment 17.

Experiment 17

In this experiment subjects were asked to make their responses on the basis of the affective content of the target stimuli whilst ignoring either affect-congruent or affect-incongruent distractors. As in the previous experiment, subjects responded to facial and verbal information in separate blocks of trials. Both video primacy and the Glaser & Glaser (1989) model predict a reverse Stroop asymmetry, that is facial expressions will interfere with affective decisions to words but not vice-versa. On the other hand, the ideas developed in the present thesis suggest that a symmetric pattern of effects might be obtained.

A second aim of this experiment was to investigate any contribution of facilitation and/or inhibition to the effects. To this end two control conditions were included in the experiment. Also the photographs of facial expressions used in Experiment 16 were replaced with schematic line drawings of faces as it was felt that the posed expressions used in the previous experiment were not particularly good representations of happy and sad expressions, indeed the rather slow overall mean RT to the expression stimuli in Experiment 16 (895 ms) is congruent with this observation.

Method

Subjects: Twelve undergraduate students participated as subjects in this experiment. All had normal hearing and normal, or corrected to normal vision.

Materials: The visual stimuli for this experiment consisted of schematic line drawings of happy and sad faces. These were prepared with black lines on a white background using MacDraw software. Two other schematic faces were prepared: a neutral face i.e. one with no facial expression and a "mixed" face i.e. a face consisting of features from the happy and sad faces arranged in a jumbled fashion (see Figure 24).

The auditory stimuli consisted of eight words expressing a "positive" emotion, eight words expressing a "negative" emotion and eight neutral words (see Appendix A). The positive and negative words were selected from the stimuli used in a pilot experiment carried out by second year undergraduate students. These received the most extreme ratings from independent judges who were asked to rate a list of emotional words on a positive/negative dimension. The neutral words consisted of a list of inanimate items considered to have no emotional value. A "mixed" word was also included to complement the mixed face. This consisted a three syllable non-word. The three syllables were extracted from the words "computer", "ecstatic" and "picturesque" to form the non-word "comstatesque". All the auditory stimuli were "clipped" using the SoundEdit software to be of comparable durations (approximately 700 ms).

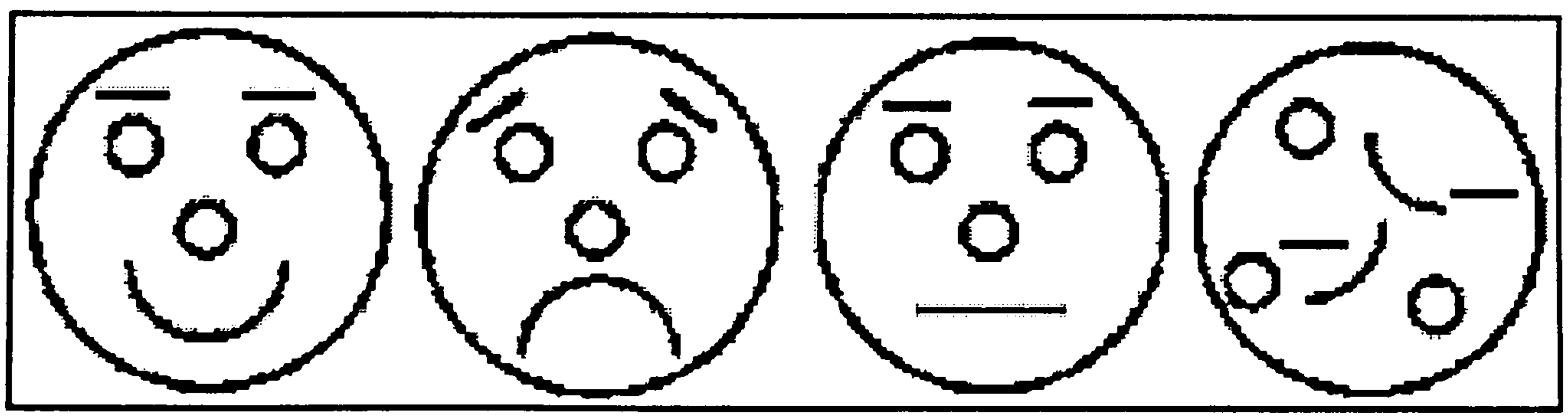


Figure 24. The "happy", "sad", "neutral" and "mix" stimuli used in Experiment 17

Design: The auditory and visual stimuli were paired together to form 4 different congruity conditions.

1) *Congruent*: e.g. happy expressions paired with positive emotional words.

2) *Incongruent*: e.g. happy expressions paired with negative emotional words.

3) *Mixed:* A mixed face paired with an emotional word (when responding to the verbal stimuli) or a mixed word paired with a facial expression (when responding to the facial stimuli).

4) *Neutral:* Here the irrelevant stimuli consisted of either a neutral word or a neutral face depending on the response condition.

The second factor was response which was made either to the expression stimuli or to the verbal stimuli. The two factors (congruity and response) were tested in a within subjects design with stimuli blocked by response. The order of presentation of response blocks was balanced across subjects. Each cell of the design contained 16 stimulus pairs giving a total of 128 trials. Half of the trials required a "positive" and half a "negative" response.

Procedure: Each trial consisted of a visual and an auditory stimulus. The onset of the auditory stimulus coincided with the presentation of the visual stimulus. Subjects' reaction times were recorded from the onset of the auditory stimulus. A 500 ms inter trial interval followed the manual response.

Subjects were told that they would be presented with a series of faces and auditorily presented words and were asked to make a binary response to either the face or the voice depending on the emotional content of the relevant stimulus. This response took the form of a manual key press to one of two keys ("A" and "L" on a standard keyboard) labelled "+" for positive stimuli and "-" for negative stimuli. Subjects used their right hand for the positive, and their left hand for the negative response and were asked to respond as quickly and accurately as possible after the onset of the stimuli. Eight question mark trials were included in each response block. As in previous experiments the presentation of such a stimuli necessitated a response using the space bar. Twelve practice trials consisting of three trials in each of the four congruity conditions immediately preceded each experimental block.

Results

Table 17 summarises the mean reaction times and error rates for each condition in the experiment. These are also displayed graphically in Figures 25 and 26. Subjects are faster and more accurate in responding to congruent as opposed to incongruent stimuli moreover this congruity effect is present for both responses to expressions and voices.

Table 17.
Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Responses to Voice and Expression Stimuli in Congruent, Incongruent, Neutral and Mix Conditions of Experiment 17.*

	Responses to Voice		Responses to Expression		Overall Mean	
	M RT	%E	*M* RT	%E	*M* RT	%E
Cong	828	5.17	658	3.00	743	4.09
Neutral	888	6.75	663	7.33	776	7.04
Mix	894	6.75	638	6.17	766	6.46
Incong	975	10.83	713	8.75	844	9.79
Mean	896	7.38	668	6.31	782	6.85

The reaction time data were entered into a 2 response x 4 congruity ANOVA. This analysis yielded a main effect of response ($F(1,11)=7.50$, $p<0.05$). Subjects were 228 ms faster to respond to the expression as compared to the voice stimuli (896 ms vs. 668 ms). The main effect of congruity ($F(3,33)=3.63$, $p<0.05$) was also a significant source of variance. Post Hoc Tukey tests established that congruent pairs received faster RT's than incongruent pairs (743 vs. 845 ms), $p<0.05$ whilst no other comparisons reached significance. The interaction between response and congruity also failed to reach significance ($F<1$).

The error scores were rather higher in this than in previous experiments. The average error score was 6.85%. The correlation between RT's and errors was 0.52 offering no evidence of a tradeoff between speed and accuracy. Because of the higher error scores the equivalent 2 x 4 ANOVA was conducted on these data. This revealed a near significant effect of congruity (F(3,33)=2.57, p=0.07). No other effects reached significance (F's<1).

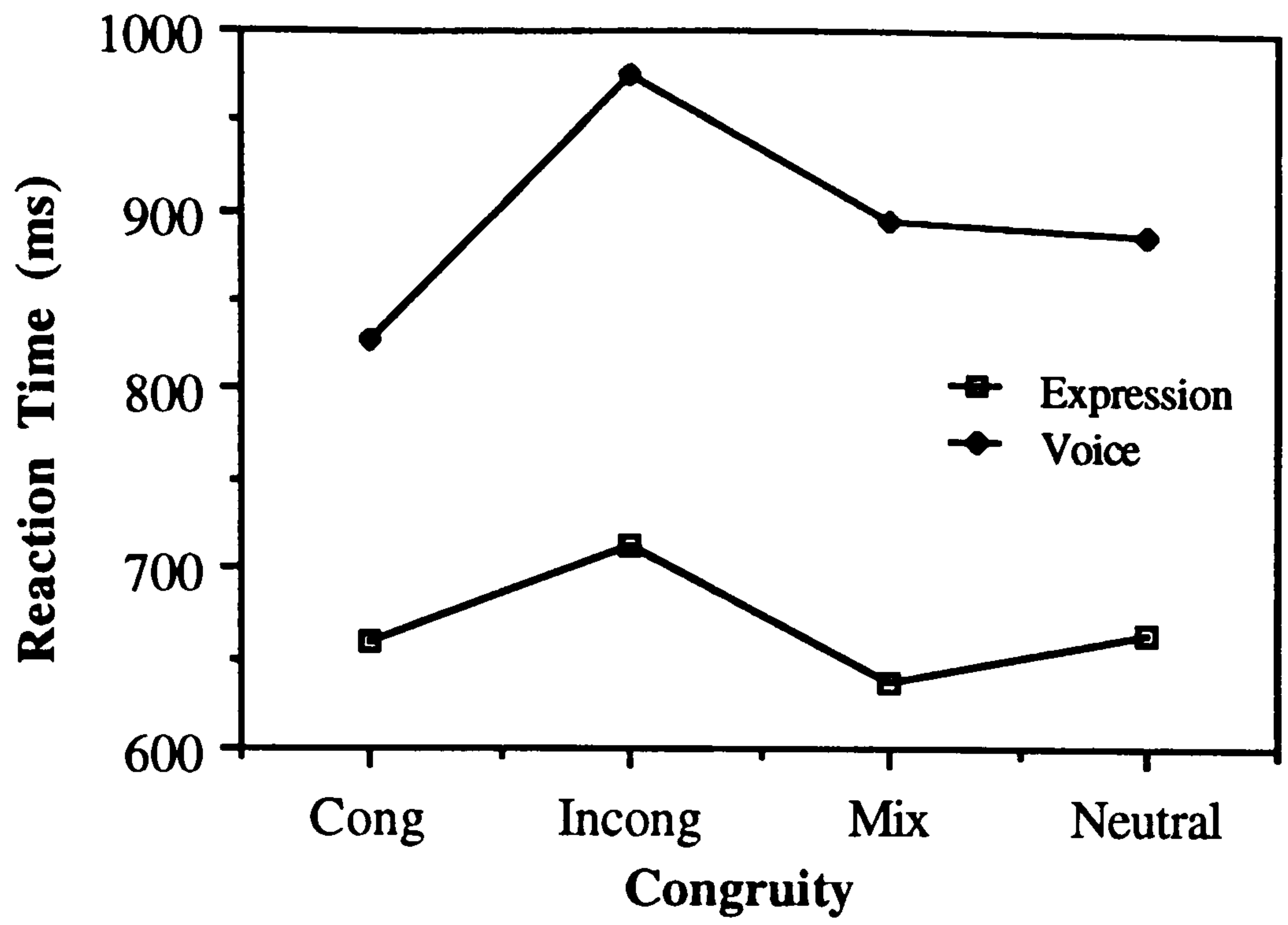


Figure 25. Graph Showing the Mean Reaction Times From Experiment 17

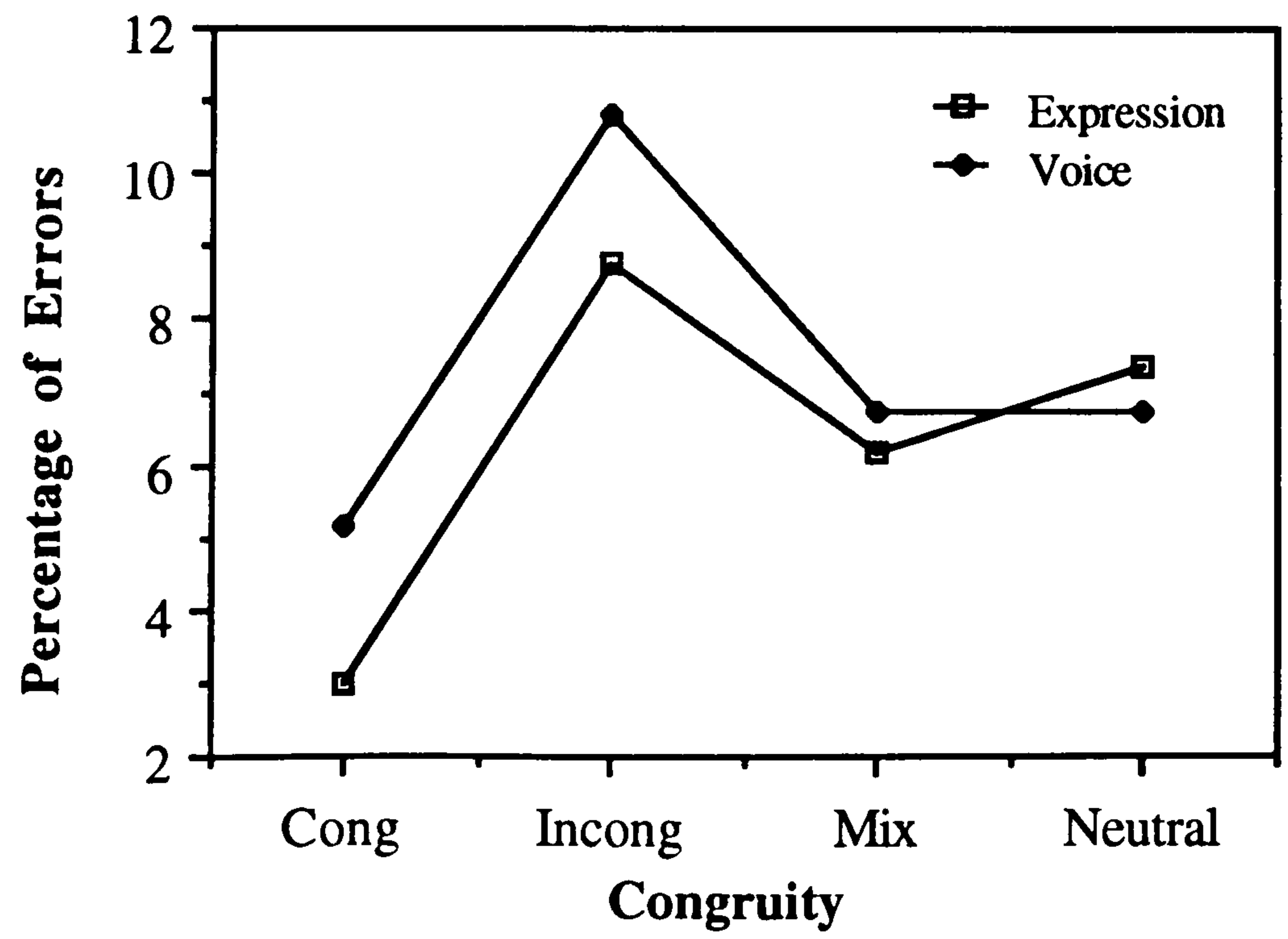


Figure 26. Graph showing Mean percentage of errors for all conditions of Experiment 17.

Discussion

The most salient of the findings from this experiment was the symmetrical congruity effect, a result not predicted by Glaser and Glaser's (1989) model of picture-word interference. Affective judgements to words were slower, and marginally less accurate, when in the presence of affect-incongruent facial expressions and notably, affective categorisations of faces were influenced in a similar way by the presence of affect-incongruent words. It is the latter result which is problematic for the model and at odds with the findings of De Houwer & Hermans (1994) who found no interfering effect of affect-incongruent words. The present findings suggest that words, as well as pictures, have direct access to affective and/or semantic information and that as a consequence, both verbal and facial information mutually interact in the processing of affect.

A second important finding was that subjects were faster to make affective judgements to the expression stimuli than to the voice stimuli. This is hardly surprising as subjects were only required to classify two face stimuli whereas 16 different verbal stimuli were used in the voice response condition. Furthermore, information from the facial stimuli will be available before the verbal stimuli as the verbal stimuli extend over time. However, it is noteworthy that in spite of the differences in overall response times, the slower dimension (words) was still able to influence judgements to the faster dimension (faces). As in earlier gesture experiments, this type of finding is problematic for models of interference based on the time of arrival of competing codes at the response selection stage of processing.

Once again the inclusion of the control conditions (neutral and mix) failed to reveal any reliable evidence of facilitation and/or inhibition although the pattern of reaction times is broadly consistent with facilitation and inhibition i.e. the mean reaction times for the neutral and mix conditions falls between the congruent and incongruent values. It would appear that the sensitivity of the experiment would have to be markedly improved to reveal any reliable facilitatory or inhibitory effects.

In summary, the results of this experiment are in line with the suggestion that facial and verbal information are mutually influential in the comprehension process. These findings offer support for neither Glaser & Glaser's (1989) model nor the video primacy hypothesis.

General Discussion

To recapitulate, the findings of Experiment 16 suggest that information extracted from facial expressions does not influence the identification and response to incongruent expression words, a result which is not predicted by the video primacy hypothesis. When subjects were asked to make an affective categorisation of the facial and verbal information in Experiment 17 however, a symmetrical pattern of results was obtained. This time facial information influenced responses to verbal stimuli whilst affective categorisation of facial expressions was influenced by the presence of affect-incongruent words. This indicates the mutual influence of facial and verbal cues in the processing of affective information.

One of the primary reasons for adopting a Stroop-like interference paradigm in the present thesis was that the patterns of interference between verbal and non-verbal dimensions could be explained by linking the findings with theoretical explanations of the Stroop effect. In this way we can begin to understand if and how information from various sources is combined in the comprehension of an utterance. The experiments reported in this chapter have been described in relation to a model proposed by Glaser & Glaser (1989). This model has recently been successfully adapted to the affective processing of words and pictures (De Houwer & Hermans, 1994). However, like most accounts of Stroop interference, this model cannot explain the symmetrical findings of Experiment 17, nor on closer inspection, the asymmetrical findings of Experiment 16. In the remainder of this section, the application of the model to the findings of Experiments 16 and 17 is discussed.

In Experiment 16 a keypress response was required contingent on the identity of either a word or a picture of a facial expression. The original version of Glaser & Glaser's model predicts that a manual responses to a picture should be immune from interference since a privileged loop exists from the perception of pictures, through the semantic system to "action on physical objects". However, in this experiment, and many others (e.g. Experiments 1-4 in the present thesis; Keele, 1972; Redding & Gerjets, 1977), words interfere with manual colour or picture responses. Recently however, Sugg & McDonald (1994) have produced a modified version of the Glaser & Glaser (1989) model. In this modification, pressing buttons labelled with words is controlled by the lexicon rather than the semantic system. A privileged loop from *reading* words to manual word responses is now established. If we assume that a similar loop exists from *hearing* speech to manual responses, then keypress responses to spoken words would be immune to interference. On the other hand, pressing word-labelled keys in response to a picture requires both semantic and lexical processing and so interference from an incongruent word is expected. So with some modification, the model could, in principle, explain the findings of Experiment 16.

The findings of Experiment 17, on the other hand, are very much more problematic. De Houwer & Hermans (1994) assume that a privileged loop exists from the perception of pictures, through the semantic/affective system to a manual response based on the affective content of the picture. No interference from affect-incongruent verbal distractors is expected and indeed De Houwer & Hermans (1994, Experiment 1) found none. In contrast the results of Experiment 17 indicate that affect-incongruent words do indeed influence the categorisation of facial expressions. This result suggests that words, as well as pictures, have direct access to affective information, perhaps stored within the semantic system.

The suggestion that affective information is stored within the semantic system is perhaps not all that controversial. Bower (1981) proposed an associative network theory linking "mood" and memory phenomena. Briefly, concepts, events and meanings are represented in memory in the form of "nodes" which are joined together by associative links to form a semantic network. When one node is activated, activation spreads, via the links, to other associated nodes. Each emotion has an identical status to other types of stored information being similarly represented as a node within the semantic system. More recently Teasdale & Barnard (1993) have introduced Barnard's (1985) "interactive cognitive subsystems" (ICS) framework to the arena of cognition and emotion. ICS aims to provide a comprehensive account of human information processing and as such is far too complex to describe here. However at the heart of the framework meaning is represented at two levels as the propositional and the implicational codes. The propositional subsystem is roughly equivalent to the semantic system described above and encodes concepts and the relationships between them in the form of propositional statements. As in Bower's (1981) offering, affective information is encoded at this level. Subjectively, meaning at this level corresponds to awareness of semantic relationships (e.g. "knowing that" the death of a loved one is a "sad" event). In contrast meaning at the implicational level is more associated with "holistic senses of knowing...or of affect" (Teasdale & Barnard, 1993, p.52). These higher level meanings actually encode some kind subjective appreciation or feeling of "sadness" at the death of a loved one. The difference in the encoding of affect at these two levels of meaning corresponds to what has been termed "hot" and "cold" processing of emotion. We are able to recall an emotional event either with (hot) or without (cold) actually re-experiencing that emotion. Whilst the ICS framework is able to incorporate both hot and cold processing, Bower was forced to propose separate hot and cold nodes within his single semantic system, clearly a somewhat ad hoc assumption. Thus the ICS approach represents a significant advancement in the study of cognition and emotion. We return to its discussion in the final chapter. For now one other important point to note is that the ICS framework provides a means whereby "structurally encoded" visual and auditory information can be combined into propositional codes and where raw sensory information can be abstracted into higher level implicational codes. This is clearly in keeping with the ideas we have developed in the preceding chapters where gestural and now facial information can be integrated with verbal material at a semantic level of analysis.

Whilst affective information may well be stored within the semantic system, the suggestion that this information can be directly accessed by both visual and verbal information is not in keeping with Glaser & Glaser's model. In their model pictures have direct access to a semantic system comprising a set of associatively interlinked concept nodes. Verbal stimuli, on the other hand, have direct access to a lexical system housing a set of word nodes, linked by linguistic rather than semantic features (e.g. pronunciation, orthography etc.). The observation of symmetrical interference effects both with affective decisions to expression/word stimuli and with naming decisions to gesture/word stimuli, rather refute these privileged access assumptions. These data are more consistent with a central, abstract and amodal semantic system to which both verbal and visual material have access, via their appropriate stored structural descriptions. As well as Barnard & Teasdale's ICS framework, those of Seymour (1973b) and more recently, Riddoch et al. (1988), share this property. Meanwhile, Glaser & Glaser's account clearly falls some way short of their goal, that is providing a comprehensive model of general reading-naming interference effects. Any general model must be able to explain not only the asymmetry of Stroop and picture-word interference, but also the symmetrical effects observed in many of the present experiments.

Overview

This chapter began with a brief introduction to expression perception and video primacy, the idea that people have a preference for visual information in the processing of emotion. However, the results of Experiment 16 indicated that, contrary to the video primacy hypothesis, visual information in the form of facial expressions did not influence the comprehension and identification of verbal material. In order to examine affective processing of visual and verbal information in more detail, a semantic classification task was introduced in Experiment 17. Predictions were derived from Glaser & Glaser's (1989) general model of interference which has been specifically applied to the affective processing of words and pictures. Consistent with the ideas developed in Chapter 4, this account places the locus of the interference effect prior to the response selection stage of processing. The symmetrical interference effects observed in Experiment 17 provide further evidence for the mutual influence of visual and verbal information in comprehension but cannot be explained within Glaser & Glaser's framework. It was suggested that frameworks containing a central amodal semantic system which also contains "cold" affective information, might be more appropriate.

Chapter Eight

Discussion, Conclusions, and Further Questions

Summary of Experimental Findings

The experiments reported in this thesis were aimed at addressing three principal questions. First, do listeners actually attend to speech related gestures in the comprehension of an utterance? Secondly, if they *are* processed, how do they interact with verbal information? Finally in what sense, if any, are gestures special? Experiments 1 and 2 revealed a symmetrical pattern of interference effects in a Stroop-type paradigm, a finding consistent with the hypothesis that listeners attend to both gestural and verbal sources of information. Auditorily presented verbal information interfered with the processing of concurrently displayed static deictic gestures and vice-versa. Furthermore the effects were found not to be due to the specific cross-modal nature of the stimuli (Experiment 3) nor to any confusion arising from the interpretation of the directional gestures (Experiment 4). Clearly listeners do attend to pointing gestures. From the perspective of dimensional interaction, this type of interference effect is thought to arise from a bi-directional exchange of information or "crosstalk" between processing channels at some level of information processing. It was suggested that the effects found in Experiments 1-4 are likely to be located post-perceptually.

The experiments reported in Chapter 4 further examined the locus of the interaction by considering a stimulus-response compatibility account of the effects. In these experiments we asked whether or not the symmetry of the effects would be maintained when the SRC between gestures and responses was disrupted. The results of Experiments 5-8 suggested that the influence of to-be-ignored gestures was either eliminated or possibly reduced under incompatible S-R conditions implicating a role for SRC. However, the interfering effects of to-be-ignored pointing gestures persisted in Experiments 9-11 despite a switch from a manual keypress to a verbal naming mode of response. Taken together, these results suggest that, whilst important, SRC is clearly not the only factor implicated in gesture-word interference. Moreover, the symmetrical nature of the interference effects is difficult to reconcile with other response selection accounts of Stroop interference, implicating a semantic locus for the interaction.

The experiments reported in Chapter 5 turned to the question of the specificity of gestures. Arrows (Experiment 12) and spatially located "dots" (Experiment 13) were paired with spoken directional words. Symmetrical interference effects were observed in both experiments raising the possibility that pointing gestures are processed spatially rather than by a system devoted to the analysis of gestures per se. The absence of any interfering effects of emblematic or iconic gestures (Experiments 14 and 15, Chapter 6) on decisions to verbal stimuli further suggested that it may be the spatial properties of the pointing gestures which is important in producing the symmetrical effects.

Finally Chapter 7 explored the interaction of facial gestures and verbal information in comprehension. Facial expressions did not interfere with keypress identification responses to spoken words in Experiment 16. The symmetry returned in Experiment 17 where subjects were asked to make affective judgements to either emotional words or schematic faces. These results were discussed in the light of Glaser & Glaser's (1989) general model of picture-word interference which places the locus of interference between response selection and perceptual encoding. Although the nature of the interference effects seems to demand just such a central locus, as with other models of Stroop-type interference, Glaser & Glaser's model fails to capture the symmetrical nature of the interference effects observed in these experiments.

To summarise, pointing gestures do seem to be processed in comprehension, producing substantial interference effects on word categorisation decisions. It was submitted that these interference effects arise because of crosstalk between processing channels at some level of analysis. Perceptual and response stage loci were rejected, logically implicating informational integration at the stage where meaning units are extracted. However, some evidence has been presented suggesting that it is the spatial properties of the pointing gestures which causes their interfering effects rather than their linguistic, or communicative status as gestures per se. In line with this there was no evidence that other non-spatial, speech related gestures were integrated with verbal information in comprehension. Nevertheless, evidence has been produced which suggests that facial expressions and verbal information can be mutually influential in *affective* processing.

Spatial Processing and Social Attention

Perrett et al. (1985, 1992) have identified populations of cells in the superior temporal sulcus (STS) region of the macaque which are sensitive to the direction in which the eyes, head and body are pointing. This work suggests that the perception of social attention involves the analysis of head and body posture, as well as eye-gaze direction. The deictic gestures used in the above experiments comprise just such head, body and gaze cues in addition to the "gestural" signal given by the hand. Furthermore, the results of the experiments reported in Chapters 4 and 5 have raised the possibility that, taken together, these cues are processed spatially. Thus, It follows that the processing of social attention may well be primarily spatial in nature.

Other evidence that the comprehension of social attention is essentially performed by a spatial system comes from a student project, partly under the author's supervision. This study was essentially a replication of the present Experiment 2, but with the gesture stimuli replaced by digitised images of heads oriented either upwards, downwards, to the left or to the right. The results of this study replicated the symmetrical effects obtained in Experiment 2 using pointing gestures, and those of Experiments 12 and 13 using arrows and dots. Importantly, the direction of the head alone was sufficient to produce a 48 ms interference effect on responses to verbal information (see Table 18). The fact the identical pattern of effects is obtained with purely spatial stimuli such as arrows and dots raises the possibility that the understanding of social attention, as indicated by the orientation of the head, involves some kind of spatial processing.

Table 18.

Mean RT's (in milliseconds) and Percentage of Errors (%E) *for Up/Down and Left/Right Decisions to Voices and Head Direction in Congruent and Incongruent Conditions.*

	Up/Down Decisions		Left/Right Decisions		Overall Mean	
Congruity	*M* RT	%E	*M* RT	%E	*M* RT	%E
			Responses to Voice			
Cong	598	4.20	613	0.52	606	2.36
Incong	651	5.20	655	13.5	653	9.35
Mean	625	4.70	634	7.01	630	5.86
			Responses to Head Direction			
Cong	530	2.10	483	1.04	507	1.57
Incong	606	7.80	518	7.30	562	7.55
Mean	568	4.95	501	4.17	535	4.56

A second student project, again partly supervised by the author, essentially replicated the usual influence of pointing gestures on responses to directional words. In addition, a further incongruent condition was included where the gesturer's head pointed in the opposite direction to the direction indicated by his hand. The effect of this modification was to significantly reduce, but not eliminate the interference effect. Reaction times to congruent, incongruent, and the moderated incongruent stimuli were 667 ms, 745 ms and 709 ms respectively, indicating a normal congruity effect of 78 ms. The size of this effect was reduced to one of 42 ms by the orientation of the head when it conflicted with the direction shown by the gesture. In order to produce this moderated interference effect, the orientation of the head must have generated a spatial code which conflicted with the spatial representation derived from the pointing gesture.

Taken together, these studies suggest that the orientation of social attention as indicated by head and possibly body posture, is processed spatially in a similar way to the arrows, dots and deictic gestures studied in earlier experiments.

The goal in the remainder of this discussion is to develop some kind of framework for understanding the patterns of interactions seen in both experimental situations, as in the Stroop and picture-word effects, as well as between sources of information which govern our communicative behaviour in everyday life. Thus, at the outset the challenge is to develop an account which can accommodate both symmetrical and asymmetrical interactions. Throughout this thesis we have seen how a number of models which provide reasonable accounts of Stroop asymmetry come unstuck when faced with symmetrical interference effects. The possibility, which is discussed below, is that symmetrical interactions are the norm rather than the exception and that asymmetries such as the Stroop effect could be accounted for by a number of additional factors which provide "bottom-up" influences on selective attention. According to this view efforts should be channeled into accounting for *symmetry* rather than asymmetry. It is suggested that the ICS framework, introduced in the last chapter, might prove useful in this regard. With its emphasis on interactivity, the ICS account may go some way in bridging the gap between the ideas of McNeill and Kendon on the processing of gesture and speech, and the information processing models of Stroop interference which we have encountered in previous chapters.

Before describing ICS in more detail, the additional factors which operate to produce asymmetry are described. In this discussion it becomes clear that these considerations have important consequences for the use and interpretation of interference effects in the study of dimensional interactions. With these thoughts in mind, the discussion returns to the key issues which motivated this work. First, we must reconsider whether or not gestures are processed, and thereafter whether, and how, gestures might interact with speech in the comprehension process.

The Capture of Attention

A number of stimulus properties are able to cause breakdowns in selective attention which are perhaps beyond the control of the subject. As will become clear, these properties might be either intrinsic to the stimuli themselves or related to their presentation. Thus an alternative approach to the study of the Stroop effect and other incidences of interference is to examine what "bottom-up" factors cause selective attention to fail. In this section it is argued that the operation of one or more of these additional factors may cause otherwise non-interacting or symmetrically interacting dimensions to interact asymmetrically. Under this view the Stroop asymmetry might be viewed as the exception rather than the rule. Two conclusions naturally follow from this standpoint. First, efforts to formulate models which describe asymmetrical effects may be misguided. The appropriate starting point may be a model or a framework which easily incorporates mutual interactions. Secondly, if the goal is to adopt an interference paradigm to investigate the interaction of pairs of dimensions then these other factors must be taken into consideration. Several of these additional factors are described below.

Dimensional Interactions

Melara & Marks consider a number of dimensional interactions as being fundamental to the cognitive system. This is because the interactions are intrinsic to the dimensions themselves. These include the so-called *integral* dimensions such as colour and brightness and *corresponding* dimensions such as position and pitch. These interactions are fundamental as selective attention to one of the dimensional attributes will inevitably be disrupted by the presence of attributes from the other dimension, that is they will produce reliable Garner interference. In addition, corresponding dimensions produce reliable Stroop-like congruity effects, indicating that these interactions are semantically based. Further evidence to suggest that these interactions are fundamental is that neither Garner nor Stroop-type interference diminish substantially with extended practice (Melara & Marks, 1990b).

Abrupt Onset

In Chapter 5 we saw how the sudden appearance of a salient visual stimulus in the periphery of the visual field can cause a mandatory shift of visual attention to the location of that event (e.g. Jonides, 1981). This type of mechanism was submitted as an explanation for the interfering effects of the irrelevant "dot" stimuli on the processing of verbal information in Experiment 13. The interference was considered to arise as a result of the encoding of some spatial information which is necessary to produce the attentional shift. This spatial code is then either in conflict or is congruent with the spatial information derived from the verbal "left/right" stimulus. In Chapter 5 we also considered the possibility that the interfering effects of irrelevant gestures are caused by a similar mechanism.

Variability

As we have seen, Garner interference has been used in the study of selective attention and relatedly, dimensional interaction. At the root of this paradigm is the idea that failures of selective attention are caused by *variability* in irrelevant information rather than simply the presence of this information per se. Garner interference measures the extent to which orthogonal trial-to-trial alterations disrupt classification performance to values on the target dimension. In everyday tasks we are able to ignore irrelevant sources of information in our surroundings. For instance, the colour of the wall in front of me does not seem to hold any consequences for my ability to attend selectively to tasks such as typing or reading. However, if the wall should suddenly change colour or move, then perhaps I might integrate this information into my current cognitive activity. In addition to registering the change, I might also appreciate the value of the dimensional change (e.g. the new colour of the wall). This example illustrates the subtle difference between the conception of Stroop and Garner interference. In the former we are sensitive to the kind of change and in the latter the change itself.

Baseline Discriminability

A further condition which seems to lead to breakdowns in selective attention has recently been revived by Melara & Mounts (1993). We have seen that selective attention can be disrupted by variability in an irrelevant dimension, however it seems as though the actual *amount* of variability dictates whether, and how extensively selective attention fails. The idea is that an irrelevant dimension will be processed obligatorily when values along it vary more than do values on the relevant dimension. In other words, a more discriminable dimension will always cause a failure of selective attention to a less discriminable dimension regardless of whether the dimensions in question can be considered as interacting or separable.

Discriminability is a rather global concept, according to Melara & Mounts (1993) it is "the time necessary to attend to, perceive, identify and respond to values varying on a single dimension" (p.640) and as such can be influenced by factors such as stimulus-response compatibility and practice (see below). The general idea is that baseline discriminability is measured within the Garner interference paradigm. Subjects are asked to make a speeded response to values on the relevant dimension whilst the irrelevant dimension remains constant. Thus in one block of baseline trials the subject might see the word "red" printed in either red or green ink and be asked to respond on the basis of the ink colour.

Melara & Mounts (1993) suggest that the classic Stroop interference is caused by mismatches in baseline discriminability between otherwise non-interacting (separable) dimensions. In their experiments, Melara & Mounts manipulated the relative baseline discriminabilities of the classic Stroop dimensions of colour and colour word. They found that when the dimensions were mismatched in baseline discriminability, the more discriminable dimension disrupted classification of the less discriminable dimension but not vice-versa. Thus a reverse Stroop effect was obtained when colours were made more discriminable than words, and a normal Stroop effect obtained under the opposite conditions. When baseline discriminabilities were matched, Melara & Mounts obtained small but symmetrical congruity effects. However, with extended practice the interference effects diminished so that the Stroop dimensions appeared to be separable, that is subjects were able to attend well to one dimension in the face of irrelevant variation in the other dimension. This is in marked contrast to corresponding dimensions such as position and pitch where the magnitude of the interference effects remains relatively constant over as many as 15 blocks each of 96 trials (Melara & Marks 1990b). Melara & Mounts (1993) suggest that it is discriminability per se, rather than the strength (e.g. Cohen et al., 1990), or the relative speed of processing of the stimulus dimensions, which underlies the Stroop effect. Moreover they claim that the crosstalk which causes Stroop interference is not inevitable or fundamental to the cognitive system as is the interaction of corresponding dimensions.

As discriminability is related to perception, identification and response processes, it embraces the notion of stimulus-response compatibility (SRC). SRC relates to the ease with which encoded stimulus attributes can be mapped onto responses (see Chapter 4). Thus if values on a dimension, such as left and right on the spatial dimension, can be easily mapped onto left/right manual responses, then these stimuli can be considered as more discriminable than in the case of an incompatible S-R arrangement. This is an important point as the SRC manipulations made in Experiments 5-11 could be construed as affecting baseline discriminabilities. This issue is discussed in greater detail in a later section.

Summary

The above discussion has hopefully illustrated that the interaction of certain pairs of dimensions is fundamental to the cognitive system. Crosstalk between these dimensions will always occur resulting in a breakdown of selective attention. However, other factors such as the presence and relative magnitude of dimensional variation or the abrupt onset of a stimulus can also result in crosstalk and consequent failure to attend selectively to one of the dimensions. It seems that asymmetries such as the Stroop and picture word effects may well represent the influence of these other factors rather than any fundamental or inevitable interaction. It may therefore be the case that many pairs of dimensions may be either fundamentally separable or symmetrically interacting.

The Use of the Interference Paradigm

Clearly the range of factors which influence selective attention must be taken into account if the interference paradigm is to be used as a tool in the study of dimensional interactions. With these considerations in mind, a prescription for its use can be formulated. First, values on the dimensions of interest are placed into conflict and the influence of the values of one dimension on responses to the other is examined. If no interference is obtained some other procedure (e.g. negative priming) should be used to determine whether or not the irrelevant item received any processing (see Chapter 6). If the answer to this question is in the affirmative then the logical conclusion is that the dimensions in question, although both processed, are non-interacting or separable. However, if interference is obtained in the first instance, then efforts should be made to: 1) determine the locus of the interaction; 2) to examine how the integration of information occurs; and 3) to examine whether the dimensions are fundamentally interacting or whether the crosstalk of information occurred for some other reason.

Implications for Gesture/Speech Comprehension

With the preceding discussion in mind we return specifically to the interaction of gestural and verbal information observed in many of the present experiments. In these studies it was claimed that the observation of a symmetrical interference effect demonstrated, first, that subjects *do* actually process non-verbal material in comprehension and second, that at the very least, information from pointing gestures is integrated with that derived from spoken material in the comprehension process. However, given the preceding discussion and the prescription for the use of interference effects described above, it might be prudent to ask whether any of these claims should be re-evaluated. We ask whether or not the evidence presented in the present series of experiments is consistent with the conclusion that gestures are processed in communication, whether gestures and speech should be considered as fundamentally interacting dimensions, and given that interference does occur (whether fundamental or not), how the proposed integration process takes place.

Are Gestures Processed?

According to the prescription described above, the first step is to determine whether or not the dimensions produce any interference. In the case of pointing gestures and spoken words this is clearly the case. The logical conclusion is that both sources of information are processed in comprehension and, it is argued, that some crosstalk or integration of information has transpired. In contrast to pointing gestures, neither emblematic nor iconic gestures exerted any interfering effects on verbal information. However, as discussed above and in Chapter 6, the conclusion that these gestures are not identified is somewhat premature. In these cases further steps need to be taken to determine whether any processing of these gestures occurred.

By eliminating perceptual encoding and response competition accounts of the interference effects between pointing gestures and words, a partial answer has been provided to the question of *where* the interaction occurs. Before asking *how* these sources of information are integrated we must first consider the possibility that pointing gestures and speech are potentially separable dimensions and that one or other of the factors described above has operated to cause the interaction and the consequent failure of selective attention.

Are Gesture and Speech Interacting Dimensions?

The ideas developed in Chapter 1 concerning the production and comprehension of gestural and verbal information would seem to predict that these dimensions should interact. The idea is that the meaning of any utterance is underspecified by either the verbal or the gestural content alone. Thus information from both sources needs to be processed in order to provide a complete representation of the speakers intended meaning. However, the above discussion has raised the possibility that several other factors might plausibly operate to promote dimensional interactions. In particular mismatches in baseline discriminability seem to be important in this regard. If mismatches in baseline discriminability can be shown to influence the interference effects we have observed then, like Stroop interference, it may be that the interaction is not intrinsic to the cognitive system. This finding would raise the possibility that the pointing gestures and verbal stimuli are from entirely separable dimensions.

In order to examine whether such an account could be applied to the present findings, measures were obtained of relative discriminability (calculated by subtracting overall gesture RT's from voice RT's) from each of 9 experiments. These included Experiments 1-6 and Experiments 9-11. Positive discriminability scores are obtained when gestures are more discriminable than words, and on Melara and Mounts' (1993) analysis (see above), these scores should correlate positively with the interfering effect of gestures on responses to words, and negatively with the interfering effect of words on gestures. It should be noted, however, that these baseline measures are not exactly equivalent to those used by Melara & Mounts.

They obtained their baselines by examining subject's choice RT's to a particular dimension (e.g. colour) whilst the irrelevant dimension remained constant (e.g. the word GREEN). Nevertheless it was felt that overall mean RT's to each dimension would provide a baseline which would be sufficient to explore the hypothesis. In line with Melara & Mounts' analysis, the discriminability scores obtained were found to correlate positively with the size of the interference effect caused by to-be-ignored gestures ($r=0.83$, $p<0.01$). In other words, as gestural stimuli became more discriminable than verbal stimuli, they caused increasing larger interference effects when subjects were asked to classify the verbal dimension. The correlation between discriminability and the magnitude of the interference effects caused by irrelevant verbal stimuli was only -0.22 (n.s.). Thus, although not as compelling, it appears that the interference effect caused by irrelevant verbal information is large when voices are more discriminable than gestures and decreases slightly as voices become less discriminable than gestures.

Thus there is some evidence that the effects might be related to the relative discriminability of the component dimensions. However, the question remains as to whether discriminability *causes* the interference or simply exacerbates an already fundamental interaction. The regression analysis described above hints at the answer to this question. Both equations derived from this analysis produce almost identical intercepts (59.22 for voice responses and 59.24 for gesture responses). This suggests that when baseline discriminabilities are matched, symmetrical interference effects of around 59 ms should be obtained. When Melara & Mounts (1993) matched the baseline discriminabilities of the Stroop dimensions, they obtained on average only 14 ms of Stroop interference in the first block of their experimental trials. This small congruity effect was rather ephemeral, diminishing with classification practice. The 59 ms congruity effect for baseline matched gesture-word stimuli predicted by the model is clearly substantially larger, perhaps suggesting a more robust interaction. However, this model remains to be empirically tested, as do the effects of classification practice. If we are to claim that pointing gestures and speech are fundamentally interacting we must demonstrate that the interference effects occur when baseline discriminabilities are matched and that these effects remain consistent over many blocks of experimental trials.

Whilst pointing gestures and speech may be fundamentally interacting, what of the emblems and iconics studied in Chapter 6. Recall that in these studies emblematic and iconic gestures failed to influence the processing of verbal information. This evidence seems to suggest that, if in fact these gestures are processed, then they do not interact with verbal information. However, substantial interference effects were evident after subjects had first learnt to identify the gestures and associate them with the appropriate response (50 ms and 45 ms in Experiments 14 and 15 respectively). From one perspective, the effects of practice might be considered as increasing the variability or discriminability of the gesture dimension, with the inevitable result of a dimensional interaction. This example illustrates the suggestion that dimensional interaction, between otherwise separable dimensions, can be induced when there is some kind of *variability* in the irrelevant dimension. Thus it matters little whether the dimensions are fundamentally interacting or separable, selective attention will always falter to some degree if some irrelevant variability exists. It is suggested that, under normal circumstances, variability is almost always present in the gesture "channel" simply because a gesture itself is a dynamic act. Therefore, even if it transpires that gesture and speech are non-interacting dimensions, the gestural variability will generally ensure that an interaction with verbal information occurs. Clearly attempts must be made to extend the scope of the interference paradigm to include stimuli with dynamic properties so that the role of movement in producing interference effects can be studied.

To summarise, whilst there is some evidence to suggest that the abrupt onset and relative discriminability of the gesture dimension may well influence the size of the interference effect, it has been suggested that these factors alone cannot account for the symmetry observed with pointing gestures and verbal information. Thus it is possible that these dimensions *are* fundamentally interacting. In contrast, the results with emblematic and iconic gestures are rather more suggestive of separability, with interference effects only emerging as discriminability is increased with continued practice. However, it was submitted that movement might be an important factor which may always ensure that speech related gestures are integrated with the ongoing processing of verbal information.

How are Gestures and Verbal Material Integrated?

The work of Thompson & Massaro (1986, 1994) bears directly on the issue of gesture/speech integration. In their most recent study, Thompson & Massaro (1994) presented 4-9 year old subjects with two objects, a ball and a doll. In an experimental trial one or other of the objects is referred to by either a verbal label, a pointing gesture or both. The subject's task is to name the object being referred to. The verbal labels were created using synthesised speech which corresponded to the sounds "ball" and "doll" or three sounds intermediate between these words. Similarly the pointing gesture was directed at either the ball or the doll or to three intermediate locations between these objects. Subjects were presented with all instances of the factorial combination of these gesture/speech pairings, and the proportion of "doll" and "ball" responses were computed. The results indicated that whilst children's judgements were more likely to coincide with speech than with gesture, they were nonetheless influenced by the gestural information. This gestural influence was greater for older children and for adults (Thompson & Massaro, 1986) than for the 4 year old children.

Thompson & Massaro (1994) concluded that their data were best described by the assumptions of the so-called Fuzzy Logical Model of Perception (FLMP). In this model the two types of information are evaluated independently and integrated with a rule best described as a multiplicative algorithm. That is, gesture and speech are combined in such a way that the least ambiguous source is more influential in comprehension. The integration process yields a single overall description which is then matched against prototype descriptions in memory. In opting for a multiplicative integration of information, the authors rejected alternative "selection models" which capture similar ideas to those of Rimé in his figure-ground model of verbal-non verbal comprehension (Rimé, 1982; Rimé & Schiaratura, 1991; see also Chapter 4, p.93). In the selection models, the utterance is again analysed into modality specific gesture-speech components but in this case, the judgement is based *only* on the dominant modality (e.g. speech) *unless* this information is not completely intelligible. In this case comprehension is made on the basis of information from the less dominant modality (gesture). Rimé's account is broadly similar although he adds that attention switches from the verbal to the non-verbal dimension when unfamiliar, bizarre or discordant information appears in the gesture channel. The important point to note is that in these latter accounts comprehension is based on only a single modality. In the FLMP information from gesture and speech is *integrated* to achieve understanding.

The idea that information is combined multiplicatively is in accord with the work on dimensional interaction described earlier. Recall that one of the factors which is thought to influence the crosstalk of information between modalities is the relative discriminability of these dimensions (e.g. Garner & Felfoldy, 1970; Melara & Mounts, 1993). When baseline discriminabilities are mismatched the more discriminable (or perhaps least ambiguous) dimension disrupts classification of the less discriminable (or more ambiguous) dimension rather than vice-versa. Thus it is becoming clear that discriminability is an important variable both in promoting interaction between otherwise non-interacting dimensions (e.g. colour and colour-word stimuli) and regulating the nature of the integration of *interacting* dimensions.

Summary

In summary, it is suggested that the symmetrical nature of the interference and the persistence of this effect in the face of the manipulation to the S-R relationship, is best explained within a model of processing where the gestural and verbal stimuli are initially analysed or evaluated in parallel. This information then interacts or is integrated at a semantic level of analysis. Comprehension and response are therefore based on the integration of the two sources of information. Moreover this integration process appears to be sensitive to the relative discriminability, or ambiguity, of the dimensions in question and may well ensure that the least ambiguous source of information contributes more "evidence" to comprehension and response. However, the evidence at present does not allow us to conclude that pointing gestures and speech are fundamentally interacting, as are the dimensions of position and pitch for example. For instance, the possibility remains that mismatches in discriminability are the sole cause of the interaction, rather than simply regulating its nature. Regardless of the answer to this question, it has been suggested that the dynamic, and hence the *variable* property of gestures may well ensure that they are always integrated with verbal information in comprehension.

The integration process included in the FLMP perhaps provides a mechanism for the integration of information described in the ICS framework (see Chapter 7 and below). Both approaches embody the ideas that visual and auditory information are perceptually evaluated in parallel and that the resulting representations are integrated subsequent to this stage.

The ICS Framework

The Interacting Cognitive Subsystems (ICS) approach was introduced in Chapter 7 as a contemporary account which not only incorporates "hot" and "cold" cognition but also seeks to provide a conceptual framework within which *any* aspect of information processing can be explained (Teasdale & Barnard, 1993). The complexity of ICS is such that there is insufficient space here to provide a comprehensive account. Instead the basic processing mechanisms are described followed by a description of the codes and the general structure of the framework.

ICS consists of a number of interrelated, semi-autonomous sub-systems, each devoted to storing and processing a particular type of informational code. Within each subsystem two basic operations occur, those of transformation and storage. First, each of the different codes recognised by ICS can be transformed into other types of code, within the boundary of certain constraints. One such condition dictates that a given transformational process can only handle one type of data stream at a time. The second operation is a process of storage. Each of the nine types of informational code is stored within a memory system specialised for material in only that particular code. This ensures that code patterns which occur regularly can be easily detected and extracted.

ICS recognises nine different informational codes at four levels of abstraction. *Visual*, *auditory* and *body state* codes encode visual and auditory sensory information (e.g. sound frequency, timbre, intensity, or the wavelength or brightness of light), as well as bodily sensations such as pressure, pain, position of the limbs as well as tastes and smells. Visual and auditory sensory codes are abstracted into intermediate structural descriptions known as *object* and *morphonolexical* codes respectively. These codes represent recurring regularities in the patterns of sensory codes, such as the sounds of words or the structures of various objects and spatial relationships. Cross-modal integration of information occurs as recurring regularities and co-occurrences in these structural descriptions are abstracted into higher level *propositional* codes. At this level, meaning is encoded as interrelated propositional statements in a semantic space, such as "dogs bark" or "birds have wings". The cross-modal integration continues as patterns of propositional meaning which have regularly co-occurred with visual and auditory sensory elements are abstracted into *implicational* codes. Meaning at this level is rather "holistic" and captures causal relationships, familiarity and a sense of affect. Finally output in the form of speech and motor activity are controlled by *articulatory* and *limb* codes. These codes and the transformational relationships between them are illustrated in Figure 27.

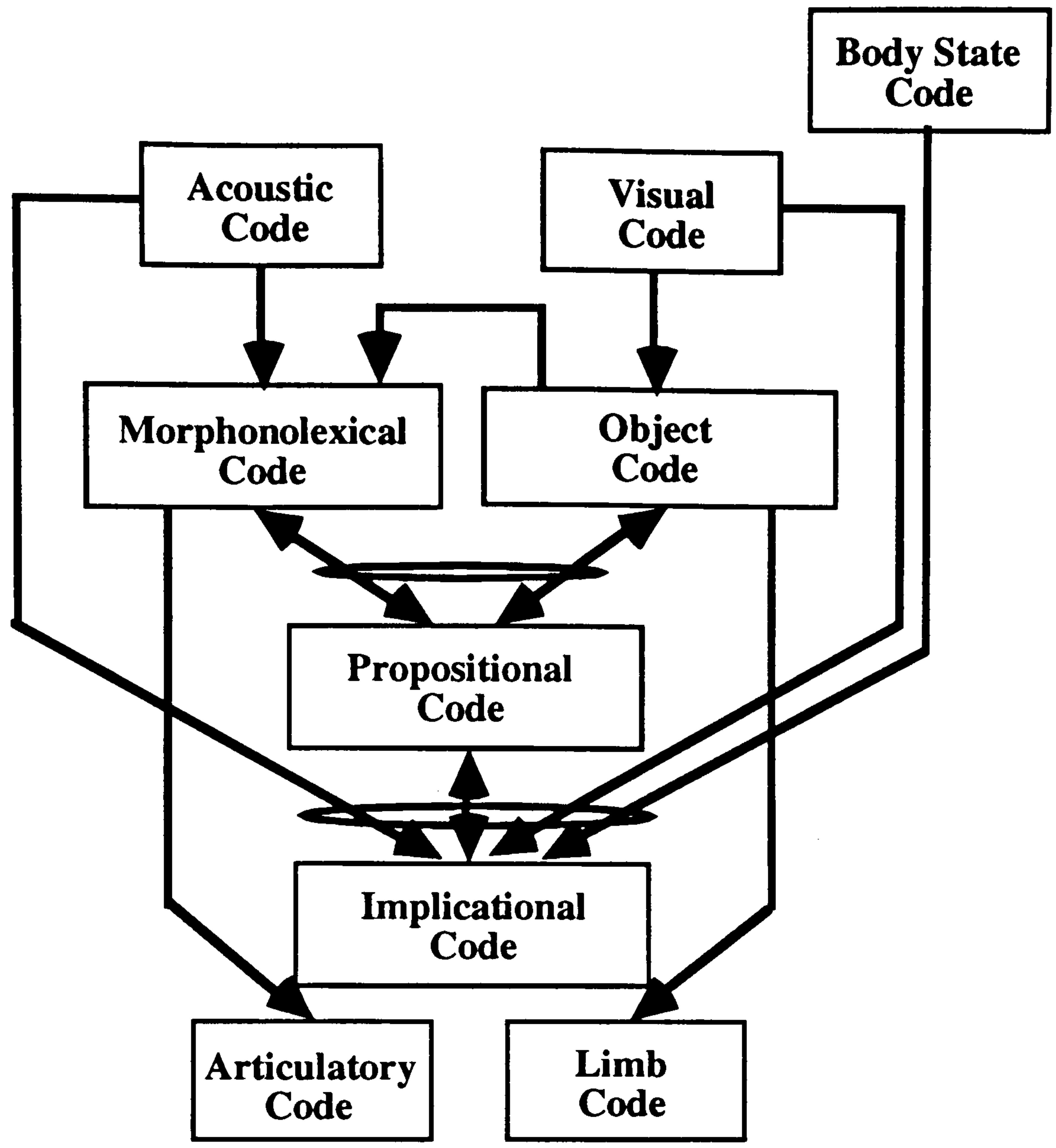


Figure 27. The basic structure and relationships between sensory, structural description, and meaning codes in ICS (the arrows linked by hoops denote that information is integrated at this point).

Informational integration appears to occur at two levels within the system. First, visual and auditory structural descriptions are combined into patterns of propositional code. Secondly, visual and auditory sensory input, along with information from the body state are integrated with propositional information into higher level implicational meanings. As an illustration, the statement "I had a really great weekend" has one meaning as an isolated proposition, but coupled with preceding propositions relating to a series of disastrous events, along with the fact that the statement was offered in a lugubrious tone of voice and with a despondent facial expression, may lead to a very different interpretation. In this situation appropriate morphonolexical representations are extracted from the acoustic code because these particular sequences of sounds have been encountered many times in the past. In a similar way an object code representing the "sad" facial expression is extracted from the memory records of the object subsystem. The outputs from the morphonolexical and object subsystems form an integrated pattern of input into the propositional subsystem. This potentially ambiguous propositional representation together with related propositions representing the context are integrated with acoustic information into an implicational code. This representation carries an appreciation of ironic content and as a result the actual intended meaning of the statement. Further propositional codes referring to sympathetic statements can then be generated from this implicational meaning. Appropriate morphonolexical codes are then generated followed by the articulatory codes perhaps describing the verbal output "ah well, never mind".

Thus ICS incorporates the notion of informational integration across both sensory and intermediate levels of analysis. In particular the idea that morphonolexical and object information can be integrated into a propositional code is similar to the position advocated for the processing and integration of gestural and verbal information. However, as suggested at the beginning of this chapter, the pointing gestures may well be processed spatially to yield information regarding the direction of social attention. According to Teasdale & Barnard (1993) this information can be specified in the object level code. This information is then integrated with verbal information into a composite propositional meaning. In more real world situations, this might correspond to a listener creating an understanding of the referent of a speaker's verbal utterance. For instance, the comprehension of the sentence "look at that penguin" uttered by one Antarctic explorer to another, is only complete if the latter understands to which of the many penguins in view the speaker is referring.

The ICS framework also fits well with the ideas of McNeill et al. (1994) described in relation to their speech-mismatched gestures experiment (see Chapter

2). Recall that when gestural and verbal information was presented in a mismatched form

> "Subjects...do not try to re-create the input to which they were exposed, but they try to form a coherent mental model and introduce changes in memory, where it is necessary to make this possible" (p.234).

The fact that subjects cannot recall the specific gestural or verbal components of the story could be explained in two ways. First, the initial comprehension of the story would involve the integration of gestural and verbal materials into propositional codes. According to the integration process described by Massaro & Friedman (1990) in the FLMP, the integrated representation "is assumed to have no "memory" of how it was obtained" (p.277). If a similar process operates in ICS the information will not be "indexed" by input channel and so will not be available for recall. Second, the initial comprehension of the story would require not only the kind of integration of information mentioned above, but also transformations of propositional codes into higher level "implicational" meanings. At this level propositional meanings are abstracted and incorporated into more holistic "schematic models" based on prior experiences. Again reference to the specific constituents of the message would be discarded in the formation of this more "holistic" level of meaning.

Thus the ICS framework, with the support of an integration process such as that proposed by Massaro & Friedman (1990) seems well equipped to account for symmetrical gesture-speech interactions as well as providing an explanation for mutual influence of verbal and facial information in affective processing (Experiment 17, Chapter 7). In addition, the framework provides us with McNeill's "common computational stage" in both gesture-speech comprehension and production.

Problems with ICS

This being said, there are a number of problems associated with ICS. A first difficulty is in explaining how symmetrical effects can occur between dimensions which require identical recoding transformations. Within ICS, any transformation process can only recode (transform) one coherent data stream at a time. Thus there should be no interference between a pointing gesture and the direction of eye gaze as the processing of both these sources of information requires a visual-object recoding (see Figure 27). The fact that only one of these transformations can occur at a time would seem to suggest either separability (i.e. no interference), or an asymmetrical interaction between such stimuli. Separability would occur if the relevant dimension was processed before the irrelevant stimulus, whilst asymmetry would be the result if the reverse were the case. This is somewhat reminiscent of the relative speed of processing explanations of the Stroop effect which were criticised in Chapter 3. However, such intra-modal interactions have been observed. For instance, in Experiment 3 (Chapter 3) the verbal and gestural stimuli were both presented visually and the symmetry still obtained. In addition, Melara & Marks (1990b; Melara, 1989) have demonstrated robust interactions between words and spatial positions, between colour and brightness, and between spoken words and pitch, all within modality pairs.

In addition to the problem of accounting for within-modality interactions, ICS, as it stands, does not seem able to account for certain cross-modal interactions. Whilst the framework allows for integrations across intermediate structural descriptions into propositional or semantic representations there seems to be no mechanism whereby earlier representations formed in one modality can influence those formed in a second modality. The McGurk effect appears to represent an example of this kind of perceptual crosstalk where the perception of speech sounds can be influenced by concurrent visual information.

A possible solution to both these problems is to divide the visual code in ICS into a number of domain specific codes representing, spatial locations, colour, depth, movement and so forth. In fact there is evidence to suggest that these early visual processes are arranged in a modular fashion (see Humphreys & Bruce, 1989). Similarly, the acoustic codes might be broken down into pitch, timbre and intensity codes etc. In this respect, it seems reasonable to propose that the visual and auditory input subsystems in ICS be further decomposed into subsystems which process and store these different visuo-spatial and auditory codes. The outputs of these subsystems might well be integrated into higher level representations. Similarly at the structural description level, separate entities have been proposed which encode faces and objects, as well as spoken and written words. Clearly if ICS is to account for the range of interactions which have been established, it must be described at a finer-grained level, incorporating additional subsystems with similar notions of interactivity.

An additional difficulty for ICS is in predicting which dimensions will be separable and which will interact. As it stands, ICS seems to integrate information which it has processed together a number of times in the past. Naturally interacting dimensions may well be those containing frequently co-occurring information. Thus, in the past, the system has learnt that certain concepts might be associated with particular verbal, facial and gestural information, and so has learnt to integrate these sources of information in comprehension. A suggestion is, therefore, that in addition to the integral and corresponding dimensions described by Melara & Marks, we might also include certain communicative components as fundamentally interacting dimensions. Whether or not gesture and speech fall into this category is an open question and an issue which is returned to shortly.

Predictions

Despite these problems, ICS nevertheless appears to make a number of predictions concerning the integration of various sources of information. For instance, according to the framework visual and acoustic codes should be integrated into high level meanings at the implicational level. Thus, judgements involving access to this level of information (perhaps affective or dispositional judgements) will be influenced by a combination of paralinguistic factors such as pitch, loudness or speed of verbal input, and low level visual information such as ambient illumination or perhaps the speed and amplitude of hand and/or body movements. Thus we might expect an interference effect if an affective judgement has to be made to a mournful voice accompanied by vigorous gesticulation and vice-versa. For the same reasons, loudness judgements made to a voice should be affected by the amplitude of hand movements and vice-versa. A similar pattern of effects would be expected if a judgement is made to a happy voice experienced in a gloomy room and vice-versa.

According to ICS, other corresponding interactions such as colour and pitch (Melara, 1989) or perhaps vertical position and pitch (Melara & O'Brien, 1987) may also be rooted in integration at the implicational level of processing. Melara & Marks (1990b) suggest that these interactions occur because subjects have difficulty assigning verbal (propositional) labels to dimensions. This confusion arises because the codes for the attribute labels are "partially amodal, that is lacking complete reference to the sensory modality from which they derive" (p.493). The generic implicational codes in ICS also have this amodal property.

The model also predicts more complex high level interactions as semantic information is integrated with sensory information into implicational meanings. So for example, the model might predict that the higher level meaning of the verbal content of an utterance (e.g. an affective judgement) will be influenced by paralinguistic information such as the pitch or loudness of the voice. Whether or not this interaction will be symmetrical is an interesting question. The acoustic code requires transformation into a morphonological code and in turn a transformation into propositional and implicational codes before a response can be made. Thus, it is possible that tone of voice will have more influence on responses to the meaning of the word than word meaning will have on judgements concerning the tone of voice. This fits nicely with data collected by Mehrabian (1972) who noted that tone of voice was more influential than verbal content in certain attitudinal judgements.

Another prediction is that judgements or interactions which rely on transformations between propositional and implicational codes will be disrupted by tasks which are considered to be demanding of central executive resources. An important property of ICS is that executive control and coordination of the system resides in continuing cycles of reciprocal transformations between implicational and propositional codes, the so-called "central engine" of cognition (Teasdale & Barnard, 1993). The task of generating either random sequence of letters or numbers is assumed to occupy central executive resources (Baddeley, 1986), that is they involve propositional-implicational/implicational-propositional transformations. Thus one might expect mutual interference between these kinds of central executive tasks, and tasks involving judgements based on implicational level meanings. This is because both involve implicational-propositional transformations and, as mentioned above, ICS can only handle one particular transformation at a time. So for instance, the judgement of mood from tone of voice and/or gesture amplitude might be disrupted by random number generation and vice-versa. On the other hand, the interaction and *manual* response to pointing gestures and words is assumed to take place at the propositional level. Because implicational-propositional transformations are not required, this kind of task should be less affected by random number generation. Recently Teasdale et al. (1995) have demonstrated that the central executive resources are also involved in the production of so-called stimulus-independent thoughts (SIT's), mental events often experienced as daydreams, which are quite unrelated to immediate sensory input. If this is the case then one might expect fewer incidences of SIT's when subjects are performing tasks involving implicational meanings (e.g. affective judgements) than when tasks are centred on the propositional subsystem (e.g. naming tasks).

Overview

The preceding discussion has attempted to illustrate how, in the study of dimensional interactions, the observation of an interference effect is really only the beginning of the investigation. It is clear that a number of additional factors should be considered all of which can promote dimensional crosstalk in one way or another. Moreover the actual locus of the interaction can vary depending on the nature of the dimensions concerned. Hopefully the discussion has also emphasised the fact that models designed to accommodate interference effects lack the complexity to account for all these variables. In this respect perhaps ICS offers the most potential. As explained earlier, this framework incorporates the ideas of informational integration across different levels of analysis. Sensory information can be combined with propositional codes into more holistic "implicational" meanings, moreover, intermediate structural descriptions encoding visual and auditory information can be integrated into propositional representations. Combining this type of framework with the idea that the integration process is based on the multiplicative combination of different sources of information provides us with a useful conceptual tool for studying the comprehension of communicative elements.

Conclusions and Future Research

In conclusion, it has been demonstrated that pointing gestures and certain facial expressions are actually processed in comprehension. Secondly, it is concluded that the information derived from these stimuli interacts with that obtained from the analysis of the verbal material, prior to response selection and probably between the identification and decision stages of processing. Moreover it has been suggested that the integration process follows a multiplicative rule, such that the least ambiguous or more discriminable dimension makes a larger contribution to the comprehension and decision process. It has also been suggested that the ICS approach offers a promising conceptual framework within which we can incorporate these mutual interactions.

As is so often the case, this research has probably raised more questions than it has answered. A number of these issues have already been raised in the above discussion, but are repeated here. For instance, are gestures and speech fundamentally interacting dimensions? What is the significance of movement in the processing of verbal and non-verbal information? What other communicative dimensions interact, under what circumstances, and at which levels of processing? In addition we might further investigate the nature of the integration process, and relatedly, explore whether gestures confer any redundancy gain when in combination with speech. These issues might be explored, perhaps within a similar interference paradigm, and perhaps using ICS as a conceptual framework. These ideas are elaborated below.

First, we must determine the extent to which pointing gestures and verbal information are fundamentally interacting, or whether the interactions we have noted are caused by other factors which influence selective attention. One possibility, suggested earlier, is to equate baseline discriminabilities and examine the nature of the interference effect over several blocks of experimental trials. Secondly, the role of movement should be examined 1) because it allows us to explore the effects of a wider variety of gestural stimuli, including those with non-spatial attributes and 2) in order to explore the role of movement in prompting dimensional interactions. It may turn out to be the case that, along with corresponding and integral dimensions, certain "communicative" dimensions naturally interact as is surely the case with the McGurk stimuli. In terms of ICS, these interactions may well emerge as values on the dimensions in question regularly "co-occur".

In addition to pointing gestures and verbal material the relationship between facial and verbal information in affective processing has also been examined. However, the inter-relationships between other aspects of verbal and non-verbal behaviour could also be explored. For instance certain non-verbal vocalisations might be expected to interact with verbal, gestural and facial information. To elaborate, prosodic cues carried by the verbal channel express emphasis and help to provide structure by signalling aspects of syntax such as the end of sentence etc. Kendon has attributed similar functions to speech marking gestures which often follow the prosodic contour of the concurrent speech. Integrating prosodic and gestural sources of information together with verbal material would seem to be of use in the syntactic processing of verbal information. So-called paralinguistic aspects of vocalisations express emotions and interpersonal attitudes independently of the content of the speech. These cues might well be integrated with information from facial expression or body posture in the processing of affect. Finally, along with pointing gestures, other cues such as gaze and body posture are important cues to the direction of social attention, important in the comprehension of referential expressions.

A further line of enquiry might be into the integration process itself. Ambiguity could be introduced into either or both of the channels and the subsequent interference effects could be measured. If the integration is multiplicative, as in the FLMP, we might expect that as gestures become more ambiguous they should exert progressively less of an effect on the classification of verbal items and vice-versa. The development of this integration process might also be studied within an interference paradigm. Thompson & Massaro's (1994) data suggest that the ability to integrate gestural and verbal information improves from the ages of 4 years to adult. Gesture and speech might well be separable dimensions for younger children who, as a result, should not experience interference effects of the kind we have described. Finally we might explore how "top down" attentional effects influence the kinds of stimulus-driven effects described above. For instance it would be interesting to determine whether the integration of verbal and non-verbal information is moderated by the orientation of visual attention.

Another issue which relates to the integration process and to some of the work which motivated the present experiments is that of the redundancy of gesture and speech. Recall that this issue is one that separates the positions of say McNeill and Rimé on gesture and speech production. Rimé argues that gesture and speech are redundant in production and so might be also be redundant in comprehension (e.g. Rimé & Schiaratura, 1991). If this is the case, subjects should be no faster at responding to a voice accompanied by a congruent gesture than at making an identical decision to a voice presented in isolation. On the other hand, if gesture and speech are integral components of an utterance, as McNeill (1985) suggests, then we might expect a redundancy gain in congruent conditions. RT's should be faster to a voice accompanied by a congruent gesture than to a voice presented alone. This issue has already been raised in Chapter 4 in the context of the contribution of facilitation and/or inhibition to the effects. The finding of a facilitatory effect (compared to an appropriate neutral condition) might be evidence for a redundancy gain. Whilst such an effect was obtained in Experiment 10 it was nevertheless absent in Experiment 11. This inconsistency was attributed to the problem of selecting the appropriate neutral or control items, a task which some have deemed impossible (Lindsay & Jacoby, 1994). For this reason the inclusion of gesture alone and voice alone conditions would be useful.

This last point has taken us back to the work of Rimé and in particular McNeill, whose theoretical ideas served as the starting point for the experiments presented in this thesis. From the standpoint that a speaker's intended meaning is expressed as a combination of his or her verbal *and* gestural performances it has been demonstrated that, as a result, a listener is likely to process and combine both sources of information in comprehension. In exploring the processes underlying what social psychologists frequently describe as "decoding" (e.g. Argyle, 1988), it becomes clear that verbal and non-verbal information interact in varied and complex ways, contrary to the ideas associated with the notion of gesture as a separate and autonomous "body language". Finally, the hope is to have convinced the reader that it is possible and profitable to study the comprehension of gestural and verbal information from within a cognitive framework.

References

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly journal of Experimental Psychology, 24*, 225-290.

Argyle, M., Lalljee, M., & Cook, M. (1968). The effects of visibility on interaction in a dyad. *Human Relations, 21*, 3-17.

Argyle, M. (1983). *The psychology of interpersoanl behaviour* (4th ed.). Harmondsworth: Penguin.

Argyle, M. (1988). *Bodily Communication* (2nd ed.). London: Methuen.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Barnard, P. (1985). Interacting cognitive subsystems: A psycholinguistic approach to short-term memory. In A. Ellis (Ed.), *Progress in the psychology of language, vol. 2* Hove, UK: Lawrence Erlbaum Associates.

Bavelas, J. B. (1994). Gestures as part of speech: Methodological implications. *Research on Language and Social Interaction, 27*, 201-221.

Beattie, G. W., & Barnard, P. J. (1979). The temporal structure of natural telephone conversations (directory enquiry calls). *Linguistics, 17*, 213-230.

Berman, H. J., Shulman, A. D., & Marvitt, S. J. (1976). Comparison of multidimensional decoding of affect from audio, video and audiovideo recording. *Sociometry, 39*, 83-89.

Birdwhistle, R. (1970). *Kinesics and context: Essays on body-motion communication.* Philadelphia: University of Pennsylvnia Press.

Bower, G. H. (1981). Mood and memory. *The American Psychologist, 36*, 129-148.

Boyle, E. A., Anderson., A. H., & Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech*, *37*, 1-20.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*, 193-209.

Bright, J. E. H., & Burton, A. M. (1994). Past midnight: Semantic processing in an implicit learning task. *Quarterly Journal of Experimental Psychology*, *47A*, 71-79.

Broadbent, D. E. (1958). *Perception and communication*. Oxford: Pergamon Press.

Bruce, V. (1986). Influence of familiarity on the processing of faces. *Perception*, *15*, 387-397.

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305-327.

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., & Rectem, D. (1983). A case of prosopagnosia with some preserved covert remembrance of familiar faces. *Brain and Cognition*, *2*, 257-284.

Butterworth, B., Hine, R. R., & Brady, K. D. (1977). Speech and interaction in sound-only communication channels. *Semiotica*, *20*, 81-99.

Butterworth, B., & Hadar, U. (1989). Gesture, speech and computational stages: A reply to McNeill. *Psychological Review*, *96*, 168-174.

Campbell, R., Landis, T., & Regard, M. (1986). Face recognition and lipreading: A neurological dissociation. *Brain*, *109*, 509-521.

Campbell, R. (1989). Lipreading. In A. W. Young & H. D. Ellis (Eds.), *Handbook of research on face processing* Amsterdam: North-Holland.

Chang, P., & Hammond, G. R. (1987). Mutual interactions between speech and finger movements. *Journal of Motor Behavior, 19*, 265-274.

Clark, H. H., & Brownell, H . H. (1975). Judging up and down. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 213-220.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332-361.

Condon, W. S. (1970). Method of micro-analysis of sound films of behavior. *Behavioral Research Methods and Instrumentation, 2*, 51-54.

Condon, W. S., & Sander, L. W. (1974). Synchrony demonstrated between movements of the neonate and adult speech. *Child Development, 45*, 456-462.

Cook, M., & Lalljee, M. (1972). Verbal substitutes for visual signals in interaction. *Semiotica, 6*, 212-221.

Corballis, M. C., & Beale, I. L. (1976). *The psychology of left and right.* Hillsdale, N.J.: Erlbaum.

Cowan, N., & Barron, A. (1987). Cross-Modal auditory visual Stroop interference and possible implications for speech memory. *Perception and Psychophysics, 41*, 393-401.

Cowan, N. (1989a). A reply to Miles, Madden, and Jones (1989): Mistakes and other flaws in the challenge to the cross-modal Stroop effect. *Perception and Psychophysics, 45*, 82-84.

Cowan, N. (1989b). The reality of cross-modal Stroop effects. *Perception and Psychophysics, 45*, 87-88.

Craft, J. L., & Simon, J. R. (1970). Processing symbolic information from a visual display: Interference from an irrelevant cue. *Journal of Experimental Psychology, 83*, 415-420.

Dalrymple-Alford, E. C., & Budayr, B. (1966). Examination of some aspects of the Stroop Colour-Word test. *Perceptual and Motor Skills, 23*, 1211-1214.

Dalrymple-Alford, E. C. (1972). Associative facilitation and interference in the Stroop colour-word task. *Perception and Psychophysics, 11*, 274-276.

Davis, S. A., Artes, R., & Hoops, R. (1979). Verbal expression and expressive pantomime in aphasic patients. In Y. Lebrun & R. Hoops (Eds.), *Problems of aphasia* Lisse: Swets & Zeitlinger.

De Houwer, J., & Hermans, D (1994). Differences in the affective processing of words and pictures. *Cognition and Emotion, 8*, 1-20.

De Renzi, E., Motti, F., & Nichelli, R. (1980). Imitating gestures: A quantitative approach to ideomotor apraxia. *Archives of Neurology, 37*, 6-10.

DePaulo, B. M., Rosenthal, R., Eisenstat, R. A., Rogers, A. L., & Finkelstein, S. (1978). Decoding discrepant verbal cues. *Journal of Personality and Social Psychology, 36*, 313-323.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*, 80-90.

Dittmann, A. T., & Llewellyn, L. G. (1969). Realtionships between vocalisation and head nods as listener responses. *Journal of Personality and Social Psychology, 9*, 79-84.

Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (submitted). Face-to-face communication and video-mediated communication:

A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied.*

Driver, J., & Tipper, S. P. (1989). On the non-selectivity of selective seeing: Contrasts between interference and priming in selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 304-314.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different colour. Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 622-639.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversation. *Journal of Personality and Social Psychology*, *23*, 283-292.

Duncan, S., & Fiske, D. W. (1977). *Face to face interaction research: Methods and theory*. Hillsdale, NJ: Lawrence Erlbaum.

Dyer, F. N. (1972). Latencies for movement naming with congruent and incongruent word stimuli. *Perception and Psychophysics*, *11*, 377-380.

Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive and response processes. *Memory and Cognition*, *1*, 106-120.

Efron, D. (1941/1972). *Gesture, race and culture*. The Hague: Mouton.

Ekman, P., & Friesen, W. V., (1967). Head and body cues in the judgements of emotions: A reformulation. *Perceptual and Motor Skills*, *24*, 711-724.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, *1*, 49-98.

Ekman, P., & Friesen, W. V. (1972). Hand movements. *Journal of Communication*, *22*, 353-374.

Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). Relative importance of face, body and speech in judgements of personality and affect. *Journal of Personality and Social Psychology, 38*, 270-277.

Ekman, P. (1982). *Emotion in the human face* (2nd ed.). Cambridge: Cambridge University Press.

Ekman, P., & O'Sullivan, M. (1991). Facial expression: Methods, means and moues. In R. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* (pp. 163-199). Cambridge: Cambridge University Press.

Ellis, A. W., & Young, A. W. (1988). *Human cognitive neuropsychology*. Hove, UK: Lawrence Erlbaum Associates.

Erikson, C. W., & Collins, J. F. (1969). Temporal course of selective attention. *Journal of Experimental Psychology, 80*, 254-261.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44*, 227-240.

Farrell, W. S. (1979). Coding left and right. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 42-51.

Feyereisen, P. (1987). Gestures and speech, interactions and separations: A reply to McNeill. *Psychological Review, 94*, 493-498.

Feyereisen, P. (1991). Brain pathology, lateralization, and nonverbal behavior. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* Cambridge: Cambridge University Press.

Feyereisen, P., & de Lannoy, J.-D. (1991). *Gestures and speech: Psychological investigations*. Cambridge: Cambridge University Press.

Feyereisen, P., & Hazan, K. (unpublished study). *Les gestes facilitent-ils la compréhension auditive des mots chez les patients aphasiques.* Cited in R. S. Feldman & B. Rimé (1991). Funamentals of nonverbal behavior. Cambridge: Cambridge University Press.

Feyereisen, P., Van de Wiele, M., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech? *European Bulletin of Cognitive Psychology, 8*, 3-25.

Fitts, P. M., & Seeger, G. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology, 46*, 483-491.

Fox, E. (1995). Pre-cuing target location reduces interference but not negative priming from visual distractors. *Quarterly Journal of Experimental Psychology, 48A*, 26-40.

Fox, L. A., Shor, R. E., & Steinman, R. J. (1971). Semantic gradients and interference in naming color, spatial direction and numerosity. *Journal of Experimental Psychology, 91*, 59-65.

Freedman, N. (1972). The analysis of movement behavior during the clinical interview. In A. R. Siegman & B. Pope (Eds.), *Studies in dyadic communication* Elmsford, NY: Pergamon.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*, 225-241.

Garner, W. R. (1981). The analysis of unanalyzed perceptions. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organisation* (pp. 119-139). Hillsdale, NJ: Erlbaum.

Gerhand, S. J., Deregowski, J. B., & McAllister, H. (1995). Stroop phenomenon as a measure of cognitive functioning of bilingual (Gaelic/English) subjects. *British Journal of Psychology, 86*, 89-92.

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 875-894.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 640-654.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, *118*, 13-42.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1990). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, *100*, 279-297.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology*, *10*, 57-67.

Graham, J. A., & Heywood, S. (1975). The effects of the elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology*, *5*, 189-195.

Greenwald, A. G. (1970). A choice reaction time test of ideomotor theory. *Journal of Experimental Psychology*, *86*, 29-25.

Hasbroucq, T., & Guiard, Y. (1991). Stimulus-response compatibility and the Simon effect: Toward a conceptual clarification. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 246-266.

Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research*, *32*, 203-218.

Hedge, A., & Marsh, N. W. A. (1975). The effects of irrelevant spatial correspondence on two-choice response time. *Acta Psychologica, 39*, 427-439.

Hintzman, D. L., Carre, F. A., Eskridge, V. L., Owens, A. M., Shaff, S. S., & Sparks, M. E. (1972). "Stroop" effect: Input or output phenomenon? *Journal of Experimental Psychology, 95*, 458-459.

Hock, H. S., & Egeth, H. (1970). Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology, 83*, 299-303.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences, 9*, 1-66.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5*, 67-103.

Humphreys, G. W., & Bruce, V. (1989). *Visual Cognition: Computational, experimental and neuropsychological perspectives*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. Long & A. Baddeley (Eds.), *Attention and performance, IX* Hillsdale, NJ: Lawrence Erlbaum Associates.

Keele, S. (1972). Attention demands on memory retrieval. *Journal of Experimental Psychology, 93*, 549-561.

Kelso, J. A. S., Tuller, B., & Harris, K. S. (1983). A "dynamic pattern" perspective on the control and coordination of movement. In P. F. MacNeilage (Ed.), *The production of speech* New York: Springer.

Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, *26*, 22-63.

Kendon, A. (1983). Gesture and speech: How they interact. In J. M. Weimann & R. P. Harrison (Eds.), *Nonverbal interaction* Beverley Hills. CA: Sage Publications.

Kendon, A. (1986). Some reasons for studying gesture. *Semiotica, 62*, 3-28.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, *77*, 576-588.

Kolb, B., & Taylor, L. (1981). Affective behavior in patients with localised cortical excisions: Role of lesion site and side. *Science, 214*, 89-91.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility: A model and a taxonomy. *Psychological Review*, *97*, 253-270.

Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of the Stroop and Simon stimuli. *Psychological Research*, *56*, 130-135.

Krauss, R. M., Apple, W., Morency, N., Wenzel, C., & Winton, W. (1981). Verbal, vocal and visible factors in judgements of another's affect. *Journal of Personality and Social Psychology*, *40*, 312-320.

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational gestures communicate. *Journal of Personality and Social Psychology, 61*, 743-754.

Kurucz, J., & Feldmar, G. (1979). Prosopo-affective agnosia as a symptom of cerebral organic disease. *Journal of the American Geriatrics Society*, *27*, 91-95.

La Heij, W., Dirkx, J., & Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Psychology, 81*, 511-525.

Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language, 24*, 133-164.

Lindsay, S. D., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 219-234.

Lockhead, G. R. (1979). Holistic versus analytic process models: A reply. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 746-755.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12*, 523-553.

Logan, G. D., Zbrodoff, N. J., & Williamson, J. (1984). Strategies in the color-word Stroop task. *Bulletin of the Psychonomic Society, 22*, 135-138.

Logie, R. H. (1985). *Visuo-spatial working memory*. Hove, UK: Lawrence Erlbaum Associates.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory and Cognition, 7*, 485-495.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 126-135.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163-203.

Maki, R. H., Grandy, C. A., & Hauge, G. (1979). Why is telling right from left more difficult than telling above from below? *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 52-67

Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review*, *97*, 225-252.

Mathews, A., & MacLeod, C. (1985). Selective attention of threat cues in anxiety states. *Behaviour Research and Therapy*, *23*, 563-569.

Maxwell, G. M., Cook, M. W., & Burr, R. (1985). The encoding and decoding of liking from behavioral cues in both auditory and visual channels. *Journal of Nonverbal Behavior*, *9*, 239-263.

McClain, L. (1983). Stimulus-response compatibility affects auditory Stroop Interference. *Perception and Psychophysics*, *33*, 266-270.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of the basic findings. *Psychological Review*, *88*, 375-407.

McGeorge, P., & Burton, A. M. (1990). Semantic processing in an incidental learning task. *Quarterly Journal of Experimental Psychology*, *42A*, 597-609.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.

McLeod, P., & Posner, M. I. (1984). Privileged loops from percept to act. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X* Hillsdale, NJ: Erlbaum.

McNeill, D. (1985). So you think gestures are non-verbal? *Psychological Review*, *92*, 350-371.

McNeill, D. (1987a). *Psycholinguistics: A new approach*. New York: Harper & Row.

McNeill, D. (1987b). So you *do* think gestures are non-verbal! Reply to Feyereisen (1987). *Psychological Review, 94*, 499-504.

McNeill, D. (1989). A straight path to where? Reply to Butterworth and Hadar. *Psychological Review, 96*, 175-179.

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

McNeill, D., Cassell, J., & McCullough, K (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction, 27*, 223-237.

Mehrabian, A. (1972). *Nonverbal communication*. Chicago and New York: Aldine-Atherton.

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General, 116*, 323-336.

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 69-79.

Melara, R. D., & Marks, L. E. (1990a). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 398-414.

Melara, R. D., & Marks, L. E. (1990b). Processes underlying dimensional interactions: Correspondences between linguistic and nonlinguistic dimensions. *Memory and Cognition, 18*, 477-495.

Melara, R. D., & Marks, L. E. (1990c). Dimensional interactions in language processing: Investigating directions and levels of crosstalk. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 539-554.

Melara, R. D., & Mounts, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode and practice. *Memory and Cognition, 21*, 627-645.

Miles, C., Madden, C., & Jones, D. M. (1989a). Cross-modal, auditory-visual Stroop interference: A reply to Cowan and Barron (1987). *Perception and Psychophysics, 45*, 77-91.

Miles, C., & Jones, D. M. (1989b). The fallacy of the cross-modal Stroop effect: A rejoinder to Cowan (1989). *Perception and Psychophysics, 45*, 85-86.

Morrel-Samuels, P., & Krauss, R. M. (1994). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 615-622.

Morris, D., Collett, P., Marsh, P., & O'Shaughnessey, M. (1979). *Gestures, their origins and distribution*. London: J. Cape.

Morton, J. (1969). Categories of interference: Verbal mediation and conflict in card sorting. *British Journal of Psychology, 60*, 329-346.

Morton, J., & Chambers, S. M. (1973). Selective attention to words and colours. *Quarterly Journal of Experimental Psychology, 25*, 387-397.

Morton, J., & Patterson, K. E. (1980). A new attempt at an interpretation, or an attempt at a new interpretation. In K. E. Patterson, M. Coltheart, & J. C. Marshall (Eds.), *Deep Dyslexia* London: Routledge and Kegan Paul.

Müller, H. J., & Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performace*, *15*, 315-330.

Nicoletti, R., & Umiltà, C. (1984). Right-left prevalence in spatial compatibility. *Perception and Psychophysics*, *35*, 333-343.

Nicoletti, R., & Umiltà, C. (1994). Attention shifts produce spatial stimulus codes. *Psychological Research*, *56*, 144-150.

O'Leary, M. J., & Barber, P. J. (1993). Interference effects in the Stroop and Simon paradigms. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 830-844.

O'Sullivan, M., Ekman, P., Friesen, W., & Scherer, K. (1985). What you say and how you say it: The contribution of speech content and voice quality to judgements of others. *Journal of Personality and Social Psychology*, *48*, 54-62.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1984). Neurones responsive to faces in the temporal cortex: Studies of functional organistation, sensitivity to identity and relation to perception. *Human Neurobiology*, *3*, 197-208.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London*, *B223*, 293-317.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organisation and functions of cells responsive to faces in temporal cortex. *Philosophical Transactions of the Royal Society of London*, *B335*, 23-30.

Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General, 112*, 515-540.

Pomerantz, J. R. (1986). Visual form perception: An overview. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines: Visual perception.* (pp. 1-30). New York: Academic Press.

Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition* Hillsdale, NJ: Erlbaum.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3-25.

Rayner, K., & Posnansky, C. J. (1978). Stages of processing in word identification. *Journal of Experimental Psychology: General, 107*, 64-80.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*, 115-119.

Redding, G. M., & Gerjets, D. A. (1977). Stroop effect: Interference and facilitation with verbal and manual responses. *Perceptual and Motor Skills, 45*, 11-17.

Ricci-Bitti, P. E., & Poggi, I. (1991). Symbolic nonverbal behavior: Talking through gestures. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* Cambridge: Cambridge University Press.

Riddoch, M. J., Humphreys, G. W., Coltheart, M., & Funnell, E. (1988). Semantic systems or system? Neuropsychological evidence re-examined. *Cognitive Neuropsychology, 5*, 3-25.

Riggio, R. E., & Friedman, H. S. (1986). Impression formation: The role of expressive behavior. *Journal of Personality and Social Psychology, 50*, 421-427.

Rimé, B. (1982a). The elimination of visible behavior from social interactions: Effects on verbal, nonverbal and interpersonal variables. *European Journal of Social Psychology, 12*, 113-129.

Rimé, B., & Gaussin, J. (1982b). Sensibilité des comportments non verbaux aux variations de la densité de communication. *L'Année Psychologique, 82*, 173-187.

Rimé, B. (1983). Nonverbal communication or nonverbal behavior? Towards a cognitive-motor theory of nonverbal behavior. In W. Doise & S. Moscovici (Eds.), *Current issues in European social psychology* Cambridge: Cambridge University Press.

Rimé, B., & Schiaratura, L. (1991). Gesture and speech. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* Cambridge: Cambridge University Press.

Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior, 5*, 172-183.

Roe, W. T., Wilsoncroft, W. E., & Griffiths, R. S. (1980). Effects of motor and verbal practice on the Stroop task. *Perceptual and Motor Skills, 50*, 647-650.

Rogers, W. T. (1978). The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances. *Human Communication Research, 5*, 54-62.

Rosenthal, R., & DePaulo, B. (1979). Sex differences in eavesdropping on nonverbal cues. *Journal of Personality and Social Psychology, 37*, 273-285.

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development, 46*, 247-253.

Rosinski, R. R. (1977). Picture-word interference is semantically based. *Child Development*, *48*, 643-647.

Rothi, L. J. G., Mack, L., & Heilman, K. M. (1986). Pantomime agnosia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *49*, 451-454.

Rutter, D. R., & Stevenson, G. M. (1977). The role of visual communication in synchronising conversation. *European Journal of Social Psychology*, *7*, 29-37.

Rutter, D. R. (1987). *Communicating by telephone*. Oxford: Pergamon.

Sapir, E. (1927). The unconscious patterning of behavior in society. In E. S. Dummer (Ed.), *The unconscious: A symposium* New York: Knopf.

Schriefers, H., & Meyer, A. S. (1990). Experimental note: Cross-modal, visual-auditory picture-word interference. *Bulletin of the Psychonomic Society*, *28*, 418-420.

Schwartz, S. P., Pomerantz, J. R., & Egeth, H. E. (1977). State and process limitations in information processing: An additive factors analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 402-410.

Seymour, P. H. (1973a). Stroop interference in naming and verifying spatial locations. *Perception and Psychophysics*, *14*, 95-100.

Seymour, P. H. K. (1973b). A model for reading, naming and comparison. *British Journal of Psychology*, *64*, 35-49.

Seymour, P. H. (1977). Conceptual encoding and locus of the Stroop effect. *Quarterly Journal of Experimental Psychology*, *29*, 245-265.

Shaffer, L. H. (1975). Multiple attention in continuous verbal tasks. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance V* London: Academic Press.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge: Cambridge University Press.

Shimada, H. (1990). Effect of auditory presentation of words on color naming: the intermodal Stroop effect. *Perceptual and Motor Skills, 70,* 1155-1161.

Sholl, M. J., & Egeth, H. E. (1981). Right-left confusion in the adult: A verbal labelling effect. *Memory and Cognition, 9,* 339-350.

Shor, R. E. (1970). The processing of conceptual information on spatial directions from pictorial and linguistic symbols. *Acta Psychologica, 32,* 346-365.

Shor, R. E. (1971). Symbol processing speed differences and symbol interference effects in a variety of concept domains. *Journal of General Psychology, 85,* 187-205.

Simon, J. R., & Rudell, A. P. (1967). The effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51,* 300-304.

Simon, J. R. (1969). Reactions towards the source of stimulation,. *Journal of Experimental Psychology, 81,* 174-176.

Simon, J. R., Sly, P. E., & Vilapakkam, S. (1981). Effects of compatibility of S-R mapping on reaction toward the stimulus source. *Acta Psychologica, 47,* 63-81.

Simon, J. R. (1982). Effects of an auditory stimulus on the processing of a visual stimulus under single and dual task conditions. *Acta Psychologica, 51,* 61-73.

Simon, J. R., & Berbaum, K. (1990). Effects of conflicting cues on information processing: The "Stroop effect" vs. the "Simon effect". *Acta Psychologica, 73,* 159-170.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General, 109,* 373-392.

Smyth, M. M., Pearson, N. A., & Pendleton, L. R. (1988). Movement and working memory: Patterns and positions in space. *Quarterly Journal of Experimental Psychology, 40A*, 497-514.

Smyth, M. M., & Pendleton, L. R. (1989). Working memory for movements. *Quarterly Journal of Experimental Psychology, 41A*, 235-250.

Stanovich, K. E., & Pachella, R. G. (1977). Encoding, stimulus-response compatibility, and stages of processing. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 411-421.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30*, 276-315.

Stirling, N. (1979). Stroop interference: An input and an output phenomenon. *Quarterly Journal of Experimental Psychology, 31*, 121-132.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Stuart, D. M., & Carrasco, M. (1993). Semantic component of a cross-modal Stroop-like task. *American Journal of Psychology, 106*, 383-405.

Sugg, M. J., & McDonald, J. E. (1994). Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 647-675.

Teasdale, J. D., & Barnard, P. J. (1993). *Affect, cognition, and change: Re-modelling depressive thought.* Hove: Lawrence Erlbaum Associates.

Teasdale, J. D., Dritschel, B. H., Taylor, M. J., Proctor, L., Lloyd, C. A., Nimmo-Smith, I., & Baddeley, A. D. (1995). Stimulus independent thought depends on central executive resources. *Memory and Cognition, 23*, 551-559.

Thompson, L. A., & Massaro, D. W. (1986). Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology, 42*, 144-168.

Thompson, L. A., & Massaro, D. W. (1994). Children's Integration of Speech and Gestures in Comprehension. *Journal of Experimental Child Psychology, 57*, 327-354.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 14*, 107-141.

Trimboli, A., & Walker, M. B. (1987). Nonverbal dominance in the communication of affect: A myth? *Journal of Nonverbal Behavior, 11*, 180-190.

Umiltà, C., & Nicoletti, R. (1992). An integrated model of the Simon effect. In D. Holender, J. Alegria, J. Junca de Morais, & M. Radeau (Eds.), *Analytic approaches to human cognition* Amsterdam: Elsevier.

Underwood, G. (1976). Semantic interference from unattended printed words. *British Journal of Psychology, 67*, 327-338.

Venus, C. A., & Canter, G. J. (1987). The effect of redundant cues on the comprehension of spoken messages by aphasic adults. *Journal of Communication Disorders, 20*, 477-491.

Verfaellie, M., Bowers, D., & Heilman, K. M. (1988). Attentional factors in the occurrence of stimulus-response compatibility effects. *Neuropsychologia, 26*, 435-444.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory and Cognition, 13*, 304-319.

Walker, S., Bruce, V., & O'Malley, C. (1995). Facial identity and facial speech processing: familiar faces and voices in the McGurk effect. *Perception and Psychophysics*, *57*, 1124-1133.

Wallace, R. J. (1971). S-R compatibility and the idea of a response code. *Journal of Experimental Psychology*, *88*, 354-360.

White, B. W. (1969). Interference in identifying attributes and attribute names. *Perception and Psychophysics*, *6*, 166-168.

Wolff, C. (1945). *A Psychology of Gesture*. London: Methuen.

Woodall, W. G., & Folger, J. P. (1981). Encoding specificity and nonverbal cue context: An expansion of episodic memory research. *Communication Monographs*, *48*, 39-53.

Woodall, W. G., & Folger, J. P. (1985). Nonverbal cue context and episodic memory: On the availability and endurance of nonverbal behaviors as retrieval cues. *Communication Monographs*, *52*, 319-333.

Young, A. W., McWeeny, K. H., Hay, D. C., & Ellis, A. W. (1986). Matching familiar and unfamiliar faces on identity and expression. *Psychological Research*, *48*, 63-68.

Zuckerman, M., Blanck, P. D., DePaulo, B. M., & Rosenthal, R. (1980). Developmental changes in decoding discrepant and nondiscrepant nonverbal cues. *Developmental Psychology*, *16*, 220-228.

Appendix A

The positive, negative and neutral words used as stimuli in Experiment 17.

<u>Positive</u>	<u>Negative</u>	<u>Neutral</u>
Ecstatic	Death	Book
Elation	Doom	Carpet
Hapiness	Evil	Chair
Joy	Grotesque	Computer
Laughter	Hatred	Paper
Love	Horror	Table
Paradise	Plague	Phone
Pretty	Torment	Watch

Appendix B

ANOVA Summary Tables

Chapter 3

Experiment 1

SOURCE: grand mean

respo	cong.	decis	N	MEAN	SD	SE
			96	788.4583	279.0274	28.4781

SOURCE: response

respo	cong.	decis	N	MEAN	SD	SE
voice			48	869.4792	248.9948	35.9393
gest			48	707.4375	286.2911	41.3226

SOURCE: cong.

respo	cong.	decis	N	MEAN	SD	SE
	cong		48	736.1667	254.3604	36.7138
	incon		48	840.7500	295.1012	42.5942

SOURCE: response cong.

respo	cong.	decis	N	MEAN	SD	SE
voice	cong		24	801.8333	233.0780	47.5769
voice	incon		24	937.1250	250.6337	51.1604
gest	cong		24	670.5000	262.4631	53.5751
gest	incon		24	744.3750	309.4422	63.1646

SOURCE: decision

respo	cong.	decis	N	MEAN	SD	SE
		u/d	48	788.1875	243.1634	35.0976
		l/r	48	788.7292	313.4334	45.2402

SOURCE: response decision

respo	cong.	decis	N	MEAN	SD	SE
voice		u/d	24	833.5833	209.5430	42.7728
voice		l/r	24	905.3750	283.0108	57.7693
gest		u/d	24	742.7917	269.4786	55.0071
gest		l/r	24	672.0833	303.7459	62.0019

SOURCE: cong. decision

respo	cong.	decis	N	MEAN	SD	SE
	cong	u/d	24	741.8333	250.1223	51.0560
	cong	l/r	24	730.5000	263.7863	53.8452
	incon	u/d	24	834.5417	231.9100	47.3384
	incon	l/r	24	846.9583	352.2684	71.9065

SOURCE: response cong. decision

respo	cong.	decis	N	MEAN	SD	SE
voice	cong	u/d	12	787.6667	220.5462	63.6662
voice	cong	l/r	12	816.0000	253.9900	73.3206
voice	incon	u/d	12	879.5000	196.3862	56.6918
voice	incon	l/r	12	994.7500	292.4595	84.4258
gest	cong	u/d	12	696.0000	278.5423	80.4082
gest	cong	l/r	12	645.0000	255.0130	73.6159
gest	incon	u/d	12	789.5833	263.5835	76.0900
gest	incon	l/r	12	699.1667	355.3569	102.5827

FACTOR:	subs	response	cong.	decision	r.t
LEVELS:	12	2	2	2	96
TYPE :	RANDOM	WITHIN	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	630180.0417	1	630180.0417	12.472	0.005	**
rs/	555807.2083	11	50527.9280			
cong.	262504.1667	1	262504.1667	27.679	0.000	***
cs/	104323.0833	11	9483.9167			
rc	22632.0417	1	22632.0417	1.243	0.289	
rcs/	200306.2083	11	18209.6553			
decisio	7.0417	1	7.0417	0.001	0.980	
ds/	113269.7083	11	10297.2462			

Experiment 2

SOURCE: grand mean

respo	cong	decis	N	MEAN	SD	SE
			112	705.6339	224.2642	21.1910

SOURCE: response

respo	cong	decis	N	MEAN	SD	SE
voice			56	749.2321	174.8884	23.3704
gest			56	662.0357	258.9335	34.6015

SOURCE: cong

respo	cong	decis	N	MEAN	SD	SE
	cong		56	677.3571	207.8692	27.7777
	incon		56	733.9107	238.0450	31.8101

SOURCE: response cong

respo	cong	decis	N	MEAN	SD	SE
voice	cong		28	712.9643	165.4835	31.2734
voice	incon		28	785.5000	179.4208	33.9073
gest	cong		28	641.7500	240.8426	45.5150
gest	incon		28	682.3214	278.7790	52.6843

SOURCE: decision

respo	cong	decis	N	MEAN	SD	SE
		u/d	56	704.6429	204.3584	27.3085
		l/r	56	706.6250	244.4155	32.6614

SOURCE: response decision

respo	cong	decis	N	MEAN	SD	SE
voice		u/d	28	730.0714	166.4908	31.4638
voice		l/r	28	768.3929	183.9133	34.7563
gest		u/d	28	679.2143	236.6667	44.7258
gest		l/r	28	644.8571	282.7599	53.4366

SOURCE: cong decision

respo	cong	decis	N	MEAN	SD	SE
	cong	u/d	28	669.1786	187.8808	35.5061
	cong	l/r	28	685.5357	229.3069	43.3349
	incon	u/d	28	740.1071	217.1719	41.0416
	incon	l/r	28	727.7143	261.1250	49.3480

SOURCE: response cong decision

respo	cong	decis	N	MEAN	SD	SE
voice	cong	u/d	14	686.7857	153.6065	41.0531
voice	cong	l/r	14	739.1429	178.3397	47.6633
voice	incon	u/d	14	773.3571	173.0321	46.2448
voice	incon	l/r	14	797.6429	191.3171	51.1317
gest	cong	u/d	14	651.5714	221.4747	59.1916
gest	cong	l/r	14	631.9286	266.8585	71.3209
gest	incon	u/d	14	706.8571	256.1913	68.4700
gest	incon	l/r	14	657.7857	307.3810	82.1510

FACTOR:	subs	response	cong	decision	r.t
LEVELS:	14	2	2	2	112
TYPE :	RANDOM	WITHIN	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	212890.0804	1	212890.0804	6.112	0.028	*
rs/	452826.2946	13	34832.7919			
cong	89552.5804	1	89552.5804	21.435	0.000	***
cs/	54312.7946	13	4177.9073			
rc	7152.0089	1	7152.0089	1.520	0.239	
rcs/	61155.8661	13	4704.2974			
decisio	110.0089	1	110.0089	0.016	0.901	
ds/	88472.8661	13	6805.6051			

Experiment 3

SOURCE: grand mean

respo	cong.	decis	N	MEAN	SD	SE
			112	660.7946	170.9127	16.1497

SOURCE: response

respo	cong.	decis	N	MEAN	SD	SE
gest			56	625.2500	157.9596	21.1082
word			56	696.3393	177.2834	23.6905

SOURCE: cong.

respo	cong.	decis	N	MEAN	SD	SE
	cong		56	623.7143	144.8452	19.3558
	incon		56	697.8750	187.5457	25.0618

SOURCE: response cong.

respo	cong.	decis	N	MEAN	SD	SE
gest	cong		28	593.4286	127.1943	24.0375
gest	incon		28	657.0714	180.4106	34.0944
word	cong		28	654.0000	157.0239	29.6747
word	incon		28	738.6786	188.8076	35.6813

SOURCE: decision

respo	cong.	decis	N	MEAN	SD	SE
		l/r	56	636.1786	164.0498	21.9221
		u/d	56	685.4107	175.5197	23.4548

SOURCE: response decision

respo	cong.	decis	N	MEAN	SD	SE
gest		l/r	28	594.6786	155.6044	29.4065
gest		u/d	28	655.8214	157.0841	29.6861
word		l/r	28	677.6786	164.4281	31.0740
word		u/d	28	715.0000	190.4311	35.9881

SOURCE: cong. decision

respo	cong.	decis	N	MEAN	SD	SE
	cong	l/r	28	597.2857	133.4907	25.2274
	cong	u/d	28	650.1429	153.1955	28.9512
	incon	l/r	28	675.0714	184.0225	34.7770
	incon	u/d	28	720.6786	191.5899	36.2071

SOURCE: response cong. decision

respo	cong.	decis	N	MEAN	SD	SE
gest	cong	l/r	14	561.3571	124.0674	33.1584
gest	cong	u/d	14	625.5000	126.4642	33.7990
gest	incon	l/r	14	628.0000	180.2882	48.1841
gest	incon	u/d	14	686.1429	182.4158	48.7527
word	cong	l/r	14	633.2143	137.2489	36.6813
word	cong	u/d	14	674.7857	177.3186	47.3904
word	incon	l/r	14	722.1429	181.8168	48.5926
word	incon	u/d	14	755.2143	200.9788	53.7139

FACTOR:	subs	response	cong.	decision	r.t
LEVELS:	14	2	2	2	112
TYPE :	RANDOM	WITHIN	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	141503.2232	1	141503.2232	1.703	0.215	
rs/	1080469.4018	13	83113.0309			
cong.	153994.7232	1	153994.7232	41.094	0.000	***
cs/	48715.9018	13	3747.3771			
rc	3097.5089	1	3097.5089	0.379	0.549	
rcs/	106270.6161	13	8174.6628			
decisio	67866.5089	1	67866.5089	4.185	0.062	
ds/	210809.1161	13	16216.0859			
rd	3972.2232	1	3972.2232	1.010	0.333	
rds/	51130.9018	13	3933.1463			
cd	367.9375	1	367.9375	0.072	0.792	
cds/	66210.1875	13	5093.0913			
rcd	10.9375	1	10.9375	0.003	0.960	
rcds/	54057.6875	13	4158.2837			

Experiment 4

SOURCE: grand mean

respo	cong	decis	N	MEAN	SD	SE
			112	665.3482	165.1297	15.6033

SOURCE: response

respo	cong	decis	N	MEAN	SD	SE
voice			56	707.1250	151.7065	20.2726
gest			56	623.5714	168.7082	22.5446

SOURCE: cong

respo	cong	decis	N	MEAN	SD	SE
	cong		56	623.0536	140.3895	18.7603
	incon		56	707.6429	177.9873	23.7846

SOURCE: response cong

respo	cong	decis	N	MEAN	SD	SE
voice	cong		28	664.6786	134.2885	25.3781
voice	incon		28	749.5714	158.4673	29.9475
gest	cong		28	581.4286	136.0932	25.7192
gest	incon		28	665.7143	189.1405	35.7442

SOURCE: decision

respo	cong	decis	N	MEAN	SD	SE
		u/d	56	669.8571	161.7776	21.6184
		l/r	56	660.8393	169.7587	22.6850

SOURCE: response decision

respo	cong	decis	N	MEAN	SD	SE
voice		u/d	28	699.1786	162.4673	30.7034
voice		l/r	28	715.0714	142.6729	26.9626
gest		u/d	28	640.5357	158.5386	29.9610
gest		l/r	28	606.6071	179.5765	33.9368

SOURCE: cong decision

respo	cong	decis	N	MEAN	SD	SE
	cong	u/d	28	625.5714	128.9423	24.3678
	cong	l/r	28	620.5357	153.3269	28.9761
	incon	u/d	28	714.1429	180.6086	34.1318
	incon	l/r	28	701.1429	178.3961	33.7137

SOURCE: response cong decision

respo	cong	decis	N	MEAN	SD	SE
voice	cong	u/d	14	650.7143	131.9880	35.2753
voice	cong	l/r	14	678.6429	140.0470	37.4291
voice	incon	u/d	14	747.6429	179.8386	48.0639
voice	incon	l/r	14	751.5000	140.7320	37.6122
gest	cong	u/d	14	600.4286	125.4939	33.5397
gest	cong	l/r	14	562.4286	148.1256	39.5882
gest	incon	u/d	14	680.6429	181.6288	48.5423
gest	incon	l/r	14	650.7857	202.0671	54.0047

FACTOR:	subs	response	cong	decision	rt
LEVELS:	14	2	2	2	112
TYPE :	RANDOM	WITHIN	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	195473.5804	1	195473.5804	7.698	0.016	*
rs/	330109.7946	13	25393.0611			
cong	200349.7232	1	200349.7232	18.132	0.001	***
cs/	143647.1518	13	11049.7809			
rc	2.5804	1	2.5804	0.001	0.979	
rcs/	45746.2946	13	3518.9457			
decisio	2277.0089	1	2277.0089	1.079	0.318	
ds/	27431.3661	13	2110.1051			
rd	17375.2232	1	17375.2232	10.816	0.006	**
rds/	20883.1518	13	1606.3963			
cd	444.0089	1	444.0089	0.307	0.589	
cds/	18821.8661	13	1447.8359			
rcd	1816.0804	1	1816.0804	0.415	0.531	
rcds/	56935.7946	13	4379.6765			

Chapter 4

Experiment 5: Responses to Voice

SOURCE: grand mean

cong	decis	N	MEAN	SD	SE
		72	734.8472	107.9804	12.7256

SOURCE: cong

cong	decis	N	MEAN	SD	SE
cong		36	718.3056	109.0640	18.1773
incon		36	751.3889	105.8056	17.6343

SOURCE: decision

cong	decis	N	MEAN	SD	SE
	u/d	36	697.5833	112.2450	18.7075
	l/r	36	772.1111	90.5383	15.0897

SOURCE: cong decision

cong	decis	N	MEAN	SD	SE
cong	u/d	18	671.9444	109.9863	25.9240
cong	l/r	18	764.6667	88.5498	20.8714
incon	u/d	18	723.2222	111.5793	26.2995
incon	l/r	18	779.5556	94.4359	22.2588

FACTOR:	subs	cong	decision	r.t
LEVELS:	18	2	2	72
TYPE :	RANDOM	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
cong	19701.1250	1	19701.1250	11.684	0.003	**
cs/	28664.1250	17	1686.1250			
decisio	99979.0139	1	99979.0139	18.969	0.000	***
ds/	89599.2361	17	5270.5433			
cd	5958.6806	1	5958.6806	4.960	0.040	*
cds/	20421.5694	17	1201.2688			

Experiment 5: Responses to Gesture

SOURCE: grand mean

cong	decis	N	MEAN	SD	SE
		72	948.5556	287.6497	33.8998

SOURCE: cong

cong	decis	N	MEAN	SD	SE
cong		36	894.7500	248.6680	41.4447
incon		36	1002.3611	316.3187	52.7198

SOURCE: decision

cong	decis	N	MEAN	SD	SE
	u/d	36	869.0278	278.4637	46.4106
	l/r	36	1028.0833	278.0216	46.3369

SOURCE: cong decision

cong	decis	N	MEAN	SD	SE
cong	u/d	18	808.5000	208.1208	49.0545
cong	l/r	18	981.0000	261.2306	61.5726
incon	u/d	18	929.5556	329.5039	77.6648
incon	l/r	18	1075.1667	293.6034	69.2030

FACTOR:	subs	cong	decision	r.t
LEVELS:	18	2	2	72
TYPE :	RANDOM	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
cong	208442.7222	1	208442.7222	12.542	0.003	**
cs/	282529.7778	17	16619.3987			
decisio	455376.0556	1	455376.0556	25.026	0.000	***
ds/	309328.4444	17	18195.7908			
cd	3253.5556	1	3253.5556	0.370	0.551	
cds/	149375.9444	17	8786.8203			

Experiment 6: Responses to Voice

SOURCE: grand mean

cong	decis	N	MEAN	SD	SE
		72	771.6111	127.3828	15.0122

SOURCE: cong

cong	decis	N	MEAN	SD	SE
cong		36	758.0000	120.9562	20.1594
incon		36	785.2222	133.8091	22.3015

SOURCE: decision

cong	decis	N	MEAN	SD	SE
	u/d	36	792.8056	149.6873	24.9479
	l/r	36	750.4167	97.9082	16.3180

SOURCE: cong decision

cong	decis	N	MEAN	SD	SE
cong	u/d	18	780.5000	135.5461	31.9485
cong	l/r	18	735.5000	103.3277	24.3546
incon	u/d	18	805.1111	165.6417	39.0421
incon	l/r	18	765.3333	92.6721	21.8430

FACTOR:	subs	cong	decision	r.t
LEVELS:	18	2	2	72
TYPE :	RANDOM	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p
cong	13338.8889	1	13338.8889	3.587	0.075
cs/	63216.1111	17	3718.5948		
decisio	32342.7222	1	32342.7222	4.391	0.051
ds/	125217.2778	17	7365.7222		
cd	122.7222	1	122.7222	0.043	0.837
cds/	48031.2778	17	2825.3693		

Experiment 6: Responses to Gesture

SOURCE: grand mean

cong	decis	N	MEAN	SD	SE
		72	905.5417	260.9296	30.7509

SOURCE: cong

cong	decis	N	MEAN	SD	SE
cong		36	865.9722	257.2154	42.8692
incon		36	945.1111	262.1700	43.6950

SOURCE: decision

cong	decis	N	MEAN	SD	SE
	u/d	36	968.2778	255.5999	42.6000
	l/r	36	842.8056	254.3344	42.3891

SOURCE: cong decision

cong	decis	N	MEAN	SD	SE
cong	u/d	18	922.9444	238.2042	56.1453
cong	l/r	18	809.0000	269.4373	63.5070
incon	u/d	18	1013.6111	270.9472	63.8629
incon	l/r	18	876.6111	241.1649	56.8431

FACTOR:	subs	cong	decision	r.t
LEVELS:	18	2	2	72
TYPE :	RANDOM	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
cong	112733.3472	1	112733.3472	6.918	0.018	*
cs/	277006.9028	17	16294.5237			
decisio	283379.0139	1	283379.0139	16.281	0.001	***
ds/	295892.2361	17	17405.4257			
cd	2392.0139	1	2392.0139	0.144	0.709	
cds/	282386.2361	17	16610.9551			

Experiment 7

SOURCE: grand mean

compa	cong	N	MEAN	SD	SE
		64	617.9063	126.2768	15.7846

SOURCE: compat

compa	cong	N	MEAN	SD	SE
comp		32	582.7500	104.2175	18.4232
incom		32	653.0625	137.8154	24.3626

SOURCE: cong

compa	cong	N	MEAN	SD	SE
	cong	32	608.1563	133.9162	23.6733
	incon	32	627.6563	119.4832	21.1218

SOURCE: compat cong

compa	cong	N	MEAN	SD	SE
comp	cong	16	558.8125	109.4947	27.3737
comp	incon	16	606.6875	96.0998	24.0249
incom	cong	16	657.5000	140.9941	35.2485
incom	incon	16	648.6250	139.0361	34.7590

FACTOR:	subs	compat	cong	r.t
LEVELS:	32	2	2	64
TYPE :	RANDOM	BETWEEN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
compat	79101.5625	1	79101.5625	2.707	0.110	
s/c	876638.8750	30	29221.2958			
cong	6084.0000	1	6084.0000	6.108	0.019	*
cs/c	29880.7500	30	996.0250			
cc	12882.2500	1	12882.2500	12.934	0.001	**
cs/c	29880.7500	30	996.0250			

Experiment 8

SOURCE: grand mean

compa	cong	N	MEAN	SD	SE
		64	542.6406	74.9223	9.3653

SOURCE: compat

compa	cong	N	MEAN	SD	SE
comp		32	545.3438	84.7819	14.9875
incom		32	539.9375	64.8437	11.4629

SOURCE: cong

compa	cong	N	MEAN	SD	SE
	cong	32	528.8438	71.6916	12.6734
	incon	32	556.4375	76.6492	13.5498

SOURCE: compat cong

compa	cong	N	MEAN	SD	SE
comp	cong	16	523.5000	74.3281	18.5820
comp	incon	16	567.1875	91.1733	22.7933
incom	cong	16	534.1875	70.9678	17.7420
incom	incon	16	545.6875	59.8562	14.9640

FACTOR:	subs	compat	cong	r.t
LEVELS:	32	2	2	64
TYPE :	RANDOM	BETWEEN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
compat	467.6406	1	467.6406	0.044	0.835	
s/c	316202.5938	30	10540.0865			
cong	12182.6406	1	12182.6406	17.704	0.000	***
cs/c	20643.7188	30	688.1240			
cc	4144.1406	1	4144.1406	6.022	0.020	*
cs/c	20643.7188	30	688.1240			

Experiment 9

SOURCE: grand mean

respo	cong	decis	N	MEAN	SD	SE
			256	572.3711	124.2130	7.7633

SOURCE: response

respo	cong	decis	N	MEAN	SD	SE
gest			128	644.6953	119.5479	10.5666
word			128	500.0469	78.3808	6.9280

SOURCE: cong

respo	cong	decis	N	MEAN	SD	SE
	cong		128	557.4922	115.2148	10.1836
	incon		128	587.2500	131.3716	11.6117

SOURCE: response cong

respo	cong	decis	N	MEAN	SD	SE
gest	cong		64	622.0156	112.8656	14.1082
gest	incon		64	667.3750	122.5827	15.3228
word	cong		64	492.9688	74.5805	9.3226
word	incon		64	507.1250	81.9792	10.2474

SOURCE: decision

respo	cong	decis	N	MEAN	SD	SE
		l/r	128	581.8594	130.8868	11.5689
		u/d	128	562.8828	116.9033	10.3329

SOURCE: response decision

respo	cong	decis	N	MEAN	SD	SE
gest		l/r	64	666.8281	119.5243	14.9405
gest		u/d	64	622.5625	116.3139	14.5392
word		l/r	64	496.8906	74.6996	9.3375
word		u/d	64	503.2031	82.3672	10.2959

SOURCE: cong decision

respo	cong	decis	N	MEAN	SD	SE
	cong	l/r	64	570.1094	124.6057	15.5757
	cong	u/d	64	544.8750	104.4490	13.0561
	incon	l/r	64	593.6094	136.8487	17.1061
	incon	u/d	64	580.8906	126.4168	15.8021

SOURCE: response cong decision

respo	cong	decis	N	MEAN	SD	SE
gest	cong	l/r	32	645.7500	121.4387	21.4675
gest	cong	u/d	32	598.2813	99.8894	17.6581
gest	incon	l/r	32	687.9063	115.6216	20.4392
gest	incon	u/d	32	646.8438	127.6682	22.5688
word	cong	l/r	32	494.4688	70.6719	12.4931
word	cong	u/d	32	491.4688	79.4026	14.0365
word	incon	l/r	32	499.3125	79.5830	14.0684
word	incon	u/d	32	514.9375	84.8440	14.9985

FACTOR:	subs	response	cong	decision	r.t
LEVELS:	32	2	2	2	256
TYPE :	RANDOM	WITHIN	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	1339082.9102	1	1339082.9102	117.028	0.000	***
rs/	354713.4648	31	11442.3698			
cong	56673.7539	1	56673.7539	21.555	0.000	***
cs/	81508.1211	31	2629.2942			
rc	15578.1602	1	15578.1602	26.681	0.000	***
rcs/	18100.2148	31	583.8779			
decisio	23047.0352	1	23047.0352	16.335	0.000	***
ds/	43737.8398	31	1410.8981			
rd	40930.3477	1	40930.3477	34.110	0.000	***
rds/	37198.0273	31	1199.9364			
cd	2506.2539	1	2506.2539	2.761	0.107	
cds/	28141.6211	31	907.7942			
rcd	597.1914	1	597.1914	0.606	0.442	
rcds/	30572.1836	31	986.1995			

Experiment 10

SOURCE: grand mean

respo	cong	decis	N	MEAN	SD	SE
			192	635.9479	113.0567	8.1592

SOURCE: response

respo	cong	decis	N	MEAN	SD	SE
voice			96	592.4688	109.8452	11.2110
gest			96	679.4271	99.0532	10.1096

SOURCE: cong

respo	cong	decis	N	MEAN	SD	SE
	cong		64	620.5000	111.4797	13.9350
	incon		64	644.4375	114.7558	14.3445
	neutr		64	642.9063	113.0920	14.1365

SOURCE: response cong

respo	cong	decis	N	MEAN	SD	SE
voice	cong		32	582.4063	110.0624	19.4565
voice	incon		32	599.2500	115.4693	20.4123
voice	neutr		32	595.7500	106.6133	18.8467
gest	cong		32	658.5938	100.7308	17.8069
gest	incon		32	689.6250	95.9888	16.9686
gest	neutr		32	690.0625	100.1743	17.7085

SOURCE: decision

respo	cong	decis	N	MEAN	SD	SE
		u/d	96	628.3542	108.8329	11.1077
		l/r	96	643.5417	117.2051	11.9622

SOURCE: response decision

respo	cong	decis	N	MEAN	SD	SE
voice		u/d	48	597.1875	113.8283	16.4297
voice		l/r	48	587.7500	106.7067	15.4018
gest		u/d	48	659.5208	94.8699	13.6933
gest		l/r	48	699.3333	100.1106	14.4497

SOURCE: cong decision

respo	cong	decis	N	MEAN	SD	SE
	cong	u/d	32	615.9375	110.0041	19.4462
	cong	l/r	32	625.0625	114.5096	20.2426
	incon	u/d	32	635.5625	110.2545	19.4904
	incon	l/r	32	653.3125	120.1828	21.2455
	neutr	u/d	32	633.5625	108.6290	19.2031
	neutr	l/r	32	652.2500	118.3709	20.9252

SOURCE: response cong decision

respo	cong	decis	N	MEAN	SD	SE
voice	cong	u/d	16	586.2500	117.5106	29.3776
voice	cong	l/r	16	578.5625	105.8055	26.4514
voice	incon	u/d	16	607.5625	122.9856	30.7464
voice	incon	l/r	16	590.9375	110.8257	27.7064
voice	neutr	u/d	16	597.7500	106.8703	26.7176
voice	neutr	l/r	16	593.7500	109.8214	27.4553
gest	cong	u/d	16	645.6250	96.5380	24.1345
gest	cong	l/r	16	671.5625	106.2613	26.5653
gest	incon	u/d	16	663.5625	91.2389	22.8097
gest	incon	l/r	16	715.6875	96.2723	24.0681
gest	neutr	u/d	16	669.3750	101.1427	25.2857
gest	neutr	l/r	16	710.7500	97.9588	24.4897

FACTOR:	subs	response	cong	decision	rt
LEVELS:	16	2	3	2	192
TYPE :	RANDOM	WITHIN	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	362964.0833	1	362964.0833	29.834	0.000	***
rs/	182490.9167	15	12166.0611			
cong	22984.2917	2	11492.1458	9.127	0.001	***
cs/	37772.5417	30	1259.0847			
rc	2908.2917	2	1454.1458	0.920	0.410	
rcs/	47443.2083	30	1581.4403			
decisio	11071.6875	1	11071.6875	6.067	0.026	*
ds/	27373.3125	15	1824.8875			
rd	29106.7500	1	29106.7500	12.705	0.003	**
rds/	34364.5833	15	2290.9722			
cd	889.1250	2	444.5625	0.463	0.634	
cds/	28818.3750	30	960.6125			
rcd	2557.6250	2	1278.8125	0.875	0.427	
rcds/	43826.5417	30	1460.8847			

Experiment 11

SOURCE: grand mean

respo	cong	decis	N	MEAN	SD	SE
			168	688.1964	147.0185	11.3427

SOURCE: response

respo	cong	decis	N	MEAN	SD	SE
voice			84	706.9524	150.2687	16.3957
gest			84	669.4405	142.1146	15.5060

SOURCE: cong

respo	cong	decis	N	MEAN	SD	SE
	cong		56	681.3214	147.7530	19.7443
	incon		56	703.6607	146.7983	19.6168
	neutr		56	679.6071	147.9251	19.7673

SOURCE: response cong

respo	cong	decis	N	MEAN	SD	SE
voice	cong		28	697.8571	148.5637	28.0759
voice	incon		28	726.1786	154.2458	29.1497
voice	neutr		28	696.8214	151.5789	28.6457
gest	cong		28	664.7857	147.7571	27.9235
gest	incon		28	681.1429	138.0370	26.0865
gest	neutr		28	662.3929	144.8563	27.3753

SOURCE: decision

respo	cong	decis	N	MEAN	SD	SE
		u/d	84	675.4762	139.1439	15.1818
		l/r	84	700.9167	154.2749	16.8328

SOURCE: response decision

respo	cong	decis	N	MEAN	SD	SE
voice		u/d	42	698.7857	144.6522	22.3203
voice		l/r	42	715.1190	157.0069	24.2267
gest		u/d	42	652.1667	130.9841	20.2113
gest		l/r	42	686.7143	152.0446	23.4610

SOURCE: cong decision

respo	cong	decis	N	MEAN	SD	SE
	cong	u/d	28	667.0714	140.1071	26.4777
	cong	l/r	28	695.5714	156.2667	29.5316
	incon	u/d	28	689.4286	140.0189	26.4611
	incon	l/r	28	717.8929	154.5065	29.1990
	neutr	u/d	28	669.9286	141.3325	26.7093
	neutr	l/r	28	689.2857	156.2208	29.5230

SOURCE: response cong decision

respo	cong	decis	N	MEAN	SD	SE
voice	cong	u/d	14	684.2857	145.5697	38.9051
voice	cong	l/r	14	711.4286	155.7334	41.6215
voice	incon	u/d	14	721.7143	153.4302	41.0060
voice	incon	l/r	14	730.6429	160.7170	42.9534
voice	neutr	u/d	14	690.3571	142.7987	38.1646
voice	neutr	l/r	14	703.2857	165.0401	44.1088
gest	cong	u/d	14	649.8571	137.6271	36.7824
gest	cong	l/r	14	679.7143	161.0047	43.0303
gest	incon	u/d	14	657.1429	122.1996	32.6592
gest	incon	l/r	14	705.1429	152.9730	40.8838
gest	neutr	u/d	14	649.5000	142.1125	37.9812
gest	neutr	l/r	14	675.2857	151.7460	40.5558

FACTOR:	subs	response	cong	decision	rt
LEVELS:	14	2	3	2	168
TYPE :	RANDOM	WITHIN	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	59100.0060	1	59100.0060	1.697	0.215	
rs/	452738.7440	13	34826.0572			
cong	20170.3929	2	10085.1964	16.078	0.000	***
cs/	16308.4405	26	627.2477			
rc	1201.6548	2	600.8274	0.911	0.414	
rcs/	17139.8452	26	659.2248			
decisio	27183.1488	1	27183.1488	23.917	0.000	***
ds/	14775.2679	13	1136.5591			
rd	3483.4821	1	3483.4821	1.489	0.244	
rds/	30406.6012	13	2338.9693			
cd	777.1548	2	388.5774	0.347	0.710	
cds/	29087.6786	26	1118.7569			
rcd	2463.8929	2	1231.9464	1.592	0.223	
rcds/	20124.2738	26	774.0105			

Chapter 5

Experiment 12

SOURCE: grand mean

respo	cong.	decis	N	MEAN	SD	SE
			96	576.6354	100.0125	10.2075

SOURCE: response

respo	cong.	decis	N	MEAN	SD	SE
voice			48	639.6667	63.3274	9.1405
arrow			48	513.6042	89.9586	12.9844

SOURCE: cong.

respo	cong.	decis	N	MEAN	SD	SE
	cong		48	562.7083	91.6777	13.2325
	incon		48	590.5625	106.8496	15.4224

SOURCE: response cong.

respo	cong.	decis	N	MEAN	SD	SE
voice	cong		24	631.2083	65.1486	13.2984
voice	incon		24	648.1250	61.6556	12.5854
arrow	cong		24	494.2083	56.0194	11.4349
arrow	incon		24	533.0000	112.3105	22.9253

SOURCE: decision

respo	cong.	decis	N	MEAN	SD	SE
		u/d	48	584.1458	86.4926	12.4841
		l/r	48	569.1250	112.3461	16.2158

SOURCE: response decision

respo	cong.	decis	N	MEAN	SD	SE
voice		u/d	24	626.7500	51.0756	10.4258
voice		l/r	24	652.5833	72.3752	14.7735
arrow		u/d	24	541.5417	94.2891	19.2467
arrow		l/r	24	485.6667	77.5733	15.8346

SOURCE: cong. decision

respo	cong.	decis	N	MEAN	SD	SE
	cong	u/d	24	565.3750	71.5753	14.6103
	cong	l/r	24	560.0417	109.7140	22.3953
	incon	u/d	24	602.9167	97.1023	19.8209
	incon	l/r	24	578.2083	116.5448	23.7896

SOURCE: response cong. decision

respo	cong.	decis	N	MEAN	SD	SE
voice	cong	u/d	12	618.1667	56.7416	16.3799
voice	cong	l/r	12	644.2500	72.6900	20.9838
voice	incon	u/d	12	635.3333	45.5439	13.1474
voice	incon	l/r	12	660.9167	74.2777	21.4421
arrow	cong	u/d	12	512.5833	37.5704	10.8457
arrow	cong	l/r	12	475.8333	66.4336	19.1777
arrow	incon	u/d	12	570.5000	123.8867	35.7630
arrow	incon	l/r	12	495.5000	89.2071	25.7519

FACTOR:	subs	response	cong.	decision	r.t
LEVELS:	12	2	2	2	96
TYPE :	RANDOM	WITHIN	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	381402.0937	1	381402.0937	23.030	0.001	***
rs/	182169.5313	11	16560.8665			
cong.	18620.5104	1	18620.5104	10.472	0.008	**
cs/	19560.1146	11	1778.1922			
rc	2871.0938	1	2871.0938	0.743	0.407	
rcs/	42489.0313	11	3862.6392			
decisio	5415.0104	1	5415.0104	1.949	0.190	
ds/	30566.6146	11	2778.7831			
rd	40057.5104	1	40057.5104	19.298	0.001	**
rds/	22832.6146	11	2075.6922			
cd	2252.3438	1	2252.3438	0.957	0.349	
cds/	25894.7813	11	2354.0710			
rcd	2137.5937	1	2137.5937	0.734	0.410	
rcds/	32048.0313	11	2913.4574			

Experiment 13

SOURCE: grand mean

respo	cong	decis	N	MEAN	SD	SE
			88	651.5795	150.6783	16.0624

SOURCE: response

respo	cong	decis	N	MEAN	SD	SE
voice			44	712.4545	114.0894	17.1996
dot			44	590.7045	159.1716	23.9960

SOURCE: cong

respo	cong	decis	N	MEAN	SD	SE
	cong		44	619.2273	127.2174	19.1787
	incon		44	683.9318	166.1614	25.0498

SOURCE: response cong

respo	cong	decis	N	MEAN	SD	SE
voice	cong		22	683.0455	123.3020	26.2881
voice	incon		22	741.8636	98.1686	20.9296
dot	cong		22	555.4091	96.9662	20.6733
dot	incon		22	626.0000	199.6623	42.5682

SOURCE: decision

respo	cong	decis	N	MEAN	SD	SE
		u/d	44	643.0000	131.6604	19.8486
		l/r	44	660.1591	168.6736	25.4285

SOURCE: response decision

respo	cong	decis	N	MEAN	SD	SE
voice		u/d	22	683.4091	112.4996	23.9850
voice		l/r	22	741.5000	110.5841	23.5766
dot		u/d	22	602.5909	139.3446	29.7084
dot		l/r	22	578.8182	179.3453	38.2365

SOURCE: cong decision

respo	cong	decis	N	MEAN	SD	SE
	cong	u/d	22	617.4091	130.7734	27.8810
	cong	l/r	22	621.0455	126.6121	26.9938
	incon	u/d	22	668.5909	130.4628	27.8148
	incon	l/r	22	699.2727	197.5356	42.1147

SOURCE: response cong decision

respo	cong	decis	N	MEAN	SD	SE
voice	cong	u/d	11	670.1818	146.3385	44.1227
voice	cong	l/r	11	695.9091	100.7377	30.3736
voice	incon	u/d	11	696.6364	69.1235	20.8415
voice	incon	l/r	11	787.0909	104.6885	31.5648
dot	cong	u/d	11	564.6364	91.4967	27.5873
dot	cong	l/r	11	546.1818	105.7647	31.8892
dot	incon	u/d	11	640.5455	170.9815	51.5529
dot	incon	l/r	11	611.4545	232.4144	70.0756

FACTOR:	subs	response	cong	decision	rt
LEVELS:	11	2	2	2	88
TYPE :	RANDOM	WITHIN	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	326107.3750	1	326107.3750	12.755	0.005	**
rs/	255674.7500	10	25567.4750			
cong	92106.9205	1	92106.9205	13.308	0.004	**
cs/	69210.2045	10	6921.0205			
rc	762.2841	1	762.2841	0.075	0.789	
rcs/	101336.3409	10	10133.6341			
decisio	6477.5568	1	6477.5568	0.786	0.396	
ds/	82411.0682	10	8241.1068			
rd	36859.1023	1	36859.1023	8.792	0.014	*
rds/	41924.0227	10	4192.4023			
cd	4023.0114	1	4023.0114	2.123	0.176	
cds/	18951.1136	10	1895.1114			
rcd	7809.5568	1	7809.5568	2.544	0.142	
rcds/	30692.0682	10	3069.2068			

Chapter 6

Experiment 14

SOURCE: grand mean

respo	rel	N	MEAN	SD	SE
		96	885.5729	230.0698	23.4814

SOURCE: response

respo	rel	N	MEAN	SD	SE
voice		48	797.7292	130.5577	18.8444
gest		48	973.4167	272.3668	39.3128

SOURCE: rel

respo	rel	N	MEAN	SD	SE
	rel	48	855.5833	206.9501	29.8707
	unrel	48	915.5625	249.6499	36.0339

SOURCE: response rel

respo	rel	N	MEAN	SD	SE
voice	rel	24	787.6250	117.0332	23.8893
voice	unrel	24	807.8333	144.6440	29.5253
gest	rel	24	923.5417	253.3451	51.7139
gest	unrel	24	1023.2917	286.7363	58.5298

FACTOR:	subs	response	rel	rt
LEVELS:	24	2	2	96
TYPE :	RANDOM	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	740786.3438	1	740786.3438	13.686	0.001	**
rs/	1244961.4063	23	54128.7568			
rel	86340.0104	1	86340.0104	9.296	0.006	**
rs/	213626.7396	23	9288.1191			
rr	37961.2604	1	37961.2604	3.169	0.088	
rrs/	275508.4896	23	11978.6300			

Experiment 15

SOURCE: grand mean

respo	cong	decis	N	MEAN	SD	SE
			192	662.8438	168.2321	12.1411

SOURCE: response

respo	cong	decis	N	MEAN	SD	SE
voice			96	658.9479	143.0007	14.5950
gest			96	666.7396	190.8458	19.4781

SOURCE: cong

respo	cong	decis	N	MEAN	SD	SE
	cong		96	640.0521	149.5065	15.2589
	incon		96	685.6354	183.0299	18.6804

SOURCE: response cong

respo	cong	decis	N	MEAN	SD	SE
voice	cong		48	646.4167	141.7419	20.4587
voice	incon		48	671.4792	144.6444	20.8776
gest	cong		48	633.6875	158.1344	22.8247
gest	incon		48	699.7917	215.3633	31.0850

SOURCE: decision

respo	cong	decis	N	MEAN	SD	SE
		b/s	96	656.8438	164.0802	16.7464
		t/s	96	668.8438	172.9363	17.6502

SOURCE: response decision

respo	cong	decis	N	MEAN	SD	SE
voice		b/s	48	655.5000	137.8818	19.9015
voice		t/s	48	662.3958	149.3245	21.5531
gest		b/s	48	658.1875	188.1551	27.1578
gest		t/s	48	675.2917	195.1089	28.1615

SOURCE: cong decision

respo	cong	decis	N	MEAN	SD	SE
	cong	b/s	48	636.0000	149.6320	21.5975
	cong	t/s	48	644.1042	150.8532	21.7738
	incon	b/s	48	677.6875	176.4660	25.4707
	incon	t/s	48	693.5833	190.9015	27.5543

SOURCE: response cong decision

respo	cong	decis	N	MEAN	SD	SE
voice	cong	b/s	24	644.4167	143.8408	29.3614
voice	cong	t/s	24	648.4167	142.6763	29.1237
voice	incon	b/s	24	666.5833	133.8016	27.3121
voice	incon	t/s	24	676.3750	157.4826	32.1460
gest	cong	b/s	24	627.5833	157.8447	32.2199
gest	cong	t/s	24	639.7917	161.5781	32.9820
gest	incon	b/s	24	688.7917	213.2469	43.5289
gest	incon	t/s	24	710.7917	221.4780	45.2090

FACTOR:	subs	response	cong	decision	rt
LEVELS:	24	2	2	2	192
TYPE :	RANDOM	WITHIN	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	2914.0833	1	2914.0833	0.088	0.770	
rs/	762525.6667	23	33153.2899			
cong	99736.3333	1	99736.3333	17.746	0.000	***
cs/	129267.4167	23	5620.3225			
rc	20213.0208	1	20213.0208	3.643	0.069	
rcs/	127604.7292	23	5548.0317			
decisio	6912.0000	1	6912.0000	4.624	0.042	*
ds/	34383.2500	23	1494.9239			
rd	1250.5208	1	1250.5208	1.552	0.225	
rds/	18535.7292	23	805.9013			
cd	728.5208	1	728.5208	0.343	0.564	
cds/	48866.7292	23	2124.6404			
rcd	48.0000	1	48.0000	0.037	0.849	
rcds/	29832.2500	23	1297.0543			

Chapter 7

Experiment 16

SOURCE: grand mean

respo	cong	N	MEAN	SD	SE
		84	857.0357	171.2440	18.6842

SOURCE: response

respo	cong	N	MEAN	SD	SE
voice		42	819.3095	102.0183	15.7418
expr		42	894.7619	214.5706	33.1090

SOURCE: cong

respo	cong	N	MEAN	SD	SE
	cong	28	822.7857	144.9358	27.3903
	incon	28	863.6071	163.4434	30.8879
	unrel	28	884.7143	200.9223	37.9707

SOURCE: response cong

respo	cong	N	MEAN	SD	SE
voice	cong	14	821.0714	112.4459	30.0524
voice	incon	14	812.7857	98.4286	26.3061
voice	unrel	14	824.0714	102.0712	27.2797
expr	cong	14	824.5000	176.0064	47.0397
expr	incon	14	914.4286	200.5771	53.6065
expr	unrel	14	945.3571	255.9404	68.4029

FACTOR:	subs	response	cong	data
LEVELS:	14	2	3	84
TYPE :	RANDOM	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	119554.2976	1	119554.2976	4.394	0.056	
rs/	353686.5357	13	27206.6566			
cong	55505.7857	2	27752.8929	5.527	0.010	**
cs/	130558.2143	26	5021.4698			
rc	55818.4524	2	27909.2262	7.147	0.003	**
rcs/	101532.2143	26	3905.0852			

Experiment 17: Reaction Time Data

SOURCE: grand mean

respo	cong	N	MEAN	SD	SE
		96	782.0729	270.2057	27.5778

SOURCE: response

respo	cong	N	MEAN	SD	SE
face		48	667.8542	206.3497	29.7840
voice		48	896.2917	279.9080	40.4012

SOURCE: cong

respo	cong	N	MEAN	SD	SE
	cong	24	742.5833	164.1039	33.4976
	incon	24	844.5833	399.9486	81.6392
	mix	24	766.0000	230.3491	47.0198
	neut	24	775.1250	235.7828	48.1290

SOURCE: response cong

respo	cong	N	MEAN	SD	SE
face	cong	12	656.7500	124.4553	35.9271
face	incon	12	713.9167	318.4526	91.9294
face	mix	12	638.0000	131.5433	37.9733
face	neut	12	662.7500	210.4256	60.7446
voice	cong	12	828.4167	157.3054	45.4102
voice	incon	12	975.2500	442.4878	127.7352
voice	mix	12	894.0000	240.6129	69.4589
voice	neut	12	887.5000	210.7362	60.8343

FACTOR:	subs	response	cong	rt
LEVELS:	12	2	4	96
TYPE :	RANDOM	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p	
respons	1252408.5938	1	1252408.5938	7.503	0.019	*
rs/	1836020.7812	11	166910.9801			
cong	138566.1979	3	46188.7326	3.626	0.023	*
cs/	420373.4271	33	12738.5887			
rc	30470.1146	3	10156.7049	0.536	0.661	
rcs/	625790.0104	33	18963.3336			

Experiment 17: Error Data

SOURCE: grand mean

respo	cong	N	MEAN	SD	SE
		96	6.8438	11.2412	1.1473

SOURCE: response

respo	cong	N	MEAN	SD	SE
face		48	6.3125	11.5724	1.6703
voice		48	7.3750	10.9964	1.5872

SOURCE: cong

respo	cong	N	MEAN	SD	SE
	cong	24	4.0833	7.6722	1.5661
	incon	24	9.7917	14.7677	3.0144
	mix	24	6.4583	10.2913	2.1007
	neut	24	7.0417	11.0394	2.2534

SOURCE: response cong

respo	cong	N	MEAN	SD	SE
face	cong	12	3.0000	4.7863	1.3817
face	incon	12	8.7500	14.5360	4.1962
face	mix	12	6.1667	11.4957	3.3185
face	neut	12	7.3333	13.6471	3.9396
voice	cong	12	5.1667	9.8796	2.8520
voice	incon	12	10.8333	15.5671	4.4938
voice	mix	12	6.7500	9.4400	2.7251
voice	neut	12	6.7500	8.2696	2.3872

FACTOR:	subs	response	cong	error
LEVELS:	12	2	4	96
TYPE :	RANDOM	WITHIN	WITHIN	DATA

SOURCE	SS	df	MS	F	p
respons	27.0938	1	27.0938	0.455	0.514
rs/	655.2812	11	59.5710		
cong	395.9479	3	131.9826	2.563	0.071
cs/	1699.1771	33	51.4902		
rc	31.1979	3	10.3993	0.178	0.911
rcs/	1931.9271	33	58.5432		